© 2015 IJSRSET | Volume 1 | Issue 5 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Predicting Teens Stress According to Behavioural Activity of Parents Using Naive Bayes Classifier

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

[1]Research Scholar, Bharathiar University, Assestant Professor, Vels University, Chennai, Tamilnadu, India
[2] Professor, Head, Department of Computer Science and Engineering, A V College of Engineering, Vilupuram, Tamilnadu, India

ABSTRACT

In today world, depression is a major issue for teenagers. Even depression leads to committing suicides or to addict for drugs or leads to any illegal activities etc,. For the successful future of teenagers parents should have more responsibility. In this paper the work is carried out to find the stress level of teenagers if parents are working (Father & Mother).Researchers are using classifier techniques in the field of Medical , Academic, Bioinformatics , Bio computing etc. Using the supervised learning techniques to evaluate the teenagers from the dataset given. Based on the Evaluation to find the teenagers stress level according to parent's behavior. This research will be useful to control or avoid stress factors among teenagers. We can also improve the parental care to teenagers. Using the classification algorithm we can predict the stress level of teenagers.
Keywords: Supervised Learning, Classification Algorithm, Naive Bayes

I. INTRODUCTION

We live in a world where large amount of data are collected daily. Analysing such data is an important task. So we need Knowledge discovery, is an essential process where intelligent methods are used to extract data from large amount of database. Knowledge discovery is known as the process of monitoring new and innovative information from database. In this proposed system, we predict the stress level of teenagers based on the attributes. The attributes are mainly focuses on parents behavior. The attributes are classified in training sets. Classification is the process of assigning new objects to predefined categories or classes. In this approach, the given data sets are learned and classified into groups. Whether which type of teenagers (mother working or housewife) is more stressful. It is a statistical approach to find the stress level. Naïve Bayes classifier is the base algorithm for this research.

II. METHODS AND MATERIAL

A. Naive Bayes Classifier

A Naive Bayes Classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naïve) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model".

In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naïve Bayes classifier considers all these properties to independently contribute to the probability that this fruit is an apple.

Depending on the precise nature of the probability model, naive Bayes classifier can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes model uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods.

An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variances of the variables for each class need to be determined and not the entire covariance matrix.

The naive Bayes probabilistic model
Abstractly, the probability model for a classifier is a conditional model

$$p(C/F1, \dots, Fn)$$

over a dependent class variable C with a small number of outcomes or classes, conditional on several feature variables F1 through Fn. The problem is that if the number of features n is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable.
Using Bayes' theorem, we write

$$P(C/F1, \dots, Fn) = p(C)p(F1, \dots, Fn/C)$$
$$p(F1, \dots, Fn)$$

In plain English the above equation can be written as

Posterior $= \dfrac{prior \times likelihood}{evidence}$

In practise we are only interested in the numerator of that fraction, since the denominator does not depend on C and the values of the features Fi are given, so that the denominator is efficiently constant. The numerator is equivalent to the joint probability model

$$p(C, F1, \dots, Fn)$$

This can be rewritten as follows, using repeated applications of the definition of conditional probability:

$$p(C, F1, \dots, Fn)$$
$$= p(C)p(F1, \dots, Fn/C)$$
$$= p(C)p(F1/C)p(F2, \dots, Fn/C, F1)$$
$$= p(C)p(F1/C)p(F2/C, F1)p(F3, \dots, Fn/C, F1, F2)$$
$$= p(C)p(F1/C)p(F2/C, F1)p(F3 /C, F1, F2)p(F4, \dots, Fn/C, F1, F2, F3)$$
$$= p(C)p(F1/C)p(F2/C, F1)p(F3/C, F1, F2) \dots p(Fn /C, F1, F2, F3, \dots, Fn-1$$

Now the "naïve" conditional independence assumptions come into play: assume that each feature Fi is conditionally independent of every other feature Fi for$j \neq i$. this means that

$$p(Fi/C, Fi) = p(Fi/C)$$

For $i \neq j$, and so the joint model can be expressed as

$$p(C, F1, \dots, Fn) = p(C)p(F1/C)p(F2/C)p(F3/C) \dots$$
$$= p(C) \prod_{i=1}^{n} p\,(Fi/C).$$

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

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

Where Z (the evidence) is a scaling factor dependent only on$F1, \dots, Fn$, 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. In practice, often $k = 2$ (binary classification) and $r = 1$ (Bernoulli variables as features) are common, and 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.

Evaluating classification algorithms

- I tell you that it achieved 95% accuracy on my data.
- Is your technique a success?

Types of errors

- But suppose that
  - The 95% is the correctly classified pixels
  - Only 5% of the pixels are actually edges
  - It misses all the edge pixels

Types of Errors

	Prediction: Edge	Prediction: Not
Edge	True Positive	False Negative
Not Edge	False Positive	True Negative

B. Data Processing for Proposed Methodology

Stress level of students can be measured using the following attributes. Each attribute, assign scaling. In this Research work for each attribute the scaling differs. The attributes are

1. Spending time with mother
2. Playing with mother
3. Sharing information regarding school
4. Sharing family problems
5. Outing with parents
6. Comparison with siblings
7. Comparison with friends
8. Parents understanding
9. Watching tv with parents
10. Family background
11. Type of family
12. Academic performance
13. Getting advices
14. Type of child
15. Favorite place (home/School)
16. Favorite person (parents / friends)
17. Future Decision

Similarly in this research work 39 attributes are used to predict the teenagers stress level. According to this pruning can be done to get a good result.

C. Proposed Methodology

Step 1 : The initial step in supervised learning is the collection of dataset. If the expertise is available then data on the informative features can be collected.
Step 2: The data preprocessing step involves reducing noise by instance selection. There are several methods available to handle missing data.
Step 3 : The training set is defined by feature subset selection , in which the irrelevant and redundant features are removed.
Step 4 : The algorithm is trained using a training data set.
Step 5 : The algorithm is tested with a test data set . Once the evaluation from the preliminary testing is satisfactory, then the classifier can be used for prediction. The classifier is evaluated based on accuracy.
Step 6 : If the accuracy is not satisfactory then the process needs to return a previous step re- examine certain attributes. The re-examine factors are

- Relevant features are missing for the problem
- Requirement of larger data set
- Selected algorithm may be inappropriate
- Attribute tuning needed

Using Naïve bayes classification algorithm the independent data are used to find the accuracy of results. The expected results are to find each teen ager stress level. The stress levels are

Types of stress	Range
Positive stress	0.7 - 1.0
Early stress	0.4 - 0.69
Negative stress	Below 0.4

Classification algorithm will be useful for supervised learning and also to predict the stress level. The computational cost and accuracy will be efficient.

III. CONCLUSION

In this research work, using classification algorithm to predict the stress level of teenagers according to behavioral activity of parents. Counseling can be provided to each teenagers and also parents to improve the performance of their children's behaviour in both physical and mental activities. It will help to avoid suicides and control illegal activities.
Further we can do this work for employee stress level according to managerial behavioral or women stress level if working.

IV. REFERENCES

[1] M. Miyakawa, "Analysis of incomplete data in a competing risks model," IEEE Trans. Rel., vol. 33, no. 4, pp. 293–296, Oct. 1984.

[2] J. S. Usher and T. J.Hodgson, "Maximum likelihood analysis of component reliability using masked system life data," Trans. Rel., vol. 37, no. 5, pp. 550–555, Dec. 1988.

[3] D. K. J. Lin, J. S.Usher, and F.M.Guess, "Exact maximum likelihood estimation using masked system data," IEEE Trans. Rel., vol. 42, no. 4, pp. 631–635, Dec. 1993.

[4] B. Reiser, I. Guttman, D. K. J. Lin, J. S. Usher, and F. M. Guess, "Bayesian inference for masked system lifetime data," Appl. Statist., vol. 44, pp. 79–90, 1995.

[5] J. O. Berger and D. Sun, "Bayesian analysis for the poly-weibull distribution," J. Amer. Statist. Assoc., vol. 88, pp. 1412–1418, 1993.

[6] C. Mukhopadhyay and A. P. Basu, "Bayesian analysis of incomplete time and cause of failure data," J. Statist. Plann. Inference, vol. 59, pp. 79–100, 1997.

[7] S. Basu, A. P. Basu, and C. Mukhopadhyay, "Bayesian analysis for masked system failure data using nonidentical weibull models," J. Statist. Plann. Inference, vol. 78, pp. 255–275, 1999.

[8] A. Sen, M. Banerjee, and S. Basu, , N. Balakrishnan and C. R. Rao, Eds., "Analysis of masked failure data under competing risks," in Handbook of Statistics. Amsterdam, The Netherlands: North-Holland, 2001, vol. 20, pp. 523–540.

[9] S. Basu, A. Sen, and M. Banerjee, "Bayesian analysis of competing risks with partially masked cause of failure," Appl. Statist., vol. 52, pp. 77–93, 2003.

[10] D. K. J. Lin and F. M. Guess, "System life data analysis with dependent partial knowledge on the exact cause of system failure," Microelectron. Rel., vol. 34, pp. 535–544, 1994.

[11] I. Guttman, D. K. J. Lin, B. Reiser, and J. S. Usher, "Dependent masking and system life data analysis: Bayesian inference for two-component systems," Lifetime Data Anal., vol. 1, pp. 87–100,1995.

[12] L. Kuo and T. E. Yang, "Bayesian reliability modeling for masked system lifetime data," Statist. Probab. Lett., vol. 47, pp. 229–241, 2000.

[13] C. Mukhopadhyay and A. P. Basu, "Masking without the symmetry assumption: A Bayesian approach," in Proc. Abstract Book 2nd Int. Conf. Math.Methods Rel., Bordeaux, France, 2000, vol. 2, pp. 784–787, Universite Victor Segalen.