Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

Predictive Analysis of Student Stress Level using Machine Learning

Dr. Anbarasi M
Assistant Professor
Vellore institute of technology
Vellore,India

Sethu Thakkilapati
Department of Computer Science
Vellore institute of technology
Vellore,India

Veeragandham Rajeev Vas
Department of computer Science
Vellore institute of technology
Vellore,India

Sarabu Venkata Bharath Viswath
Department of computer Science
Vellore institute of technology
Vellore,India

***Abstract*--Mental pressure is a significant element that influences our solid life.Stress is a common experience for many students and can be caused by a variety of factors. Some of the main causes of stress in student life include academic pressure, social expectations, financial concerns, and personal issues. Academic pressure is a major source of stress for students, as they often face demanding coursework, tight deadlines, and high expectations from teachers, parents, and peers. This can lead to feelings of overwhelm, anxiety, and self-doubt, which can further exacerbate stress levels. Conventional pressure identification technique utilize eye to eye interviews, it takes a lot of time and arduous assignment. Scholar of the present time are perpetually exposed to huge measure of pressure, the contributing variables for this are in bounty. Numerous scholars can't adapt up to the difficult and unpleasant climate and neglect to get help in the correct manner, hence directing a relentless harm to their ways of life. We propose an answer for the instructive association where the specialists can foresee the pressure of the scholar utilizing Machine Learning. Two ML models like Random Forest (RF) and Decision Tree (DT) are proposed to anticipate the feeling of anxiety of the scholars. For this study the dataset of 2958 scholars has been gathered by some designing schools of the northern India learning at graduation level. The information is gathered utilizing on the web and disconnected polls.**

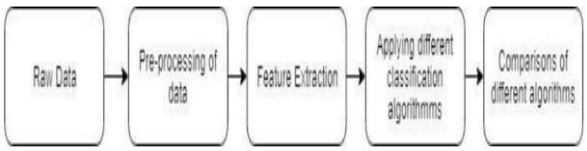
Keywords--Machine learning,Stress Prediction,Stress Level,Random Forest(RF) and Decision Tree(DT) models,Visualization Tools, Performance, Stressed, Unstressed.

I .INTRODUCTION

Stress is a term habitually used equivalently with negative valuable encounters or life events. Legitimate exploration on tension and disquiet offers various perspectives on the issue [1]. The progressions in way of life, school environment, academic pressure, profession pressure, different model of educating, increment of work, inn life and so on can be different explanations behind the improvement of mental pressure. For the most part above factors are the circumstances which get expeditious changes the existence of an understudy. Gradually the impact of these elements become basic to them and begin making pressure [2]. Additionally, the increment feeling of anxiety begins harming their wellbeing and mental harmony. There are different cases that comes before us in everyday life, where understudies end it all or on the other hand begins ingesting medications [3].

II .PROPOSED APPROACH

The work followed for the execution of our investigation process are shown. To achieve the prediction of scholar stress detection system using questionaries' and answers. The data from the various questions like academic grade, ID, stress data and the personal data of the person are detected



"Fig. 1. Steps followed for implementation"

Step insightful interaction to accomplish the proposed cycle:

1. Data Analysis Phase: This stage investigates information and its boundaries to check any repetition in information esteems that might influence expectation results.

2. This stage channels information to eliminate all unfilled/excess qualities.

3. Data Filtration Phase: This stage channels information to eliminate all vacant/excess qualities.

4. Train-Test Split Phase: This stage divides information into preparing and testing information subsets. For instance, information are separated into two sections for each a proportion of 80% preparing information and 20% test information.

5. Data-Scaling Phase: Before information are passed to the model, the information are scaled by model prerequisites. By large normalization or standardization is utilized for this. The objective to do this undertaking is to make the information limited to the particular reach. Thusly, this stage

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

reshapes information to make them more reasonable for the model.

6. Model-Building Phase: In this stage we are utilizing sklearn bundle of python which contains many bundles for characterization and relapse task. Here we are utilizing three generally utilized classifiers RF and DT and contrast their exactness with picked the best model. It is seen that RF performs better compared to the DT model.

7. Prediction Phase: In this stage we test our model with the test input information and make the expectation. Then, at that point, that yield is contrasted with testing information with ascertain misfortune and precision.

III .MODULES DETAILS

Source data regularly incorporate numerous flaws like missing worth, redundancies and irregularities. Hence pre-handling is needed to create a clean dataset that will guarantee that a ML strategy can assemble and prepare a model flawlessly with no mistakes [6]

"Data Gathering" Information gathering is an extremely essential module and the first process towards the concept. It generally deal with the gathering of the right dataset of personal and academic details of the scholars. Information gathering also makes it difficult to improve the dataset by it are outside to add more information that. Our academic/personal details attributes are mainly comprises of the id, academic level of the scholar and the question with answer details. We have given all the details of the scholar, survey details that are used for each class data.

"Dataset" Scholar information gathered from the school/college management of the students were utilized to make the method of 1025 subtleties. The informational index contained 2958 scholars details, of whom 1253 were stressed and 103 were not stressed one. The dataset included 19 credits and two specials, to be specific the scholar ID and the more details, which contains the two classes stressed or not. They were posed fundamental inquiries about their sentiments in circumstances that they could have experienced somewhat recently and their responses to it [9].

"Data Pre-processing" Information pre-handling is a piece of ML. Raw information is typically, conflicting or fragmented and normally contains numerous blunders. The information pre-handling includes looking at for missing data, checking for clear cut values, separating the dataset into preparing and test set ultimately do a component scaling to limit the area of details with the goal that they can measure up on normal environs. In this paper we have utilized is to remove invalid value using null () process for really looking at null values and label Encoder () for changing data into non numerical data into numerical data.

Clear cut Data - Clear cut information are factors that contain name values instead of numeric qualities. Along these lines, here we have addressed harmless cells as worth 0 and dangerous cells as worth 1.

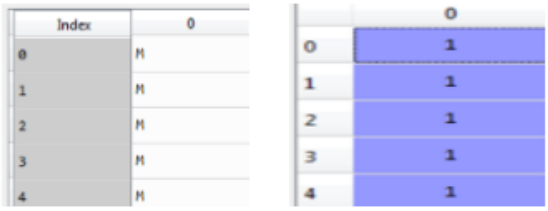


Fig 2: Data Conversion

Missing Data - Missing information incorporate void qualities or qualities not viable with the information design. For instance, highlights with mathematical organizations should comprise of numbers just, and ca exclude any images or alphabetic characters. The least complex methodology is to dispose of or eliminate such information all together. Anyway these information focuses could be significant; subsequently, we utilize the most extreme probability approach [4]. All missing information were supplanted with directly added values.

Data Sampling - Since the classes are intensely imbalanced, we increase the preparation information to get adjusted circulation among the classes. We reflect and turn the information to make new increased informational index. In examining, there can be a larger number of informational collection chose than expected to work with. Execution of More information can result in more prominent computational and memory prerequisites. A more modest positioned test of the chose information can be taken for thought, which will be much faster to find and demonstrate the arrangements prior to considering the whole informational index

Data Splitting - For every investigation, we split the entire dataset into 80% planning set and 20% test set. We used the arrangement set for resampling, hyper limit tuning, and setting up the model and we used test set to test the introduction of the pre-arranged model. While isolating the data, we showed an inconsistent seed (any sporadic number), which ensured comparable data split each time the program executed.

A dataset used for ML should be separated into three subsets — getting ready, test, and endorsement sets.

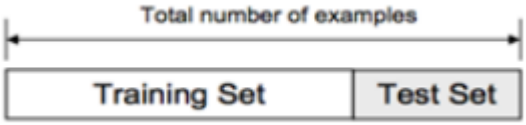


Fig 3: Data Splitting

Planning set: - A data scientist uses an arrangement set to set up a model and portray its optimal limits — limits it requirements to acquire from data.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

Test set: - A test set is expected for an evaluation of the pre-arranged model and its capacity for theory. The last choice suggests a model's ability to recognize plans in new unnoticeable data following having been arranged over a planning data. It's essential to include different subsets for planning and testing to avoid model over fitting, which is the deficiency for theory we referred to beforehand.

"Decision tree" In this model is as a flowchart where the node on inside hub addresses the dataset credits and the external branches are the result. Decision Tree is picked because of the fact that they are quick, solid, simple to explain and very little information for process is required. In Decision Tree, the classification of class label begins from foundation of the tree. The worth of the root credits is differentiated with record's credits. On the outcomes of differentiation, the relating branch is followed to that worth and bounce is made to the following next hub node.

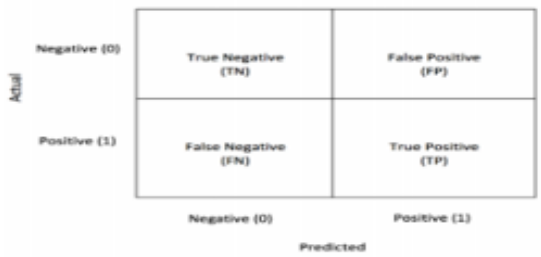
Random Forest- This model are utilized for prediction as well as retrogression process. It makes a tree for the information and makes classification using that. Random Forest [5] methods can be utilized on enormous datasets and can deliver a similar outcome in any event, when huge sets record values are absent. The created test data from the decision tree can be saved with the goal that it tends to be utilized on different information. In random forest there are two phases, initially make an random forest then, the prediction is done using the random forest classifier made in the primary stage [6].

Validation result- The results got by applying RF and DT are shown in this portion. The estimations used to finish approval aftereffects of the models is Accuracy score.

$$\textbf{Accuracy} = (TP+TN) / (P + N)$$

Approving precision is the significant boundary that we used in this work. Accuracy can be portrayed, using condition. The worth of accuracy will show the overall accomplishment of the ML models.

Confusion matrix- All anticipated genuine positive and genuine negative partitioned by all certain and negative. Genuine Positive (TP), True Negative (TN), False Negative (FN) and False Positive (FP) anticipated by all calculations are introduced in fig 2.



"Fig. 4. Confusion matrix"

Genuine positive (TP) shows that the positive class is anticipated as a positive class, and the quantity of test positive classes was really anticipated by the model.
Misleading negative shows (FN) that the positive class is anticipated as a negative class, and the quantity of negative classes in the example was really anticipated by the model.
Misleading positive (FP) shows that the negative class is anticipated as a positive class, and the quantity of positive classes of tests was really anticipated by the model.
Genuine negative (TN) demonstrates that the negative class is anticipated as a negative class, and the quantity of test negative classes was really anticipated by the model.

IV. RESULT AND OBSERVATION

In the exploration the pre-dealt with dataset is used to finish the assessments and the recently referenced ML models are researched and applied. The recently referenced approval estimations are gotten using the disarray network. Disarray network makes sense of the result approval of the model.

Table 1. "Values obtained from confusion matrix "

Algorithm	True Positive	False Positive	False Negative	True Negative
DT	124	01	17	108
RF	123	02	05	120

The disarray network gained by the proposed model for different techniques is shown in Table 1. The precision score procured for Random Forest and Decision Tree techniques is shown in Table 2.

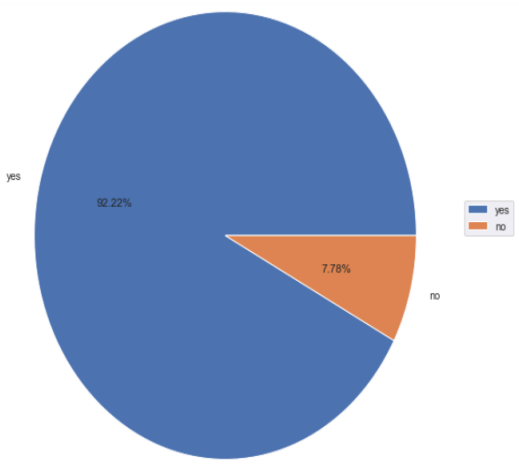
Algorithms	Training accuracy	Testing Accuracy
DT	92.80	99.46
RF	100	97.20

Table 2. "Accuracy for both models"

Dataset

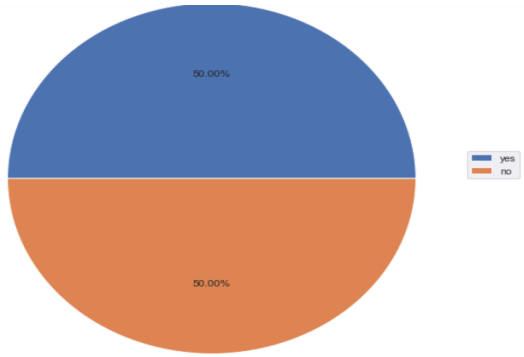
```
df2['Q5-Stressed about Academic issues'].value_counts()

Yes    2728
No      230
Name: Q5-Stressed about Academic issues, dtype: int64
```

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

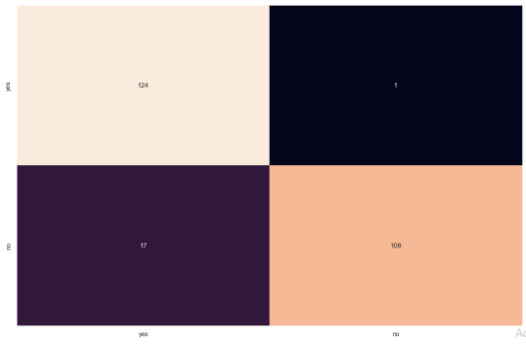


After Sampling:

```
0    500
1    500
Name: Q5-Stressed about Academic issues, dtype: int64
```

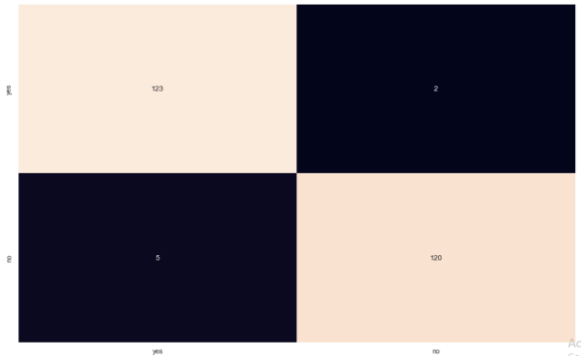


Decision Tree Confusion Matrix




```
the accuracy on testing data 0.928
the accuracy on training data 0.9946666666666667
```

Random Forest Confusion Matrix




```
the accuracy on testing data 0.972
the accuracy on training data 1.0
```

The below figure shows that the student is classified as stressed.

V. CONCLUSION

In this study an effort has been made to develop a model that could predict stress level in students The model uses Decision Tree and Random Forest Machine . Students' stress level can be perfectly predicted using Decision Tree algorithms by studying the key variables affecting the problem at hand. Data from student surveys on a variety of topics, including the Financial position,mental stability,interest in public welfare,interests in different activities etc., provide a strong foundation for this process. A thorough comparison of numerous ML models, including Random Forest and KNN etc., has been conducted with the suggestion of the best algorithm for stress level Analysis.

Experiment was performed using the dataset of 513 students collected from various students by using online and offline questionnaire. Results of the experiments shown in the form of confusion matrix, Sampling, overall accuracy . The accuracy of the DT model is 99.46% and RF model is 97.20%. In the future, the study may be improved by including more machine learning algorithms and by increasing the size of dataset, also by introducing some new features with the help of some feature engineering technique.

The most frequent cause of poor student performance is mental health issues. Mental illness can influence students' motivation, focus, and social connections—all of which are crucial components of their academic success. Numerous universities and colleges around the world have turned to online learning as a result of the recent coronavirus pandemic. Emergency remote learning (ERL) was widely

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

used in higher education during the COVID-19 epidemic, but little is known about the factors that affect student happiness and stress levels in such a novel learning environment.

VI. REFERENCES

[1] Jung, Yuchae, and Yong Ik Yoon. "Multi-level assessment model for wellness service based on human mental stress level." Multimedia Tools and Applications 76.9 (2017): 11305-11317
[2] Towbes, L.C. and Cohen, L.H., 1996. Chronic stress in the lives of college students: Scale development and prospective prediction of distress. Journal of youth and adolescence, 25(2), pp.199-217.
[3] Ghaderi, A., Frounchi, J. and Farnam, A., 2015, November. Machine learning-based signal processing using physiological signals for stress detection. In 2015 22nd Iranian Conference on Biomedical Engineering (ICBME) (pp. 93-98). IEEE.
[4] García S, Luengo J, Herrera F. Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. Knowledge-Based Syst. 2016;98:1–29
[5] Khosrowabadi, Reza, Chai Quek, Kai Keng Ang, Sau Wai Tung, and Michel Heijnen. "A Brain-Computer Interface for classifying EEG correlates of chronic mental stress." In IJCNN, pp. 757-762. 2011.
[6] Subhani, Ahmad Rauf, Wajid Mumtaz, Mohamed Naufal Bin Mohamed Saad, Nidal Kamel, and Aamir Saeed Malik. "Machine learning framework for the detection of mental stress at multiple levels." IEEE Access 5 (2017): 13545-13556.
[7] Evgeniou, T. and Pontil, M., 1999, July. Support vector machines: Theory and applications. In Advanced Course on Artificial Intelligence (pp. 249-257). Springer, Berlin, Heidelberg.
[8] Verma, G. and Verma, H., 2019. Predicting Breast Cancer using Linear Kernel Support Vector Machine. Available at SSRN 3350254.
[9] Nisha Raichur, Nidhi Lonakadi, Priyanka Mural, “Detection of Stress Using Image Processing and Machine Learning Techniques”, IJET, vol 09, No 3S, July 2017.
[10] Ahmad Rauf Subhani, Wajid Mumtaz, Mohamed Naufal Bin Mohamed Saad, Nidal Kamel, Aamir Saeed Malik, “Machine Learning Framework for the Detection of Mental Stress at Multiple Levels”, IEEE Access, Volume 05, July 2017.
[11] Dr.K.Rameshwaraiah, A.Ramakanth, “Detecting Stress Based on Social Interactions in Social Networks”, IJRTER, ISSN (online): 2455-1457,August 2017.
[12] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell. SmartGPA: how smartphones can assess and predict academic performance of college students. In UbiComp’15, 2015.