JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# "A Full-fledged Real Time Application to Find IT Employee Stress using Efficient Machine Learning Algorithms"

**Kavya P,**
Vidya Vikas Institute Of Engineering and Technology, Mysuru

**Abstract**

Today's world is full of tension. All are under stress in some situation due to one or the other reasons. There are numerous factors which effects stress in an human being. In IT sectors employees are more common to go under stress because of work pressure, overload, higher employees domination etc.. If a person is under stress then that may lead to so many mental health problems such as depression, anxiety, somatization, lack of concentration etc.. some times even it any lead to death. So it is required to identify human stress at early stages, so that proper solutions can be given and can get rid of stress. Many research works done on this stress prediction. Most of the research papers uses ML techniques for stress prediction and many papers uses IOT based sensors to extract the features required for stress prediction. Many papers just presented an idea of stress prediction, but no implementation is done. There are some search papers where implementation is done. These implementation papers uses some ready tools such as WEKA tool, R Tool, Rapid Miner or Programming languages such as PYTHON or R Language. Using these ready tools and languages it is easy to predict the stress as it supports ready libraries for stress prediction. Data science techniques are efficient to process training datasets and can predict human stress in less time with better results.

**Keyboard:** Data Science, Machine Learning, Stress, IT profession.

## Introduction

Today's world is full of IT and IT companies have more scope and demand. With changing lifestyle and work cultures, there is an increase in the risk of stress among the employees. Systems finds factors that strongly determine the stress levels. Stress were identified based on gender, family history and availability of health benefits in the workplace. By identifying the stress of employees, we can come up with some approaches to reduce stress and create a much comfortable workplace for their employees.

Many research works uses many parameters such as gender, age, family history, e_provided health benefits, share about illness, tech company, tech role, acquiring leave etc.. Research works uses machine learning algorithms or AI algorithms to find stress of an employee. Finding the risk factors which affects the employee mental health is the major objective of all these research papers.

**Datasets:**

**Datasets download from**

1.https://www.kaggle.com/datasets

2.https://github.com/awesomedata/awesome-public-datasets

3.https://data-world.com./

4.https://archive.ics.uci.edu/ml/datasets.php

5.https://datasetsearch.research.google.com/

**Size: More then 2500 records used.**

**Parameters List**

Gender - 1-Male,2-Female

Age - numeric

Financial_Issues - 0-No,1-Yes

Family_Issues - 0-No,1-Yes

working_Hours - numeric (in hrs)

Learning_Method - 1-Fair,2-Not Good

Health_Issues - 1-Spectacles wearer, 2-Migraine Headache

Partiality_Fix - 0-No,1-Yes

Collegue_Issue - 0-No,1-Yes

Pressure - 0-No,1-Yes

Regular - 1-Regular,2-Irregular

Interaction - 1-Poor,2-Good, 3-Better, 4-Best

Result - 0-Stress Free, 1- Under Stress.

**Related Works**

**1. Title: Classification Algorithms based Mental Health Prediction using Data Mining**

**Authors:** Vidit Laijawala, Aadesh Aachaliya, Hardik Jatta, Vijaya Pinjarkar

**Year of Publications:** 2020

**Description :**The emotional, psychological and social welfare of a person is revealed by their mental health. It influences how an individual will think, feel or handle a situation. Numerous factors contribute to mental health issues which lead to mental illness like stress, social anxiety, depression, obsessive compulsive disorder, drug addiction, workplace issues and personality disorders.

*Disadvantages*

- Small datasets used which leads to less accurate results.
- Data mining techniques used for mental health prediction.
- Huge data is required.

**2. Title: Machine Learning Techniques for Stress Prediction in Working Employees**

**Authors:** U SRINIVASULU REDDY, ADITYA VIVEK THOTA, A DHARUN

**Year of Publications:** 2020

**Description :** Stress disorders are a common issue among working IT professionals in the industry today. With changing lifestyle and work cultures, there is an increase in the risk of stress among the employees. In this paper, we would like to apply machine learning techniques to analyze stress patterns in working adults and to narrow down the factors that strongly determine the stress levels..

*Disadvantages*

- Less parameters used for stress prediction
- Boosting algorithm is not suitable for real time application.
- Ready Tools used for stress prediction.

**3. PREDICTIVE ANALYSIS OF STUDENT STRESS LEVEL USING NAÏVE BAYESIAN CLASSIFICATION ALGORITHM**

**Author** : Monisha S, Meera R,Vijay Swaminath.R, Dr.Arun Raj L

**Year**: 2020

**Description :** Overall academic performance and social compulsion have created a pressurized psychological state for students.Frequently stated stress factors should be reduced to help the students to excel in their academic performance and activate the social activities, thereby decreasing individual health issues like migraine headaches, spectacle wearing and so on..

*Disadvantages*

- This concept predicts stress of only college students, cant be applicable for working employees.
- Algorithms used here takes more time processing.
- Less efficient results.

**Difference between Existing Works and Proposed Work**

- Many existing research works presented an idea but no implementation is done.
- In many existing works implementation is done but Algorithms used were not programmed, they have used ready libraries for algorithms and tools used for algorithms. But in the proposed system we program the algorithm means we write our own logic for the algorithm and results will be tested.
- All Existing works used static datasets, but in the proposed system we build real world project with ML model for dynamic datasets.
- Many research works uses less amount of training data-sets, in the proposed system we use huge data-sets for processing.
- All existing works uses PYTHON or R Language or Ready Data science tools for prediction and which works for static datasets, but in the proposed system we implement the concept for dynamic datasets (real time application).
- All existing works are just model development, can't be used in real time. Here we build this concept as real time application using front end technology as "*visual Studio*" and back end technology as "*SQL Server*" and C# or C++ as programming language.

- ❖ Proposed system is a real world application with model using Microsoft technologies.
- ❖ Proposed system uses more parameters, all existing works used less number of parameters.
- ❖ Proposed system is a GUI based software for an IT company which is not done before. All existing works are just machine learning models.
- ❖ Proposed system predicts employee stress and also stress levels and provide solutions to reduce stress. Existing works just classifies employee into stress or stress free but no stress level prediction and also no solution.

**Proposed System**

1. Proposed system meant for business sector.
2. Systems finds factors that strongly determine the stress levels.
3. Stress were identified based on gender, family history and availability of health benefits in the business sector.
4. By identifying the stress of working employees, we can come up with some approaches or solutions to reduce stress and create a much comfortable working place.
5. System uses many parameters such as Gender, Age, Financial_Issues, Family_Issues, Working_Hours, Learning_Method, Health_Issues, Partiality_Fix, Colleagues_Issue, Pressure, Regular, Interaction etc..
6. System uses machine learning algorithms or AI algorithms to find stress of an working employee.
7. System can be developed as an real time application which is useful for companies. As *Visual Studio* and *SQL Server* is more supportive with real time application, we use these technologies for application development.
8. Finding the risk factors which affects the working employee mental health is the major objective of the system.
9. System also gives suggestions for the working employees based on the stress levels.
10. System aims at identifying the factors and predicting the student stress levels and to come up with solutions to reduce the stress level so as to improvise the working employees performances.

**Parameters List**

---

Gender - 1-Male,2-Female

Age - numeric

Financial_Issues - 0-No,1-Yes

Family_Issues - 0-No,1-Yes

working_Hours - numeric (in hrs)

Learning_Method - 1-Fair,2-Not Good

Health_Issues - 1-Spectacles wearer, 2-Migraine Headache

Partiality_Fix - 0-No,1-Yes

Collegue_Issue - 0-No,1-Yes

Pressure - 0-No,1-Yes

Regular - 1-Regular,2-Irregular

Interaction - 1-Poor,2-Good, 3-Better, 4- Best

**Methodology**

**Overview of the work:**

Today's world is full of IT and IT companies have more scope and demand. With changing lifestyle and work cultures, there is an increase in the risk of stress among the employees. Systems finds factors that strongly determine the stress levels. Stress were identified based on gender, family history and availability of health benefits in the workplace. By identifying the stress of employees, we can come up with some approaches to reduce stress and create a much comfortable workplace for their employees.

System uses many parameters such as gender, age, family history, e_provided health benefits, share about illness, tech company, tech role, acquiring leave etc.. System uses machine learning algorithms or AI algorithms to find stress of an employee. System can be developed as an real time application which is useful for IT companies.As Visual Studio and SQL Server is more supportive with real time application, we use these technologies for application development. Finding the risk factors which affects the employee mental health is the major objective of the system.

**Step 1 : Data Collection (Stress Data)**

This is the first step in the employee stress prediction process where we collect stress data. Data collected from many sources which contains parameters such as Gender, Age, Financial_Issues, Family_Issues, working_Hours, Learning_Method, Health_Issues, Partiality_Fix, Colleague_Issue, Pressure, Regular, Interaction etc…

**Step 2: Data Preparation**

Here stress data analyzed and only relevant data extracted. The data required for processing extracted and segmented according to the requirement. Required data extraction is done because entire data not required for processing and if we input all data, it requires too much of time for processing, so data processing is done.

**Step 3: Specify Constraints**

Stress parameters used for stress level prediction are fetched. Parameters such as Gender, Age, Financial_Issues, Family_Issues, working_Hours, Learning_Method, Health_Issues, Partiality_Fix, Colleague_Issue, Pressure, Regular, Interaction etc..

**Step 4: ML Algorithms - Supervised Learning**

**Supervised learning** is an approach to machine_learning that is based on training data that includes expected answers.

**KNN Algorithm and NaiveBayes or Decision Tree**

"*KNN Algorithm and Naive Bayes or decision tree*" is used for stress prediction because of the following reasons;

1. efficient classifier

2. works fine for less number of parameters as well as more number of parameters.

3. Works fine for small data-set as well as big data-set.

4. more accurate results

**Step 5: Stress Prediction**

System predicts the stress of working employee based on the parameters using machine learning algorithm. We use 1 algorithm for stress prediction "KNN algorithm and Naive Bayes algorithm".

**Step 6: Results**

Here we find the accuracy of the algorithm by dividing the training datasets into training and testing datasets. 90% considered as training datasets and 10% considered as testing datasets.

**Step 7: Visual Representation**

Outputs displayed for the users on GUI.

**Experiment Results**

**Stress using NB Algorithm Results**

| Constraint | Naive Bayes |
|---|---|
| **Accuracy** | 95 % |
| **Time (milli secs)** | 19006 |
| **Correctly Classified (precision)** | 95 % |
| **Incorrectly Classified (Recall)** | 5 % |

**Discussion**

Here we build a real time application useful for the society. This project build using Microsoft technologies. Stress Training datasets trained using Naive Bayes algorithm and we got very good results. Naive Bayes algorithm is programmed in such a way that, it works for dynamic datasets. Naive Bayes algorithm logic is written and it's our own library. We are getting around 95% of accurate results and it takes around 1000 milli seconds for prediction.

**Conclusion**

Gender, family history of illness, and whether an employer gives mental health benefits to their employees had more importance than other parameters in determining whether a person can develop mental health issues. From our study, it was found that people working in a tech company were slightly more at the risk of developing stress even if their role was not tech-based. These insights can be effectively used by corp-orates to frame better HR policies for their employees.

**Reference**

[1] Bhattacharyya, R., & Basu, S. (2018). India Inc looks to deal with rising stress in employees. Retrieved from 'The Economic Times'

[2] OSMI Mental Health in Tech Survey Dataset, 2017 from Kaggle

[3] Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS medicine, 2(10), e267.

[4] Shwetha, S, Sahil, A, Anant Kumar J, (2017) Predictive analysis using classification techniques in healthcare domain, International Journal of Linguistics & Computing Research, ISSN: 2456-8848, Vol. I, Issue. I, June-2017

[5] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266.

[6] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.

[7] Gender and Stress. (n.d.). Retrieved from APA press release 2010