Book of Proceedings of the 1st Faculty of Science International Conference FSIC 2022 Vol. 2, 1-7
Air Force Institute of Technology, Kaduna Nigeria.
ISBN:9789787946589

A Stress Based Prediction Model for University Student Using Support Vector Machine and Grid-Search-CV for Parameter Turning

***A. Jibrin[1], J.K. Alhassan[2], S. A. Adepoju[3]**
[1]Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, Minna.
[2]Department of Cyber Security Science, Federal University of Technology, Minna.
[3]Department of Computer Science, Federal University of Technology, Minna.

Corresponding Author

Aisha Jibrin
Department of Computer Science,
School of Information and Communication Technology,
Federal University of Technology,
Minna. Nigeria.
ayshermuserh@gmail.com

Abstract

The current academic system consists of various mental struggles ranging from family, peers, lecturers and the academic system generally. However, high level of stress on university students negatively affect their academic performance. In this paper, we describe how to efficiently select the best parameters to develop the proposed model. The Grid-Search-CV techniques is adopted to fine-tune the Support Vector Machine(SVM) classifier with different parametric combination, the best parameter configuration that provides the highest prediction accuracy is selected for training our model. Hence, the proposed student stress prediction model has shown a high degree of prediction accuracy (99%).

Keywords: *Finetuning, Grid-Search-CV, Mental Health stress, Student, Support Vector Machine.*

INTRODUCTION

In today's educational system various countries have ranked education at the top of their agenda, in other to meet and satisfy their social demands (Nazari & Far, 2019). However, the prediction and evaluation of academic performance is crucial for assessing the quality of education and the degree at which the educational goal is achieved.

Stress is considered a psychological and hyperarousal state in humans (Pankajavalli *et al.*, 2021). Stress has been identified as a major pandemic of the twenty-first century by the World Health Organization (WHO). Generally, the imbalance triggered by the differences between situation demands and individual inability to effectively handle challenging circumstances can result to stress. Stress can occur due to varieties of reason, this includes work pressure, event traumatic stress, sadness and a lot more. Productivity can be reduced at work due to several negative or bad emotional feelings. Stress is considered as the feeling of acquiring more knowledge and the simultaneous perception of student not having enough time to acquire that knowledge. Factors surrounding stress pathology may include bad utilization of time, lack of necessary academic skills or knowledge to compete with classmates (Nazari & Far, 2019).

The stress dataset was downloaded from Kaggle repository. The data is the Heart Rate variability values taken from subjects to predict if they are stressed or not. the data is explored and scaled between the range of '0' and '1'. The data is divided into training (80%) and testing set (20%). 80% of the data set is used to train the model developed using SVM and fine tuned using Grid Search-CV to obtain the best parametric combination. The testing data set is used to evaluate the performance of the model.

A. Research problem

Its identified in the work of (Nazari & Far, 2019) that poor academic performances are basically the cause of academic stress. Moreover, the stress manifested in students could be a psychological stress, physical stress, chronic loss of energy, mental stress, and sleepless night. Hence, its essential to develop a more sophisticated and highly accurate model in detecting and managing stress levels in student. This can be achieved using best stress factors for prediction stress level (Park et al., 2020).

B. Research Goal

This study is proposed to develop a Stress prediction model using Support Vector Machine (classifier). Below includes the methodological step in achieving our goals

i. Gathering of Heart Rate Variability (HRV) data sample.
ii. Fine tuning the SVM using Grid-Search-CV
iii. Develop the model using the best parametric combination
iv. Evaluate and test the proposed model.

REVIEWED WORK

Zamkah *et al*., (2020) identifies that majority of the physiological based emotional and stress management system are developed based on sweat or skin conductivity. The researcher carries out a comprehensive review on the current state of human stress emotional markers. In today's Information technology organizations, stress disorders is one of the frequently raised issues. Reddy et al., (2018) proposed a machine learning based techniques to identify traces of stress in IT professionals and minimize the factors that strongly identify stress levels. Furthermore, Barker et al., (2018) reveal that the most common experiences within the university students are the self-reported depression. The researchers examine the depressive symptoms of each student raging from September to April. The findings have implications for knowing when and who are more likely to encounter depression symptoms among college students. Pankajavalli et al., (2021) also identifies that wearable technology, recent studies are concentrating on creating non-invasive methods to predict stress. Because stress patterns are very subjective and differ from person to person, the models created for stress prediction typically do not produce accurate results.

Although students can use the Internet to gain learning tools and engage in thought-provoking discussions with others, information, both good and bad, can have a harmful impact on students' mental health (Shen, 2021). The quality of the family environment has a direct impact on children's physical and mental health. There is no doubt that the main body of family education plays an important role in students' mental health education (Jing Li 2021). The pressure of social competition is extremely high in the new era (Shen, 2021). Because of technological advancements and the obsession that comes with the use and reliance on social media, it provides opportunity and anonymity for spreading negativity with no consequences.

PROPOSED STRESS BASED PREDICTION MODEL USING SVM AND GRID-SEARCH-CV

A. Proposed Model Diagram

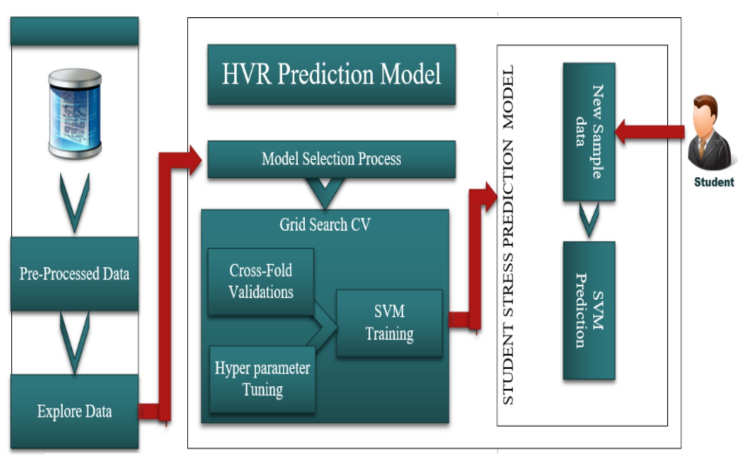

Figure 1. Conceptual diagram of the proposed stress prediction model

Figure 1 illustrates the conceptual design process of the proposed stress prediction model using support vector machine (SVM) learning algorithm and Grid-Search-CV for parameter tuning (this identifies the best classification accuracy using cross fold validation and reveal the best SVM parameter combination for training). Heart Rate Variability (HRV) data sample is firstly preprocessed and transformed into a suitable format for machine learning prediction. The pre-processed data sample are fed into the Grid-Search-CV to explore the dataset further, this is done by performing cross-fold validation and parameter tuning to identify the best parameter combination that yields the best accuracy. Then the model is trained using SVM classifier with the best parameter identified by Grid-Search-CV. Finally, the classification model is made available for update and prediction of university student stress level.

The proposed approach is illustrated in a simple step below as follows;

STEPS:

1. Importing of HVR dataset.
2. Data preprocessing and cleaning.
3. Then dataset is transformed and standardized for efficient prediction
4. Splitting of dataset into training and testing set
5. Training of model using Support Vector Machine with varieties of SVM parameter using Grid-Search-CV
6. Train the model with the best parameter turning
7. Testing of dataset using standard evaluation metrics (Accuracy, recall, precision and F1-score)

B. Support Vector Machine

The Support Vector machine is grouped as a non-probabilistic model (Shitole, 2021). the internal working is based on the principle of structural risk management to identify the appropriate or best hyperplane distinguishing two or more classes in an n-dimensional space. The hyperplane is express mathematically below.

$$w. x - b = 0 \qquad \text{............................ eq (1)}$$

$$\mathbf{c\ (x, y, f\ (x))} = \begin{cases} 0, \text{ if } y * f(x) \geq 1 \\ 1 - y * f(x), \text{ else} \end{cases} \qquad \text{...eq (2)}$$

Furthermore, for multiclass problem the SVM can be trained to differentiate a class from the rest of the classes. The support vector machine can fine tune the value of different kernel in other to efficiently distinguish each class (Awad & Khanna, 2015).

Mathematical definition of an SVM multiclass problem with Mc_i classes, with and input vector *x*,

$$\sum_{i=1}^{M} P(c_i | x) = 1 \qquad \text{.................. eq (3)}$$

The probability of correctly classifying class =

$$P_c = \sum_{i=1}^{M} P(x \in R_i, c_i) = \sum_{i=1}^{M} P(c_i) \int_{R_i} p(x|c_i)\, dx,$$

$$\text{... eq (4)}$$

R*i* represent the region of the *N* feature space, by definition region R is represented as

$$P_c = \sum_{i=1}^{M} \int_{R_i} P(x|c_i) p(x)\, dx \geq \frac{1}{M} \sum_{i=1}^{M} \int_{R_i} p(x)\, dx,$$

$$\Rightarrow P_c \geq \frac{1}{M}; \qquad \text{---eq (5)}$$

Finally, the probability of the multiclassification error =

$$P_e = 1 - P_c \le 1 - \frac{1}{M} = \frac{M-1}{M}.$$ …. eq (6)

C. Grid-Search-CV

The Grid search CV is refered to as an exhaustive search based mechanism based on a predefined subset of hyper-parameter space (Sulistiana & Muslim, 2020). The study will be adopting this approach to fine tune the support vector machine hyper parameters

ALGORITHM: Grid Search for parameter C on SVM

Initialize list of C candidates
FOR every c in list of C candidates
    Train SVM with c on TrainingSet
    Evaluate SVM classification on ValidationSet
    IF accuracy > MaxAccuracy
        THEN save MaxC = c
    ENDIF
ENDFOR
RETURN MaxC

MODEL TRAINING

The section diligently illustrates how the proposed model is developed ranging from data collection to model training, evaluation and result discussion.

1. Tools used

Jupyter lab is an Integrated development environment tools for data science. It provides the convenience and easy installation of necessary tools or dependency to perform operations such as; data importation, preprocessing, training and development of machine learning models. Jupyter lab version 3.4.3 is considered and python version 3.9.0 is selected as the choice of programming language.

2. Data collection

The dataset available for developing the proposed model is downloaded from Kaggle repository. Kaggle is an online data science repository providing vast amount of dataset for machine and deep learning models (Casper et al., 2020). The HVR dataset downloaded contains 41033 data sample, 35 independent variables with 1 dependent variable.

3. Data preprocessing

Immediately after data importation into the jupyter lab environment, its essential to explore the dataset in other to get meaningful insight about the HVR data samples. Considering this study, the data preprocessing stage includes encoding the independent variable into numerical forms and standardization of data value using min max scaler techniques, this ensures that the datapoint are uniformly distributed. Finally, the dataset is split into training (model development) and testing (model evaluation) set.

DATA PREPROCESSING

```
[ ]
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import LabelEncoder
```

```
lb = LabelEncoder()
dataset['condition'] = lb.fit_transform(dataset['condition'])
dataset.condition
```

```
0        1
1        2
2        1
3        1
```

Figure 2. Data Preprocessing.

Data Standadization

```
min_max = MinMaxScaler()
# train_x = min_max.fit_transform(train_x)
# test_x = min_max.fit_transform(test_x)

for c in dataset.columns:
    dataset[c] = min_max.fit_transform(dataset[[c]])

dataset.head()
```

	MEAN_RR	MEDIAN_RR	SDRR	RMSSD	SDSD	SDI
0	0.225313	0.184864	0.088443	0.324583	0.324578	
1	0.382444	0.287996	0.058163	0.654295	0.654295	
2	0.530982	0.395639	0.196938	0.751429	0.751430	
3	0.358287	0.286304	0.168890	0.296569	0.296546	
4	0.270276	0.203066	0.217693	0.371941	0.371880	

5 rows × 34 columns

Figure 3. Data standardization

4. Model Training

Figure 2 illustrated how the dataset independent variable is converted into numerical format and figure 3 shows how standardization is carried out using the min max scaler method.

RESULT AND DISCUSSION

The Grid-Search-CV is adopted in training the university stress prediction model using different parameter tuning or configuration, based on the figure 4, the support vector machine is trained with parameter configuration of ‘gamma = 10’, ‘C = 1, 10 and 20 as value’ and ‘kernel = rbf or linear’. However, the resulted cross fold validation scores and mean test scores after training are visualized in a tabular form, the essential columns and best parameter configuration is specified in figure 5

```
1 comparison_result.iloc[4]
```

```
mean_fit_time                          0.8077
std_fit_time                         0.008853
mean_score_time                      0.356274
std_score_time                       0.003528
param_C                                    20
param_kernel                              rbf
params               {'C': 20, 'kernel': 'rbf'}
split0_test_score                    0.997076
split1_test_score                    0.995249
split2_test_score                    0.994881
mean_test_score                      0.995735
std_test_score                        0.00096
rank_test_score                             1
Name: 4, dtype: object
```

Figure 4. SVM training using Grid-Search-CV

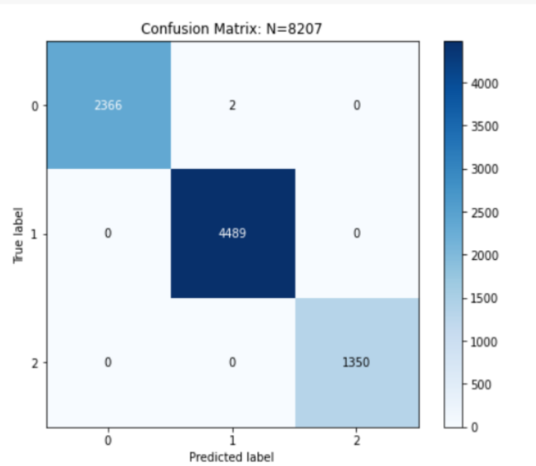

Figure 5. Evaluating the SVM Grid-Search-CV Results

Based on the result analysis on figure 5, the Grid-Search-CV reveals that the row 4 combination of parameter yields the best result, by using the 20 for param_c, and ‘rbf’ for param_kernel a mean test score of 0.9957 % accuracy is gotten. And this will be used in evaluating the stress prediction model.

	precision	recall	f1-score	support
0	98.50	99.00	1.00	2368
1	99.50	98.00	1.00	4489
2	99.51	98.00	1.00	1350
accuracy			99.99	8207
macro avg	1.00	1.00	1.00	8207
weighted avg	1.00	1.00	1.00	8207

Figure 6. Classification Report

```
[ ] 1 comparison_result[['param_C', 'param_kernel', 'mear
```

	param_C	param_kernel	mean_test_score
0	1	rbf	0.986109
1	1	linear	0.635677
2	10	rbf	0.995613
3	10	linear	0.640307
4	20	rbf	0.995735
5	20	linear	0.639576

Figure 7. Confusion Matrix

Figure 7 shows the correctly predicted datapoint against the datapoint that are miss classified. The diagonal value in the matrix indicate the numbers of correctly classified datapoint, while the other data in the confusion matrix denote the misclassified data.

Furthermore, (Jin et al., 2021) carried out quantitate comparison between existing work on stress prediction using the Heart Rate Variability (HRV) measure. In this study the existing models, (Jin et al., 2021) model and the proposed model will be compared. Using the table below

Table 1. Result Comparison Analysis

S/N	Approach	Accuracy (%)
1	Random Forest	93%
2	SVM	94 %
3	MO-SVR	95.8%
4	DCNN	95.2%
5	MOR-DCNN (Jin et al., 2021)	98.2%
6	Proposed Model	99.9 %

(Jin et al., 2021)

Based on the accuracy comparisons of the existing models and the proposed model, its proofed that the proposed model with (99.9 %) accuracy perform better than the existing approach.

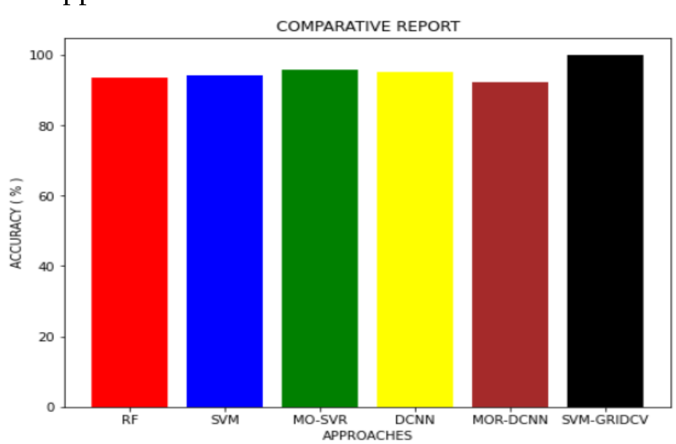

Figure 8. Comparison of existing model and proposed model

CONCLUSION AND RECOMMENDATION

This study provides an efficient and more accurate prediction model for detecting stress level using the Heart Rate Variability of university students. The proposed approach has proven better in comparison to the existing study that is evaluated. However, the developed model can be adopted by other sectors apart from the educational sector, sectors such as the financial, agricultural and medical sectors to predict their workers' stress levels.

REFERENCE

Awad, M., & Khanna, R. (2015). Efficient learning machines: Theories, concepts, and applications for engineers and system designers. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, *January*, 1–248. https://doi.org/10.1007/978-1-4302-5990-9

Barker, E. T., Howard, A. L., Villemaire-Krajden, R., & Galambos, N. L. (2018). The Rise and Fall of Depressive Symptoms and Academic Stress in Two Samples of University Students. *Journal of Youth and Adolescence*, *47*(6), 1252–1266. https://doi.org/10.1007/s10964-018-0822-9

Casper, B., Bojer, S., & Meldgaard, J. P. (2020). *Learnings from Kaggle ' s Forecasting Competitions*.

Chrousos, G. P. (2009). Stress and disorders of the stress system. *Nature Reviews Endocrinology*, *5*(7), 374–381. https://doi.org/10.1038/nrendo.2009.106

Jin, N., Zhang, X., Hou, Z., Sanz-Prieto, I., & Mohammed, B. S. (2021). IoT based psychological and physical stress evaluation in sportsmen using heart rate variability. *Aggression and Violent Behavior*, *September 2020*. https://doi.org/10.1016/j.avb.2021.101587

Jing, L. (2021) Optimal Modeling of College individuals' Mental Health Based on Brain-Computer Interface and Imaging Sensing. *Proceedings of the Fifth*

International Conference on Intelligent Computing and Control Systems (ICICCS 2021,) IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2

Nazari, N., & Far, D. M. (2019). The relationship between teaching skills, academic emotion, academic stress and mindset in university student academic achievement prediction: A PLS-SEM approach. *Journal of Intellectual Disability - Diagnosis and Treatment*, *7*(3), 119–133. https://doi.org/10.6000/2292-2598.2019.07.03.9

Pankajavalli, P. B., Karthick, G. S., & Sakthivel, R. (2021). An Efficient Machine Learning Framework for Stress Prediction via Sensor Integrated Keyboard Data. *IEEE Access*, *9*, 95023–95035. https://doi.org/10.1109/ACCESS.2021.3094334

Park, K., Jung Kim, M., Kim, J., Cheon Kwon, O., Yoon, D., & Kim, H. S. (2020). Requirements and Design of Mental Health System for Stress Management of Knowledge Workers. *International Conference on ICT Convergence*, *2020-Octob*, 1829–1832. https://doi.org/10.1109/ICTC49870.2020.9289464

Reddy, U. S., Thota, A. V., & Dharun, A. (2018). Machine Learning Techniques for Stress Prediction in Working Employees. *2018 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2018*, *May*. https://doi.org/10.1109/ICCIC.2018.8782395

Shen, Y. (2021). Analysis on Mental Health Education Strategy of Higher Vocational College individuals Based on Information Technology. *2021 2nd International Conference on Education, Knowledge and Information Management (ICEKIM), 978-1-7281-6834-0/21/$31.00 ©2021 IEEE DOI 10.1109/ICEKIM52309.2021.00151*

Shitole, A. (2021). *Twitter Sentiment Analysis Using Supervised Machine Learning*. *March*. https://doi.org/10.1007/978-981-15-9509-7

Sulistiana, & Muslim, M. A. (2020). Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy. *Journal of Soft Computing Exploration*, *1*(1), 8–15.

Zamkah, A., Hui, T., Andrews, S., Dey, N., Shi, F., & Sherratt, R. S. (2020). Identification of suitable biomarkers for stress and emotion detection for future personal affective wearable sensors. *Biosensors*, *10*(4). https://doi.org/10.3390/bios10040040