International Journal of Engineering & Technology, 7 (2.33) (2018) 206-208

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A machine learning driven approach to improve efficiency of classification algorithm using prediction of affliction

V. Sumalatha [1] *, Dr. R.Santhi [2]

[1] Research Scholar, Bharathiar University, Asst. Professor, VISTAS, India
[2]Research Supervisor, Bharathiar University, Coimbatore, India
**Corresponding author E-mail: sumalathaphd2014@gmail.com*

Abstract

Machine learning plays a key role in a wide range of applications such as data mining, natural language processing and expert systems. It provides a solution in all domains for further development when large data is applied. Supervised learning is consist of mathematical algorithm to optimize the functions with given inputs. Machine learning solves problems that cannot be solved by numerical values. In this research paper, a model is developed to improve classification algorithm using anxiety of juvenile. Prediction and classification are made using data. A machine learning tool is used for pre-processing and first level of model is data preparation and ranking prototype used for filtration of data . Then Probabilistic estimation hypothesis is to find the hypothesis value based on statistical functions and classification of anxiety predictor model is used for prediction and classification. Comparison of Algorithm and experimental are done using machine learning software. According to the experiment, the model is more efficient and accurate compared with other classification algorithm as results shown.

Keywords: *Machine Learning; Bayes Classification; Anxiety.*

1. Introduction

Machine learning used in scientific discipline with the design and development of algorithm that allows the computer to identify the behaviour of large and historical data. Knowledge based data is also used in machine learning [1]. The data is collected from experts or large database. It is used to make effective and systematic outcome from past experience. Testing and training data is used to solve the given problem [3] [10]. Classification algorithm have been used for different learning based model. The different types of machine learning algorithm are:
Supervised Learning – it generates a function based on training dataset that maps input to labels or classes
Unsupervised Learning – it seeks similarities between points to from cluster
Semi – Supervised Learning – it is Combination of labelled and unlabelled data to generate classifier.
Reinforcement Learning - it is learning by Environment Interaction
Knowledge discovery is a process consist of several stages inclusive of many techniques like data pre processing, analysis, evaluation and interpretation of the result. Machine learning is a direct approach like classification of data. Recent development in computational techniques enable researcher to move from classical probabilistic model for advanced approaches using machine learning techniques. It can be defined as the automated process of extracting useful knowledge and information including classification, cluster and structures in large data set [3].

2. Naïve bayes classifier

Bayes classification technique is based on assumption of independence among predictors. The presence of a particular class feature is unrelated to the presence of any other feature [10]. The feature depends on each other or upon existence of other features [3]. All attribute probabilities independent with each other. The bayes classification is useful for very large data base. It is a high sophistication classification method [4]. Naïve Bayes classification provides the way of calculating posterior probability P(c|x) from P(c), P(x) and the equation is

P(c |x) = P (x|c) P(c)

P(x)

Where P(c|x) is the posterior probability of class(c, target) given predictor(x, attributes. P(c) is the prior probability of class. P(x|c) is the likelihood which is the probability of predictor given class [8]. P(x) is the prior probability of predictor.

$$p(C/F1,\dots,F-n) = \frac{1}{z} p(C) \prod_{i=1}^{n} p(Fi/C)$$

Where Z (the evidence) is a scaling factor dependent only onF1… Fn, i.e., a constant if the value of the feature variables are known. Models of this form are much more manageable, since they factor into a so-called class prior p(C) and independent probability distribution p(Fi/C).If there are k classes and if a model for each p(Fi/C=c) can be expressed in terms of r parameters, then the corresponding naive Bayes model has (k-1)+ nrk parameters. so the total number of parameters of the naive Bayes model is $2n + 1$.

Where n is the number of binary features used for classification and prediction.
The applications of naïve bayes classification is,
Real time prediction: Learning classifier and prediction in real time
Multiclass prediction: Predict the probability of multi classes of target variable.

3. Research methodology

The proposed model in this paper for prediction and classification is used to improve the accuracy and efficacy of classification algorithm. The model is tested with juvenile anxiety based on paternity and digital usage dataset. The steps are used to develop the research is as follows
Step 1: Data Preparation according to juvenile anxiety level.
Step 2: Data Pre processing is used to reduce noisy data, missing values and repeated values.
Step 4: In pre - processing, Scaling of each attributes differs depends upon the factors, Ranking for selection of attributes and also pre processing is used to reduce noisy data and missing values.
Step 5: Using the pre processed data, the data is divided into Train dataset and Test dataset.
It will be useful to test our model based on Bayes classification algorithm. Using pre processing tool the statistical reports are generated.
Step 6: For the pre processed data set the mean and standard deviation is calculated.

Mean(x) = 1/n * sum(x)

Standard deviation (x) = sqrt (1/n *sum (xi - mean(x) ^2))

Step 7: Using the CAPM Model, preprocessed dataset is applied to find ranking prototype, probabilistic estimation hypothesis and finally classifications and prediction model for calculating accuracy of the research model effectively. The research model is developed based on Bayes classification algorithm with classification and predictor class. The independent attributes are applied over the model.

4. Experiment & result

In this research paper , Using Machine learning tool the data are pre-processed , after pre-processing the attributes are selected using ranking, further the attributes applied over the existing algorithm and classification model (CAPM) . Scaling of attributes vary from each attributes depends upon the factors Fig-1.shows the statistical report for required for each attributes. The ranking is used for PEH and CAPM. The statistical report are generated for each attribute and analysis are made between attributes. The model has given a accurate time sequence for Ranking prototype, and execution time for classification algorithm Fig-2. Fig-3 Evaluation result for classification algorithm and CAPM. Using tool, various classification algorithm is applied with the data set to develop model with efficiency. The developed model is comparatively more efficient than other classification algorithms. Fig-4 Time for classification algorithm shows the comparison between accuracy of different classification.

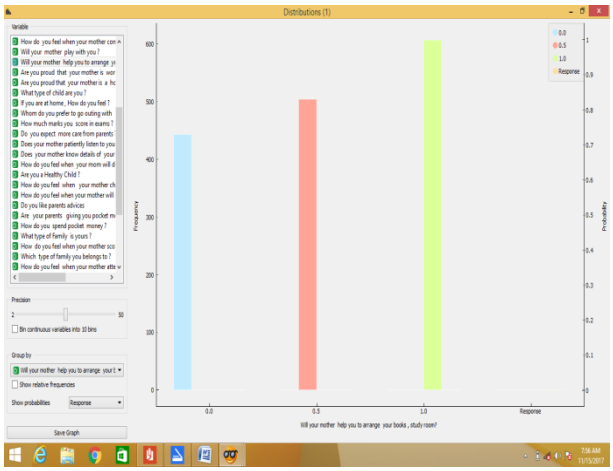

Fig. 1: Statistical Report for Each Attribute.

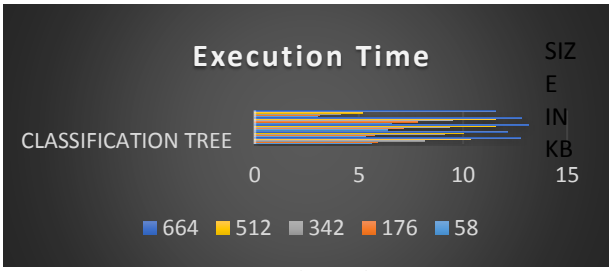

Fig. 2: Execution Time of Classification Algorithms.

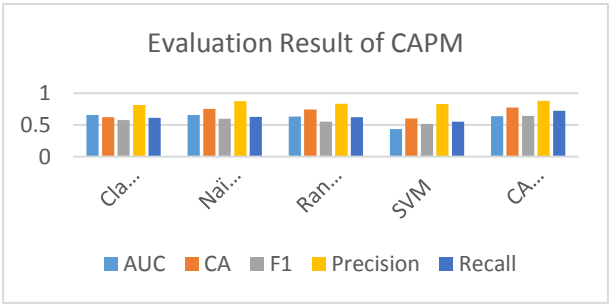

Fig. 3: Evaluation Results.

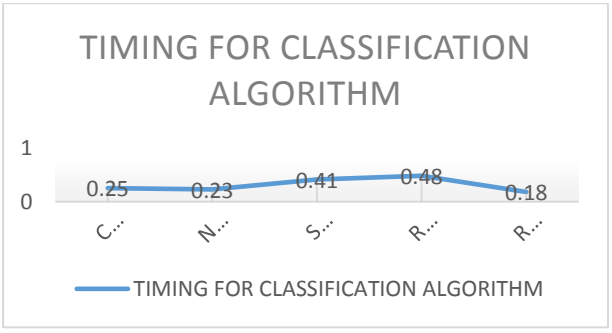

Fig. 4: Time Difference For Ranking Prototype.

5. Conclusion and future enhancement

The research paper is to develop a model for improving efficacy of classification algorithm and tested with juvenile affliction based data set. In this research, machine learning tools used for preprocessing and validation of the model. Based on ranking prototype, the attributes are selected for further development of the model. Then the estimation hypothesis calculated to find the probabilistic functional value. The CAPM model is developed based on the result of PEH values. According to the research, the developed models are more accuracy and efficient than other classification algorithm as shown in results. Further the research can improve with unsupervised learning algorithm also.

References

[1] Clarke B, Bayes Model averaging and stacking when model approximation error cannot be ignored, Journal of Machine Learning Research (2003), 683-712.

[2] Dean J, Ghemwat S, Map Reduce: A flexible data preprocessing tool, Communication of the ACM 53(2010), 72-77.

[3] V.Sumalatha, Dr.R.Santhi, "Efficacy of Improving Classification Algorithm to Develop a Probabilistic Model Based Approach" in International Journal of Pure and Applied Mathematic (Scopus Indexed). ISSN 1314-3395. Vol 116 No .21, 2017.

[4] V.Sumalatha, Dr.R.Santhi, "Deployment of Classification Approach for Sensing Stress of Juvenile Using CAPM Model" in Jour of Adv Research in Dynamical & Control Systems, 15-Special Issue, October 2017 (Scopus Indexed). ISSN 1943-023X.

[5] Liu Haifeng, Liu Shousheng , et al, 2013 improved K nearest neighbor algorithm based on spatial distribution of the training sample JCSSTI, 32(1):80-85

[6] Dong Yuanyuan, Tang Xiaoxia, et al, 2012 Text Classification based on BP-Adaboost[J], Network Security Technology, Vol 3: 42-45.

[7] Hua Jin, et al, 2013 short Text classification based on rule and information entrophy[A],Asia-Pacific Computational intelligence and information Technology conference (APCIIT).

[8] Domingos P 2000, Bayesian averaging of Classifier and the over fitting problem. ICML Conferences.

[9] Moo-Young Lee and Su-sum Park, 2016 Influence of family strength on the Psychological well being of adults-focusing on mediation effect of social capital, Indian Journal of Science and Technology, Vol 9(26).

[10] Guttman, D.K.J. Lin, B.Reiser and J.S Usher, "Dependent masking and system life data analysis.bayseian index for two components system.

[11] Flajolet P and Martins G.N 1995 Probabilistic counting algorithm for data base applications. Journal of computer and system sciences, 904: 23-37.