



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 10    Issue: VII    Month of publication: July 2022**

**DOI: <https://doi.org/10.22214/ijraset.2022.46082>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Detection and Analysis of Stress in IT Professionals by Using 5ML Techniques

Bindu K N<sup>1</sup>, Dr. Siddartha B K<sup>2</sup>, Dr. Ravikumar G K<sup>3</sup>

<sup>1</sup>PG Student, <sup>2</sup>Assoc. Prof. & HOD, <sup>3</sup>Professor & Head(R&D), Dept. of CSE, BGS Institute of Technology, Adichunchanagiri University, BG Nagar, Karnataka, India-571448

**Abstract:** *Our project's primary objective is to use ML techniques to detect stress in IT employees. Our approach is an advancement over prior stress recognition methods that lacked personal counseling, but it now includes an analysis of employees and the recognition of occupational stress in them, as well as providing them with appropriate stress management remedies via a survey form that is sent out on*

*a regular basis. Our system is primarily focused on stress management and creating a healthy and creative job atmosphere in order to get the most out of them through work time.*

*This research focuses on the construction of an intelligent system that uses ML to determine if a person is stressed or not stressed. The data for this study was collected from more than 600 male and female volunteers between the ages of 18 & 50. The acquired data consists of five (5) distinguishing traits (i.e. systolic blood pressure, diastolic blood pressure, glucose and gender). Employing Python IDE and sci-kit learn ML libraries, an autonomous system was constructed using ML techniques for categorization such like Linear Regression (LR), K-Nearest Neighbor (KNN), Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM). To find the ideal settings for each algorithm, Jupyter Notebook was used to improvements in service delivery using Grid search. The most important features connected to a person's stress condition are identified using feature selection technique. With an optimised training-testing average accuracy of 95.00 percent - 96.67 percent, the results we predict if one individual is stressed or not stressed after optimization.*

**Keyword:** *Stress, Systolic, Diastolic, Machine Learning, Decision Tree.*

## I. INTRODUCTION

By providing new technologies and services, the IT company is actively setting a new benchmark in the marketplace. Employee perceived stress were also found to raise the bar higher in this research. Despite the fact that many companies offer psychological health facilities to the employees, the problem remains out of control. In this research, we attempt to go deeper into the issue by attempting to detect stress in working employees in businesses. We would want to use machine learning approaches to assess stress and focus down the reasons that have a substantial influence on stress levels.

ML, which is an application of AI, gives the system the ability to instantly learn and grow from self-experiences something without being even being explicitly designed (AI). ML is a programming language that allows computers to retrieve information and knowledge for themselves. Using ML, explicit programming creates a mathematical model based on "training examples" to do the task based on projections or judgements.

The purpose of this study is to develop an autonomous system that can use physiological signs and machine learning algorithms to assess whether or not a human is stressed. A data capture unit including glucose, systolic blood pressure, diastolic blood pressure, and gender is included in the suggested system. After the dataset has been collected, the advantage of emerging model will be trained using ML methods for classification such as LR, KNN, RF, DT, and SVM, using the Python IDE and Sci-kit learn machine learning libraries. The results are then compared using default characteristics versus pre-processed characteristics.

## II. RELATED WORK

Mike's paper [1] is an excellent example. To improve TensiStrength accuracy, the barrier uses WSD innovation as a pre-processing stage, followed by lexicon-based anxiety or relaxation method. This study employed a dataset of thousand tweets using the word "Fine," which displays equivocal meaning in various sentence situations. This paper also removes unnecessary words from the tweet, such as prepositions, conjunctions, interjections, and articles, before working on the remaining words and categorizing them in a -5 to +5 range based on definitions of words.

A comprehensive hybrid approach that merges a factor graph model with CNN is used in the paper [2] to investigate the relationship

between user mental stress levels and their interpersonal relationships.

This is because Convolutional neural networks can learn a single latent feature from multiple modalities, whereas FGM excels in modeling co-relations. To derive user material parameters using tweet-level characteristics, a CNN with cross auto-encoder(CAE) was utilized, and also a partly labeled factor graph to incorporate user-level social interactions to determine stress. To increase accuracy, we compared the findings using the Twitter and SinaWeibo datasets, as well as comparison methods such as LR, SVM, gradient boosted decision tree, and DNN. As a result, we've developed a method for recognizing mental anguish states in frequent social media data from users, however the constraint is that we can only find users who really are disturbed but not on social media. The TensiStrength: Lexicon-based method is highlighted by the author Mike The Wall [3]. TensiStrength detects stress and relax in social networking messages using a lexical method and a system of regulations. It outperforms a similar emotion analysis software by a small margin. When matched to a typical ML approach, TensiStrength performs admirably. And which to use is determined by the nature of the text to be analyzed as well as the task's objectives. Twitter, a dataset, and a vocabulary are utilized in this paper to provide numerical strength ratings ranging from -5 to +5. The tasks of detecting sentiment and tension or relaxing are similar but not identical. TensiStrength's results in tweets are quite accurate when it comes to human programmers, but less accurate when compared to sentiment classification programmers and less effective when matched to ML methods optimized and trained on the same dataset. Because stress and relaxing are such significant components of everyone's lives, software must be developed that can accurately assess them and assist in the development of smart potential apps.

The research[5] concentrates in using social networks to identify and detect stress in a person; network sites distinguishing qualities that might be beneficial or negative. Twitter is being used to assess and predict depression in this study. To begin, they used crowdsourced to identify users who had been diagnosed with depression and reported it. Crowd workers can also opt-in to have their data harvested and evaluated using the computer software on their Twitter profiles. Here, a survey is used as the major technique for determining the amount of depression among group of employees. Here, we examine the properties of depressed and non-depressed users and draw certain findings, such as the fact that people with depression have fewer social activities, have more negative feelings, have a higher self- attentional concentration, and so on. I worked on an SVM classification to detect depression in a single individual who exhibits depressive symptoms in the evenings and at night.

### III. EXISTING SYSTEM

We employ an existing algorithm that is based on human facial expression and mindset to manually predict if a person is stressed or not. Some people's faces or facial expressions may be the same all the time, making them appear stressed. Some people may feign a smile or a frown. It may vary due to human predictions in manual prediction.

### IV. PROPOSED SYSTEM

In Our Proposed system based on human the clinical data and their life style and relationship data which we collect from corporate, hospitals and doctors of IT Employees in csv format, with the help of glucose, systolic and diastolic values we are going to predict whether the person has stress or not.

### V. METHODOLOGIES

The first phase of depression is considered stress. Anxiety can be caused by a variety of causes, including economics, career, relationships, and others. Professionals in the corporate sector are oblivious to the dangers of their jobs. Chronic stress is common, especially among IT professionals. Companies used to offer corporations a poll application form, then use that analysis to calculate stress levels based on the survey form, which included clinical values. Since the documents would have to be delivered manually, it required not only a long time and also a lot of effort. Utilizing anticipatory stress reduction systems aimed at lowering stress and improving employee health, the Stress Detection System aids employees in coping with issues that create stress. We developed a system at work that will photograph individuals at frequent intervals and then supply them with traditional survey forms. The physical effort will be reduced, and time will be saved. Using our services, carefully prepared Survey questions, this organizational strategy can be used to assist improve workplace stress.

The data set will be saved in CSV format and will include glucose, systolic blood pressure, diastolic blood pressure, glucose gender, and many other variables. These datasets will be used to train the LR, KNN, RF, DT, and SVM ML algorithms. We will train the dataset using default values with feature pre - processing phase to build the most intelligent model. The dataset will next be pre-processed and feature-selected to determine the most important characteristics among the three independent variables. Finally, the improved parameters of the LR, KNN, RF, DT, and SVM will be used to train the selected and pre-processed features. The



classification accuracy will be used as a performance metric. Here we will also try with Deep Learning using MLP Classifiers.

**A. Data Collection**

A survey of IT professionals was employed as the dataset in this case. It primarily contains information such as age, gender, place of employment, type of employment, whether he or she is self-employed, and clinical values such as systolic, diastolic, and glucose? And there are plenty more. Doctors and big organizations are two sources of information.

**B. Data Pre-Processing**

Then, in order for ML to work, we must first eliminate non-essential information such as comments and timestamps. Data cleansing is the term for this procedure. Data cleaning is the process of removing unnecessary data from a dataset so that it can be used for further research. Incorrect format, capture problems, and missing data are all considered garbage data. Because several of the properties contain blank input values, default values will be set to them. It is 0 for integers, 0.0 for floats, and NaN for strings.

We've now converted the gender property to a standardized way by replacing all unknown inputs with standard input. The data must then be encoded. After then, the data is double-checked to see if any data is missing. The dataset is then scaled and fit. ML approaches are now being used and compared to see which one best fits the dataset.

**C. Algorithm**

LR is the first algorithm. The LR function is a sigmoid expression featuring an S-shaped arc that receives any real integer and converts them to a range between 0 and 1. The equation is  $y = e(b_0 + b_1 * x) / (1 + e(b_0 + b_1 * x))$ .

The KNN classifier is the second algorithm. The training dataset is used to predict new occurrences, and similar K examples are found.

The decision tree is the third algorithm. A DT classification approach is one in which data is segmented according to a set of parameters. The leaves are referred to as the final outcome. Data is separated at nodes, which are called decision points.

Random forest is the fourth strategy. The output of RF is the mean or mode of classes, and it is an SVM that works by generating a huge volume of decision trees throughout the training process.

SVM (Support Vector Machine) is the 5th method. It's a supervised machine learning approach that could be used for analysis as well as segmentation.

The abbreviation MLP Classifier stands for Multi-layer Perceptron Classifier, which is related to a Neural Network. Unlike other classification methods such as SVM or Naive Bayes Classifier, MLP Classifier uses an underlying Neural Network to do classification.

**VI. WORKING OF THE SYSTEM**

- 1) Step 1: Login to the system.
- 2) Step 2: Enter the values in the input fields and click on submit.
- 3) Step 3: Entered values will be sent in array format to the evaluating model.
- 4) Step 4: Based on the entered value model will give the result.
- 5) Step 5: Result would contain either the person is stressed or not.

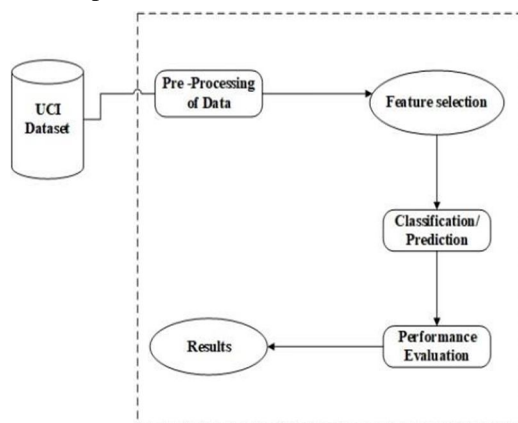


Fig-1: Workflow of the System

### VII. EXPECTED RESULT

We tested the system with both Stressed and non- stressed Values to ensure that the algorithm utilized in the system is robust. As predicted, the algorithm spotted weather the person is stress or not.

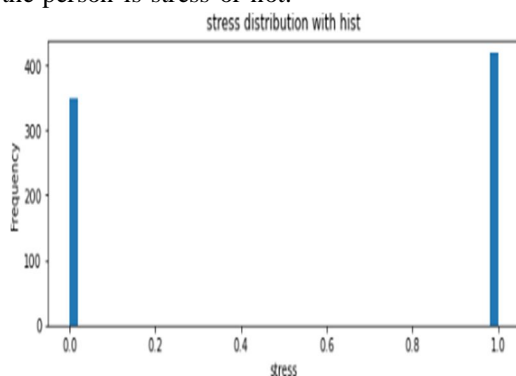


Fig2. Stress distribution

This graphic depicts the pressure distribution in our given dataset; we have a higher prevalence of stressvalue 1 in our dataset than a no-stress number, implying that datasets with pressure are more prevalent in our database.

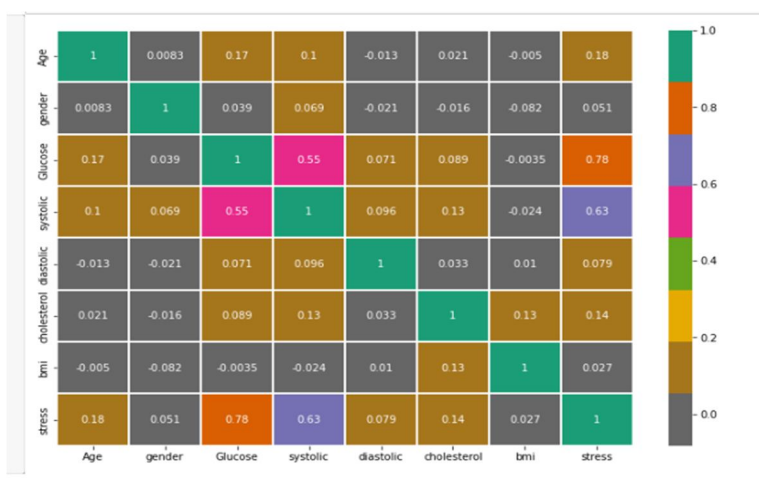


Fig3. Heat Map

The green representation, which has a value of 1, is highly correlated, while the block colors, which have a value of grey, are much less so. The color block orange indicates the relationship between Glucose and anxiety, which is slightly higher than the pressure with arterial blood pressure and strain with distention, which are color blocks accompanied by purple and brown.

#### A. Equations for evaluating the outcomes

The accuracy is determined using the confusion matrix, which is then used to evaluate the True Positive, True Negative, False Positive, and FalseNegative.

True Negative (TN): Whenever a case was found to be negative and was projected to be negative.

FP stands for False Positive, which occurs when a case is negative but projected to be positive.

Whenever a case proved positive and was projected to be positive, it was referred to as a True Positive.

False Negative (FN): When a case was positive but the outcome was projected to be negative.

TN = confusion matrix of [0,0]FP = confusion matrix of [0,1]TP = confusion matrix of [1,1]FN = confusion matrix of [1,0]

The formula for calculating accuracy is  $accuracy = (TP + TN) / (TP + FP + TN + FN) * 100$ , whereas the formulae for precision, recall, and f1 score are  $precision = TP / (TP + FP)$ ,  $recall = TP / (TP + FN)$ , and  $f1\ score = 2 * precision * recall / (precision + recall) * 100$ , and  $specificity = TN / (TN + FP) * 100$ , respectively.

**B. Logistic Regression**

```

*****Start Evaluate Logistic Regression*****
Testing score of trained LR model Is : 98.7012987012987
Confusion Matrix:
[[67 1]
 [ 1 85]]

Accuracy : 98.7012987012987
Precision: 0.9883720930232558
Recall : 0.9883720930232558
F1 score : 98.83720930232558
Specificity : 98.52941176470588

Classification Report:
      precision    recall  f1-score   support

     0         0.00      0.00      0.00         0
     1         1.00      0.56      0.72        154

   accuracy          0.56        154
  macro avg          0.50      0.28      0.36        154
 weighted avg          1.00      0.56      0.72        154

*****End Evaluate Logistic Regression*****

```

Table1. The training model's testing score for logisticregression accuracy is 98.70 percent.

**C. Decision Tree**

```

*****Start Evaluate Dession Tree*****
Testing score of trained Dession Tree model Is : 97.40259740259741
Confusion Matrix:
[[66 2]
 [ 2 84]]

Accuracy : 97.40259740259741
Precision: 0.9767441860465116
Recall : 0.9767441860465116
F1 score : 97.67441860465115
Specificity : 97.05882352941177

Classification Report:
      precision    recall  f1-score   support

     0         0.00      0.00      0.00         0
     1         1.00      0.56      0.72        154

   accuracy          0.56        154
  macro avg          0.50      0.28      0.36        154
 weighted avg          1.00      0.56      0.72        154

*****End Evaluate Dession Tree*****

```

Table2. The training model's accuracy testing score forDecision Trees is 97.40 percent.

**D. KNN**

```

*****Start Evaluate KNN*****
Testing score of trained KNN model Is : 98.7012987012987
Confusion Matrix:
[[67 1]
 [ 1 85]]

Accuracy : 98.7012987012987
Precision: 0.9883720930232558
Recall : 0.9883720930232558
F1 score : 98.83720930232558
Specificity : 98.52941176470588

Classification Report:
      precision    recall  f1-score   support

     0         0.00      0.00      0.00         0
     1         1.00      0.56      0.72        154

   accuracy          0.56        154
  macro avg          0.50      0.28      0.36        154
 weighted avg          1.00      0.56      0.72        154

*****End Evaluate KNN*****

```

Table3. The training model's test score for KNN accuracy is 98.70 percent, as can be seen.

E. Random Forest

```

*****Start Evaluate Random Forest*****
Testing score of trained Random Forest model Is : 98.7012987012987
Confusion Matrix:
[[67  1]
 [ 1 85]]

Accuracy : 98.7012987012987
Precision: 0.9883720930232558
Recall   : 0.9883720930232558
F1 score : 98.83720930232558
Specificity : 98.52941176470588

Classification Report:
      precision    recall  f1-score   support

     0       0.00      0.00      0.00         0
     1       1.00      0.56      0.72        154

   accuracy          0.50      0.28      0.36        154
  macro avg          0.50      0.28      0.36        154
 weighted avg          1.00      0.56      0.72        154

*****End Evaluate Random Forest*****

```

Table4. We can see that the trained model's Random Forest accuracy testing result is 98.70 percent.

F. SVM

```

*****Start Evaluate SVM*****
Testing score of trained SVM model Is : 98.7012987012987
Confusion Matrix:
[[67  1]
 [ 1 85]]

Accuracy : 98.7012987012987
Precision: 0.9883720930232558
Recall   : 0.9883720930232558
F1 score : 98.83720930232558
Specificity : 98.52941176470588

Classification Report:
      precision    recall  f1-score   support

     0       0.00      0.00      0.00         0
     1       1.00      0.56      0.72        154

   accuracy          0.50      0.28      0.36        154
  macro avg          0.50      0.28      0.36        154
 weighted avg          1.00      0.56      0.72        154

*****End Evaluate SVM*****

```

Table4. We can see that the training model's SVM accuracy testing score is 98.70 percent.

G. Comparison Graph

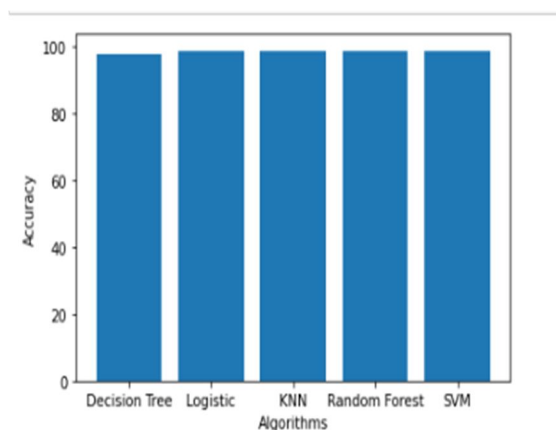


Table5. compares five algorithms in terms of performance

The above image depicts a performance comparison of five algorithms: Decision Tree, Logistic Regression, KNN, SVM, and Random Forest. The decision tree algorithm scored 97.40 percent accuracy, while the other four algorithms scored around 98.70 percent. We chose the KNN model as the final prediction model, despite the fact that the other three algorithms had similar accuracy.

### VIII. CONCLUSION

Psychological stress has been shown to be harmful to one's health. During face-to-face interviews, conversations, and other activities, stress is identified in the existing system. a situation in which one person evaluates two or more people. Regarding work environment, marital status, and clinical measures such as systolic, diastolic, and glucose, this research provides a systematic framework for evaluating users' mental stress states.

### IX. FUTURE WORK

We utilized a jupyter notebook to construct the software, and it was a success. In Python, our project has been successfully tested. We also looked into the project's uses and future scope. Our solution can be by using the API we can build a mobile application where employees can check whether they are stressed or not easily.

### REFERENCES

- [1] Reshmi Gopalakrishna Pillai, Mike Thelwall, Constantin Orasan, "Detection of Stress and Relaxation Magnitudes for Tweets", International World Wide Web Conference Committee ACM, 2018.
- [2] Huijie Lin, Jia Jia, Jiezhon Qiu, "Detecting stress based on social interactions in social networks", IEEE Transactions on Knowledge and Data Engineering, 2017.
- [3] Mike Thelwall TensiStrength, "Stress and relaxation magnitude detection for social media texts", Information Processing & Management Science direct, 2017.
- [4] Jiawei Han, Micheline Kamber, Morgan Kaufman "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, 2003.
- [5] Munmun De Choudhury, Michael Gamon, Scott Counts, Eric Horvitz, "Predicting Depression via Social Media", Seventh International AAAI Conference on Weblogs and Social Media, 2013.
- [6] Thelwall Mike, Kevan Buckley, Paltoglou, "Sentiment strength detection for the social Web", Journal of the American Society for Information Science and Technology, 2013
- [7] Thelwall M., Buckley, Paltoglou, Cai, Kappas, "Sentiment strength detection in short informal text". Journal of the American Society for Information Science and Technology, 2010.
- [8] Dr. G V Garje, Apoorva Inamdar, Harsha Mahajan, "Stress Detection And Sentiment Prediction: A Survey", International Journal of Engineering Applied Sciences and Technology, 2016.
- [9] Alok Ranjan Pall, Diganta Saha, "Word Sense Disambiguation: A Survey", IJCTCM Vol.5 No.3, July 2015.





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)