



STRESS LEVEL PREDICTION USING MACHINE LEARNING

Prof. Swati Tyagi, Prashant Malkoti, Srishti Kandpal

Dronacharya Group of Institutions, Greater Noida

swatityagi.dgi@gmail.com, kandpalsrishti411@gmail.com, malkotiprashant@gmail.com

1878

ABSTRACT: In this modern world where life is accelerated with top speed stress become very usual in mass. No age group of people is unaffected with this problem even study says that college going students are more prone to stress due to various factors. Predicting stress merely with few physical features is not easy and delays in precautions and remedies. Hence various ways have been used by researchers to predict stress. Through this paper we are putting efforts to contrast various studies made on stress prediction using machine learning. Best technique according to accuracy and prominent factors of stress specially in college/university students.

Keywords: Stress, machine learning, SVM(support vector machine), Nave-Bayes, Random Forest.

Number: 10.14704/nq.2022.20.7.NQ33236

Neuro Quantology 2022; 20(7):1878-1886

1 INTRODUCTION

The current educational system and the tough competition inclines the anxiety and stress amongst students. Other factors which contribute towards the mental disparity amongst students include parental pressure, peer pressure, health issues, financial conditions. Moreover, most of the students have some personal issues like relationships or family problems which they are not very comfortable to share with anyone, affecting the mental wellbeing of the student. The additive has been the pandemic of the corona virus, dysfunctioning the normalcy of student's lives andsuscepting them to more pressure thus leading to ill performance. Pandemic impact on economics has placed huge stress on students, especially if parents are going thorough financial instability.

Stress amongst students affects their health, performance and overall mental well-being. A survey was conducted among students in a questionnaire format to analyse the condition of students between the age of 16-25.

Automation for student stress prediction in institutes and educational organizations has been very minimal. Observing each student and his or her profile is a huge task. This responsibility lies under human interaction and that is why our work paves way for the automatic stress prediction of each student succumbing under various parameters and proposes the solution to each student rightly. This is done with the help of Machine learning and data science techniques. Keeping a check of

each student's stress levels, and monitoring it closely, helps to heighten their performance in an organization.

2 LITERATURE SURVEY

Y. Hou, J. Xu, Y. Huang, and X. Ma(2017) stated that stress has been a havoc mental stress on the students and which is causing many problems to them in their academics and as well as in their personal life. A survey on the basis of graphical analysis of data has reported that 15 students out of 91 test record to have normal or nearly no stress .26 students are identified with the range of 1% to 25%of mental stress and 29% found to having the range from 25% to 50% of stress, 20 students have 50 % and above stress level.

A. Hasanbasic, M. Spahic, D. Bosnjic, H. H. Adzic, V. Mesic, and O. Jahic (2019) studied for predictive analysis of stress over the students and the author had used the method of the KNN classification and used this algorithm to find the stress level among the students so that they can easily help the students to get overcome from that and help them .in this modal they have the accuracy of 94 % and in the accuracy the correct data is 95.4% and the incorrect data is 94.5%. the main thing is that the implementation of this KNN classification is self-coded.

A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam, and A. Chakrabarty(2019), they have focussed on the university students and their main focus is on university students because it will help many of the students to get



overcome from this. In this they have used the KNN, SVM and the decision-making tree to make the prediction correct and so that they can help the university students by the following algorithms and the methods used. This methodology is true when the survey has taken place and this survey is helping them in many ways.

Entrepreneurial Competency in University Students | Kaggle(2021), they are focussing on the university students with the help of the machine learning techniques and this help them to find the correct the data of the students and on the basis of that data they can guide the students to do this thing or not to get rid from the stress and make physical and mental health good. by using the MCC techniques they have founded the 200 instances and 8 attributes.

]Gao Huang, Kilian Q. Weinberger & Alice X. Zheng Zhixiang (Eddie) Xu (2019), they focussed on the students on the criteria of the financial pressure, educational pressure and for that they have used the raw data for this and this help them to find the data on the basis of the raw material and make the survey efficient and by this survey they have found that among the 200 students 160 have the stress level due to their financial status and by this we can ask the government to do something for this type of problems.

H. Elmunsyah, R. Mu'awanah, T. Widiyaningtyas, I. A. E. Zaeni, and F. A. Dwiyanto , they focused on predicting the mental health problems of the higher studies students and for this we are using the logical reasoning, regression method so that we can find the stress on these students also as they have the pressure of the placement and they are majorly focussed on their studies and they do not look over their own growth. The study has MATLAB STUDY tool box and the SVM has achieve the best result with the accuracy of 90 %.

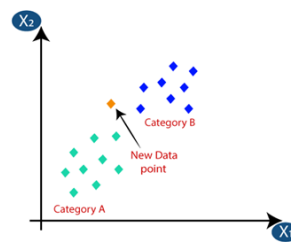
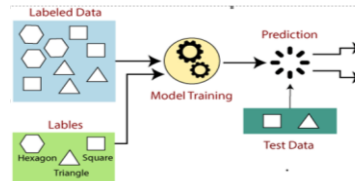
A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola (2020), focus on stress prediction of professional students using the machine learning (native bayes etc.) and with the help of this model they have the accuracy of 89.9% and with help of this they can easily guide the students for their best.

3 Techniques/Models of Machine Learning

Machine learning uses programmed algorithms that learn and optimise their operations by analysing input data to make predictions within an acceptable range. With the feeding of new data, these algorithms tend to make more accurate predictions. There are 3 ways of machine learning process: supervised, unsupervised and semi supervised.

3.1 Supervised learning

In such learning models are trained using labelled dataset, where the model learns about each type of data. Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output. Learning stops when the algorithm achieves an acceptable level of performance.



Some supervised learning algorithms are as the following:

3.1.1 K-Nearest Neighbour:

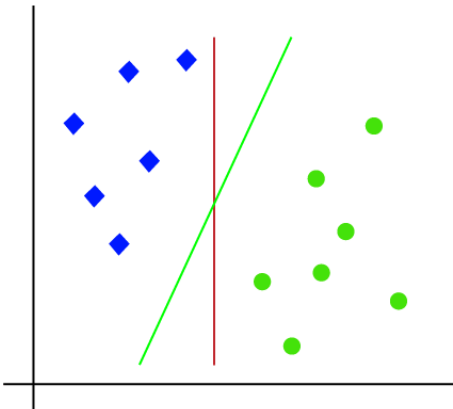
K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. Firstly, number of K neighbours are selected. Then the Euclidean distance of K number of neighbours is predicted nearest neighbours as per the calculated Euclidean distance is taken. Among these k neighbours, number of the data points in each category is counted. New data points are assigned to that category for which the number of the neighbour is maximum.

3.1.2 Support Vector Machine:

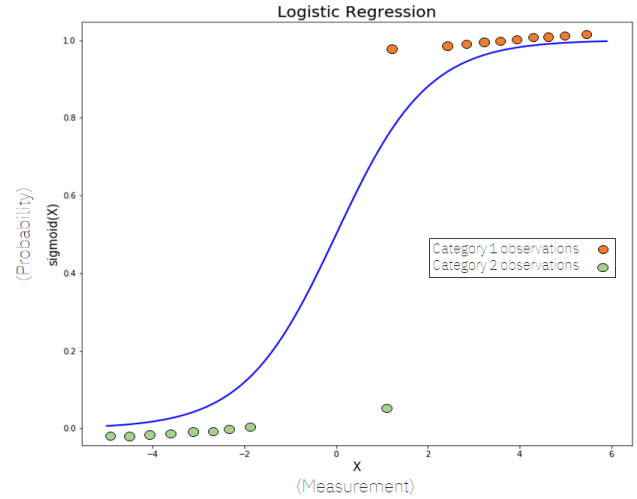
Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. It first maps each data



item into an n-dimensional feature space where n is the number of features. It then identifies the hyperplane that separates the data items into two classes while maximising the marginal distance for both classes and minimising the classification errors. The marginal distance for a class is the distance between the decision hyperplane and its nearest instance which is a member of that class. More formally, each data point is plotted first as a point in an n-dimension space (where n is the number of features) with the value of each feature being the value of a specific coordinate. To perform the classification, we then need to find the hyperplane that differentiates the two classes by the maximum margin.

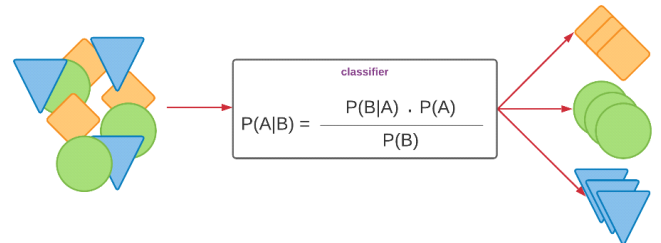


1. **Logistic Regression:** Logistic regression is a process of modelling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on. Multinomial logistic regression can model scenarios where there are more than two possible discrete outcomes. Logistic regression is a useful analysis method for classification problems, where you are trying to determine if a new sample fits best into a category. As aspects of cyber security are classification problems, such as attack detection, logistic regression is a useful analytic technique.



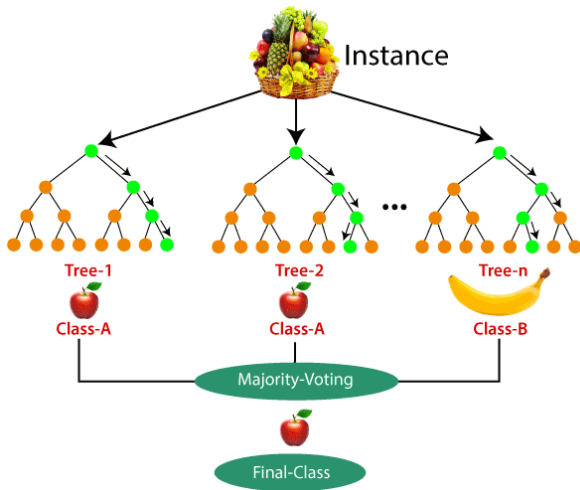
2. **Naïve Bayes:** Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Data is first converted to a frequency table. Then likelihood tables are created by finding probabilities. Then using Naïve Bayesian equation to calculate the posterior probability for each class, class with highest posterior probability is obtained.

Naive Bayes Classifier



Random Forest: Random Forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. Firstly, n number of random records are taken from dataset having k number of records. Then, individual decision trees are constructed for each sample. Then each decision tree generates an output. Final output is considered based on majority voting or averaging for classification and regression respectively.





3. **Decision Tree:** It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node. For the next node, the algorithm again compares the attribute value with the

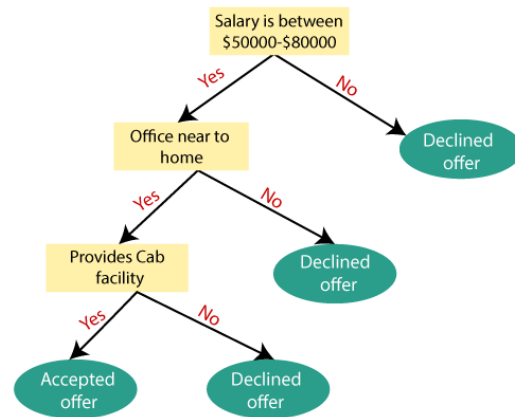
4. COMPARITIVE STUDY

Stress among students has been affecting them adversely in a wrong way. It is concluded that different kind of techniques can be used to detect stress level.

Study	Research Problem / Applications	Techniques used	Results	Suggested Future Work
[1]	Predicting level of generalized anxiety disorder of coronavirus pandemic among college age students using artificial intelligence technology	7 supervised machine learning models	Best performance based for AUC is AdaBoost (0.93) followed by neural networks (0.936). Highest accuracy and F1 were for neural network (0.754)	None
[2]	Classification Algorithms based mental health prediction using data mining	Decision Tree, Random Forest, Naïve Bayes	Decision tree has highest accuracy out of 3 algorithms with 82.2%. For random forest and naïve bayes, both have accuracy of 70.3% and 78.7% respectively	Create a system to predict specific mental illness that the person suffers
[3]	The BMI and mental illness nexus: A machine	Linear Regression	The accuracy of the model is 0.69. This	Relationship can help psychiatrists

other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree.

4.



3.2 UNSUPERVISED LEARNING

It is a type of learning in which we are using the artificial intelligence algorithms to identify the patterns in data set containing data points that are neither classified nor labelled . in our whole study we are not using any method of unsupervised learning we are only using the method supervised learning .



	learning approach		shows the is a strong relationship between BMI and mental illness.	easily find the causes for mental disorder
[4]	Classification of employee mental health disorder treatment with K-nearest Neighbour Algorithm	K-nearest neighbours	Test with variation value of K, highest accuracy is 0.8515 with value of K=17. As for the category of requiring mental health treatment the model correctly classifies 32 data out of 38	None
[5]	Machine learning techniques for stress prediction in working employees	Logistic regression, KNN classier, Decision tree, Random Forest, Boosting	Implementation of ensemble learning (boosting) has given the highest accuracy which is 75.13%	Naïve bayes, convoluted neural networks can be used
[6]	MIDAS: Mental Illness Detection and Analysis via social media	Twitter API, TF-IDF, Radom Forest	Used 10-fold cross validations to evaluate our models. Applying only the TD-IDF features, precision of 9.6% for both the BP and BPD models was achieved	None
[7]	Predicting mental health problems in adolescence using machine learning techniques	Random forests, support vector machines, neural network, XGBoost, Logistic Regression	Model performance was tested using AUC score. The best model is random forests with AUC score of 73.90% followed by SVM with 73.60%, neural network with 70.50%, logistic regression with 70.00% and XGBoost with 69.20%	None
[8]	Prediction of mental health problems among higher education students using machine learning	Linear regression, Logistic regression, KNN, Support vector machines, Naïve bayes	The prediction accuracy of the two regression methods are similar nut in smaller relative error (0.2-0.15) range, the prediction accuracy of	Expand the sample size to improve accuracy of the classifier



			logistic regression model varies from 0.779 to 0.838, showing better predictive effects. For classification algorithms, in terms of accuracy, the naïve bayes classification based on polynomial model and SVM classification have the highest accuracy, and the accuracy rate is about 0.82 when the data is 5500	
[9]	Recognition of stress levels among students with wearable sensors	Support vector machines, linear discriminant analysis (LDA) Ensemble, KNN, Decision tree	This study uses MATLAB machine learning toolbox for classification. Based on the results, SVM obtained the best accuracy score with 91% followed by LDA with 90.7%. Ensemble with 89.5%, KNN with 87.6% and decision tree with 86.1%	None
[10]	Prediction of depression in Bangladeshi undergraduate using machine learning	KNN, Random Forest, Support vendor machine	The model performance was tested using accuracy, precision, recall, f-measure, and AUC score to determine best model. The best model is Random Forest with 75% accuracy, 70% precision, 53% recall and 60% f-measure. In terms of AUC score, SVM gained the best score with 80.20%. There is not much difference between random forest and SVM but the KNN model is definitely not suitable to be used in	Use larger dataset and running more algorithms to see if they can get better accuracy and lower false negatives. Moreover, they are also working on finding out the optimal features and using them to predict depression more accurately



			this research	
[11]	Stress prediction of professional students using machine learning	Machine learning techniques (Naïve bayes, Logistic Regression, kappa, statistics measure, MCC, Mean absolute error, ROC)	Model performance was tested using accuracy and the result are on the basis of experimental result. Correctness are bayes net - 88.3495%, Multiplayer Neleprov - 85.394%, Logistic regression - 86.4076%	None
[12]	Prediction of mental stress level based on machine learning	By using signal pre-processing, image and video processing SVF and RF algorithm	Measure the performance evolution in terms of various parameters like accuracy, precision which is 92.38%	Helpful for predicting using conversation and social media data
[13]	Predictive analysis of student stress	KNN Classification Algorithm	They have model accuracy and algorithm of 94%, correctly classified 94.5%, incorrectly 5.945%	None
[14]	Prediction of metal health among university students	Mainly three algorithms (KNN, SVM, Decision tree)	Decision tress showed accuracy of 0.64, KNN showed of 0.59, SVM of 0.44	None
[15]	Prediction of stress using machine learning in university students	By using the method of MCC	In this data set we have founded the 200 instances and 6 attributes	It is helpful for the university students to overcome from this
[16]	The level of stress among college students, in case of education, financial and etc.	In this we are using the raw data and using the machine learning	Like there is unfair grading system in college -80 %. And unfair percentage section distribution - 60%	None



5. DISCUSSION

As from all the papers we have seen that on a large scale the students whether they are university students, professional students, school students it was found that the 90 % of the students are being suffering from the stress and majorly the factor on which they are suffering the stress are (peer stress, family stress, financial stress, friends/relationship stress, social anxiety and etc.)

As due to these parameters it is affecting the student's personal life as well as their mental and physical health and this result in the miscommunication and can cause many problems in their academics. And to have to support this there should be a good parental skill so that they can easily communicate over these topics and the students also feel free to express their views on these topics and this will help them in many ways like this will help in making a good friendship between the child and the parents.

A combination of various machine learning techniques, as well as deep learning techniques, is used to find the exact stress level and also used to predict the disease. various machine learning techniques such as K- near neighbours. The final class label is based on the majority of the number of votes decided.

And we have also kept in mind about the different circumstances and the situation of the students and on the basis of that we are trying to help the students so that this can help then in living a stress-free life and they can do the things in a better way without doing the thing in a wrong manner so we have used the different methods to predict the stress level of the students.

6. CONCLUSION

Stress is a problem which is increasing day by day affecting the individual physical and psychological health. The classification techniques we have used are naive bayes, logical regression, decision tree, KNN, SVM the accuracy of various techniques is calculated and compared. According to this paper we observe the kappa, statistic f- measure, MCC, mean absolute error, ROC area, false positive, true positive, RMSE and recall bayes net classifier gives the highest accuracy of 88%. To classify

and predict the mental stress on obtained data by SVM and RF classifiers, it can be observed That SVM gives better accuracy than RF algorithm. The proposed designed is helpful for predicting the mental stress of user using the conversation and social media data.

Student are the victim of massive amounts of stress and this is increasing at an alarming rate with the ever-growing competition on educational grounds. This surfaces the other parameters also. Our work is an aid which helps in predicting the stress level due to various parameters consider through our upcoming survey.

REFERENCES

- [1] H. Alharthi, "Predicting the level of generalized anxiety disorder of the coronavirus pandemic among college age students using artificial intelligence technology," in Proceedings - 2020 19th Distributed Computing and Applications for Business Engineering and Science, DCABES 2020, Oct. 2020, pp. 218–221, doi: 10.1109/DCABES50732.2020.00064.
- [2] V. Laijawala, A. Aachaliya, H. Jatta, and V. Pinjarkar, "Classification Algorithms based Mental Health Prediction using Data Mining," Jul. 2020, pp. 1174–1178, doi: 10.1109/icces48766.2020.9137856
- [3] R. Reya Pillai, S. Saravanan, and G. Krishna Shyam, "The BMI and mental illness nexus: A machine learning approach," in Proceedings of the International Conference on Smart Technologies in Computing, Electrical and Electronics, ICSTCEE 2020, Oct. 2020, pp. 526–531, doi: 10.1109/ICSTCEE49637.2020.9277446.
- [4] H. Elmunsyah, R. Mu'awanah, T. Widiyaningtyas, I. A. E. Zaeni, and F. A. Dwiyanto, "Classification of Employee Mental Health Disorder Treatment with K-Nearest Neighbor Algorithm," in ICEEIE 2019 - International Conference on Electrical, Electronics and Information Engineering: Emerging Innovative Technology for Sustainable Future, Oct. 2019, pp. 211–215, doi: 10.1109/ICEEIE47180.2019.8981418.
- [5] U. S. Reddy, A. V. Thota, and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," Dec. 2018, doi: 10.1109/ICCIC.2018.8782395.



- [6] E. Saravia, C. H. Chang, R. J. De Lorenzo, and Y. S. Chen, "MIDAS: Mental illness detection and analysis via social media," in Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016, Nov. 2016, pp. 1418–1421, doi: 10.1109/ASONAM.2016.7752434.
- [7] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, "Predicting mental health problems in adolescence using machine learning techniques," PLoS One, vol. 15, no. 4, p. e0230389, 2020, doi: 10.1371/journal.pone.0230389.
- [8] Y. Hou, J. Xu, Y. Huang, and X. Ma, "A big data application to predict depression in the university based on the reading habits," in 2016 3rd International Conference on Systems and Informatics, ICSAI 2016, Jan. 2017, pp. 1085–1089, doi: 10.1109/ICSAI.2016.7811112.
- [9] A. Hasanbasic, M. Spahic, D. Bosnjic, H. H. Adzic, V. Mesic, and O. Jahic, "Recognition of stress levels among students with wearable sensors," May 2019, doi: 10.1109/INFOTEH.2019.8717754.
- [10] A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam, and A. Chakrabarty, "Predicting Depression in Bangladeshi Undergraduates using Machine Learning," in Proceedings of 2019 IEEE Region 10 Symposium, TENSYPMP 2019, Jun. 2019, pp. 789–794, doi: 10.1109/TENSYPMP46218.2019.897136
- [11] "Entrepreneurial Competency in University Students | Kaggle." Stress prediction of professional students using machine learning" 1886
- [12] Gao Huang, Kilian Q. Weinberger & Alice X. Zheng Zhixiang "Prediction of mental stress level based on machine learning "(Eddi" , 13 Jan 2019, p. 10.)
- [13] E. Chesney, G. M. Goodwin, and S. Fazel, "Predictive analysis of student stress" KNN classifications.
- [14] D. Banerjee, J. R. Kosagisharaf, and T. S. Sathyanarayana Rao, Prediction of mental health among university students
- [15] D. K. Cheung, D. K. Y. Tam, M. H. Tsang, D. L. W. Zhang, and D. S. W. Lit, Prediction of stress using machine learning in university students
- [16] M. Shafi, J. Liu, and W. Ren," The level of stress among college students, in case of education, financial and etc."