ISSN NO: 1133-598X

Modeling Consumer Behavior Prediction Using a Machine Learning Algorithm

Y Prasad Reddy[1,] N J Pramod Dinakar 2, S Khaja Kizer 3, A Ramprakash Reddy4

[1,2,3,4] Asst. Professor, Department of CSE, K.S.R.M College of Engineering(A), Kadapa

Abstract –

The ability of machine learning algorithms to accurately forecast future outcomes has boosted their significance. The volatility of particular customer scenarios makes performance predictions difficult. Many different algorithms have been created to do the same thing. AODE, Naive Bayes, and AODEsr Bayes were all analyzed here. We utilized the WEKA program to implement these strategies and develop a new, more precise model. Throughout the course of development, we have worked diligently to clean up the data and make it more dependable; now we must filter out the unnecessary details. This process will assign the value Wj to the newly filtered data. The mistake is denoted by E (j, k), where j is either a constant or an assumption. Your k-related objectives are the deciding factor. An alternative function for describing noise is N = E + Wj.

INTRODUCTION

A subfield of AI, machine learning entails programming a computer to make accurate predictions. During this training, we provide the machine with a set of guidelines or patterns to follow. Therefore, Machine Learning uses database knowledge to produce input data. We need an algorithm and pattern to get the necessary information since we are designing our system to make predictions or extract useful information from an incoming data set. Following the completion of these two stages, the machine will be able to finish the

Actions to be taken

Collect the necessary data, sort it, and synthesize it. Infer outcomes by using analytical evidence Determine the likelihood of individual impacts changing in response to a given development.

TYPES OF MACHINE LEARNING

Machine learning algorithms are basically used to recognize patterns and subsequently generate a solution. Machine learning.algorithms.are classified as:

SUPERVISED LEARNING

In this type of learning, information is available in advance. In order to ensure adequate allocation of data to groups of algorithms, that should be explained. In other words, the system learns on the basis of input and output power. In supervised learning, the program manager, who acts as a type of teacher, gives the correct amount of feedback. The purpose is to train the method in the perspective of sequential input and output calculations and establish communication. The Naive Bayes is a model of probabilistic distinctions, based on the concept of autonomy. Though, in numerous real-world mining applications, this statement is often dishonored. In response to this statement, scholars have done a great deal of testing the correctness of NB by abating the quality of their stability. Webb et al. [1] have proposed an idea named Averaged One- Dependence Estimators (AODE) that decreases the independent predictive value by sampling all prototypes from a constrained class of dependent classifiers. Inspired by this research, we rely on that passing on diverse value to these different classifiers can lead to greater enhancement. We have experimentally verified our algorithm with Weak tool [2], using Super Market data sets and briefly defined a comparative study between Naive ayes, AODE and Aiders. The investigational outcomes indicate that proposed algorithm meaningfully leave behind all the other algorithms used to compare.

UNSUPERVISED LEARNING

In this learning scheme, values are not available previously. Basically it is used for clustering purposes. The machine tries to organize and filter the information entered according to specific features. For example, a machine can learn that coins of different colors can be arranged according to a different "color" to arrange them.

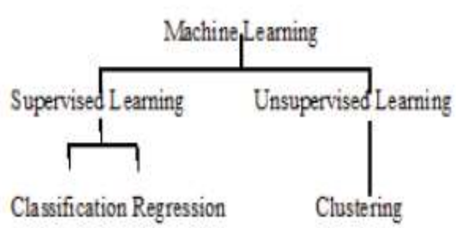

Fig. 1

NAIVEBAYS ALGORITHMS

Classification is first and foremost important thing in the area of data mining and machine learning. Learning about Bayesian classification is the process of making a different classifier from a training set by class labels.

Take Xi, i.=.1,.2,. ..., n,.are the values of the values xi, i.=.1,.2,...., n correspondingly. These are the attributes which will be used jointly to forecast the value of E of the study.

Therefore, the Bayesian classifier [9] can be demarcated as:

Arg.max.P(c) P(x1, x2, xn | c) where c.__C Consider that entire features are autonomous given the class,

then the resultant classifier is named Naive ayes: arg.max P(c) _ ______ _ _ __ __where c_C

Correctly Classified Instances	2948	63.71%
Incorrectly Classified Instances	1679	36.29%
Kappa statistic	0	
Mean absolute error	0.4624	
Root mean squared error	0.4808	
Relative absolute error	100%	
Root relative squared error	100%	
Total Number of Instances	4627	
Time Taken to Build	0.02 seconds	

Fig. 2

Detailed Accuracy By Class			
			Weighted Avg.
TP Rate	1	0	0.637
FP Rate	1	0	0.637
Precision	0.637	0	0.406
Recall	1	0	0.637
F-Measure	0.778	0	0.496
ROC Area	0.499	0.499	0.499
Class	low	high	

Fig. 3

Confusion Matrix

a	b	← classified as
2948	0 |	a = low
1679	0 |	b = high

AODE

The most recent development project for NaiveBays is named Averaged One-Dependence Estimators, or simply AODE [1]. In AODE, a set of fixed-income students learn and a prediction is generated by guessing the guesses of all those students who once relied on one. For simplicity, a single dependency class is created first for each character, where the attribute is set to be the parent of all other attributes. Subsequently, AODE reaches directly to the aggregated scale containing most of the unique trees obtained by the Bayes construct. AODE divides the test model using Equation [9]:

$$\text{Arg} \quad \max\left(\sum_{i=1 \wedge T(x_i)\geq m}^{n} P(x_i, c) \prod_{j=1, j<>i}^{n} P(xj \mid x_i, c)\right/$$

!"#_$%&!') Where T (xi) is a calculation of the number of training sessions that have the value of the xi and is used to impose the limit on which they place on the support required to accept possible conditional limitations. The nonparent is the number of root symbols, fulfilling the condition that the training conditions contain more than m examples of the values of the parent attribute Ai. In the present study they use m = 30. In addition, AODE measures the probability basis P (xi, c) and P (xj | xi, c) as follows:

$$P(x_i, c) = T(a_i, c) + 1 / N + v_i * k$$
$$P(x_j \mid x_i, c) = T(x_j, x_i, c) + 1 / T(x_i, c) + v_j$$

The median reliability estimation algorithm works the same way as the Naive Bayesian class, but allows for two dimensional dependencies within the input test while continuing ignoring the complex dependency relationships Involving three or more values. AODE performs well with a large number of data objects. However, because all input price pairs are considered by the integrative method, it is not possible to use the AODE algorithm with high dimensional values. When there are multiple input values, it may be reasonable to use only dependency estimates in those cases where the dependency is proven or at least suspected.

Correctly Classified Instances	2957	63.91%
Incorrectly Classified Instances	1670	36.09%
Kappa statistic	0.0943	
Mean absolute error	0.492	
Root mean squared error	0.4923	
Relative absolute error	106.39%	
Root relative squared error	102.40%	
Total Number of Instances	4627	
Time Taken to Build	1.06 seconds	

Fig. 4

Detailed Accuracy By Class			
			Weighted Avg.
TP Rate	0.9	0.182	0.639
FP Rate	0.818	0.1	0.558
Precision	0.659	0.507	0.604
Recall	0.9	0.182	0.639
F-Measure	0.761	0.268	0.582
ROC Area	0.687	0.68	0.687
Class	low	high	

Fig. 5

Confusion Matrix

```
   a      b          <- classified as
 2652   296 |          a = low
 1374   305 |          b = high
```

Subsumption Resolution (AODEsr) one-dimensional dependency ratios a certain type of dependency between Symbols results in a singular value of the other. For example, consider Gender and Pregnancy as two signs, and

Baby = yes means Gender = woman. Therefore, gender = female is the width of pregnancy = yes. As such, Pregnancy = no standard sex = male. Where one value xi is a combination of the other, xj, P (y | xi, xj) = P (y | xj). As a result dumping the most common value from any calculation should not hurt any post merger equations, while assuming the independence between them may be. Motivated by this observation, Sub gumption Resolution (SR) [2] identifies two values so that one can appear to complete one and remove the norm.

Correctly Classified Instances	2959	63.95%
Incorrectly Classified Instances	1668	36.05%
Kappa statistic	0.0962	
Mean absolute error	0.4919	
Root mean squared error	0.4923	
Relative absolute error	106.38%	
Root relative squared error	102.39%	
Total Number of Instances	4627	
Time Taken to Build	2.23 Seconds	

Fig. 6

Detailed Accuracy By Class			
			Weighted Avg.
TP Rate	0.899	0.184	0.64
FP Rate	0.816	0.101	0.557
Precision	0.659	0.509	0.605
Recall	0.899	0.184	0.64
F-Measure	0.761	0.27	0.583
ROC Area	0.687	0.687	0.687
Class	low	high	

Fig. 7

Confusion Matrix

a	b	← classified as
2650	298	a = low
1370	309	b = high

PROPOSED MODEL

Many inaccurate features may be existing in the data to be extracted. Therefore we need to identify and remove such type of inaccurate data. There are numerous mining procedures those are not good for large numbers of attributes. So, the feature selection techniques need to be used before the introduction of any type of mining algorithm. The basic purposes of feature selection are to simplify overload and optimize model performance and deliver quicker and more accurate representations. The biggest task with overloading and machine learning is that we don't know how well our model will work on new data until and unless we test it on the data set.. To do this, we can split our initial dataset into training and test subsets separately. We'll train and tune our model with training set and then will Apply to test set. Once the data is filtered, it will be stored in other file and will be allotted a weight wj. This data will be used to train our model. Data may come continuously in large volume. So, every data will go under this process and will be allotted with a constant value wj. An error can be demarcated as a function E(j, k) where j._.J or it is hypothesis and k is the goal function. Similarly noise can be defined by another function N= E + Wj.

ALGORITHM FOR PROPOSED MODEL

1: begin

2: Insert the data values with their attribute

3: Check the noise and find out the error rate

4: Filter the data on the basis of error rate and categories them.

5: Assign a weight Wi to filtered and corrected data

6: Use this weighted data to test a model

7: Select the model which is classifying correctly

8: end

	NaiveBays	AODE	AODEsr	Proposed Model
TP Rate	1	0.9	0.899	1
FP Rate	1	0.818	0.816	1
Precision	0.637	0.659	0.659	0.699
Recall	1	0.9	0.899	0.789
F-Measure	0.778	0.761	0.761	0.751
ROC Area	0.499	0.687	0.687	0.711
Correctly Classified Instances	0.63713	0.639075	0.639507	0.6425
Incorrectly Classified Instances	0.36287	0.360925	0.360493	0.3575

Fig. 8 *Comparison of NB, AODE, AODEsr and Proposed Model*

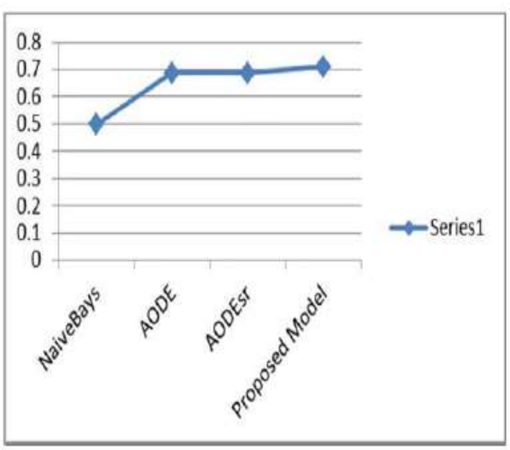

Fig. 9 *Graphical Representation of performance with Proposed Model*

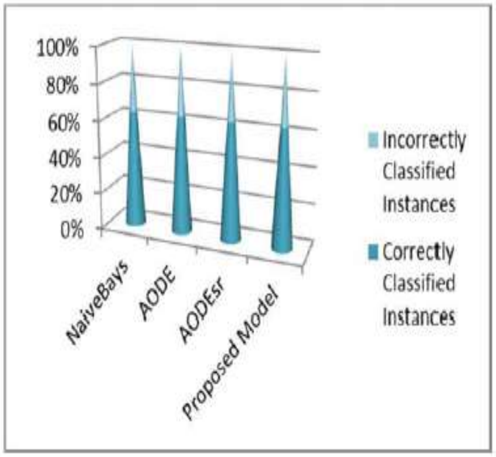

Fig. 10 %age of classification of NB, AODE, AODEsr and Proposed Model

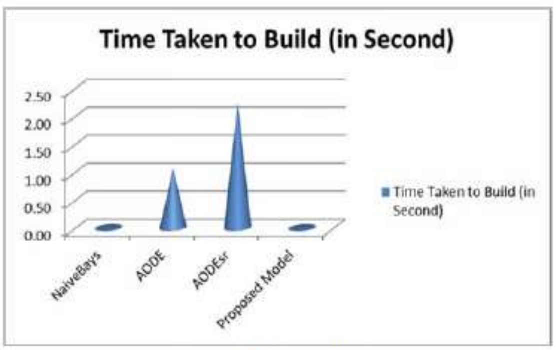

Fig. 11 Time taken to build the model

CONCLUSION

The experimental evidence suggests that the proposed method outperforms Naive Bays, AODE, and Aiders in terms of accuracy. Comparatively, it has a shorter learning curve than AODE and Aiders. The approach proposed in this study is a complicated one-dimensional model, and as such is not comparable to high-dimensional data...

REFERENCES

Webb, G.I., and Bought on, J., Wang, and Z.: Not so naive bays: Aggregating one-dependence estimators. Learning Machines 58 (2005), p. zbMATHCrossRefInternet search engine Google Scholar

Highly Scalable Attribute Selection for Averaged One-Dependence Estimators. [2] Chen, S., A.M. Martinez, and G.I. Webb. 2014. Journal Article by Tseng, Ho, and Zhou. Knowledge Discovery and Data Mining: Future Prospects, edited by A.L.P. Chen and H.Y. Kao. PAKDD 2014. The 8444th issue of Lecture Notes in Computer Science. Cham: Springer.

[3] Data mining: Practical Machine Learning Tools and Techniques, with Java Implementation. Witten, I.H., and Frank, E. http://prdownloads.sourceforge.net/weka/datasets-UCI.jar, Morgan Kaufmann, San Francisco, 2000. Internet search engine Google Scholar

Probabilistic Reasoning in Intelligent Systems, by J. Pearl [4]. S.F. : Morgan Kaufmann, 1988. Internet search engine Google Scholar

Databases for machine learning at UCI: work by Merz, Murphy, and Aha [5]. To be found at http://www.ics.uci.edu/mlearn/MLRepository.html (Department of ICS, University of California, Irvine, 1997).

Induction of selective Bayesian classifiers. [6] Langley, P. and Sage, S. Published as pages 339-406 in Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence (1994). Internet search engine Google Scholar

Scaling up the accuracy of naive Bayes classifiers using a decision tree-based approach [7] Kohavi, R. Knowledge Discovery and Data Mining (KDD 1996), Volume 2: Proceedings, Pages 202-207. (1996) AAAI Press, Menlo Park. Internet search engine Google Scholar

Machine Learning, Volume 29, Issues 131-163 (1997); Friedman, Geiger, Goldszmidt, Bayesian Network Classifiers. zbMATHCrossRefInternet search engine Google Scholar

Weightily Averaged One-Dependence Estimators. (2006) Jiang L., Zhang H. Q. Yang and G. Webb (eds.) Recent Advances in Artificial Intelligence (PRICAI 2006). Lecture Notes on Computer Science, Volume 4099, PRICAI 2006. This is a Springer publication; it is published in both Berlin and Heidelberg.

Learning Bayesian networks is NP-Complete, according to Chickering and colleagues [10]. Learning from Data: Artificial Intelligence and Statistics V, edited by David Fisher and Herbert Lenz, pages 121–130. Heidelberg, Germany: Springer (1996) Internet search engine Google Scholar

Fey Zheng and Geoffrey I. Webb, "Effective Lazy Elimination for Averaged-One Dependence Estimators," arXiv:1703.04050, 2017. 2006, pages 1113-1120, In Proceedings of the 23rd International Conference on Machine Learning.