International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

2023IJIEMR.Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 06th Feb 2023. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=ISSUE-02

DOI: 10.48047/IJIEMR/V12/ISSUE 02/09

Title **An Approach To Identify Accurate Machine Learning Model To Build Human Stress Level Prediction System**

Volume 12, Issue 02, Pages: 60-65

Paper Authors
Ramu Y, Ramachandra Rao Kurada,Sunil Pattem

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

An Approach To Identify Accurate Machine Learning Model To Build Human Stress Level Prediction System

Ramu Y
Assoc. Prof., Dept. of CSE ,Shri Vishnu Engg. College for Women, Bhimavaram
yramumail@gmail.com
Ramachandra Rao Kurada
Prof., Dept. of CSE, Shri Vishnu Engg. College for Women, Bhimavaram
ramachandrarao.kurada@gmail.com
Sunil Pattem
Asst. Prof., Dept. of CSE, Shri Vishnu Engg. College for Women, Bhimavaram
sunilpattem1979@gmail.com

Abstract— In the present fast growing world, people are busy with their personal and professional lives, where everyone is expected to do multitasking to meet their aspirations or work timelines, which leads to negative stress sometimes. If individuals fail to recognize and handle this stress, then it causes to several physical and mental health disorders as well as fatalities also. In this paper, prediction of stress levels was made by using machine learning algorithms on a data set with 916 samples with features such as age, sex, blood pressure, heart beat, body temperature and stress level number. After applying various machine learning algorithms, their performance is evaluated with metrics such as MAE (Mean Absolute Error), RMSE (Root Mean Square Error) and RMSLE (Root Mean Square Logarithmic Error) to identify an accurate model. From our experimental results, we identified that Decision Tree as an accurate model to predict the stress levels by using the most effective evaluation metric, RMSLE (Root Mean Square Logarithmic Error), which is robust metric even when outliers exist in the dataset. And, it is used to build an application to predict the stress levels of individuals in real-time, which helps to recognize the stress at early stage to take necessary healthcare precautions in guided approach.

Keywords — Stress Level Prediction, Health Care, Machine Learning, Decision Tree, RMSLE (Root Mean Square Logarithmic Error).

Introduction

Stress is the response of either human body or brain in an uncomfortable situation or emotion perceived, resulting in physical, emotional and intellectual responses. In general, stress is defined as a feeling of strain and pressure, which occurs from any event or thought that makes people feel frustrated, angry, or nervous. In the present competitive environment, people are trying to manage their personal and professional lives, where everyone is expected to do multi-tasking to meet their aspirations or work timelines.

These scenarios cause to negative stress in people, who are not able to manage and deal the things with lack of maturity or awareness. Stress leads to a several chemical reactions in our body and brain, which causes people to respond in "fight-or-flight" mode, which in-turn increase heart-beat rate, body-temperature and blood-pressure. As of the current situations, many people have succumbed to stress especially in the adolescent age group and with professional pressures. Nowadays, stress is one of the major reasons for heart attacks, depression and suicides. To prevent this, we need to focus on recognizing stress levels to initiate the stress management through self-care practices and/ or with the help of a health-coach like a doctor, psychologist or stress-management coach. In this work, we build an application to predict the stress levels of individuals using different machine learning algorithms, which helps to recognize the stress at early stage to take necessary healthcare precautions in guided approach.

Related work

Diagnostic parameters like EEG, GSR, EMG and SpO2 were used to measure the

automated stress prediction in individuals [1]. This measurement recorded by using various external and internal sensors which is always not feasible to individuals because of the need of several medical equipment's to record these parameters. Stress prediction using various parameters such as gender, academic year, academic grades, regularity to course attendance, and perceived physical problems related to medical students revealed few the key observations [2] like (i) that students' academic score or regularity to attend classes was not significantly associated with the stress level (ii) Students experience high stress in the first three years of the course (iii) Physical problems are associated with high stress levels. Estimation of stress level based on tweets posted by users in social networking platforms revealed few key observations like – the language used by individuals in Twitter platform, that scores higher in stress indicates the low mental health, inadequate services and/or facilities as well as poor socioeconomic status in the sample space chosen. It is a kind of observation about stress levels at individuals or region or country level [3].

Studies on mental health, human behavior and stress levels through machine learning approaches facilitates the health experts for their diagnosis. Different levels of anxiety, depression as well as stress were predicted used various classification methods and noticed that Random Forest as best model for the data gathered for measuring [4]. Stress classification methods were evaluated with Deep Belief Networks (DBN), which shown better results like 66.23% of accuracy and 75.32% of specificity, which was better than support vector machine, naive Bayesian classifier, and random forest methods used in the study [5]. Support vector machine, naive Bayesian classifier, and random forest methods shown better results compared to logistic regression model to predict Chronic stress. And, random forest method provided results with 99% of sensitivity and 79% of positive predictions [6]. Fuzzy clustered Bayesian model was developed to assist professionals in the process of diagnosing the basic mental health problems [7]. In addition to these, researchers also explored to study the changes in mental health due to long period of home quarantine connected with sentiment analysis [8] and personalizing the machine learning models in a principled approach to have a meaningful mood prediction [9].

Methodology

The objective of our work was to develop a machine learning application to predict the stress levels of human beings, which is highly convenient to provide input parameters or symptoms; ease to operate; and accurate in predictions. To identify an accurate machine learning application, we have adopted a pilot model in building the real-time machine learning application (MLop) as shown in the Fig. 1.

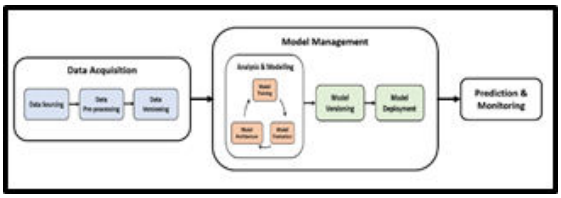

Fig 1. A pilot model to predict human stress levels with accurate machine learning algorithm in real-time machine learning application (MLop).

Data Acquistion

In this stage, data is collected from one or more relevant sources as per the requirements of the application, and merged to store it as a data version, after necessary data pre-processing tasks like data cleaning, noise removal and data transformations. Data versioning represents the data set or subset of the dataset as per feature engineering in machine learning, which is also a key aspect in the model versioning of the model management at a later stage.

Model Management

Machine learning models are expected to be, accurate, scalable, and robust to the business or application requirements. To fulfil this, we adopt model training, model evaluation and model architecture, in a cyclic approach to meet the changing needs of the application in real-time. Once the model architecture is completed, we build the model version as per the latest data version to initiate the model deployment.Model management is highly essential to mitigate problems related to

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

overfitting, under fitting, bias, and performance. It is highly essential to develop ML application which are easy to trace, complain as per regulations.

Model Deployed

The model deployed as part of model management is used to predict the output as per input parameters. The deployed model is also monitored about its performance and it depends on model type, performance measures, and data distribution. Monitoring is very crucial and critical when we integrate machine learning model in real-time application for production, so that production teams can reveal the potential issues before they impact the business or application objectives.

Experimental results

Dataset

In our work, we have used a dataset related to STRESS, which consists of 916 rows (or instances) and 6 features such as age, sex, blood pressure (bp), heart beat (hb), body temperature (temp) and stress level number (num). In this data set, stress level number (num) is in the range of 0-4. A correlation matrix helps to understand the correlation coefficients between different attributes in the dataset. It is most popular tool used to summarize data; data analysis & diagnostic analysis from basic to advance levels. Correlation coefficients represents the strength of the relationship between two attributes.

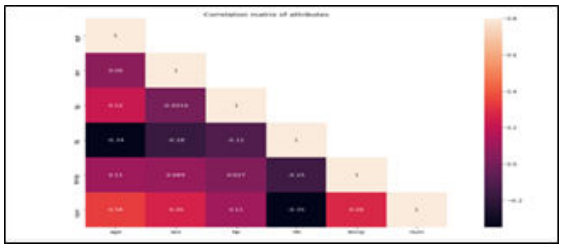

Fig 2 Correlation matrix of different features in STRESS dataset.

We have following three cases related to Correlation coefficients. Case-1: If coefficient value = zero, then it indicates no linear relationship. That is, if one variable increases, there is no change in the other variable (increase or decrease). Case-2: If coefficient value is in between 0 and +1/-1, then it represents there is a relationship As coefficient value approaches close to either -1 or 1, the asset of the relationship increases and the data points will be portrayed as a line or closer to a line. Case-3: If coefficient value indicates the extreme values of such as -1 and 1, then it represents a perfect linear relationship between attributes. In this case, all of the data points fall on a line. From the correlation matrix as shown in Fig. 2, we can observe the correlations like age & stress level number (num) is 0.34; body temperature (temp) & stress level number (num) is 0.28; sex & stress level number (num) is 0.26; and age & blood pressure (bp) is 0.22. These correlation coefficients show low positive relationship among attributes.

Table 1: List of evaluation metrics for model performance.

Evaluation Metric	Formula	Description
MAE (Mean Absolute Error)	$MAE = \frac{\sum_{i=1}^{n}\lvert y_i - \hat{y}_1\rvert}{n}$	It represents the sum of all the absolute error (or difference) between the actual and observed (or predicted) parameters
RMSE (Root Mean Square Error)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2}$	It represents the standard deviation of residual error of parameters and it is always positive. It indicates better performance of the model if it is low value. RMSE is sensitive in it's performance when outliers exist in the data set
RMSLE (Root Mean Square Logarithmic Error)	$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(y_i+1) - \widehat{(y+1)})^2}$	It is calculated by using logarithmic scale of the predicted and actual parameters after transformation, by adding 1 as a constant value to both. It is a robust metric even when outliers exist in dataset.

Model Training

In our work, we have used various regression models such as Linear Regression, Polynomial Regression, Decision Tree, and Random Forest for prediction of stress level. These models are trained with various combinations (or ratios) of training and testing data splits such as 60: 40, 70: 30 and 80: 20.

Model Evaluation

In general, regression-based machine learning model generates the predicted value. This predicted value is expected to be same as the actual value. And, the difference between these predicted and

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

actual value is known as residual error. We need an accurate machine learning model which gives prediction with residual error close to ZERO value or ZERO, which is highly essential for the overall performance of the machine learning application in real-time. In our work, the performance of various algorithms used in the model training phase are evaluated about its accuracy with three different metrics such as MAE, RMSE, and RMSLE (Root Mean Square Logarithmic Error), which are listed in Table-1 are popular to reveal the occurrence of residual error in model prediction [4]. These values are recorded for each algorithm in each combination (or ratio) of train-ing and testing data splits such as 60: 40, 70: 30 and 80: 20 are shown in the Fig. 3, Fig. 4, and Fig. 5.

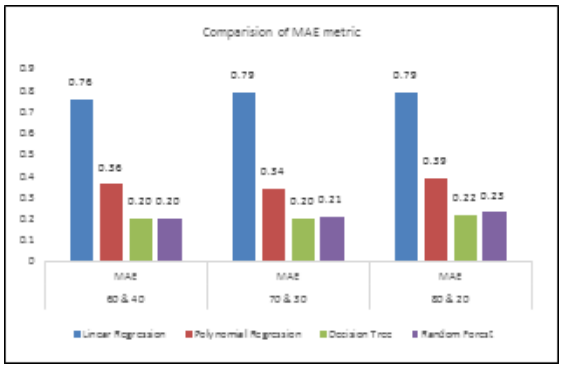

Fig 3. Comparison of MAE metric on regression models with different training & testing data splits.

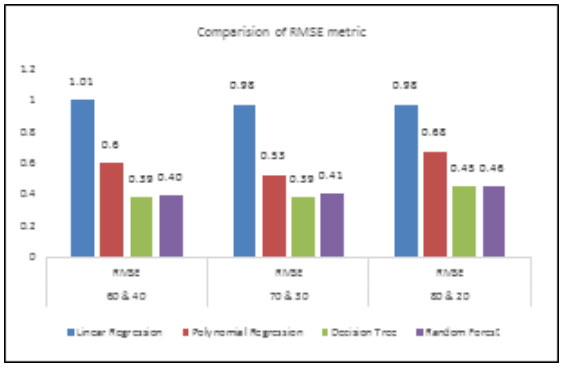

Fig 4. Comparison of RMAE metric on regression models with different training & testing data splits.

Model Architecture

As per the observations from the model evaluation phase, we preferred RMSLE (Root Mean Square Logarithmic Error) as a selective metric to choose the best fit model in terms of its accuracy for implementation, because RMSLE is robust metric even when outliers exist in dataset. From the overall experimental results shown in Table 1, Fig. 6, and Fig. 7, we noticed that Decision Tree model is providing more accurate prediction because it's residual error values such as MAE, RMSE and RMSLE are very low, when it is compared with other models.

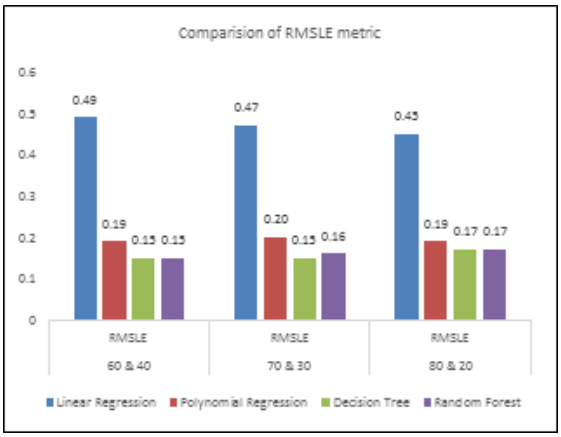

Fig 5. Comparison of RMSLE metric on regression models with different training & testing data splits.

Table 2 :Comparison of performance of regression models with different evaluation metrics, training & testing data splits

Ratio of Training and Testing data Splits	Evaluation Metric	Linear Regression	Polynomial Regression	Decision Tree	Random Forest
60:40	MAE	0.76	0.36	0.20	0.20
	RMSE	1.01	0.60	0.39	0.40
	RMSLE	0.49	0.19	0.15	0.15
70:30	MAE	0.79	0.34	0.20	0.21
	RMSE	0.98	0.53	0.39	0.41
	RMSLE	0.47	0.20	0.15	0.16
80:20	MAE	0.79	0.39	0.22	0.23
	RMSE	0.98	0.68	0.45	0.46
	RMSLE	0.45	0.19	0.17	0.17

As per RMSLE value (RMSLE = 0.15), Decision Tree model is providing best results in the ratio of 70: 30 for training and test data splits. So, the best fit model

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

which is selected for the architecture in the real-time application for prediction is, Decision Tree with a combination of 70: 30 for training and test data splits.

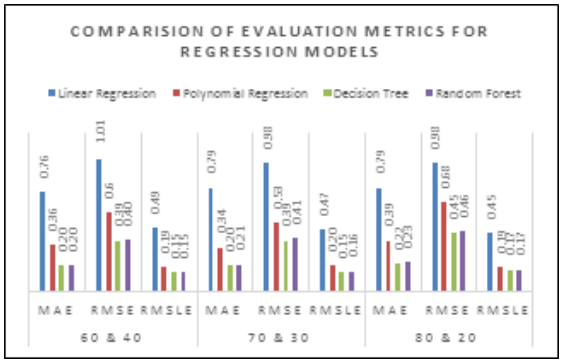

Fig 6. Comparison of performance of regression models with different

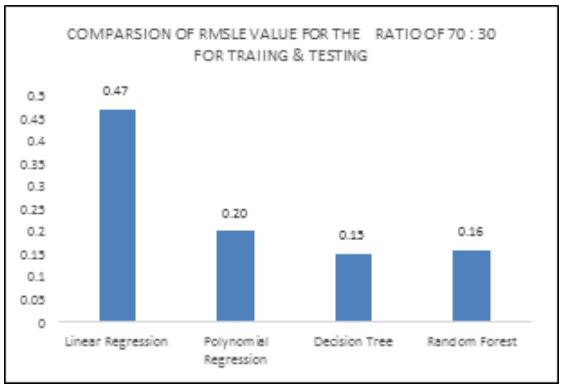

Fig 7. Comparison on RMSLE value for selection of best fit model

Model Versioning and Development

The first version of the model to predict the stress levels is to use Decision Tree machine learning algorithm with training and test data splits as 70: 30. This version may be enhanced to higher versions as per the changes in data versions and / or application requirements. The first version of the best fit model is deployed by the code written in Python programming language.

Table 3: Stress Levels - Sustainable & Suggested for Doctor Consultation

Age Group (in Years)	Sustainable Stress Levels	Stress Levels Suggested for Doctor Consultation
10 – 35	1, 2, & 3	4
36 – 65	1 & 2	3 & 4
66 – 75	1	2, 3 & 4
Above 75	None	1, 2, 3 & 4

Prediction and Monitoring

In our work, stress level number (num) is in the range of 0-4 and its categories. Stress levels are classified 0 to No Stress, 1 to Low Stress, 2 to Moderate Stress, 3 to High Stress and 4 to Very High Stress which is going to be predicted by using machine learning algorithms. We have classified these stress levels which are sustainable to different age groups as shown in the Table 3. In this case, a person with age = 40, can tolerate stress levels 1 and 2, and suggested to meet the doctor for consultation if it crosses the stress level 2, that is stress level 3 & 4. The working scenarios of web application system developed to predict human stress levels, which is deployed with the decision tree machine learning model are shown in Fig. 8.

Conclusion & Future Work

From our work, we found that the Decision Tree model is an efficient machine learning algorithm to provide more accurate prediction with less residual error as indicated by RMSLE. Prediction of stress levels is feasible with minimum set of features like age, sex, blood pressure (bp), heart beat (hb), body temperature (temp). These features data acquisition is very easy with minimal and affordable medical equipment such as B.P. monitor, pulse meter and thermo meter. Stress prediction application with minimum I.T. infrastructure & medical devices is feasible for small to large scale organizations and educational institutions to monitor and predict the stress levels of their stake holders like employees, staff and students. It helps to monitor and recognize the stress levels at early stage to take necessary healthcare precautions in guided approach to avoid or reduce several physical and mental health disorders as well as fatalities. This application is a pilot model of the stress prediction system which may be enhanced based on the real-time learning; requirements of end-users; inputs from model monitoring to develop an “improved model version and data version” - for the more effective and accurate real-time machine learning application (MLop) to predict the stress levels.

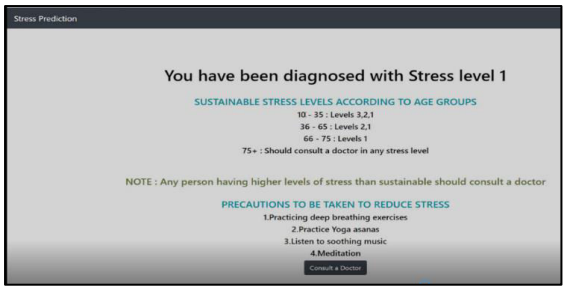

Fig 8. Stress Prediction Result Page and Suggestive Remedies

References

[1] Qianli Xu, Tin Lay New, Cuntai Guan 2015 Cluster-based analysis for personalized stress evaluation using physiological signals IEEE Journal of Biomedical and Health Informatics 19 (1) 275–281.

[2] Hamza M. Abdulghani, Abdulaziz A. AlKanhal , Ebrahim S. Mahmoud, Gominda G. Ponnamperuma, and Eiad A. Alfaris 2011 Stress and its effects on medical students: a cross-sectional study at a college of medicine in Saudi Arabia Journal of Health, Population and Nutrition 29 (5) 516-522.

[3] Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes Eichstaedt , Lyle H. Ungar 2019 Understanding and Measuring Psychological Stress Using social media Proceedings of the International AAAI Conference on Web and Social Media 13 (01) 214-225.

[4] Anu Priya , Shruti Garg, Neha Prerna Tigga 2020 Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms Proceedings of International Conference on Computational Intelligence and Data Science (ICCIDS 2019) 167 1258–126.

[5] Se-Hui Song, Dong Keun Kim 2017 Development of a Stress Classification Model Using Deep Belief Networks for Stress Monitoring Healthcare Informatics Research 23 (4) 285-292.

[6] Bozorgmehr A, Thielmann A, Weltermann B 2021 Chronic stress in practice assistants: An analytic approach comparing four machine learning classifiers with a standard logistic regression model. PLoS ONE 16 (5) e0250842 https://doi.org/10.1371/journal.pone.0250842.

[7] Sumathi M.R, Poorna B.2017 Prediction of Mental Health Problems Among Children using Fuzzy Clustered Bayesian Model International Journal of Control Theory and Applications 10 (23) 83 – 93.

[8] Imrus Salehina, Sadia Tamim Dip , Iftakhar Mohammad Talha , Ibrahim Rayhan , Kanij Fatema Nammi 2021 Impact on Human Mental Behavior after Pass through a Long Time Home Quarantine Using Machine Learning I. J. Education and Management Engineering 11 (1) 41-50.

[9] Natasha Jaques, Ognjen (Oggi) Rudovic, Sara Taylor, Akane Sano, Rosalind Picard 2017 Predicting Tomorrow's Mood, Health, and Stress Level using Personalized Multitask Learning and Domain Adaptation Journal of Machine Learning Research 66 (2017) 17-33.

[10] Akhilendra, Evaluation Metrics for Regression models- MAE Vs MSE Vs RMSE vs RMSLE https://akhilendra.com/evaluation-metrics-regression-mae-mse-rmse-rmsle/.