MIER Journal of Educational Studies Trends & Practices
November 2022, Vol. 12, No. 2, pp. 301 - 319
DOI: 10.52634/mier/2022/v12/i2/2280

An Ensemble Machine Learning Model for Automatic Prediction of Perceived Personal Well-Being of Indian University Students During COVID-19 Lockdown

Kavita Pabreja, Shubham Arya and Parichay Madnani

COVID-19 has impacted personal well-being globally in a disruptive manner. Frequent lockdowns have slowed down dramatically the economy of every nation. There is a fear of future insecurity cropping up in the minds of the people. The paper aims to restructure the popular Personal Well-being Index (PWI) according to the relevant indicators that impacted students' life in India during the second wave of COVID-19. The students at Delhi state university participated in the research. The researchers use various machine learning algorithms such as Lasso Regressor (LR), Support Vector Regressor (SVR), and Decision Tree Regressor (DTR) to predict the perceived PWI. The R-squared value for LR, SVR and DTR are 0.9103, 0.9159 and 0.5339. Mean squared errors are 0.0034, 0.0035 and 0.0105 respectively. The five most influential determinants of perceived PWI were extracted. An ensemble model of the three mentioned base learners was designed to remove the overfitting and underfitting problems. The algorithm has demonstrated impressive performance, with an R-squared value of 0.9839 and MSE of 0.0014. A GUI-based prediction model was implemented in Python that triggered the ensemble model at the back end to predict PWI based on five questions only, along with recommendations for the respondents.

KEYWORDS: Perceived Personal WellBeing, Support Vector Regressor, Decision Tree Regressor, Lasso Regressor, Ensemble Model, COVID19

Kavita Pabreja ✉
Associate Professor, Department of Computer Applications, Maharaja Surajmal Institute, Delhi, India.
Email: kavitapabreja1@gmail.com. ORCID: https://orcid.org/0000-0001-9856-0900

Shubham Arya
Associate Analyst, Deloitte, India.
Email: shubhamarya406@gmail.com. ORCID: https://orcid.org/0000-0002-7427-7024

Parichay Madnani
Scholar, Department of Computer Applications, Vellore Institute of Technology, India.
Email: parichayman2@gmail.com. ORCID: https://orcid.org/000-0003-1355-5622

Introduction

India accounts for the highest percentage of the youth population globally which is a 34.33% share of youth in the world's population, as per a report by the Ministry of Statistics and Programme Implementation, Government of India (2017). According to the Office of the Registrar General and Census Commissioner, India; there are 189 million people in the age group years 15 to 21 which constitutes 18.4% of the total Indian population. India's government spends a meagre share of its GDP on public healthcare which is just 1.8% of the GDP and India's rank is 175^{th} amongst 189 countries in terms of prioritizing healthcare, as mentioned by PRS Legislative Research report (2021). Indian union budget 2021-22 has set aside only 0.81% of the "Ministry of Health and Family Welfare and Social Justice and Empowerment budget" for Mental Health. It is documented by UNICEF (2021) that there is a social stigma surrounding mental health issues in Indian children and youth, and they feel reticent to seek support for any psychological or mental stress-related issues. Moreover, the excruciating and alarming fact is that 14% of 15-24 old Indians are reported to be depressed.

The impact of losses in the Indian economy on account of mental health conditions during 2012-2030 is estimated to be US$1.03 trillion which needs immediate attention and thoughtful action. Moreover, COVID-19 has affected India adversely, with multiple lockdowns, job losses, and school closures; the school and college students feel insecure about their future, and this has deteriorated their mental well-being. There has been an increase in the number of deaths due to suicide which is 21% more in 2020 as compared to 2019.

All these facts and figures urgently require attention, and this study is an attempt to develop an artificial intelligence-based solution for interpreting and predicting the subjective well-being (SWB) of students. SWB measures quality of life, based on questions related to satisfaction with life (Cummins, 2013, 2018), and the measure used all across the world to reveal this multidimensional aspect of quality of life is the Personal Well-being Index (PWI) (Van Beuningen & De Jonge, 2011).

Often, SWB is considered the same as happiness which is not the case. Happiness refers to moment-to-moment feeling and does not include the broader concept of satisfaction with life, having a feeling of future security, cohesive personal relationships, physical health and safety, positive feelings towards achievements in life, and connectedness with society (Cummins, 2013). All these dimensions contribute toward calculating PWI. Well-being becomes even more important during COVID-19 pandemic times that have affected severely every human being across the globe.

Digital technologies like artificial intelligence, big data analytics, and machine learning have been deployed extensively in health care systems on a global scale (Badawi et al., 2014; Jones et al., 2012) . Mental health monitoring systems based on ubiquitous sensing technologies and machine learning have been developed (Flesia et al., 2020; Garcia-Ceja et al., 2018) for monitoring stress. Detection of stress among university students has been done (Ahuja & Banga, 2019; Ghaderi et al., 2015) using machine learning techniques. For early diagnosis of diseases also, big data analytics technologies have been deployed (Kaur et al., 2018). The prediction of opioid overdose risk amongst medical policy beneficiaries using data mining algorithms (Lo-Ciganic et al., 2019) has been proposed.

This study addresses a social aspect as understanding the personal well-being of an individual helps the person in improvising his /her life. Our artificial intelligence-based predictive module can sense the well-being with a minimal set of questions being extracted with the help of machine learning techniques and provides recommendations to enhance the quality of life so as to live life joyfully without getting into any kind of mental health issues like depression, stress, and suicide. The motivation for the application of the Machine Learning (ML) technique and then inventing a hybrid model lies in the global acceptability of digital technologies in healthcare sectors as mentioned in the Introduction section.

The paper is structured as follows. The next section covers the Related background work, followed by the methodology adapted which includes Data Pre-processing and Exploratory data analysis. The next section discusses the results of the application of various Machine Learning algorithms and extraction of the most influential features that determine perceived PWI. This section is followed by the Development of the Ensemble model of Machine learning algorithms and the Development of a Graphical User Interface for instant prediction of PWI. Finally, Results and discussions, Conclusions, Limitations, Managerial implications, and future scope are written.

Review of Literature

Research publications have been studied and analysed from two different dimensions. The first is to understand various types of well-being indices in practice globally so that a new framework based on psychological traits, may be developed in the Indian context and the second is to exploit the usage of advanced machine learning techniques for the prediction of this formulated well-being index.

WHO links health directly with well-being and there are two dimensions to it viz. objective and subjective. Objective well-being focuses on food, education, and safety and uses measures like life expectancy and mortality rates whereas the basis of subjective well-being (SWB) is parameters like life satisfaction, positive emotions, future security, relationship, etc.

SWB can be calculated based on questions related to satisfaction feeling towards life as a whole (Cummins, 2013). Improving well-being is an objective of numerous countries' governments and a large number of standardized well-being scales have been designed and adapted to measure the SWB. Measures like WHO-5 (Hermanns, 2007), MHC-SF (Keyes, 2009) are based on positively worded questionnaires and are not designed to assess any mental illness. A shorter version of WEMWBS named SWEMWB (Haver et al., 2015) focuses on the psychological functioning of individuals and is found to be reliable and valid. Measures like PANAS (Francis, 2014) , Oxford Happiness Questionnaire (Watson & Clark, 1994) , The Everyday Feeling Questionnaire (Uher & Goodman, 2010) and Office of National Statistics Wellbeing Measure Beaumont and Lofts (2013) have both negative and positive worded questions. The International Wellbeing Group (Cummins, 2013) has developed the Personal Wellbeing Index (PWI) to measure SWB which is a subjective dimension of quality of life. PWI is constructed to measure seven domains of satisfaction, viz. Standard of Living, Personal Health, Achieving in Life, Personal Relationships, Personal Safety, Community-Connectedness, and Future Security Cummins (2013). The score on a scale of 0 to 10 is averaged to calculate a measure of SWB. PWI is well accepted in many countries viz. Australia, Canada, China, Greece, Romania, Italy, and Spain (Cummins, 2013; Misajon et al., 2016) . The normative range of PWI score is 70 to 80 for western countries and the average PWI score is 72.98. For Australia, the PWI normative range is a little higher and it is from 74.2 to 76.9 (Team, 2018). For Romania (Baltatescu & Cummins, 2006), the PWI score stands at 69.7 for the general population of 18+ age. Authors (Baltatescu & Cummins, 2006) have found that males usually score higher on SWB in Romania, also with increasing age, PWI is stated to be decreasing. It is documented that the PWI score in China and Asia is approximately 10 points lower than in western countries and the range is between 60 and 70 (Chen & Davey, 2008; Cummins, 2013) . Since PWI has demonstrated good validity and reliability in both western and non-western countries, it has been considered a suitable tool in the Indian context as well (Mcintyre et al., 2020) . In this study on the Indian population during pre-COVID times the authors calculated the PWI score as 74.43. A new well-being measure has been developed by ICICI Lombard for the Indian population, and the score is quantified at 58.3 only during the year 2018 (Mcintyre et al., 2020) .

Digital technologies and artificial intelligence have been used extensively in the field of mental health and well-being. Mental stress detection in university students has been done by Ahuja and Banga (2019) using the ML algorithms like Support Vector Machine, Random Forest, and Linear Regression. Stress has also been predicted in students at a Delhi state university by Pabreja et al.

(2021). The authors have emphasized the importance of stress prediction by linking its early detection and avoiding suicide in the case of youngsters in the age group of 15 to 29 years (Ahuja & Banga, 2019) . Predictive analytical ML models have been used for the identification of short-term and long-term risks for suicide (Walsh et al., 2017) .

After understanding many well-being measures validated across the world and reviewing the application of Artificial Intelligence and Machine Learning techniques globally, it has been decided to assess the personal well-being of Indian students during the COVID-19 lockdown period (May, 2021) in India. With the widespread COVID pandemic and so many variants rolling in recurrently, India specifically Delhi has witnessed many lockdowns in 2020 and 2021. It is very much relevant to assess SWB in these periods of uncertainty.

Objectives of the Study

Numerous research publications have been studied and to the best of our knowledge, it has been found that none of the papers have focused on the prediction of PWI of Indian students' populations using advanced digital technologies like ensemble machine learning that too during COVID-19 pandemic times when the overall quality of life was severely impacted in terms of future security, social connectedness, and physical and mental health. Hence, the following objectives have been framed to address this matter of grave concern.

1. To restructure the framework of necessary indicators of life quality to reflect multiple dimensions of well-being with a specific focus on aspects affecting students' life during the COVID-19 lockdown period of 2021 in India.

2. To apply and compare the performance of different machine learning algorithms for regressing the perceived PWI value.

3. To extract important features that determine PWI by application of feature selection techniques available with machine learning algorithms.

4. To develop a hybrid ensemble of machine learning algorithms to predict PWI with a minimal set of features and higher accuracy than individual machine learning algorithms so as to detect vulnerable students early.

5. To develop a Graphical User Interface that prompts users to answer just five extracted important determinant questions of PWI and predicts the PWI with very high accuracy.

Research Methodology

A critical analysis of various research works in the field of well-being and the application of digital techniques in psychology and the medical domain was done. PWI scale comprises seven domains of life satisfaction, to measure SWB. It is stated (Michaelson et al., 2012) that though PWI performs fairly well to assess SWB, key measures can be supplemented with additional well-being measures to explore the situation-specific well-being of individuals. Hence, under a few key indicators of PWI, additional questions have been added to observe the impact of the lockdown and the deadly situation that wreaked havoc during the COVID-19 pandemic-imposed lockdown period in New Delhi. The questions for each domain are mentioned in Table 1. Each question is a concise measure compliant to be used in the survey to understand the psychological traits of respondents to find perceived well-being.

Table 1

Domains and Corresponding Questions Used in The Questionnaire To Understand The Perceived Personal Well-Being.

Domains	Questions
Standard of Living	Has this one-year lockdown affected your standard of living?
Personal Health	Has this lockdown made you lethargic? Has this lockdown affected your physical health? Has Covid affected your diet? Has online education badly affected your eyesight? Have you gained weight during social isolation?
Achieving in Life	Has online learning deteriorated your ability to learn and achieve in life?
Personal Relationships	Has lockdown improved your personal relationships with your family and friends?
Personal Safety	Are you overall satisfied with your output in terms of professional, academic, and personal growth during this long lockdown period?
Community-Connectedness	Has COVID affected your ability to interact with others?
Future Security	Has COVID-19 affected you mentally and made you feel insecure? Has COVID-19 affected your future plans?

There was a complete lockdown in New Delhi from April 19, 2021, to May 31, 2021. The highest number of daily COVID positive cases were reported in India on May 7, 2021, which was 4,14,188 (East & Region, 2021). In New Delhi, the positivity rate was between 20 and 30 percent. This survey was conducted in the last week of May 2021 when the overall situation was uncertain, and people were fearful about their future and had witnessed the deaths of many of their near and dear ones. The study included many steps and a diagram showing the methodology adopted to achieve the objectives, is shown in Figure 1.

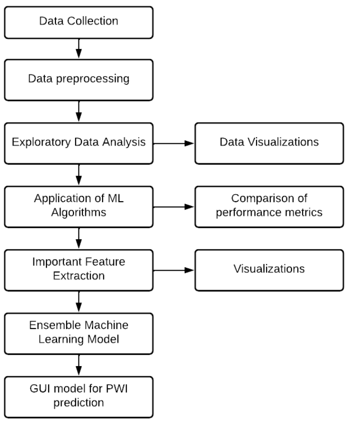

Figure 1. Framework of the Methodology Followed in the Study.

Sample of the Study

Non-probability convenience sampling was used for selecting respondents from a pool of undergraduate programs at a Delhi state university. Data was collected from 194 respondents in the age group of 17 to 21 years. They were reassured that the data will be kept confidential so that honest responses are gathered. To explore the variation of well-being among the target group according to gender, basic demographic data was also collected in addition to well-being questions. The sample comprises 104 males (58% of total respondents) and 76 Females (42% of total respondents). The questionnaire was a self-administered one and hence without any intervention, it was completed online using google forms. The respondents were asked to select either the “Yes” or “No” option for each question. The data pre-processing, exploratory data analysis, development of machine learning algorithm, a hybrid model of ensembles, and development of Graphical user interface predictor model- all have been implemented using Python language.

Data Cleaning And Data Transformation

As per the recommendation of the PWI manual (Cummins, 2013) , the respondents whose responses were all Yes or all No meant lack of understanding or invalid data, consequently, 14 records were deleted from the dataset. The final data set had 180 responses, 104 males and 76 females. The positive responses to questions were transformed to 1 and negative to 0. Accordingly, PWI was calculated by totalling and averaging responses to all 12 questions. The PWI hence calculated had the values between 0 and 1 which were divided into three ranges 0 to 0.33 as Low: 0.34 to 0.66 as medium and 0.67 to 1.0 as High level of well-being.

Exploratory Data Analysis

The perceived well-being index has been calculated and since the time of data collection was full of insecurity that affected the overall quality of life, the results also reflect the same. The PWI score was 41.58 for male respondents and for females, it was 37.39. Though as per ICICI Lombard, the PWI for Indians is generally at 58.3 the results revealed by this study mirrored the fearfulness and uncertainty that prevailed during the lockdown period in India. Further, the categorization of males and females according to the category of PWI has been done and shown in Figure 2. Accordingly, it is found that more percentage of male respondents (19.23%) demonstrated a high PWI as compared to females (11.84%). More than 50 percent of females reported low PWI score.

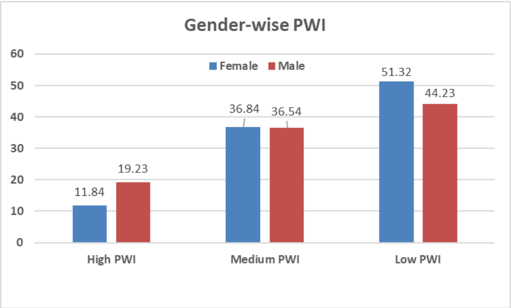

Figure 2. PWI Category-Wise Distribution Of Respondents As Per Gender.

To find out the correlation between all predictors and PWI, the correlation matrix was generated. It was found that all predictors have a correlation value of less than 0.5 between them and hence none of the questions looked redundant primarily. The questions that belonged to domains "Standard of Living" and "Personal Health" showed a correlation value of more than 0.5 with the PWI. The correlation matrix is shown in Chart 1 and the heatmap generated using Python language is shown in Figure 3.

Question	Lockdown Standard of Living	Made Lethargic	Affected Health	Deteriorated Learning Ability	Affected Diet	Affected Eyesight	Made Future Insecure	Gained Weight	Improved Personal Relationship	Affected Future Plans	Affected Interaction Ability	Overall Satisfaction	Personal_WellBeing_index
Lockdown Standard of Living	1.00	0.24	0.17	0.22	0.16	0.16	0.11	0.11	0.19	0.10	0.04	0.42	0.55
Made Lethargic	0.24	1.00	0.16	0.00	0.17	0.23	0.19	0.09	0.02	0.11	0.07	0.24	0.46
Affected Health	0.17	0.16	1.00	-0.02	0.25	0.15	0.10	0.27	0.16	0.02	0.18	0.23	0.50
Deteriorated Learning Ability	0.22	0.00	-0.02	1.00	0.08	0.15	0.10	0.06	0.03	0.11	0.23	0.23	0.41
Affected Diet	0.16	0.17	0.25	0.08	1.00	0.16	0.21	0.10	0.13	0.14	0.07	0.23	0.51
Affected Eyesight	0.16	0.23	0.15	0.15	0.16	1.00	0.20	0.26	0.08	0.19	0.16	0.24	0.55
Made Future Insecure	0.11	0.19	0.10	0.10	0.21	0.20	1.00	0.09	0.01	-0.03	-0.01	0.11	0.37
Gained Weight	0.11	0.09	0.27	0.06	0.10	0.26	0.09	1.00	0.03	0.13	-0.09	0.08	0.41
Improved Personal Relationship	0.19	0.02	0.16	0.03	0.13	0.08	0.01	0.03	1.00	0.01	-0.04	0.17	0.33
Affected Future Plans	0.10	0.11	0.02	0.11	0.14	0.19	-0.03	0.13	0.01	1.00	0.14	0.23	0.38
Affected Interaction Ability	0.04	0.07	0.18	0.23	0.07	0.16	-0.01	-0.09	-0.04	0.14	1.00	0.07	0.34
Overall Satisfaction	0.42	0.24	0.23	0.23	0.23	0.24	0.11	0.08	0.17	0.23	0.07	1.00	0.60
Personal_WellBeing_index	0.55	0.46	0.50	0.41	0.51	0.55	0.37	0.41	0.33	0.38	0.34	0.60	1.00

Chart 1: **Correlation Matrix Between All Predictors and Predictand**

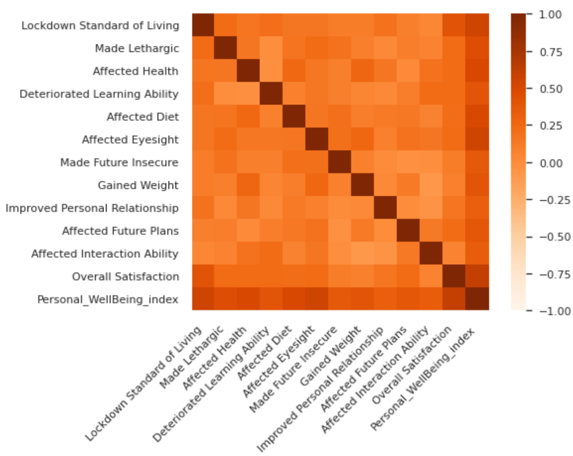

Figure 3. Heatmap Showing Correlation Values Between All Predictors and Predictand.

Application of Independent Machine Learning Algorithms

An effort has been made to predict perceived well-being by leveraging many independent weak learners. The performance of any machine learning model depends upon its bias and variance. High bias and low variance algorithm viz. Lasso Regressor; and high variance and low bias algorithms viz. Decision Tree Regressor and Support Vector Regressor have been selected as base models and their performance for the prediction of PWI has been recorded. The responses to the questionnaire served as input features (predictors) and the calculated perceived well-being score was considered as the output variable (predictand) for the algorithm. All these three algorithms are supervised machine learning algorithms. The dataset was split into two parts viz. 70% for training the algorithm and 30% was set aside for testing the same.

Lasso regressor was chosen as it reduces the problem of overfitting by including fewer variables as predictors that result in high prediction accuracy (Bonaccorso, 2017; Raschka & Mirjalili, 2019) . Lasso regressor is an extension of Linear regressor which is a weak learner and with an increase in bias, this algorithm results in a decrease in variance and hence reduces overfitting.

Decision tree regressor is one of the most commonly used basic machine learning algorithms. The benefit of selecting this model is that its performance is not deteriorated with the non-linearity in the data, however, to reduce mean squared error, it learns even from the noise present in the input data which results in overfitting (Bonaccorso, 2017; Raschka & Mirjalili, 2019) .

Support Vector Regressor is a complex algorithm that suffers from high variance due to its high sensitivity to input data. This model tends to learn from the noise present in the training data and hence results in overfitting (Bonaccorso, 2017; Raschka & Mirjalili, 2019) . Figure 4 shows the residual plots and scatter plots for actual vs. predicted values of PWI generated for all these three learners. It is evident from the plots that Lasso and Support Vector Regressor has performed quite well and better than decision tree regressor.

The quantitative performance metrics have also been calculated. The coefficient of determination (R-squared) is more than 0.9 for Lasso and Support Vector Regressors and is quite low i.e., 0.53 for decision tree regressor. Similarly, the mean squared error between actual and predicted values are lesser in the case of Lasso and Support Vector Regressors as is evident from Table 2.

Table 2

The Performance Metrics for Lasso, Support Vector and Decision Tree Regressors.

Performance Metric	Base Learner		
	Lasso Regressor	Support Vector Regressor	Decision Tree Regressor
R-SQUARED	0.9103	0.9159	0.5339
Mean Squared Error	0.0034	0.0035	0.0105

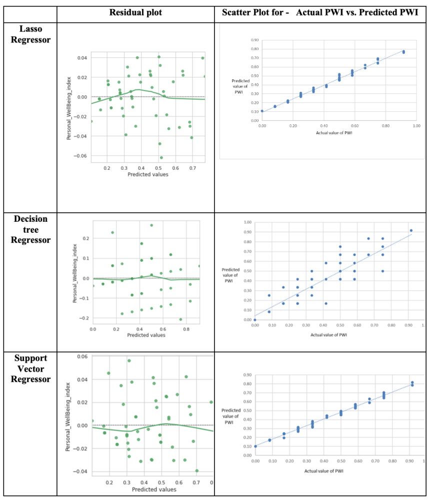

Figure 4. Plots to Depict Performance Metrics of Lasso Regressor, Decision Tree Regressor.

Discovery of Ranking of Influential Features that Determine Perceived PWI

The most contributing factors on which the output field (perceived personal wellbeing index) depends, have been found by finding weights of the input features by applying the feature_importances function of the machine learning algorithm. Hence the ranking of features has been obtained as per their importance and influence. The motivation behind this step is to extract only the five most influential features to reduce overfitting, as redundant and less influential features can be removed. This resulted in very few questions to ask respondents to assess their personal well-being index. Figure 5 shows feature importance in the form of a bar chart and the description of each feature along with its weight is given below.

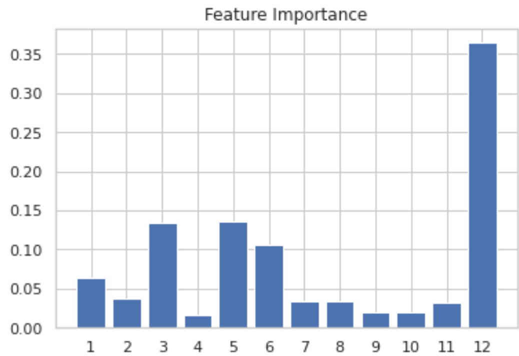

Figure 5. Bar Plot Showing the Importance of Features.

1. Lockdown Standard of Living = 0 06388998928957515
2. Made Lethargic = 0 037193847835135790
3. Affected Health = 0 1337754519690603
4. Deteriorated Learning Ability = 0 016331088603470250
5. Affected Diet = 0 13674187880938785
6. Affected Eyesight = 0 10589023359570725
7. Made Future Insecure = 0 034836663526551306
8. Gained Weight = 0 03458635548067141
9. Improved Personal Relationship = 0 019970957573124178
10. Affected Future Plans = 0 01955758163368047
11. Affected Interaction Ability = 0 032679776617268876
12. Overall Satisfaction = 0 36454617506636705

It is evident that the features viz. 1,3,5,6, and 12 are the most influential determinants of perceived PWI during the COVID-19 lockdown period in India. This step has essentially given us an in-depth understanding that the Indian students were severely affected in terms of their "Standard of Living", "Health", "Diet", "Eyesight" and "Overall Satisfaction".

Development of Ensemble Model of Machine Learning Algorithms Using Only the Important Five Features

Though the independent weak learners have produced an impressive R-squared value that validates the accuracy of these models, since all these are weak learners, they may not be robust for unseen new cases. Lasso Regressor is a high bias and low variance algorithm whereas decision tree regressor and Support vector regressor are high variance and low bias algorithms. So, to derive a trade-off between bias and variance an ensemble model has been designed whose building blocks are the mentioned three weak learners viz. decision tree, lasso, and support vector regressors. This model is developed to work only on the five important features that have been extracted.

This ensemble model relies on a voting regressor that fits these three weak learners on the dataset and produces the average of individual predictions to generate the final prediction (Zhang & Ma, 2012).

The performance of the hybrid model improved enormously as the bias-variance trade-off was handled excellently by combining the three weak learners in the ensemble and hence a strong learner was designed named as DeSuLa algorithm. Figure 6 shows the residual plots and scatter plots for actual vs. predicted values of PWI generated for DeSuLa hybrid model. The R-squared value and mean squared error for the same are mentioned in Table 3.

Table 3

The Performance Metrics for Ensemble Machine Learning Model-DeSuLa.

Performance Metric	Ensemble Model
R-SQUARED	0.9839
Mean Squared Error	0.0014

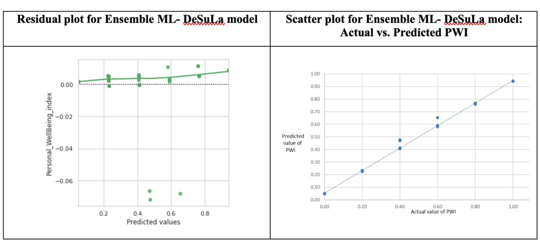

Figure 6. Plots to Depict Performance Metrics of Ensemble Machine Learning Model-DeSuLa.

Development of Graphical User Interface for Instant Prediction of PWI

The performance metrics of the hybrid machine learning model were impressive, and it was selected for integration with a front end based on Graphical User Interface (GUI). The motivation behind the development of this GUI model was to ask a minimal number of questions, that represented five important features extracted earlier in this study. Through an Application Programming Interface, the responses were passed to the back-end machine learning algorithm. This model predicted the perceived PWI category depending on the response to the questions.

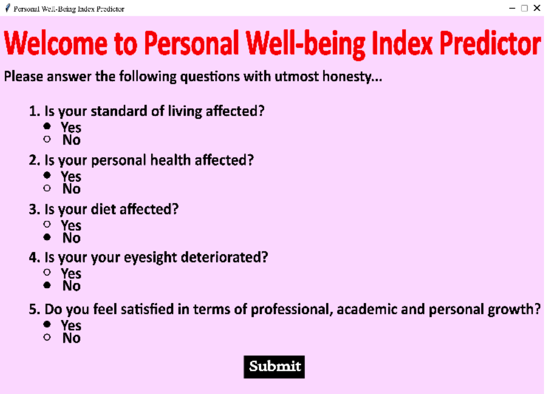

Figure 7. First Frame of GUI Based Perceived PWI Predictor.

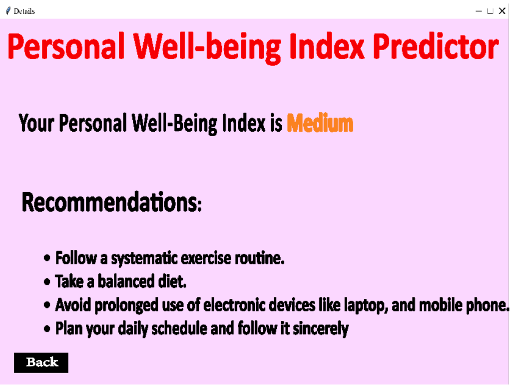

Figure 8. Window Displaying Stress Score and Recommendation for Participants.

Tkinter module of the python programming language was used for the development of GUI. The screenshots of the first window that receives responses to questions and the next window that displays the PWI category with customized recommendations have been shown in Figure 7 and Figure 8 respectively.

Results and Discussions

After performing a detailed study of research papers from the year 2006 to 2021, it was observed that none of the papers have a focus on the prediction of PWI of Indian students' populations and accordingly the objectives were framed. With this thorough study, all the decided objectives have been met. To quantify SWB, the PWI scale includes seven different life satisfaction domains. According to Michaelson et al. (2012) , although PWI does a decent job of assessing SWB, important measures can be complemented with additional well-being measures to investigate a person's situation-specific well-being. Therefore, extra questions have been added under a few important PWI indicators to track the effects of lockdown on students' life during the COVID-19 lockdown period of 2021 in India and the restructured framework has been mentioned in Table 1.

Different machine learning algorithms for regressing the perceived PWI value have been experimented with and their performance has been compared. It has been found that Lasso and Support Vector Regressors performed better than decision tree regressors both in terms of mean squared error and coefficient of determination (R-squared), as shown in Figure 4 and Table 2.

The five most important features that determine PWI have been extracted by the application of feature selection techniques available with machine learning algorithms. The motivation behind this step was to extract only the five most influential features to reduce overfitting, as redundant and less influential features can be removed. These are shown in Figure 5.

A hybrid ensemble of machine learning algorithms comprising three weak learners viz. decision tree, lasso, and support vector regressors has been developed to predict PWI with a minimal set of features and higher accuracy than individual machine learning algorithms to detect vulnerable students early. The performance of this novel DeSuLa model is explained in Figure 6 and Table 3.

Finally, to facilitate the calculation of PWI, a Graphical User Interface that prompts users to answer just five extracted important determinant questions has been developed in the Python tkinter module that predicts the PWI with very high accuracy, shown in Figure 7 and Figure 8 respectively.

Conclusions

The purpose of this piece of research is to reveal the perceived subjective well-being index of students at an Indian university during the lockdown period of the second wave of COVID-19 in India. The basic PWI questionnaire was supplemented with additional questions to explore more about the health and personal safety of the respondents. The questionnaire was administered online using a google form to get insight into the subjective well-being of the respondents. Various exploratory and explanatory data analytics have been applied and the findings are as follows. Perceived PWI for males is higher than for females. The majority of the respondents belong to the category "Medium PWI" which essentially evidences the insecurity feeling of the youth. The prediction of PWI was done by leveraging artificial intelligence-based machine learning algorithms based on all twelve questions that capture the essence of wellbeing. Since these algorithms viz. Decision tree, Support Vector Machine, and Lasso Regressor suffer from problems of overfitting or underfitting and may not perform well when the training data changes. Hence, it was decided to design a novel ensemble machine learning model that used the mentioned three models as base learners. This hybrid model, DeSuLa demonstrated very high accuracy for the prediction of PWI, with a Mean squared error of 0.0014 and an R-squared value equal to 0.9839. Moreover, the DeSuLa is designed and developed to work based on only five most important contributing factors on which PWI depends. The feature importance function of the machine learner was applied for extracting these five important predictors.

Following this, a GUI-based prediction model was developed that is based on a very short questionnaire comprising of five questions only. It predicted the perceived PWI by invoking the novel hybrid model, DeSuLa at the back end. Additionally, the prediction model also gave recommendations for improvement of subjective well-being depending upon the category of the PWI outcome.

Hence it is proved that the machine learning ensemble model has immense potential to become a powerful instrument for predicting subjective wellbeing.

Limitations

There is a wide difference in the cultural, social, and emotional status of people all across the globe. The psychological needs of the youth also vary accordingly. This study focuses on Indian youth who are enrolled in a regular undergraduate program of a technical stream and hence cannot be generalized to other countries and other academic profiles. A study on the data of those youth who are not in a comparable economic status group would reveal more insights. Future longitudinal studies should collect samples of a larger size and from different specialized streams of courses from multiple universities.

Managerial Implications And Future Scope

Various studies by Government agencies have declared that COVID may become endemic and hence such judgment on personal wellbeing using machine learning algorithms may help avoid serious psychological issues that may lead to mental illness. The main deliverable of this study is the extraction of five important contributing factors for calculating well-being. Hence a much smaller questionnaire can be designed and floated from time to time to assess the well-being of the students. This may also be offered as a feature in the mobile app of academic institutions and students' well-being may be tracked regularly. The responses to the questions may be integrated with sensor data from smartwatches and fitness bands to augment the accuracy level of detection. Further automatic integration from sentiment analysis of the posts and tweets on social media by the subject can also be done to get even more accurate insight into personal well-being. With all these strategies, vulnerable youth may be identified early and can be connected to a counsellor for professional advice. In addition to this, well-being workshops may be organized in educational institutions to improve lifestyle in case low PWI is observed. The developed ensemble model in this research may also be adapted by insurance companies to understand future mental health issues and accordingly plan the premium of their policies.

References

Ahuja, R., & Banga, A. (2019). Mental stress detection in university students using machine. *Procedia Computer Scienc*, *152*, 349-353. https://doi.org/10.1016/j.procs.2019.05.007

Badawi, O., Brennan, T., Celi, L. A., Feng, M., Ghassemi, M., Ippolito, A., … Zimolzak, A. (2014). Making Big Data Useful for Health Care: A Summary of the Inaugural MIT Critical Data Conference. *JMIR Medical Informatics*, *2*(2). https://doi.org/10.2196/medinform.3447

Baltatescu, S., & Cummins, R. (2006). Using the Personal Wellbeing Index to explore subjective wellbeing of high-school and college students in Romania. In *7th ISQOLS Conference:"Prospects for Quality of Life in the New Millennium" Grahamstown.*

Beaumont, J., & Lofts, H. (2013). *Measuring national well-being - health, 2013*. Office for National Statistics,.

Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

Chen, Z., & Davey, G. (2008). Normative life satisfaction in Chinese societies. *Social Indicators Research*, *89*(3), 557-564.

Cummins, R. A. (2013). *Personal wellbeing index-adult (pwi-a)*. Melbourne: Centre on Quality of Life, Deakin University.

Cummins, R. A. (2018). The golden triangle of happiness: Essential resources for a happy family. *International Journal of Child, Youth and Family Studies*, *9*(4), 12-39. https://doi.org/10.18357/ijcyfs94201818638

Flesia, L., Monaro, M., Mazza, C., Fietta, V., Colicino, E., Segatto, B., & Roma, P. (2020). Predicting perceived stress related to the covid-19 outbreak through stable psychological traits and machine learning models. *Journal of Clinical Medicine*, *9*(10), 1-17. https://doi.org/10.3390/jcm9103350

Francis, L. J. (2014). Oxford Happiness Questionnaire. *Encyclopaedia of Quality of Life and Well-Being Research*, 4548-4551. https://doi.org/10.1007/978-94-007-0753-5_4071

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Torresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, *51*, 1-26. https://doi.org/10.1016/j.pmcj.2018.09.003

Ghaderi, A., Frounchi, J., & Farnam, A. (2015). Machine learning-based signal processing using physiological signals for stress detection. In *22nd iranian conference on biomedical engineering (icbme)* (p. 93-98). IEEE.

Haver, A., Akerjordet, K., Caputi, P., Furunes, T., & Magee, C. (2015). Measuring mental well-being: A validation of the short Warwick–Edinburgh mental well-being scale in Norwegian and Swedish. *Scandinavian journal of public health*, *43*(7), 721-727.

Hermanns, N. (2007). WHO-5-Well-Being-Index. *Der Diabetologe*, *3*, 464-465. https://doi.org/10.1007/s11428-007-0179-2

Jones, S. S., Heaton, P. S., Rudin, R. S., & Schneider, E. C. (2012). Unraveling the IT Productivity Paradox - Lessons for Health Care. *New England Journal of Medicine*, *366*(24), 2243-2245. https://doi.org/10.1056/nejmp1204980

Kaur, P., Sharma, M., & Mittal, M. (2018). Big Data and Machine Learning Based Secure Healthcare Framework. *Procedia Computer Science*, *132*, 1049-1059. https://doi.org/10.1016/j.procs.2018.05.020

Keyes, C. L. M. (2009). *Atlanta: Brief description of the mental health continuum short form (MHC-SF).*

Lo-Ciganic, W. H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwoh, C. K., … Gellad, W. F. (2019). Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. *JAMA Network Open*, *2*(3).

https://doi.org/10.1001/jamanetworkopen.2019.0968

Mcintyre, E., Saliba, A., & Mckenzie, K. (2020). Subjective wellbeing in the Indian general population: a validation study of the Personal Wellbeing Index. *Quality of Life Research*, *29*(4), 1073-1081. https://doi.org/10.1007/s11136-019-02375-7

Michaelson, J., Mahony, S., & Schifferes, J. (2012). Measuring wellbeing: a guide for practitioners. London: New Economics Foundation.

Misajon, R. A., Pallant, J., & Bliuc, A. M. (2016). Rasch analysis of the Personal Wellbeing Index. *Quality of Life Research*, *25*(10), 2565-2569. https://doi.org/10.1007/s11136-016-1302-x

Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with python, scikit-learn, and tensorflow 2*. Packt Publishing Ltd.

Uher, R., & Goodman, R. (2010). The everyday feeling questionnaire: the structure and validation of a measure of general psychological well-being and distress. *Social psychiatry and psychiatric epidemiology*, *45*(3), 413-423.

UNICEF. (2021). *The State of the World's Children 2021; On My Mind: promoting, protecting and caring for children's mental health.* . Retrieved from https://www.unicef.org/india/press-releases/unicef-report-spotlights-mental-health-impact-covid-19-children-and-young-people

Van Beuningen, J., & De Jonge, T. (2011). *Personal wellbeing index: Construct validity for the netherlands*. Netherlands.

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, *5*(3), 457-469.

Watson, D., & Clark, L. A. (1994). Panas_X. *FEBS Journal*, *277*(18), 3622-3636.

Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer Science & Business Media.