

# A Machine Learning Approach to Predict Juvenile Stress level with Parental Behavior and usage of Gadgets using Bayes Classification

<sup>[1]</sup> Sumalatha.V , <sup>[2]</sup> Dr.Santhy.R

<sup>[1]</sup> Research Scholar, Bharathiar University, Asst.Prof, Vels University, Tamil Nadu,

<sup>[2]</sup> The Principal. PSBB Millennium School ,OMR, Tamil Nadu.

**Abstract:** This research paper focusing to identify and reduce the stress level of juvenile by developing a model based on classification algorithm. **Methods/Analysis :** The tools used for data collection are standard 39 attributes for Paternity Behavior and 23 attributes for Usage of Digital gadgets among juvenile. Selection of attributes are analyzed depending upon the descriptive statistical analysis using machine learning tool. To find the best performance of each attribute analysis are also prepared in data preprocessing . Confusion matrix and Test Score for each algorithm is also generated. **Findings :** To consider all these constraints and to improve the quality of juvenile, this research paper will design and develop a model using naïve bayes algorithm to predict the stress level of juvenile. Two different type of dataset is used to predict stress level of juvenile that is parental behavior or care towards juvenile and usage of digital gadgets among juvenile. The naïve bayes classification model is used to classify and predict the stress level. Further this research paper will shows the accuracy and efficiency of the new generated model. **Improvement/Applications :** This research will be useful to improve the performance of juvenile in academic or various activities by reducing the stress level of juvenile. Further it is also useful to reduce criminal activities, committing suicides, addicting to drugs and discontinuing studies among juvenile in the society.

**Keywords:** Classification Algorithm, stress, Naïve Bayes , Machine Learning, juvenile

## I. INTRODUCTION

In real world situation stress is a common cause among people. In this research paper, juvenile stress is focused because if stress level is more for them it leads to committing suicides, addicting to drugs, doing illegal activities in society. Analyzing large amount of data is an important task. So we need knowledge discovery using machine learning algorithm. In this research paper two different data set collected, one data set is based on parental care and the other one is using digital gadgets among juvenile. Using the attributes in the data set, a model is designed and developed to classify and predict stress level. In each data set some set of attributes are used to predict stress. The attributes are classified as test data and trained data. The trained data is used to train a model but test data will give actual accuracy of data. The algorithm naïve bayes <sup>1,2</sup> is a probabilistic machine learning concept for classifying and predicting and this algorithm is used as a base algorithm for the research. The developed model will give a accurate and efficient result comparatively to all other classification algorithm.

## II. MACHINE LEARNING

Machine learning is a branch of artificial intelligence, is a scientific discipline with the design and development of algorithms that allow the computer to identify the behavior of empirical data <sup>3</sup>. Knowledge base is also used in machine learning. It is collected from experts or large database. Machine learning algorithms are used to make effective system from past experience. The given problem is solved by using past experience or to use training data. Effective algorithm have been used to discover for specific learning type. Machine learning refers to change in system that performs task with artificial intelligence <sup>4</sup>. The machine learning algorithm types are :

- Supervised Learning - generate a function based on training dataset that maps input to labels or classes.
- Unsupervised Learning – it is not provided with classification. Instead seek out similarities between points to form cluster.
- Semi-supervised learning – Combines both labeled and unlabeled data to generate classifier.

- d. Reinforcement learning – learn by interacting with environment

### III. NAIVE BAYES CLASSIFIER

It is a simple probabilistic classifier. It is based on bayes theorem with strong independence assumption. naive bayes is also called as “independent feature model”. It is trained very efficiently in a supervised learning approach. Bayes classification is outperformed by other classification approach<sup>5</sup>. The advantage of naive bayes classifier is , it only requires a small amount of training data to estimate the attributes for classification.

Model of Naïve bayes Classifier:

Abstractly, the probability model for a classifier is a independent 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<sup>6</sup>. We therefore reformulate the model to make it more tractable.

Using bayes theorem we can write

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

In practice 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

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

### IV. DIFFERENT TYPES OF STRESS

In this research paper , juvenile stress has predicted depends upon the behavior activity of paternity and more use of digital gadgets by juvenile. Stress can effect human health either physically or mentally. Behavior change or physical change in human is a symptoms of stress<sup>7</sup>. Stress is harmful to our body. Generally the stress is classified as three different types psychologically<sup>8,9</sup>.

- A. Acute Stress - Negative

It is a very common stress for humans and easily recoverable .A immediate reaction to new challenge, event or demand. Ex: Minor accidents or Mistakes , Argument with Friends and Family members.

- B. Episodic Acute Stress – Partial Stress

This type of stress will happen for short duration and in design and development of best response to future situation.

Ex : Often Headache, Feeling Tired, sleepless

- C. Chronic Stress – Positive stress

It is very serious stress occurs for very long time. This type of stress needs medical treatment and it will take time to recover<sup>10</sup>. It leads to abnormal diseases or health risk .Ex : Cancer, Heart diseases, Lack of Concentration, Suicides , Etc

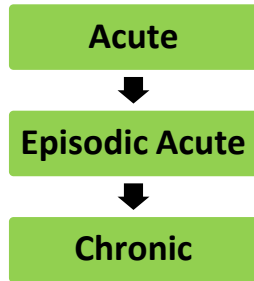


Figure 1 : Different Types of stress

### V . THE METHODOLOGY

The process of the proposed methodology is as follows:

Step 1 : the first step of proposed methodology is to collect two different types of data set. The data set is collected from juvenile through online and also in offline.

Step 2 : Data preprocessing involves reducing noisy data , missing values in the data set.

Step 3: Selection of attributes depends using preprocessing techniques.

Step 4: The training set is defined by selection of attributes to identify irrelevant, redundant data are removed for future development of model.

Step 5: The selection of algorithm is important . and it is trained with training data set.

Step 6 : The algorithm is tested with test data set. If the test is satisfactory. The classifier is used for prediction. It is calculated based on accuracy.

Step 7 : If the accuracy is not satisfactory then process go to previous step to check attribute selection. and it is reexamined.



Figure 2 . Research Methodology

The proposed algorithms is used to predict the juvenile stress according to the usage of digital gadgets and behavior of paternity. The efficiency and accuracy of this algorithm is comparatively better than the existing classification algorithm.

### VI Data Preprocessing and Comparison of Algorithm

The research paper consist of two different data set. The first data set used to prediction of stress for juvenile is parental care or paternity behavior towards juvenile Table 1. The second data set is Usage of digital gadgets among juvenile Table 1. In each data set some specific attributes are used for prediction after data preprocessing. Then the data set is applied with classification algorithms.

Table 1 . Description of Data set

S.No	Behaviour of Paternity	Gadget Usage
1	Spending Time with juvenile	Social Media Accounts
2	Discussing personal problems	Time duration of using digital gadgets
3	Favourite Place to spend - Home/School/College	Sleepless / Late Night Sleep
4	Favourite Person – Parents/Friends	Feeling lazy / drowsy / sleepy always
5	Decision Maker	Medical Treatment
6	Performance in Academic	Academic Interest
7	Character of juvenile	Often giving immediate response in social media

8	Comparison with siblings	Unhealthy
9	Comparison with friends	Using high power spectacles/contact lens
10	Background of Family	Behaviour in winning/loosing the game
11	Understanding between parents	Misunderstanding with parents/siblings regarding Digital gadgets
12	Family Type	Updating photos , status
13	Interaction with parents	Posting daily activities in social media regularly

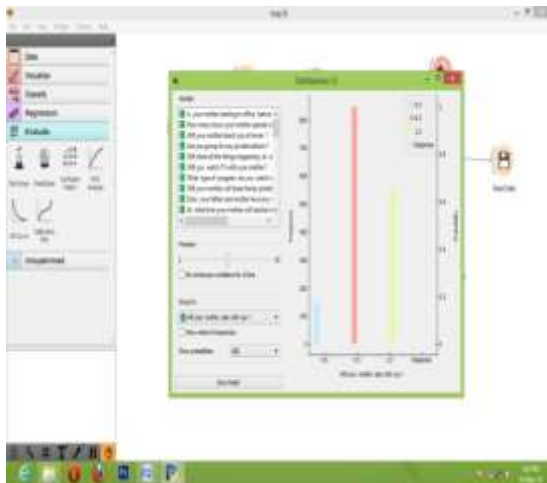


Figure 3. Statistical Report for attribute

In this research paper scaling vary from one attribute to others. This technique is used for making the model effectively and to get accurate result. The probabilistic approach<sup>11,12</sup> is used as a prediction of anxiety level. The Figure 4 shows comparison of algorithm and its test score result. The classification ,SVM algorithms have the result of 0.906 and naïve bayes algorithm have the result of 1.000 compare with all algorithms bayes algorithm is more effective and accurate.

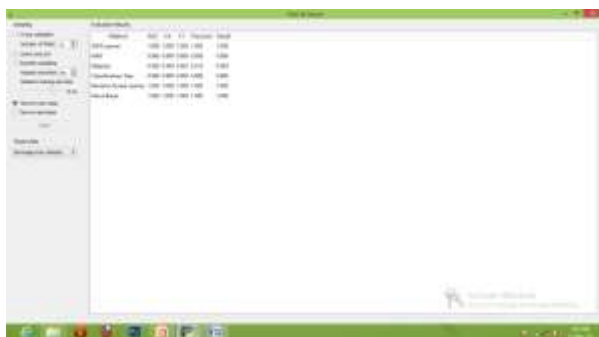


Figure 4. Algorithms Evaluation Result

### VII. Conclusion and Result

In this research paper the stress factor of juvenile is measured by using the model. The dataset is used for predicting stress is digital gadget usage and paternity behavior towards juvenile. The model is used for classification and prediction of stress. If stress level is predicted for juvenile then further it will be refer to medical counseling or psychological treatment or to improve parental care. The model has given a accurate and efficient result among all other classification algorithm. The confusion matrix Figure 5 shows the prediction of stress for selected attributes.

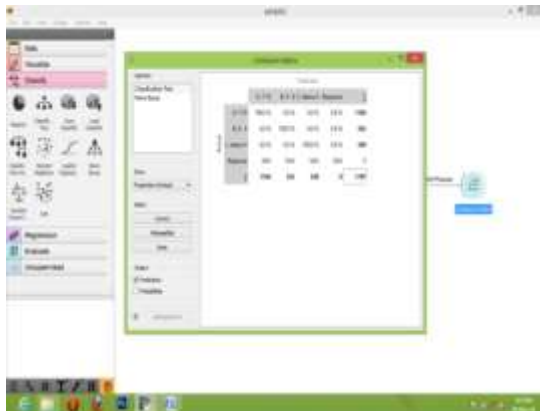


Figure 5. Confusion Matrix

### VIII. FUTURE WORK

The Research work has mainly focused for teen students to analyze the stress level according to parental behavior and more usage of digital gadgets. In future we can also focus to find the stress level of Kids and also applicable to find the stress of Employers in industries. The same algorithm will also apply to find the anxiety level . by comparing with the new algorithm , Artificial Intelligence based algorithms will also be suitable to apply.

### References

- 1.Data Mining and knowledge Discovery with Evolutionary Algorithm . A.A.Springer-Verlag. 2002.
2. T.Karthikeyan , P.Thangaraju. Genetic Algorithm based CFS and naïve bayes algorithm to enhance the predictive accuracy, Indian Journal of Science and Technology, Vol 8(27), Oct 2015
- 3.Liu Haifeng, Liu Shousheng , et al, improved K nearest neighbor algorithm based on spatial distribution of the training sample [J].JCSSTI, 2013,32(1):80-85
4. Dong Yuanyuan, Tang Xiaoxia, et al, Text Classification based on BP-Adaboost[J], Network Security Technology, 2012(3): 42-45.
5. Hua Jin , et al, short Text classification based on rule and information entropy[A],Asia-Pacific Computational intelligence and information Technology conference (APCIIT 2013).
6. P.Domingos , Bayesian averaging of Classifier and the over fitting problem. ICML Conferences 2000.
7. Moo-Young Lee and Su-sum Park, 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), July 2016.
- 8.Shinhong Min, The Relationship between the resilience, prenatal stress and Confidence for Child Birth: Focused on pregnant woman, Indian Journal of Science and Technology, Vol 9(43), Nov 2016.
9. Chang Seek Lee, Jong Min Koh, Hyang Ran Myoung ,Level of the Psychological Factors Affecting the Donation Acts and Its Relationship with young people's Characteristics in Korea, Indian Journal of Science and Technology, Vol 9(43), Nov 2016.
10. Feldman T. Techniques and applications for sentiment analysis[J]. Communication of the ACM,2013 56(4): 82-89.
11. J.Dean and S.Ghemawat Map Reduce. A flexible data processing tool. Communication of the ACM, 53:72-77,2010
12. P.Flajolet and G.N.Martin Probabilistic counting algorithm for data base applications. Journal of computer and system sciences, 904: 23-37,1995