STRESS PREDICTION OF STUDENTS USING MACHINE LEARNING

DISHA SHARMA¹, NITIKA KAPOOR² & DR. SANDEEP SINGH KANG³

¹Research Scholar, ME-CSE-Department, Chandigarh University Gharuan, India
²Assistant Professor, CSE-Department, Chandigarh University Gharuan, India
³Professor, CSE-Department, Chandigarh University Gharuan, India

ABSTRACT

On the daily basis, people are suffering from the stress illness because of many factors which include the social factor, external stimulus or environment factors and internal factors. In healthcare, vast development have been made with the use of machine learning. Stress is a fatal disease causing a considerable number of fatalities across the world. The machine learning enable the prediction of the possibility of stress prediction in the under studies of students like, graduate, under graduate, post graduate and professional students. In this paper we analyze the performance of machine learning techniques to reduce the risk of stress prediction resulting in early treatment of the understudies’ students. The data set was collect from university with the help of pss scale and it made up of more than 200 student’s data. Different types of classification algorithms Naive Baye’s, Linear Regression, Multi-layer perceptron, Bayes Net, J48 and random forest are using and also we calculate their accuracy with the help of performance parameter like TP, FP, ROC, F-Measure etc. In this research Random forest classifier gives high accuracy of 94.73%.

KEYWORDS: Stress Prediction, Feature Extraction, Machine Learning Technique, Dataset.

Received: Jun 10, 2020; Accepted: Jun 30, 2020; Published: Jul 30, 2020; Paper Id.: IJMPERDJUN2020534

I. INTRODUCTION

Nowadays, stress became a major issue in these days. Sometimes its response is positive as well as negative. These days the term stress is considered to be one of the major factors learning to various health problems (1993, Gmelch,).[1] when it crosses a certain level, it complexes the day by day in our lives and powers people to digress from the typical public activity.[8] The growing pace of lives and cantered lifestyles suggest that pressure is a vital piece of human life. A human being in a state of changing in accordance with pressure exhibits direct.[2][8].
There are different types of sources of stress but three are commonly found in everyone that are external stress, environmental stress, physical stress.

**The Environment Stressor:** The external stress can be occurred because of the environmental stressor when the person is unable to respond to the external and internal stimulus or situation it causes stress.[8] The examples include disturbance in the environment, crowding, cold and hot weather, traffic, a high number of crimes in the societies, pollution, and pandemic viruses.[8][23]

**Social Stressor:** Every individual lives in a society and interact with many people in their day to day life. The external stressor can become the source of stress an individual's life which includes- Hot and cold climate, natural disasters, criminal offenses, contamination, and death. These kinds of stressors happen to the earth and humans have no control over these kinds of stressors. [11]

**Physiological Stressor:** Every individual suffers from a different kind of stressful situation in their lives. Every individual also plays a social role in their lives by playing multiple characters with different people. Each person does their jobs to live in a society and with family, such as brother, parents, friends, boss, life partners, and many more social roles they play. [8][2]

Because of stress there might be other medical problems like weight, respiratory failure, diabetes, asthma and so on. Because of stress, there might be other medical problems like weight, respiratory failure, diabetes, asthma and so on. Every day, many students commit suicide in different parts of the country.[2] according to the lancet report in 2012 our nation has reported a large number of suicide cases age between fifteen to twenty eight.. In the year 2015, 8,934 number of student suicide cases were formed. In 2010-2015, 39,775 students were commits suicide due to stress.[2][1].

Hans Selye was the father of stress illness and have given different definitions of stress in investigating. In 1956 his view was the "Stress is a tendency of eager or physical weight.”[1]

There are three types [24] of stress -:

![Figure 1.2: Stress Types](image_url)

Acute stress is known as short term stress. The 'acute' word meaning the symptoms grow quickly still does not generally long-lasting. Episodic acute stress is intense stress that occurs during a certain period. E.g. parents pressure, college exam, assignments, etc. chronic stress is very dangerous and harmful stress. Ongoing stress full situation that stressful but not deadly. It stays for a long time like months, years, etc. E.g. like relationships, family issues, illness, etc.
II. RELATED WORK

G. Harrison et al. [8] stated that as globalization is increasing, the innovation is also increasing and in our everyday life is taking the advantage of this and one of the innovations which became so important part in our day to day life is Cell phones. According to Occam's in 2017, reported that 94% of the adolescent population in the UK are using personal cell phones and more than the three fourth of the same population have a smart phones. According to the study of Deloitte's portable purchaser in 2016, they stated that 33% advanced mobile phones are not used for voice calls but they are more used for scrolling emails, online shopping, screeching for the information and updated news, downloading music, using the social media sites, getting the location by maps. As innovation becomes an important part of our daily lives, we are more likely to depend on innovations. When we don't have or use our smart phones, we felt that something important in our lives is missing. [8] According to the evidence, many individuals felt uneasiness and frustration when they are not using their smart phones.

In paper [16], the authors used a decision tree algorithm which is then applied to the data which is collected from the two tests to say that these tests are unsatisfactory. The stress level among the students at the starting of the term or semester and the end of the semester. The study reveals that the level of stress among students at the starting of the semester was less and at the end of the semester the level of stress is increased.

In paper [22], authors measure stress by using different modes like EEG, GSR, EMG, and SpO2. Different type of methods is used to recognition algorithms are being used to record or measure automated stress detection. The measurement recorded from all the external and internal sensors is checked with the value of the index which is used for detecting the stress.

In paper [27], the author measured stress by using different kinds of methods such as heart rate recorder, EMG, Respiration recorder, and GSR hand and foot data. Recorder and measurement of respiration is important elements in the process to predict stress.

In paper [28], authors used sensor signals to predict the level of stress in an individual and they found 75% accuracy. Reddy et al. [7] Stress becomes a major problem for employees who are working in companies like IT.

Employees who are working in these kinds of companies suffer from the drastic change in their personal and social life as well as in their work-life and these kinds of major changes in the life become the cause of stress which not only affect the physical health of the employees but also the mental health. As per their aim of the paper, they used OSMI (Open Sourcing Mental Illness) survey dataset of 2017, from the tech industry. Various machine learning strategies used different modes such as Boosting, Packing, Random timberland and Decision tree. They use different kinds of characteristics like Orientation of the students, Age, Family, Sexual, E- produce medical advantages and so on. Studies reported that 75% of employees working in the tech industries had the highest number of chances to adapt to stress or pressure. In paper [29], authors use ECG (Electrocardiogram) signals to predict the level of stress in an individual’s [7].

III. DATA SET AND PREPROCESSING

The self- generated questionnaires dataset was taken from more than 200 students of the University. Data set were collected through a primary and secondary source like Gmail, Google drive survey form. The collecting dataset was used on the PSS scale. In PSS scales there are different types of questions related to emotion, past month history .The dataset was collected for PSS, a test which incorporates various inquiries by large including the whole enthusiastic inquiry.[10][20]
This type of question has five option answers like (a) Never, (b) Almost never, (c) often, (d) very often, (e) sometimes. Every option has its weight age for calculation of stress. They were asked some basic questions about their feelings, emotion and present situations that they might have encountered in the last few months and their response to it [10]. The stress divided into basically three categories like Normal, moderate, high. The obtained raw dataset is converted into numeric data. The dataset is split into training and testing data in the ratio of 80:20.

**PSS Scale**: The Perceived Stress Scale (PSS) is a stress pressure assessment instrument. This was established by a physiological researcher by Sheldon Cohen in 1983. [10] This scaled show presents about your emotions, feelings during the previous many month and present. [10]

**ADULT ADHD Self-Report Scale**: Adult Adha is a self-report scale is known as a stress assessment instrument. There are different types of questions ask on this scale which is related to emotions, feelings, history, during previous many month and present. [20]

**Weka**: It is an Open Source Data Mining programming included (https://www.cs.waikato.ac) as part inside Pentaho that gives techniques to pre-handling content, It is very helpful for mining of various opinions of people, so it is highly used in the opinion mining field also used for other areas, for example, drawing data from a database and perusing CSV records, and a lot of programmed learning calculations.

**IV. PERFORMANCE PARAMETER**

In this research we are using different types of parameters are:-

**True Positive Rate**: It is also know as Sensitivity (TP). [8] It calculates the values which are correctly classified the accuracy and formula is:-

\[
TP = \frac{\text{TRUE POSITIVE}}{\text{TRUE POSITIVE + FALSE NEGATIVE}}
\]

**False Positive**: It is also known as Specificity [7]. It calculates the values which are incorrectly classified as true and formula is:-

\[
TP = \frac{\text{TRUE POSITIVE}}{\text{TRUE POSITIVE + FALSE NEGATIVE}}
\]

**Recall**: Normally, we can use recall for wholeness what percentage of positive tuple did the classifier label as positive. In recall 1.0 is best Score. In others Words we can say that Inverse connection between recall and precision.

\[
\text{RECALL} = \frac{\text{TRUE POSITIVE}}{\text{FALSE POSITIVE + TRUE POSITIVE}}
\]

**F-Score**: F-score is also known as F – measure. It is used to calculate by detecting out the harmonic mean of recall and precision and formula is:-

\[
F - \text{SCORE} = \frac{2 \times \text{PRECISION} \times \text{RECALL}}{\text{PRECISION + RECALL}}
\]

**MCC-**: MCC is known as Matthews’ correlation coefficient. The Matthews correlation coefficient is introduced by biochemist Brian W. Matthews in 1975. It is used in machine learning technique and used to measure the quality of binary (two-class) classifications. It is determined by using mulling over the True positive, false positive, True Negative and False Negative qualities. It generally gives the value between + 1 and -1.
**Precision:** Precision compares the quantity of effective magnificence predictions that truly belong to the high quality elegance. [7] Formula is:

\[
\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} - \text{False Positive}}
\]

**Receiver Operating Characteristic Curve (ROC):** This is a chart showing the efficiency of a classification algorithm at some classification threshold.

**V. PROPOSED METHODOLOGY**

In this research, we are using different types of Machine Learning technique to predict the stress level of the students. In the questionnaires dataset, we recognize the student’s different ways circumstances and conditions. This proposed model includes PSS and adult adha questionnaires dataset collection, pre-processing, cleaning, feature extraction and comparison on the basis of their performance parameter as shown in below figure.

![Overall Methodology](image)

**Figure 6.1: Overall Methodology**

**VI. RESULT AND EXPERIMENT**

In this research we can used our own develop questionnaire dataset from survey form. Data set holds more than 200 instances and has 6 attributes. The data set has: Instances = 210 and Attributes = 6

![Raw Data](image)

**Figure 7.1: Raw -Data**
Figure 7.2: Numeric Data

All data set are visualized in the figure.

Figure 7.3: Visualized Data
Table I: In this table we classified accuracy, correctly instances and kappa static.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Correctly</th>
<th>Kappa static</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes Net</td>
<td>90.35%</td>
<td>0.82</td>
</tr>
<tr>
<td>Multilayer perceptron</td>
<td>90.43%</td>
<td>0.84</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>85.96%</td>
<td>0.75</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>86.8%</td>
<td>0.72</td>
</tr>
<tr>
<td>J48</td>
<td>90.32%</td>
<td>0.83</td>
</tr>
<tr>
<td>Random forest</td>
<td>94.73%</td>
<td>0.90</td>
</tr>
</tbody>
</table>

In second table we classified TP (true positive), FP (false positive), precision, Recall.

Table II

<table>
<thead>
<tr>
<th>Classifier</th>
<th>MAE</th>
<th>RMSE</th>
<th>RAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes Net</td>
<td>0.07</td>
<td>0.23</td>
<td>18.8</td>
</tr>
<tr>
<td>Multilayer perceptron</td>
<td>0.13</td>
<td>0.22</td>
<td>20.0</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.12</td>
<td>0.26</td>
<td>30.8</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.15</td>
<td>0.27</td>
<td>40.99</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.06</td>
<td>0.15</td>
<td>16.93</td>
</tr>
<tr>
<td>J48</td>
<td>0.83</td>
<td>0.24</td>
<td>54.5</td>
</tr>
</tbody>
</table>

In third table we classified TP (true positive), FP (false positive), precision, Recall.

Table III

<table>
<thead>
<tr>
<th>Classifier</th>
<th>TP</th>
<th>FP</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes Net</td>
<td>0.90</td>
<td>0.9</td>
<td>0.89</td>
</tr>
<tr>
<td>Multilayer Perceptron</td>
<td>0.89</td>
<td>0.71</td>
<td>0.83</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.93</td>
<td>0.89</td>
<td>0.86</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.92</td>
<td>0.17</td>
<td>0.86</td>
</tr>
<tr>
<td>J48</td>
<td>0.96</td>
<td>0.17</td>
<td>0.87</td>
</tr>
<tr>
<td>Random Forest</td>
<td>1.0</td>
<td>0.18</td>
<td>0.93</td>
</tr>
</tbody>
</table>

We have values of F-score, Recall, MCC, and ROC Area.

Table IV

<table>
<thead>
<tr>
<th>Classifier</th>
<th>F- Score</th>
<th>Recall</th>
<th>MCC</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes Net</td>
<td>0.915</td>
<td>0.93</td>
<td>0.88</td>
<td>0.911</td>
</tr>
<tr>
<td>Multi layer</td>
<td>0.917</td>
<td>0.96</td>
<td>0.80</td>
<td>0.897</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.87</td>
<td>0.905</td>
<td>0.716</td>
<td>0.918</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.855</td>
<td>0.91</td>
<td>0.762</td>
<td>0.938</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.95</td>
<td>1.00</td>
<td>0.898</td>
<td>0.99</td>
</tr>
<tr>
<td>J48</td>
<td>0.917</td>
<td>0.80</td>
<td>0.897</td>
<td>0.802</td>
</tr>
</tbody>
</table>

With the help of accuracy and classifier, we can generate an accuracy graph.
Figure 7.4: Accuracy Graph

Figure 7.5: TP Rate

Figure 7.6: FP Rate
In above figure Machine-learning framework compared with classifiers using accuracy, TP, FP.

VII. CONCLUSIONS

Stress is a problem which is increasing day by day which also affects the individual physical and psychological health. We have developed and presented a proposed model for stress prediction based on the Students dataset and also using PSS (perceived Stress scale) for calculate stress level on student dataset. The classification techniques we have used Baye’s Net, Logistic Regression, Naive Bayes, J48, Random Forest and Multilayer perceptron techniques for the prediction of stress. With the help of Weka tool, the accuracy of different techniques is calculated and compared. According to this paper we observe the Kappa statistic, F-measure, mean absolute error, MCC, ROC Area, False positive(FP), True positive(TP), RMSE and Recall and Random Forest classifier gives the best accuracy of 94.73%.

VIII. FUTURE SCOPE

The proposed work can be extended further. Let us consider the case of a university. After a few months, the stress prediction should be taken in college to know how well their students are. This survey will help us to know that the students are not facing any high-stress levels and studying properly.

REFERENCES


