"USING BAYESIAN CLASSSIFIER FOR CLASSIFICATION OF CITIES"

[1]SUSHMA A. BAKHADE
Shree. H.V.P.Mandal's, P.G. Department of Computer Science & Technology, D.C.P.E. An Autonomous College, Amravati.
sushamabakhade03@gmail.com

[2]RUPALI S. PIMPLE
Shree. H.V.P.Mandal's, P.G. Department of Computer Science & Technology, D.C.P.E. An Autonomous College, Amravati.
rupalipimple7@gmail.com

[3]PROF. S. Y. THAKUR
Shree. H.V.P.Mandal's, P.G. Department of Computer Science & Technology, D.C.P.E. An Autonomous College, Amravati.
sythakur@rediffmail.com

ABSTRACT: *Naïve Bayes is among the simplest probabilistic classifier. In the process of this classifier with the known structure class probabilities and conditional probabilities are calculated using training data. and then values of these probabilities are used to classify new observations. Results show that naïve bayes is the best classifiers against several common classifiers (Such as decision tree, neural network, and support vector machines) in term of accuracy and computational efficiency.*

Keywords: Naïve Bayes Classifier, Bayesian Classifier, city classifier etc.

1. INTRODUCTION

Bayesian network introduced by pearl(1988) are high level representations of probability distribution over a set of variable x={x1,x2,…….xn}that are used.[4]But In this paper the main aim of the Bayesian classifier is to classify cities whether it is a "Smart, Metro, or Normal city" according to their various parameter. It is necessary to define the parameter for the classification. Here we classify the cities mainly on the basis of the area. Naïve Bayes has proven effective in many practical applications, including text classification, medical diagnosis and system performance management. Naive classifier uses statistical as well as a supervised learning method for classification. It is based on application of Bayes Theorem with naïve independence assumption. Classification can be done by finding rules that partition the given data into disjoints groups. Bayesian classification it is classification technique based on 'Bayes Theorem' with an assumption of independence among predictors. There are two type of classification. First is Decision tree classifier and second is Bayesian classifier. In this paper Decision tree classifiers which perform by constructing a tree based on training instances with leaves having class labels. Decision tree classifier and they work better in the case of missing & null attribute values .The naïve byes text classifier has been widely used because of its simplicity in both the training and classifying stage.

2. NAIVE BAYES CLASIFIER IS ALSO USEFUL FOR FOLLOWING PURPOSE

2.1 Selective Naïve Bayes Classifier

The Naïve independence assumption can harm the performance when violated. In order to better deal 9with highly correlated variables the selective Naïve Bayes approach of Longley and Sage (1994) exploits wrapper approach to select subset of variables which optimizes the classification accuracy. The selective Naïve Bayes approach performs quite well on datasets with a reasonable numbers of variables; it does not scale on very large datasets with hundreds of thousands of instances and, thousands of variables such as in marketing application.[4]

Assumption and Notation of Selective Naïve Bayes Classifier

Let X = (X1, X2……..Xk) be the vector of the k explanatory variables and Y the class variable Let λ1, λ2,……λj be the J class labels of Y.
Let N be the number of instances and D = {D1,D2,……DN}the labeled database containing the instances Dn = (X(n),Y(n).
Let M ={Mm}be the set of all potential selective naïve bayes models.

The Naïve Bayes classifier assigns to each instances the class value having the highest conditional probability. The naïve bayes classifier is poor at predicting the true class conditional probabilities since the independence assumption is usually violated in real data application [4]

2.2 Document Classification

Document classification is a growing interest in the research of text mining. Correctly identifying the documents into particular category is still presenting challenge because of large and vast amount of features in the datasets Naïve Bayes is potentially good at serving as a documents classification

model due to its simplicity. With the explosive growth of the textual information from the electronic documents and World Wide Web. Recently numerous research activities have been conducted in the field of documentation. Classification is particularly applying in spam filtering, e-mail categorization, website classification, formation of knowledge repositories and ontology mapping,[3]

2.3 Optimization Models

The optimization is used the learning process. The learning process of this classifier with known structure, class probabilities, and conditional probability are calculated using training data and then values of these probabilities are used to classify new observation. [2]

In this paper, we proposed use of Bayesian classifier for classification of cities. For this purpose we use Naïve Bayes classifier algorithm. For this research we can collect the information of cities in term of city names, area, population, transportation, water management system, waste management system, industries, hospital, literacy, government services, & public safety to classify city whether it is smart city, normal city, or metro city. By using such details of cities we collect data only for 100 cities & these 100 cities will be classify on the basis of area. Only by using naïve bayes algorithm. And then these 100 cities will be evaluated in the terms of smart, normal &metro cities

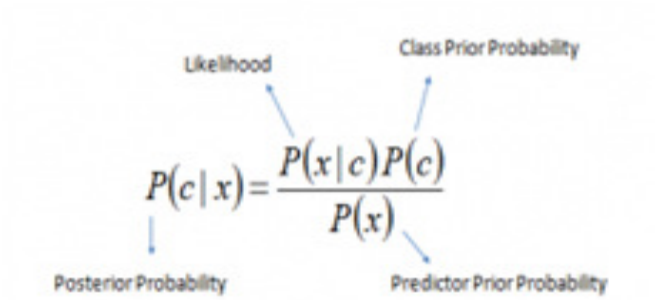
3. METHODOLOGY USED

NAÏVE BAYES: The Naïve Bayes classifier is based on Bayesian probability model. If a class is provided, Naive Bayes classifier assumes that the value of one feature is independent of any other feature. is based on the mathematical principle of conditional probability If n attributes the given independent assumptions made by the naïve bayes classifier is 2n! A conditional probability model is given as P(Ci/x)

3.1. NAIVE BAYES ALGORITHM

In this paper the classification technique based on Bayes Theorem with an assumption of independence among predictors. In simple term , a naïve bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

The Equation Below



$$P(c|X) = P(x_1|c)\times P(x_2|c)\times\cdots\times P(x_n|c)\times P(c)$$

Where;

P(c/x) is the posterior probability of class c with given predictor.
P(c) is the prior probability of class.
P(x/c) is the likelihood which is the probability of predictor given class.
P(x) is the prior probability of predictors**.**

3.2. HOW NAÏVE BAYES ALGORITHM WORK

In this paper we can use Naïve Bayes algorithm lets follow the steps to perform it.

Step1: Convert the data set into a frequency table.

Step2: Create likelihood table by finding the probabilities.

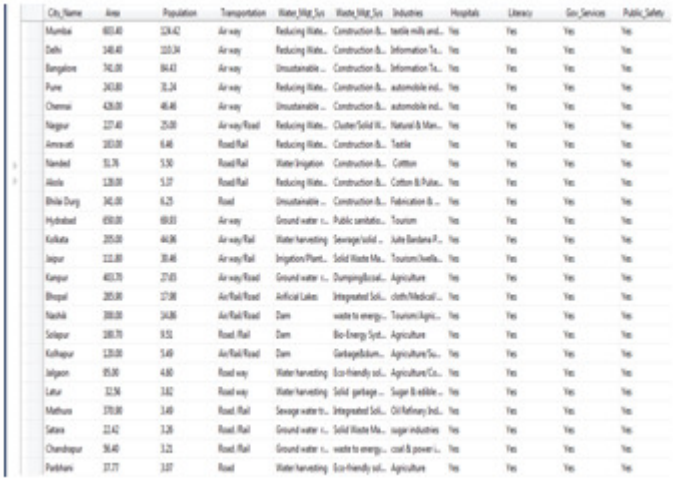
Step3: Now use Naïve Bayesian equation to calculate the posterior probability for each class.

4. PARAMETERS OF CITIES

Following parameters are used for the classification of cities.

1. City name
2. Area
3. Population
4. Transportation
5. Water Management System
6. Waste Management System
7. Industries
8. Hospitals
9. Literacy
10. Government Services
11. Public Safety

5. DATA SET FOR CITIES



City_Name	Area	Population	Transportation	Water_Mgt_Sys	Waste_Mgt_Sys	Industries	Hospitals	Literacy	Gov_Services	Public_Safety
Mumbai	603.40	124.42	Air way	Reducing Wate...	Construction &...	textile mills and...	Yes	Yes	Yes	Yes
Delhi	148.40	110.34	Air way	Reducing Wate...	Construction &...	Information Te...	Yes	Yes	Yes	Yes
Bangalore	741.00	84.43	Air way	Unsustainable ...	Construction &...	Information Te...	Yes	Yes	Yes	Yes
Pune	243.80	31.24	Air way	Reducing Wate...	Construction &...	automobile ind...	Yes	Yes	Yes	Yes
Chennai	426.00	46.46	Air way	Unsustainable ...	Construction &...	automobile ind...	Yes	Yes	Yes	Yes
Nagpur	227.40	25.00	Air way/Road	Reducing Wate...	Cluster/Solid W...	Natural & Man...	Yes	Yes	Yes	Yes
Amravati	183.00	6.46	Road/Rail	Reducing Wate...	Construction &...	Textile	Yes	Yes	Yes	Yes
Nanded	51.76	5.50	Road/Rail	Water Irrigation	Construction &...	Cotton	Yes	Yes	Yes	Yes
Akola	128.00	5.37	Road/Rail	Reducing Wate...	Construction &...	Cotton & Pulse...	Yes	Yes	Yes	Yes
Bhilai Durg	341.00	6.25	Road	Unsustainable ...	Construction &...	Fabrication & ...	Yes	Yes	Yes	Yes
Hyderabad	650.00	69.03	Air way	Ground water r...	Public sanitatio...	Tourism	Yes	Yes	Yes	Yes
Kolkata	205.00	44.86	Air way/Rail	Water harvesting	Sewage/solid ...	Jute Bardana P...	Yes	Yes	Yes	Yes
Jaipur	232.80	30.46	Air way/Rail	Irrigation/Plant...	Solid Waste Ma...	Tourism/textile...	Yes	Yes	Yes	Yes
Kanpur	403.70	27.83	Air way/Road	Ground water r...	Dumping&Locat...	Agriculture	Yes	Yes	Yes	Yes
Bhopal	285.88	17.88	Air/Rail/Road	Artificial Lakes	Integrated Soli...	cloth/Medicall...	Yes	Yes	Yes	Yes
Nashik	268.00	14.86	Air/Rail/Road	Dam	waste to energy...	Tourism/Agric...	Yes	Yes	Yes	Yes
Solapur	180.76	9.51	Road /Rail	Dam	Bio-Energy Syst...	Agriculture	Yes	Yes	Yes	Yes
Kolhapur	120.00	5.49	Air/Rail/Road	Dam	Garbage&dum...	Agriculture/So...	Yes	Yes	Yes	Yes
Jalgaon	85.00	4.80	Road way	Water harvesting	Eco-friendly sol...	Agriculture/Co...	Yes	Yes	Yes	Yes
Latur	32.56	3.82	Road way	Water harvesting	Solid garbage ...	Sugar & edible ...	Yes	Yes	Yes	Yes
Mathura	170.00	3.49	Road /Rail	Sewage water tr...	Integrated Soli...	Oil Refinery Ind...	Yes	Yes	Yes	Yes
Satara	22.42	3.28	Road /Rail	Ground water r...	Solid Waste Ma...	sugar industries	Yes	Yes	Yes	Yes
Chandrapur	56.40	3.21	Road /Rail	Ground water r...	waste to energy...	coal & power i...	Yes	Yes	Yes	Yes
Parbhani	37.77	3.07	Road	Water harvesting	Eco-friendly sol...	Agriculture	Yes	Yes	Yes	Yes

5.1 Frequency Table

City	Count value
Normal city	**30**
Metro city	**61**
Smart city	**9**
Total city	**100**

5.2 Likelihood Table

City	Count value
Normal city	**30=30/100=0.3**
Metro city	**61=61/100=0.61**
Smart city	**9=9/100=0.9**
Total city	**100**

6. ADVANTAGES

1. Very simple, easy to implement and fast.
2. If the NB conditional independence assumption holds, then it will be converge quicker than discriminative models like logistic regression
3. Even if the NB assumption doesn't hold, it works great in practice.
4. Need less training data.
5. Highly scalable. Its scale linearly with the number of predictors and data points.
6. Can make probabilistic predictions.
7. Handles continuous and discrete data.
8. Can be used for both binary and multiclass classification problems.

7. APPLICATIONS

There are four applications as follows

Real time Prediction: Naive Bayes is an eager learning classifier and it is sure fast .Thus, it could be used for making predictions in real time.
Multi class Prediction: This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.

Text classification/ Spam Filtering/ Sentiment Analysis: Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments.

Recommendation System: Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a given resource or not.

8. CONCLUSION

In this paper, classify the cities using Bayesian Classifier. For this purpose Naïve Bayes classifier methodology is used. Because, this methodology is easy to calculate and easy to understands than the other classifier method like Decision-tree classifier, SVM etc. This classifier also used for the Data mining approach. So, it will better predict and model the behavior of the segmentation. Segmentation with the help of data mining from various existing system is very important exercise & a must for effective business development. Presently, using Bayesian classifier for a classification of cities can be done only for 100 cities but, in the future apply bayesian classifier for classification of cities India level.

9. REFERENCES

[1] Bhosale Dipali &Ade Roshani: "Feature Selection based classification using Naïve Bayes, J48 and Support Vector Machine", International Journal of Computer Applications, Volume 99, No.16, August 2014.

[2]Taheri Sona & Mammadav Musa: "Learning the Naïve Bayes Classifier with Optimization Models", In J. Appl. Math.Computer.Sci Vol.23, No4, p.p 787-795, 2013.

[3] Ting S.L. Ip, W. H. Albert H.C. Tsang: "Is Naïve Bayes a Good Classifier For Document Classification?",International Journal Of Software Engineering And Its Applications Val. 5, No. 3,p.p 37- 46 July 2011.

[4] Boulle Marc "Compression- Based Averaging Of Selective Naïve Bayes Classifiers", Journal Of Machine Learning Research, 8 p.p1659 -1685, 2007.

[5] Rish I.: "An Empirical Study of the Naïve Bayes Classifier", T .J. Watson Research Center, 30 Saw Mill River Road, Hawthorne.

[6] J. Hilden, Statistical diagnosis based on Conditional independence does not require it. Comput.Biol.Med. 14(4) : 429-435, 1984.

[7] P. Langley, W.Iba, and k. Thompson. An analysis of baysian classifier. In proceeding of the Tenth National Conference of on Artificial Intelligence, pages 399- 406, San Jose, CA 1992.AAAI press.

[8] P. Domingos and M. Pazzani, on the optimality of the simple Baysian Classifier under zero- one loss. Machine Learning, 29: 103- 130, 1997.

[9] TOM M. Mitchell. Machine Learning. McGraw- Hill, 1997

[10] I. Rish, J. Hellerstein, and T. Jayram. An analysis of data characteristics that affect Naïve Bayes Performance. Technical Report RC21993,IBM T.T. Watson Center ,2001.