Prediction of Stress Level on Indian Working Professionals Using Machine Learning

Kavita Pabreja, Maharaja Surajmal Institute, GGSIP University, India*
https://orcid.org/0000-0001-9856-0900
Anubhuti Singh, Deloitte, India
Rishabh Singh, Deloitte, India
Rishita Agnihotri, Deloitte, India
Shriam Kaushik, Prague University of Economics and Business, Czech Republic
Tanvi Malhotra, Deloitte, India

ABSTRACT

Stress levels amongst the Indian employees have increased due to a variety of factors and are a matter of great concern for organizations. This study is based on Indian working professionals and real data has been collected by using non-probability convenience sampling. A questionnaire was drafted based on 18 factors affecting the mental health of professionals. This study addresses two dimensions. The first is to identify the important influential features that trigger stress in the lives of working professionals, and the second is to predict the stress levels. Various supervised machine learning algorithms have been experimented with, and of all these algorithms, the support vector machine regressor model showed the best performance. The main contribution of the paper lies in the identification and ranking of 10 important stress triggering features that can guide organizations to develop policies to take care of their employees. The other deliverable is the development of a GUI-based stress prediction software based on machine learning techniques.

KEYWORDS
Data Transformation, Data Visualization, Explanatory Data Analysis, Exploratory Data Analysis, Feature Selection, Random Forest Regressor, Stress, Supervised Learning, Support Vector Machine

INTRODUCTION

Modern society is witnessing a continuous deterioration in the occupational health of working professionals in India. According to the latest reports for the year 2020, by ADP Research Institute, 70 percent of the Indian workforce is suffering from stress at their workplace which is a matter of grave concern. Survey reports by 1to1help, an Employee Assistance Program provider in India have...
found that there has been a large increase in the number of employees who are highly depressed or are vulnerable to suicides due to an increase in stress in their lives. Another surveyor, Optum affiliated with Nasscom, observed that almost half of the Indian employees suffered from some type of stress or mental health issue. They surveyed eight lacs employees in seventy major organizations, each with a minimum staff of four thousand and five hundred. The results were alarming and stated that the ratio of employees at high risk of taking their lives due to stress had increased from 4% to 8% in two years. They also found that family, money, and job were the most common stressors amongst the employees.

Parenting, pregnancy, change, caregiving, and social isolation were some more factors that led to stress. Some employees also fear for their job security in the restructuring of the organization even though they are better paid. It is specifically true for managers who are in a middle-level position and have the responsibility of kids as well as loans. Employees are also found to be struggling with stress in their personal lives due to broken relationships or marriages. Newer factors like social media also contribute to the intense peer pressure that expects people to have a certain lifestyle or be a certain image to qualify as successful. This undue and unmanaged stress leads to mental health issues viz. anxiety and depression, and damage to physical health by causing frequent headaches, elevated blood pressure, chest pain, upset stomach, and problems sleeping.

All these facts and figures provide enough motivation to find out the most influential factors that are the major contributors to stress amongst Indian working professionals and more importantly find the ranking of these factors so that one can work towards managing the most important stress contributors instead of all factors in general. In past, many authors have applied various statistical approaches to find the relation between different stress-causing factors and done critical analysis of the same.

Our proposed methodology is based on a machine-learning algorithm for the prediction of the stress level of employees in advance and hence mitigating the risks associated with undiagnosed stress levels that may trigger serious health complications. The most distinguishing feature of this piece of research is the extraction of factors in order of importance as contributing factors that cause stress in Indian working professionals. A Graphical User Interface has been developed by using Python tkinter for the prediction of stress levels. Thereafter, the strategies to reduce stress can be applied to make the lives of working professionals comfortable. This study utilizes some useful tips from Misra, 2020 that have elaborated the complete procedure from selection of the topic of research to writing conclusion section in a very systematic manner.

The paper comprises of following sections: Background, methodology for research (data preprocessing, exploratory data analysis, development of predictive model using machine learning algorithm, extraction of important stress-causing features using machine learning, GUI designing), conclusions, limitations of the study, and finally recommendations and future scope.

BACKGROUND

In past, many descriptive, and inferential approaches have been applied by researchers all over the world to understand the stress-causing factors from various dimensions. These factors viz. workplace environment, workload, job security, relationship with family members, have been studied by researchers with a statistical approach or data mining approach. Many such papers have been critically studied and analyzed to derive a systematic methodology for the prediction of stress based on minimum features. These papers are based on data of employees of different organizations like sugar mills, agricultural institutes, banks, and educational institutes, all sectors being from various countries like India, Malaysia, Vietnam, and China. Most of the authors have applied a statistical approach in their studies for finding a correlation between various stress-causing factors and regression analysis for understanding cause-effect relationships. Many of them have checked the reliability of their questionnaire by using Cronbach’s alpha coefficient for each stress-causing factor. A summary of different papers has been presented that built the basis of selection of influential factors and finally
minimizing the set of features so that a machine learning algorithm can make a prediction even before the actual serious impact of stress creeps in.

Prasad et al. (2015) have tried to find out the causes of stress in the employees of the International Agricultural Research Institute, Hyderabad, and its effects on their performances. They also wanted to find out the proficiency of the management techniques used at the workplace to reduce stress and evaluate how workplace stress affects them physiologically. The authors collected the information of 200 employees via a structured questionnaire and checked its authenticity with the help of Cronbach’s alpha coefficient for each variable. In this paper, the performances of the employees were checked based on absenteeism, poor work relations, reduced productivity, low self-esteem, and feeling of boredom at the place of work. They found that stress existed on a medium level amongst the employees and it negatively affected their performance. The authors used Regression analysis to find the cause-and-effect relationship between the factors and the performances of the employees. Karl Pearson’s correlation coefficient was also calculated to find the relation between the various stress-causing factors and the performance.

With a focus on stress-causing factors that fall under just one category, Pattnaik and Mishra (2016) looked into organizational factors as a reason for bringing in stress. They tried to find out people’s thoughts towards the relation between stress and competition in the workplace. The data of 72 employees were collected via a structured questionnaire, and telephonic, personal, and online interviews. The data were interpreted with the help of categorizing of data, tables, and pie charts and it was found that the employees do experience stress which has both negative and, in some cases, even positive effects. The two prime factors were excessive workload and organizational conflicts. The men felt frustrated and also felt a lack of concentration when stressed whereas the women felt frustrated and exhausted when stressed and both experienced lifestyle imbalance. Both female and male employees also faced physiological effects. The employees believed that stress is more in a highly paid job and they also felt that the facilities to manage stress at the workplace is not enough. It was observed that the female employees were more vocal in comparison to men in terms of sharing the problems related to stress.

A similar study has been conducted by Saravanan and Muthu Lakshmi (2019) to calculate the stress level of the employees of a Nationalized Bank in Trichy city. The sample size of the employees was 100 and a structured questionnaire was utilized to collect the data. The questionnaire was built with the help of a statistical package of Social Sciences. The answers of the questionnaire lay on a Likert scale of 1 (Strongly disagree) to 5 (Strongly disagree) for stress management. The data was then studied with the help of graphs and statistics and it was found that major stressors were interpersonal conflict and work pressure. They also concluded that more than fifty percent of employees can manage stress. They recommended that stress can be managed by meditation, yoga, exercise, different types of therapy, networking, and hobbies. Next, the teaching staff of an educational institute was studied by Rawal & Mhatre, 2018, to identify the factors that cause stress. The authors investigated the impact of stress on their productivity and performance and suggested recommendations to tackle stress. They conducted their study with the help of a questionnaire formed with help of primary and secondary data. It was found that the professors faced work-related stress along with family-related stress. They often had an overload of work and would neglect their homes to meet deadlines at the organization. But they feel that the measures taken by the organization to help manage stress do work well and they can comfortably manage a social life. This study too focussed on a general approach towards reducing stress.

Another study was done on a sample size of 100 participants that focussed on three major features viz. workload, job security, and shift duty concerning the performance of the employees (Vijayan, 2017). The authors studied the relation of the stressors and their impact on performance at the place of work by applying different statistical methods viz. T-test, Chi-square, correlation, and regression. It was found that the mentioned three features are positively correlated and have a huge impact on
employees’ performance. Of these features, the workload was found to be impacting the most on the performance of the employees.

Kuong and Yen (2016) conducted a study in Vietnam by collecting data from a sample size of 378 employees, using a structured questionnaire with a Likert scale of 1 to 5. They utilized SPSS software for exploratory factor analysis to enhance the dependability and authenticity of the measured variables and to find the relation between them. It was done with the help of Kaiser-Meyer-Olkin (KMO) and Bartlett’s test of sphericity. They also made use of Varimax Rotation for independent and dependent variables. Path analysis and multiple regression were used to find the relation between the dependent and independent variables. The study was mainly done on five major features viz. work overload, role ambiguity and role conflict, working relationship, career development, working environment, and their effect on the job stress and job performance was analyzed. They found that these factors affect job performance through job stress directly and indirectly. They concluded that the working environment was the major factor that affected job stress, followed by working relationships and career development whereas working relationships came out to be the major factor that affected job performance followed by working environment, job stress, and career development. These three features were related to job stress positively. They also found that work satisfaction played a big role in stress levels and performance. The authors were able to rank the stressors but could not do any prediction of stress levels.

In yet another work, the authors (Murali et al., 2017) aimed to find the effects of mainly four factors viz. workload, the pressure of time, lack of motivation, and role ambiguity on the performance of employees. The study was done on employees of a Malaysian organization and the sample size used was 136 respondents who were between the age of 20 to 60 years and had 30 years of experience in the field. Data was collected through questionnaires with a Likert scale of 1-5 whose normality was checked with the help of Kurtosis and Skewness and reliability was checked with the help of Cronbach Alpha. Descriptive analysis, correlation analysis, and regression analysis were applied to interpret the data. They found that the major factors that affected the performance of the employees were time pressure and role ambiguity. They also found that less support from the organization adds to the job stress and job dissatisfaction. There are little different findings by Sahoo, 2016 based on Health and safety executive model, and the main contributors for stress have been identified as job content, workload, and workplace, working hours, participation at the place of work, authority, career development, pay and position in the organization, interpersonal relationships, organizational culture, work-home interface. The effect of stress has been analyzed from two dimensions viz. on individuals and organizations. The effects of work stress on individuals were physical, emotional, intellectual, behavioral. Effects on the organization came out to be high staff turnover and recruitment costs, high absenteeism and presenteeism levels, reduced productivity levels, increased health and safety issues, litigation, reputational damage, increased training costs. The author has also listed the resources that could help manage, reduce and prevent stress to help the employees.

An interesting and unique study has been performed by Lawanont and Inoue (2018) that made use of multiple sensors, Raspberry Pi, and Arduino to collect data. Three force sensors were used to study the working behavior, the first sensor in the cushion of the seat, the second in the table mat, the third in the mouse pad. The amount of pressure applied can be related to the stress levels of a person. Temperature and humidity sensors, ambient light sensors, etc. were used to collect data about the working environment. The Arduino board was connected to all sensors and the pre-processing of the data was done via the Raspberry pi. A 10-question survey known as PSS (perceived stress scale) was used to calculate the stress levels of a person for a month. This study was performed on 7 participants. Unsupervised learning was utilized on the data frame to form clusters to find the relation between the level of stress and the working behavior of each feature. They made use of k-means clustering and hierarchical algorithm for the clustering of data. These clusters were then used to analyze and interpret the data. In both the algorithms, the environmental features were alike between the two clusters. The characteristics of the clusters also indicated information about working behavior that
was related to PSS score such as, where working behaviors in cluster A (higher stress) were probable to have fewer changes in active working whereas cluster B represents more active working. Most of the cluster’s members showed that it was also related to the lower PSS score. Results showed cluster B consisted of the working behavior that led to the lower stress level. With this approach too, the prediction model could not be developed.

A study by Reddy et al. (2018), is based on the survey conducted by OSMI (Open Sourcing Mental Illness) during the year 2017. It consisted of 750 responses from the employees. The dataset at first contained 68 features from both personal and work life. The authors used many machine learning techniques like Logistic Regression, KNN Classifier, Decision Trees, Random Forest Classifier, Boosting, and Bagging. To evaluate the models, they used Classification Accuracy, False Positive Rate, Precision, AUC (Area under Curve) Score, Cross-validation AUC. They found that boosting achieved the highest accuracy of 75.13, thus performing better than other models when accuracy, false-positive rate, and precision are taken into account. Whereas with cross-validation AUC, Random forest classifier achieved the higher score. They also found that women were more mentally stressed than men and the employees at a tech organization had a higher risk of developing stress and mental illness. To achieve a better performance response, an ensemble of the basic machine learning algorithms has been experimented with by Odusami et al. (2021). The authors developed a hybrid machine learning model, K_LoRD, that is based on four different algorithms viz. K Nearest Neighbor, Logistic Regression, Random Forest and Decision Tree. This hybrid model outperformed the ordinary individual models in terms of Accuracy and Receiver Operating Curve for the prediction of customer churn in the Telecommunication industry.

To provide relaxation to respondents, the data from 32 participants were collected via both, questionnaires after the usage of the Virtual environment to collect demographic, usability, and user experience; and by studying their behavior and reactions during and after the session (Thoondee & Oikonomou, 2017). The Unity 3D 5.3 was used to build the Virtual Reality application which was coded with C#. The environment was displayed using Oculus Rift DK2. The virtual environment consisted of features of a peaceful forest such as a Sunflower Garden, Pond, Horses, Day/Night cycle, Particle system, Shelter, Sound, Interaction. All the tests were completed well and the session was given positive feedback and a high rating. 87.5% of participants supported the implementation in the office, 71.9% felt relaxed in the session and 93.3% said they would try it in the future when it is available. The project was thus successful in relaxing the participants. The approach was to relax all participants irrespective of whether they are stressed or not as there was no identification or prediction of stress levels.

Fei and Bing (2012) performed an interesting study that applied exploratory factor analysis and further authenticated the results by performing confirmatory factor analysis. The findings relate the dynamic development of the working state of the workers with their psychological contract content with the organization. The authors analyzed two groups of participants where the first group had 169 participants from Beijing, Tianjin, and Langfang, and the second group had 219 participants from Beijing, Tianjin, Qingdao, and Shanghai. They conducted this study based on their background, work stress calculations consisting of eight features, and psychological contract calculations which comprises of nineteen features. To analyze the data, the authors made use of regression analysis and found that at various levels of educational background there are various and diverse effects of transactional contracts for workers on work stress. Like for example, people tend to go for rewards equivalent to how much effort their job expects from them. In the model of this study, three angles of the scalene triangle are the dimensions of psychological contract notable by the worker. A point inside the scalene triangle is the work stress that is in line with the policies to manage the psychological contract, which forms the best balance of work stress fitting in with the psychological contract structure. In particular, as the dynamic development of the working state of the workers, their psychological contract content also changes, which in turn affects their perception of investment symbols given by the organization.
A study by Cox et al. (2000) considered the pros and cons of extending the existing risk management standards in the field of physical threats to cover psychosocial threats. This study discussed the control cycle approach to the management of stress at the workplace and includes details of learning and training at the workplace which may lead to additional benefits to the organization as a whole. According to some basic principles of involvement, Stress management programs have been categorized as objective (prevention, timely reaction, or rehabilitation), agency (organization and/or employees), and target (organization and/or individuals). The research also suggests that the most beneficial for both the organization and the employee might just be organizational-level interventions or programs that target both. It is important for the proper management of the stress management programs that they be evaluated from time to time. Unfortunately, they found that there was a discrepancy between the requirements from the studies and the actual practice in the organizations. A review stated that most of these programs were inadequate and dealt only with the individual’s issues, very few were properly designed and evaluated to benefit both the organization and the individual. Thus ultimately, in terms of the effectiveness, three kinds of interventions were categorized as primary, secondary, and tertiary. Thus, the results tell us that programs that deal with stressors in the workplace have better results.

Burman and Goswami (2018) have reviewed the work of various authors that focussed on stress at the workplace and categorized them based on certain factors. The authors divided all sources and literature into four categories as journals, author profile, research methodology, and type of industries/research. They discovered that after reviewing the conceptual and descriptive research material on work stress, they could draw the basics and understanding of it. The study also found that there were a greater number of empirical studies than exploratory and longitudinal studies and that the bulk of the articles were not by professionals but rather by academicians. The top three contributors to these research papers and materials were from the US, UK, and India in that order. They also discovered that even though India has contributed to this research, based on both practical and conceptual knowledge related to stress, and its effects but the scientific findings have not been put to much use to deal with the actual problem. Adequate management techniques and tools are not being used to reduce stress levels; they are still only on paper. The research material also shows that stress from a workplace affects their performance and productivity at the organization as well as their physical and psychological well-being. Work stress can also lead to various diseases like coronary heart disease, blood pressure, depression, anxiety, nervousness, etc.

A study chose a sample size of 130 participants from various private institutes in Hyderabad, India by Stratified Random Sampling (Rahoo et al., 2017). The objective of the study was to find out the causes of work stress amongst the faculty members and how satisfied the members were in a private institute. They discovered that several stressors were disturbing the teachers. Demands at work, work relationships, role, position, pace and intensity of change in the workplace, and lack of support were found to be a major stressors. It was found that Age, Gender, and Marital status directly affect stress due to major differences found in the analysis of each factor and its relationship with stress. After combining the results of analyzing all the factors including the main cause of tension between the participants, poor peer relations, lack of rest, long working hours, harassment by the staff, lack of communication, lack of wage employment, and limited training, were major stressors and it was also witnessed that workplace stress plays a major role in the stress in an individual’s lives. Stress and job burnout outgrow when there is role erosion, role overload, resource inadequacy, role isolation, and role ambiguity at the workplace. Here the focus was related to roles played by employees in various spheres of their jobs.

Another research work made use of ex post facto research design for recognizing reasons and impact of stress amongst the employees of the sugar mills (Malarvizhi & Jeyarathnam, 2016). This design is a quasi-experimental study that tests the reliability of the questionnaires used for analyzing the symptoms of stress, health problems, and coping mechanisms. The Cronbach alpha reliability test was used and the result was found to be 0.77, which is at the higher end thus being reliable enough.
To analyze and interpret the data, descriptive Analysis and percentage analysis such as Average Score Analysis, Analysis of Variance, t-Test, Garrett Ranking Technique, and Simple Regression Analysis were utilized to achieve the aims of the study. They also found that twelve significant coping strategies were: yoga, meditation, physical exercise, entertainment, being away from a stressful environment, sleep, speaking with like-minded persons, playing with pet animals, prayer, meditation, positive thinking, time management, and tour. Through a review, the respondents were asked to rate coping strategies they used to find the most significant and common strategy that worked on most people. They discovered that “positive thinking”, prayer and speaking with like-minded persons are significant contributors to coping with the stress.

A case study in China had a sample size of 240 respondents, their data was collected via a questionnaire consisting of twenty-four questions (Meng & Wang, 2018). The prime statistical methods ANOVA and multifactor line regression were used for the analysis of the work stress of the university faculty. It was observed that the levels of stress of university faculty are significantly affected by professional status, age, and period of teaching. Along with them, administrative affairs, structural constraints, and personal characteristics play a major role in the work stress of the faculty members. This empirical study discovered that the mechanism used for evaluation which is based on quantitative performance indicators has significantly increased the occupation-related stress on the teachers.

Another research by Bhui, et al. (2016), aimed to recognize the stressors at the workplace and also the personal, individual, and organizational management programs or techniques that employees use to reduce their stress in public organizations, private organizations, and NGOs. Their sample size was 51 participants from various organizations. Qualitative interviews were conducted to collect the data from the participants. Results showed that unfavorable working environments and management practices were major and common stressors in the workplace. Along with those factors, unrealistic requests of work, lack of support, bad or unfair treatment, low decision-making freedom, lack of appreciation, effort–recognition imbalance, conflicting roles, lack of transparency, and poor communication within the organization were some more factors that led to increase in the stress levels of the employees. They found that organizational programs were found to be very helpful in reducing stress levels if there was an improvement in the management styles, which included physical exercise, breaks in between, and making sure that there is enough time for planning work tasks. Also, to prevent and reduce stress, personal interventions were very significant.

Shapiro et al. (2005) have gone a step ahead by suggesting a stress management program, for health care professionals that would decrease the level of stress and improve their quality of life. Frank et al. (2015) also studied the effectiveness of a similar type of mindfulness-based stress reduction (MBSR) program on educators. They observed the improvement in the quality of sleep of the respondents. These authors also could not provide any insight into understanding and predicting stress in advance.

One similar study by Rawat and Sultana (2021) demonstrates the application of Machine Learning algorithms for advanced resource planning in Hospital Emergency Departments. The authors have developed their model called light gradient boosted machines (LGBM) and found that its accuracy and time taken for prediction is better than other ordinary ML models like decision tree and GBM.

After reviewing all these research work by various authors globally, it was observed that none of the studies was offering a complete solution from the identification of stress-causing factors to the selection of a minimal subset of important stressors to the prediction of stress. Hence, the following three research gaps were observed:

1. It was found that none of the studies is based on all factors that surround the lives of working professionals viz. Organisation Specific, Personal Factors, Work Environment Specific and Job Profile Specific.
2. It was also observed that none of the studies focusses on the identification of the most influential stress-causing factors in a ranked manner.
3. Also, none of the studies is able to predict the stress level of the respondent.

This paper fills these gaps with the help of machine learning algorithms for ranking stress-causing factors and comparing and then applying the most accurate algorithm for the prediction of stress. Machine learning algorithms can provide a reasonably good accuracy of prediction with the number of records that are equal to more than ten times that of the number of features or questions.

Finally, A GUI is designed and developed by using the widgets available in the tkinter framework of Python where the respondent was required to answer a few situation-based questions that were mapped to various situations that the working professionals face in day to day life. These questions capture the level of stress of respondents indirectly. The front end (of GUI) uses an API to make use of the trained algorithm at the back end to predict the level of stress of the respondent.

**METHODOLOGY FOR RESEARCH**

A critical analysis of various research papers and articles was done, and important influential factors that cause stress were identified. The approach of the research is quantitative so that a predictive model using machine learning algorithms can be developed by observing the big picture first and narrowing it down by a selection of influential features that are more important contributors to causing stress. A detailed questionnaire was designed with 18 questions to understand all dimensions of stress-causing factors in the lives of the working professionals. The questions were grouped under four categories viz. Organization-Specific, Personal Factors, Work Environment Specific and Job Profile Specific. The answers were on a five-point Likert scale. As there were 209 respondents, so the sample size was technically enough to train the machine learning algorithm for the prediction of stress because it is more than ten times the number of questions/ features. The mode of data collection was online by using a questionnaire developed in google forms. The categories and questions under each of them are mentioned in Table 1.

This survey was conducted from February to early May 2020 which was just the beginning of the COVID pandemic period in India. By using non-probability convenience sampling for the selection of respondents, we gathered data from 209 participants. They were asked to provide demographic data viz. name, email address, age, gender, organization name, years of experience, marital status, location, sector, which are some of the features that may play a role as to how much stress affects a person. Respondent name, organization name, and email addresses were asked only to eliminate incorrect and redundant data to help with the pre-processing step of Data Science. The sample characteristics are shown in Table 2.

To understand the effect of stress-causing factors, 18 questions were asked from each participant. The system framework in the form of a Data Flow Diagram has been shown in Figure 1. Exploratory and explanatory data analysis have been done in detail as shown in Fig. 2 to Fig. 7 and Table 3 to Table 8. Popular machine learning language Python has been used for data cleaning, data reduction, data visualization, feature extraction, and application of machine learning algorithms. The tkinter module has been used to develop a Graphical User Interface to predict the stress level based on situation-based questions.

**Data Pre-Processing**

The collected data had some missing fields and some outliers; hence data cleaning was done. Finally, the dataset was reduced to 197 entries. Demographic fields from the dataset were also eliminated before the application of the machine learning algorithm.

Depending on the answers to 18 questions, each answer was given a numeric value from 1 to 5 depending on whether the answer on the Likert scale reduces or increases the stress levels. For example, if the respondent says that he/she “Always” gets “Support from family”, then the transformed
Table 1. Details of all variables corresponding to each question used in the study

| Variable                          | Description                                             | Category                     | Transformed Data                                                                 |
|----------------------------------|----------------------------------------------------------|------------------------------|----------------------------------------------------------------------------------|
| Improper Management              | Improper management of the organization                |                               | Negligible=1<br>Very little extent=2<br>Little extent=3<br>Great extent=4<br>Very great extent=5 |
| Unpleasant Organisation Culture  | Unpleasant culture in Organisation                      | Organism Specific            | Strongly disagreed=1<br>Disagreed=2<br>Neutral=3<br>Agreed=4<br>Strongly Agreed=5 |
| Hospitality Issues               | Lack of Hospitality                                      |                               | Very great extent=1<br>Great extent=2<br>Little extent=3<br>Very little extent=4<br>Negligible=5 |
| Complex Demands                  | Unrealistic/Complex Demands from the Management          |                               | Strongly disagreed=1<br>Disagreed=2<br>Neutral=3<br>Agreed=4<br>Strongly Agreed=5 |
| Job Security Concern             | Sense of safety regarding job                           |                               | Very great extent=1<br>Great extent=2<br>Little extent=3<br>Very little extent=4<br>Negligible=5 |
| No/less Organisational Support   | Less Support from Organisation                          |                               | Negligible=1<br>Very little extent=2<br>Little extent=3<br>Great extent=4<br>Very great extent=5 |
| Work-life Imbalance              | Finding the right balance between work and personal life |                               | Really Good=1<br>Good=2<br>Fair=3<br>Bad=4<br>Really Bad=5                     |
| No/less Family Support           | Support from family                                      |                               | Always=1<br>Often=2<br>Sometimes=3<br>Rarely=4<br>Never=5                      |
| Low Morale                       | Feeling very low/down                                    | Personal Factors             | Never=1<br>Rarely=2<br>Sometimes=3<br>Often=4<br>Always=5                      |
| Health Issues                    | Poor Health                                              |                               | No impact=1<br>Little extent=2<br>Some extent=3<br>Great extent=4<br>Very great extent=5 |
| Disturbed sleep pattern          | Disturbed sleep pattern due to smartphones/televsions/laptops etc. |                               | Strongly disagreed=1<br>Disagreed=2<br>Neutral=3<br>Agreed=4<br>Strongly Agreed=5 |

continued on following page
value for Always is 1 (Low stress) and if he says he “Never” gets “Support from family”, then the transformed value for “Never” is 5 (High stress). Corresponding to answers to all 18 questions, a score was calculated and it was transformed to Low stress (<1.33), medium stress (1.34 to 2.66), or a high stress (>2.66).

Exploratory Data Analysis
To get deep insight into collected data in terms of sample characteristics, observing the patterns, identifying outliers, various bar charts and heatmaps were generated by using popular plotting libraries of Python as given below:

- **Matplotlib**: Low level, provides lots of freedom.
- **Pandas Visualization**: Easy to use interface, built on Matplotlib.
- **Seaborn**: High-level interface, great default styles.

Bar chart for observing stress levels among participated males and females showed that the majority of both genders experienced medium-level stress. It was noticed that 56.8% of males and 48.61% of females fall in the category of medium-level stress. This is shown in Figure 2 and the corresponding data is shown in Table 3.
Bar chart to get an insight of participants, with respect to their marital status which made evident that major percentages of single as well as married people were medium level stressed, shown in Figure 3 with corresponding data in Table 4. Also, an important observation was that the stress level is higher in married professionals than in singles.

Bar chart for participants with respect to their age, made a clear observation that the majority of participants experience medium-level stress for all age intervals up to 40 years. For later ages, the majority of them observe low-level stress, as depicted in Figure 4 with corresponding numbers in Table 5.

Bar chart was also crafted to study the level of stress of participants with respect to their professional fields, shown in Figure 5 and corresponding data in Table 6. Since a majority of the participants belonged to the IT and Teaching and Education Sector, an insight was taken into these fields. It was observed that the Teaching and Education sector seemed to experience greater high-level stress than the IT sector. The majority of IT sector participants observed medium-level stress which was higher than the Teaching and Education sector percentage level.

Bar chart for studying the stress level of participants with respect to their work experience in terms of years showed that for participants having 0-10 years of work experience, majority of them observed medium-level stress. As the years of experience increase, the level of stress reduces. This is depicted in Figure 6 and Table 7.

Lastly, a correlation matrix was formed to find the Pearson’s correlation values among each of the features with the target variable. It was found that all features had a positive correlation with the stress value and the features having a correlation value of more than 0.6 are shown in Table 8.

### Table 2. Sample characteristics in terms of demographics data

| Gender       | Number of respondents |
|--------------|-----------------------|
| Male         | 131                   |
| Female       | 77                    |
| Transgender  | 1                     |

| Marital Status | Number of respondents |
|----------------|-----------------------|
| Divorced       | 1                     |
| Married        | 86                    |
| Separated      | 1                     |
| Single         | 121                   |

| State          | Number of respondents |
|----------------|-----------------------|
| Assam          | 1                     |
| Chandigarh     | 1                     |
| Chhattisgarh   | 1                     |
| Delhi          | 99                    |
| Gujarat        | 2                     |
| Haryana        | 28                    |
| Karnataka      | 12                    |
| Maharashtra    | 19                    |
| Mizoram        | 1                     |
| Odisha         | 1                     |
| Tamil Nadu     | 7                     |
| Telangana      | 7                     |
| Uttar Pradesh  | 23                    |
| Uttarakhand    | 2                     |
| West Bengal    | 5                     |

| Sector                        | Number of respondents |
|-------------------------------|-----------------------|
| Finance and Accounts          | 17                    |
| Business development and sales| 25                    |
| Creative arts and design      | 3                     |
| Engineering and Manufacturing | 10                    |
| Pharmaceuticals and Healthcare| 6                     |
| Hospitality and Event Management| 6                    |
| Information Technology        | 64                    |
| Law                           | 3                     |
| Marketing and media relations | 7                     |
| Recruitment and HR            | 3                     |
| Teaching and Education        | 37                    |
| Others                        | 28                    |
Surprisingly, the feature, ‘Lack of participant’s family support’ had the correlation value of just 0.399163, hence it was eliminated from the further study, and the rest of the seventeen features were retained. For a visual understanding of correlation values among features, a heatmap is generated where darker colors represent a high positive correlation value and lighter means lower positive correlation, as shown in Figure 7.

Table 3. Categorization of males and females according to the level of stress

| Stress Level | Gender          |
|--------------|-----------------|
|              | Male            | Female         |
| High         | 19(14.4%)       | 17(22.22%)     |
| Medium       | 74(56.8%)       | 37(48.61%)     |
| Low          | 38(28.8%)       | 23(29.17%)     |
Figure 2. Percentage of male and female under three different categories of stress

Table 4. Categorization of respondents as per marital vs. level of stress

| Stress Level | Single   | Married  |
|--------------|----------|----------|
| High         | 18(14.65%) | 19(21.52%) |
| Medium       | 70(57.76%) | 41(47.37%) |
| Low          | 33(27.59%) | 26(29.11%) |

Figure 3. Percentage of single and married working professionals under three different categories of stress
To make a predictive model for stress analysis, Machine Learning has been used. Since the data was labeled, therefore supervised learning method has been applied. An effort has been made to fit an algorithm that learns the mapping function to derive output as a function of seventeen input features. Through this process, the best mapping function has been found that makes sure that for a new similar

Table 5. Stress level according to different age groups

| Stress Level | 20-30 | 31-40 | 41-50 | 51-60 |
|--------------|-------|-------|-------|-------|
| High         | 18(14.53%) | 6(22.22%) | 6(17.86%) | 6(29.41%) |
| Medium       | 74(58.97%) | 14(51.85%) | 11(35.71%) | 10(52.94%) |
| Low          | 33(26.5%) | 7(25.93%) | 15(46.43%) | 3(17.65%) |

Figure 4. Distribution of professionals according to age under three different categories of stress

Table 6. Level of stress according to profession type

| Stress Level | Information Technology | Teaching/ Education |
|--------------|------------------------|---------------------|
| High         | 5(8.06%)               | 5(12.5%)            |
| Medium       | 36(56.46%)             | 18(50%)             |
| Low          | 23(35.48%)             | 14(37.5%)           |

Development of Predictive Model Using Machine Learning Algorithm
To make a predictive model for stress analysis, Machine Learning has been used. Since the data was labeled, therefore supervised learning method has been applied. An effort has been made to fit an algorithm that learns the mapping function to derive output as a function of seventeen input features. Through this process, the best mapping function has been found that makes sure that for a new similar
Figure 5. Distribution of IT professionals and Teaching professionals under three different categories of stress

Table 7. Relation between years of experience and level of stress

| Stress Level | 0-5            | 6-10          | 11-15         | 16-20         | 21-25         | 26-30         |
|--------------|----------------|---------------|---------------|---------------|---------------|---------------|
| High         | 17 (14.02%)    | 3 (18.75%)    | 7 (36.84%)    | 3 (16.67%)    | 2 (18.18%)    | 1 (12.5%)     |
| Medium       | 71 (57.94%)    | 10 (62.5%)    | 5 (26.32%)    | 6 (33.33%)    | 4 (36.36%)    | 7 (87.5%)     |
| Low          | 34 (28.04%)    | 3 (18.75%)    | 7 (36.84%)    | 9 (50%)       | 5 (45.49%)    | 0             |

Figure 6. Distribution of working professionals according to number of years of experience, under three different categories of stress
data being input to the algorithm, the best prediction of the output variable is made. For this purpose, the dataset has been divided into two parts in the ratio of 3:1, for training and testing respectively.

From various supervised regression learning models, Linear regression, Lasso regression, Decision trees regression, SVM regressor model, and Random Forest regressor have been applied. In linear regression, the model tries to fit a linear equation between the dependent and independent

| S.No | Feature                              | Correlation value |
|------|--------------------------------------|-------------------|
| 1    | Unpleasant organisation culture      | 0.756670          |
| 2    | Lack of hospitality from the organisation | 0.755765          |
| 3    | Complex demands from the management  | 0.742613          |
| 4    | Working Conditions of the organisation | 0.752460          |
| 5    | Lack of organisational support to the employees | 0.669368          |
| 6    | Low morale of the participants       | 0.685482          |
| 7    | Improper management by the organisation | 0.723241          |

Figure 7. Heatmap showing correlation among all input features
variables to find out the relation between the two. But in this dataset, the results showed overfitting, and to prevent it, regularization techniques have been used. In this technique, Ridge Regression and Lasso Regression are used to adding a penalty term to the objective function. With our datasets, Ridge regression failed to show the desired results. Thus, it was decided to opt for lasso regression since it shrinks the coefficient to zero.

In the decision tree regressor, the dataset is divided into smaller subsets to build a tree-type structure in the model where the final subset is also known as decision or leaf nodes. This type of regressor can deal with both the numerical type of data and also the categorical type of data. Leaf or decision nodes represent attributes via branches of the model.

The Random Forest Regressor combines the predictive analysis from various machine learning algorithms to build a more accurate prediction model. For classification and regression, it utilizes the ensemble learning method and it is much more accurate than any one single model.

We also applied Support Vector Regressor that recognizes the non-linearity in the dataset and gives us a well-trained prediction model. It is almost similar to Support Vector Machine as it utilizes the same technique with very small changes. The algorithm considers the points that lie within the boundary line and the line that fits best is the line hyperplane with the maximum number of points. It tries to reduce the errors as much as possible.

Finally, for the evaluation of the performance of the models, the R squared value and Mean Squared Error was calculated. R-squared value can be defined as the statistical measure of how close the data is to the fitted regression line, also known as the coefficient of multiple determination for multiple regression. Whereas the mean square error (MSE) is the average of the square of the difference between the observed and predicted values of a variable and is used as a measure of quantitative assessment to calculate the performance of an algorithm. After the evaluation of all the models, the results are presented in Table 9.

According to Table 9, the Support Vector Regression model gives the best Mean squared error and reasonably good R-squared value and hence is giving the best performance, followed by the Linear Regression model and Random Forest regression model. Decision tree Regression on the other hand showed poor results giving a high MSE and a low R-squared value.

For visualization of errors, residual plots and prediction-error plots were made to get proper insights of all the models except Decision tree regression. Residual plots are shown in Table 10 that shows the line that fits the residual errors lies in the range -0.025 to +0.025 and supports the application of machine learning algorithms for the prediction of stress levels. Scatter plots showing the relationship between actual and predicted values of stress are visualized in Table 11. Ultimately, according to the plots shown, and performance metrics as mentioned in Table 9, Support Vector Regression performed the best.

**Extraction of Important Stress Causing Features Using Machine Learning**

The Random Forest model has been used to find the importance of every feature in terms of causing an impact on stress level. It is calculated by the node impurity weight. The node probability can be

| Model                      | R-Squared | MSE    |
|----------------------------|-----------|--------|
| Linear Regression          | 0.9897    | 0.006596 |
| Support Vector Regression  | 0.9881    | 0.006194 |
| Random Forest Regression   | 0.8890    | 0.045211 |
| Lasso Regression           | 0.8761    | 0.064594 |
| Decision Tree Regression   | 0.6309    | 0.262037 |
Table 10. Residual plots for different machine learning algorithms

Lasso Regression algorithm

Linear Regression algorithm

Random Forest Regression algorithm

Support Vector Regression algorithm

Table 11. Actual vs Predicted stress values scatter plots for different algorithms

Lasso Regression algorithm

Linear Regression algorithm

Random Forest Regression algorithm

Support Vector Regression algorithm
calculated by the number of samples that reach the node, divided by the total number of samples. The most important feature is the one that has a higher node impurity weight. For finding the importance of each feature, rfe() method has been used so that the importance of features in predicting the stress level can be obtained. Figure 8 lists down each factor along with its feature importance value and a bar chart for these factors is shown in Figure 9.

From Figure 8, it is found that the following are the top ten most important features in decreasing order of their importance in causing stress in the working professionals:

1. Unpleasant organization culture.
2. Lack of hospitality from the organization.
3. Complex demands from the management.
4. Low morale of the participants.
5. Working Conditions of the organization.

Figure 8. Importance of each feature as given by RFE function for feature selection

| Feature Description                        | Importance Value |
|--------------------------------------------|------------------|
| Workload                                   | 0.013181700813699 |
| Time Pressure                              | 0.029191076212916144 |
| Income Dissatisfaction                      | 0.016651255130472924 |
| Poor Peer Relations                         | 0.012324955779628808 |
| Working Conditions                          | 0.051956280477543616 |
| Working Hours                              | 0.01971977303412803 |
| No/Less Organisational Support              | 0.04934892189927965 |
| Complex Demands                            | 0.113564694462298291 |
| Improper Management                        | 0.041923156052802904 |
| Unpleasant Organisation Culture             | 0.2750023195373936 |
| Hospitality Issues                         | 0.13708240792271248 |
| Job Security Concern                        | 0.024286539624781407 |
| Work-life Imbalance                         | 0.0393388571477908 |
| Low Morale                                 | 0.06016974930779748 |
| Poor Health Issues                         | 0.02903316919793979 |
| Tiredness from Screen                      | 0.0302599414184524562 |
| Disturbed sleep cycle                      | 0.04742512244743475 |

Figure 9. Visualization of the importance of the seventeen input features used in the questionnaire
6. Lack of organizational support to the employees.
7. Disturbed sleep pattern due to smartphones/televisions/laptops etc.
8. Improper management by the organization.
9. Work-life imbalance.
10. Tiredness from screen.

In this research, it is observed that ‘Unpleasant organization culture’ is the most important feature i.e. major stress is caused due to this factor. ‘Poor peer relation’ is the least important feature.

From the above result, we can conclude that the topmost features that contributed towards stress mostly belong to the organization-specific category, as it is visible that features 1 to 3 belong to the said category. These factors play important role in the stress of a working professional. Thus, it can be concluded that organizational support and working environment affect the stress majorly.

Following this once again all four selected machine learning algorithms except Decision tree Regression were trained and tested with the most influential 10 features shortlisted with the help of the rfe() function. The results obtained are shown in Table 12.

All four predictive models viz. Linear Regression, Support Vector Regression, Random Forest Regression, and Lasso Regression have given a reasonably good value for performance measures R-squared and mean squared error. We have selected the Support Vector Regression model to be used at the backend of our working Predictive design in the next section of Graphical User Interface (GUI) designing as its performance is the best.

**GUI Designing**

After the analysis of the data, the results of the study were utilized to build a Graphical User Interface (GUI) to help other working professionals in similar generation and working backgrounds assess their stress levels. Through this interface, the professionals can find out their stress levels and seek out proper help if required.

The GUI contains 10 situation-based questions, 5 options in each question that have already been assigned ordinal values in increasing order. GUI has been designed using the tkinter module. In the interface, the first window as shown in Figure 10, pops up, asks for the age category, and based on the selection, the appropriate Test window is formed with the help of the best prediction model which is support Vector Regression for this study on Millenials. The study on Generation Z was done earlier by Pabreja et al., 2021 and the frontends for both are integrated.

If the user clicks the button “Millenial”, then the GUI for this research brings the Introduction page that contains some simple guidelines and a start button to start the test. The page is made with the help of buttons and labels, as shown in Figure 11.

When the user clicks the start button, the test begins and the window containing the first question pops up. As a sample, the window of the second question is shown in Figure 12.

Pagination is used to transition between the multiple pages of the test. Pagination is defined as segregating content into separate pages. For each question, a different frame was created. To enable the next window, each frame is looped through, and the user can attempt the test and finally submit.

### Table 12. Comparison of Performance metrics for the dataset with reduced features

| Model                   | R-Squared | MSE    |
|-------------------------|-----------|--------|
| Linear Regression       | 0.9160    | 0.042532 |
| Support Vector Regression | 0.9164   | 0.042184 |
| Random Forest Regression | 0.8624   | 0.082615 |
| Lasso Regression        | 0.9166    | 0.042681 |
Figure 10. First frame of GUI based Stress Prediction model

![First frame of GUI based Stress Prediction model](image1)

Welcome to the stress prediction model, kindly select your age category

- Millennial
- Generation Z

Figure 11. Introduction frame of GUI based Stress Prediction model

![Introduction frame of GUI based Stress Prediction model](image2)

STRESS ANALYSIS

Stress is called "THE INVISIBLE" as it is often overlooked but one of the major reason behind deteriorating mental health is stress. It affects one's mental and physical well being.

- Answer carefully, you cannot jump back to the previous question.
- Select the best suited option as per real life scenarios.
- Answer the questions with utmost honesty.

To test your stress level take up this test.

Start Test

Figure 12. Window to accept a response from the participant

![Window to accept a response from the participant](image3)

2. You are to submit your project in a week. Your team is not working as per the requirement and things are not going according to the deadline. How is this situation going to affect the completion of your project?

- Motivate your team to complete the project within the deadline
- Make a priority list and complete the work accordingly.
- Complete most of the tasks partially.
- Take help from someone who is on another team and try to complete the work.
- Try to complete the task all by yourself.
The button called ‘show result’ helps fetch all the input values using a get function and calculate the stress levels. The calculated value is then compared with the range values which ranged as: < 1.33 as low-level stress, 1.34-2.66 as medium-level stress, and >2.66 as high-level stress. These are the same range values that were used before in the questionnaire. When the show result button is clicked the final report is shown. The report shows the stress level and recommendations to cope with stress, as shown in Figure 13.

CONCLUSION

The survey was conducted online to collect the data on eighteen well-designed questions to capture the views of respondents concerning the stress-causing factors. More than two hundred working professionals of various organizations, companies, educational institutions of higher studies all across India responded. The questions belonged to four categories viz. Organization Specific, Personal Factors, Work Environment Specific and Job Profile Specific.

Extensive data visualization techniques were utilized for exploratory data analysis, a description of important findings is given below:

- The majority of both males and females observed medium-level stress.
- Major percentages of single as well as married people were medium level stressed. Also, an important observation was that the stress level is higher in married professionals.
- The majority of participants experience medium-level stress for all age intervals up to 40 years. At later ages, the majority of them experience low-level stress.
- It was observed that Teaching and Education sector professionals seemed to experience greater high-level stress than the IT sector workforce. The majority of IT sector participants observed medium-level stress which was higher than the Teaching and Education sector percentage level.
- Participants who had 0-10 years of work experience observed medium-level stress. Participants having 11-15 years of experience experienced high-level stress.
- It is observed that the majority of the people experience medium-level stress i.e. 53.8%. Also, in the High-level stress category, there were 17.26% of the respondents and 28.94% had low-level stress.

Figure 13. Window displaying stress score and recommendation for participants
Following this, for explanatory data analysis, various machine learning algorithms were applied viz. Linear regression, Lasso Regression, Decision trees regression, SVM regressor model, Random Forest regressor for prediction of stress levels.

Next, after extraction of 10 most influential stress-causing factors using rfe() function was done and, again all four machine learning algorithms were applied. Once again, Support Vector Regression showed the best accuracy and it was further chosen for integration with the GUI.

The main deliverable from this study is the identification of ten important features that are more influential features out of seventeen features that trigger stress in the lives of working professionals. The application of various machine learning algorithms made this identification and prediction of stress possible. In this study, the top ten factors affecting the stress level in working professionals were found to be Unpleasant organizational culture, Lack of hospitality from the organization, Complex demands from the management, Low morale of the participants, Working Conditions of the organization, Lack of organizational support to the employees, Disturbed sleep pattern due to smartphones/televisions/laptops, etc, Improper management by the organization, Work-life imbalance, and Tiredness from the screen.

It is evident from the results that the factors from the category ‘Organization Specific factors’ dominated the most as far as the stress is concerned. Finally, a Graphical User Interface (GUI) which contained the situation-based questions corresponding to each of the ten short-listed factors was developed. Each question along with the options had been framed after a thorough study on different stress calculators. Each question had a few options, and each of the options was allotted a score that helped in calculating the final score of stress level. The final score enabled the user to observe and compare his/her level of stress among other people.

LIMITATIONS OF THE STUDY

The data from working professionals from fifteen different states of India were collected in the months of late February to early May 2020 which was the initial phase of the covid pandemic in India. The study was done based on this data only, thus, it should not be generalized. One major limitation was observed that most of the participants had 0-5 years of experience and belonged to the IT or Education sector. For the rest of the intervals, enough data entries could not be collected.

RECOMMENDATIONS AND FUTURE SCOPE

With the identification of these important features, the organization can develop a policy to take care of their employees and develop stress management programs for them to manage these factors. Also, by knowing the stressors in advance, the organizations can be careful at their end that pleasant organizational culture is maintained, there are realistic demands from the management, the employees are supported well in times of their needs which in turn would boost their morale. The success of any organization is directly proportional to the mental well-being of its employees. The employees can perform to the best of their capacity if they are mentally motivated and this, in turn, helps them take care of their personal lives as well. Thus, this study addresses an important social aspect of a working professional’s life too.

In the future, a mobile App or Web App may be developed that becomes a module of the Employee Management Software of the organization and detects the stress levels so that it can be prevented before it harms the employees. In fact, with new technological innovations, real-time monitoring of the stress levels can be done by monitoring the vitals of employees with the help of smart medical alert bracelets and correlating these with the results of our GUI-based Stress Prediction working model. By integrating this health monitoring band with the Internet of things, automatic notifications about stress levels can be generated. Finally, with the knowledge of the ranking of stressors and the correlation of stress-causing factors, appropriate stress detection rules can be triggered.
REFERENCES

Bhui, K., Dinos, S., Galant-Miecznikowska, M., de Jongh, B., & Stansfeld, S. (2016). Perceptions of work stress causes and effective interventions in employees working in public, private and non-governmental organisations: A qualitative study. BJPsych Bulletin, 40(6), 318–325. doi:10.1192/pb.bp.115.050823 PMID:28377811

Burman, R., & Goswami, T. G. (2018). A systematic literature review of work stress. International Journal of Management Studies, 5(3-9), 112-132.

Cox, T., Griffiths, A., & Rial-González, E. (2000). Research on work-related stress. European Communities.

Employee wellness by Optum, affiliated with NASSCOM. (n.d.). https://nasscom.in/employee-wellness/

Frank, J. L., Reibel, D., Broderick, P., Cantrell, T., & Metz, S. (2015). The effectiveness of mindfulness-based stress reduction on educator stress and well-being: Results from a pilot study. Mindfulness, 6(2), 208–216. doi:10.1007/s12671-013-0246-2

Khuong, M. N., & Yen, V. H. (2016). Investigate the effects of job stress on employee job performance—a case study at Dong Xuyen industrial zone, Vietnam. International Journal of Trade. Economics and Finance, 7(2), 31.

Lawanont, W., & Inoue, M. (2018, January). An unsupervised learning method for perceived stress level recognition based on office working behavior. In 2018 International Conference on Electronics, Information, and Communication (ICEIC) (pp. 1-4). IEEE. doi:10.23919/ELINFOCOM.2018.8330700

Malarvizhi, V. R., & Jeyarathnam, M. (2016). Stress and Coping Techniques among Employees of Sugar Mills in Tamilnadu. Amity Journal of Training and Development, 1(1), 58–76.

Meng, Q., & Wang, G. (2018). A research on sources of university faculty occupational stress: A Chinese case study. Psychology Research and Behavior Management, 11. 597–605. doi:10.2147/PRBM.S187295 PMID:30573995

Misra, S. (2020, November). A Step by Step Guide for Choosing Project Topics and Writing Research Papers in ICT Related Disciplines. In International Conference on Information and Communication Technology and Applications (pp. 727-744). Springer.

Muraale, S., Basit, A., & Hassan, Z. (2017). Impact of job stress on employee performance. International Journal of Accounting and Business Management, 5(2), 13–33.

Odusami, M., Abayomi-Alli, O., Misra, M., Abayomi-Alli, A., & Sharma, M. (2020). A Hybrid Machine Learning Model for Predicting Customer Churn in the Telecommunication Industry. In International Conference on Innovations in Bio-Inspired Computing and Applications (pp. 458-468). Springer.

Pabreja, K., Singh, A., Singh, R., Agnihotri, R., Kaushik, S., & Malhotra, T. (2021). Stress Prediction Model Using Machine Learning. In Proceedings of International Conference on Artificial Intelligence and Applications (pp. 57-68). Springer. doi:10.1007/978-981-15-4992-2_6

Panigrahi, C. M. A. (2016). Managing stress at workplace. Journal of Management Research and Analysis, 3(4), 154–160.

Pattnaik, L., & Mishra, A. (n.d.). Effect of Workplace Stress: A Study in Indian Context. Proceedings of XXIII Annual International Seminar “New Directions in Higher Education”, 551-572.

Prasad, K. D. V., Vaidya, R., & Anil Kumar, V. (2015). A study on causes of stress among the employees and its effect on the employee performance at the workplace in an International Agricultural Research Institute, Hyderabad, Telangana, India. International Journal of Management Research and Business Strategy, 4(4), 68–82.

Rahoo, L. A., Raza, S. A., Arain, M. W., & Memon, M. (2017). A study on occupational stress among faculty members in Private Institutes of Hyderabad, Sindh. Research on Humanities and Social Sciences, 7(1).

Rawal, A., & Mhatre, S. (2018). A study on work stress and its impacts on employee productivity with respect to teacher (self-financing). IOSR Journal of Business and Management, 15-23.
Rawat, S. S., & Sultana, R. (2021). Advance Resource Planning in Hospital Emergency Departments Using Machine Learning Techniques. *International Journal of Human Capital and Information Technology Professionals, 12*(3), 74–86. doi:10.4018/IJHCITP.2021070105

Reddy, U. S., Thota, A. V., & Dharun, A. (2018, December). Machine learning techniques for stress prediction in working employees. In *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (pp. 1-4). IEEE. doi:10.1109/ICCIC.2018.8782395

Saravanan, K., & Muthulakshmi, K. (2017). A study on stress management among employees in nationalized bank, Trichy City. *International Journal of Trend in Scientific Research and Development, 1*(6), 1220–1230. doi:10.31142/ijtsrd5795

Shapiro, S. L., Astin, J. A., Bishop, S. R., & Cordova, M. (2005). Mindfulness-based stress reduction for health care professionals: Results from a randomized trial. *International Journal of Stress Management, 12*(2), 164–176. doi:10.1037/1072-5245.12.2.164

The Workforce View. (2020). *ADP Research Institute.* https://www.adpri.org/wp-content/uploads/2020/10/06223715/COLL_WFV_Vol2-Print_US_2020_570908_98571_FV.pdf

Thoondee, K. D., & Oikonomou, A. (2017, July). *Using virtual reality to reduce stress at work.* In *2017 Computing Conference.* IEEE.

Vijayan, M. (2017). Impact of job stress on employees job Performance in aavin, coimbatore. *Journal of Organisation and Human Behaviour, 6*(3), 21.

Xiang, F., & Liu, B. (2012, September). A study on work stress of real estate industry knowledge workers on the basis of psychological contract. In *2012 International Conference on Management Science & Engineering 19th Annual Conference Proceedings* (pp. 1272–1280). IEEE.
Kavita Pabreja is working as an Associate Professor in the Department of Computer Applications at Maharaja Surajmal Institute, an affiliate of GGS Indraprastha University, New Delhi, India. She received her Ph.D. in Computer Science from Birla Institute of Technology & Science, Pilani. She holds more than 25 years of experience which includes teaching experience of 20 years and industry experience of over 5 years with Indian as well as USA MNC. She has authored four books, two with International publishers - “Application of Artificial Intelligence tools – Impact on Weather Prediction”, “Object-Oriented Programming Using C++”, and two with National publishers- “Learning Visual Basic.Net Programming”, “Front End Design Tool VB.NET”. She has contributed over sixty papers in International Journals / Book/ International conferences of repute. She is a recipient of a Best Research Paper Award from EMC data storage systems and Amity for All India Competition in Data Research; another Best Research Paper award for developing An Android Application to illustrate the Perspective of Community on Blood Donation at the National Conference organized by IITM, an affiliate of GGSIP University, India.

Anubhuti Singh is a recent graduate of Bachelor of Computers Applications from Maharaja Surajmal Institute, an affiliate of Guru Gobind Singh Indraprastha University with 9.5 CGPA. She has various online certifications in Data Analysis and Visualization using Python Programming. Possessing a great interest in the field of Machine Learning, she is looking forward to pursuing a master’s in data science. Presently she is working as an Associate Analyst with Deloitte India.

Rishabh Singh is an Associate Analyst at Deloitte India (Offices of the US). He pursued his UG in BCA from Maharaja Surajmal Institute GGSIP University, Delhi, India. He along with his team worked on the frontend and integration of the Stress Prediction model, research paper of this project is published in Springer Books. He is a trained professional in Machine learning using python. Currently, he is a part of the agile workforce in Deloitte India and works on SAP ERP and Oracle Fusion Cloud as a technical consultant.

Rishita Agnihotri is an Associate Analyst at Deloitte India (Offices of the US). She pursued her UG in BCA from Maharaja Surajmal Institute GGSIP University, India. She along with her team worked on the machine learning and integration of the Stress Prediction model. The research paper of this project is published in Springer Books. She is a trained professional in Data Analysis and Machine learning using python. Currently, she is a part of the agile workforce in Deloitte and works on SAP ERP, Oracle Fusion Cloud, and Oracle EBS as a technical consultant.

Shriam Kaushik is a recent graduate of Bachelors of Computer Applications from Maharaja Surajmal Institute, an affiliate of GGS Indraprastha University, New Delhi, India. She is currently pursuing a Masters's in Business Administration from Prague University of economics and business. She was the Public Relations Representative in the IEEE EXECOM of her college and has done various courses such as Python Programming. She, along with her team, has written a research paper on ‘Stress Analysis using Python and ML’ which is published in the Springer book series. The paper has also been presented at the International Conference of Artificial Intelligence and Applications 2020.

Tanvi Malhotra is a recent graduate of Bachelors of Computer Applications from Maharaja Surajmal Institute, an affiliate of GGS Indraprastha University, New Delhi, India. She is trained in Python and has hands-on experience with a live project NetImpact. She has proficient knowledge of Javascript and Django. She, along with her team, has written a research paper on ‘Stress Analysis using Python and ML’ which is published in the Springer book series. The paper has also been presented at the International Conference of Artificial Intelligence and Applications 2020. Presently, she is working as an Associate Analyst with Deloitte India.