



## Employee Stress Prediction using ML Techniques

<sup>1</sup>Dr. P Vijayakarthishik, <sup>2</sup>Dr. Sheshappa SN, <sup>3</sup>Pamidi Mohammad Sameer, <sup>4</sup>Sameer Ahmed Gadwale, <sup>5</sup>Shravan N, <sup>6</sup>Sumit

<sup>1</sup>Professor and HOD, Information Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India

<sup>2</sup>Associate Professor, Information Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India.

<sup>3,4,5,6</sup> Students, Information Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India

### ABSTRACT

The present world is brimming with pressure. All are under pressure in a circumstance because of either reason. There are various variables that impact pressure in a person. In IT areas, representatives are more likely to go under pressure on account of work pressure, over-burden, higher worker mastery, and so on. 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. It can even be fatal at times. 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 has been done on this stress prediction. Most of the research papers use ML techniques for stress prediction, and many papers use IOT based sensors to extract the features required for stress prediction. Many papers just present an idea of stress prediction, but no implementation is done. There are some search papers where implementation is done. These implementation papers use 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 stress as they support ready libraries for stress prediction. Data science techniques are efficient for processing training datasets and can predict human stress in less time with better results.

Keywords: Data Science, Machine Learning, Stress, IT Profession

### Introduction

The present world is loaded with IT, and IT companies have more extensions and requests. Representatives are bound to encounter pressure because of the changing way of life and working societies. Frameworks recognise components that fundamentally impact feelings of anxiety. Stress was recognised in light of orientation, family ancestry, and accessibility of medical advantages in the work environment. By recognising the pressure of representatives, we can concoct a few ways to deal with it and make a much more agreeable work environment for their representatives.

Many research works utilize numerous boundaries, for example, orientation, age, family ancestry, gave medical advantages, share about ailment, tech organization, tech job, gaining leave and so on. Research work utilises AI calculations or AI calculations to track down the pressure of a worker. Finding the gambling factors that influence the worker's emotional wellness is the significant goal of every one of these exploration 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 than 2500 records used.**

#### List of Parameters

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
Colleague_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: *Data Mining based Classification Algorithms for Mental Health Prediction*

**Authors:** Aadesh Aachaliya, ViditLaijawala, Hardik Jatta, and Vijaya Pinjarkar.

**Publication year:** 2020

**Description:** The mental health of an individual reveals their emotional, psychological, and social well-being. It influences a person's thoughts, feelings, and reaction to situations. Stress, social anxiety, depression, OCD, drug addiction, problems at work, and personality disorders are a few of the factors that affect mental health issues and mental illness.

#### Disadvantages:

- Results are less accurate because of the use of small datasets.
- Data mining methods for predicting mental health.
- A lot of data is needed.

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

**Authors:** U Srinivasulu Reddy, Aditya Vivek Thota, A Dharun

**Publication year:** 2020

**Description:** In today's industry, stress disorders are a common problem for working IT professionals. Because of shifting lifestyles and workplace cultures, employees are now more likely to experience stress. In this article, we'll use machine learning techniques to examine stress patterns in working adults and highlight the factors that have the biggest effects on stress levels.

#### Disadvantages:

- For stress prediction, fewer parameters are used.
- The boosting algorithm is not suitable for real-time use.
- Prediction of stress using Ready Tools.

#### 3. *Predictive Analysis of Student Stress Level Using Naïve Bayesian Classification Algorithm*

**Authors:** Monisha S, Meera R, VijaySwaminath.R, Dr.Arun Raj L

**Publication year:** 2020

**Description:** The combination of overall academic performance and social pressure has put students under pressure psychologically. In order to help students succeed academically and engage in social activities, it is important to lessen the stressors that are frequently cited. This will help people experience fewer personal health problems like migraine headaches, wearing glasses, and other issues.

#### Disadvantages:

- This idea only predicts the stress of college students; it is not applicable to working people.
- The algorithms here require more processing time.
- Less effective outcomes.

### Proposed System

1. The proposed system is intended for the commercial sector.
2. Systems pinpoint variables that have a big impact on stress levels.
3. Based on gender, family history, and the presence of health benefits in the corporate world, stress was found.
4. By understanding the stress experienced by employees, we may develop strategies or solutions to lessen stress and improve the working environment.
5. The system takes into account a variety of factors, including gender, age, financial concerns, family concerns, working hours, learning methods, health concerns, issues with colleagues, pressure, regularity, and interaction, among others.
6. The system looks for employee stress using machine learning or AI algorithms.
7. The system can be created as a real-time application that benefits businesses. Since they are more real-time application supportive, we choose

Visual Studio and SQL Server for application development.

8. The main goal of the system is to identify the risk factors that have an impact on the mental health of working employees.
9. Based on the stress levels, the system also makes recommendations to the working staff.
10. The system's objective is to pinpoint the root causes and project future levels of student stress. We conceive of methods to lower stress levels so that the working employees' performances can be improved.

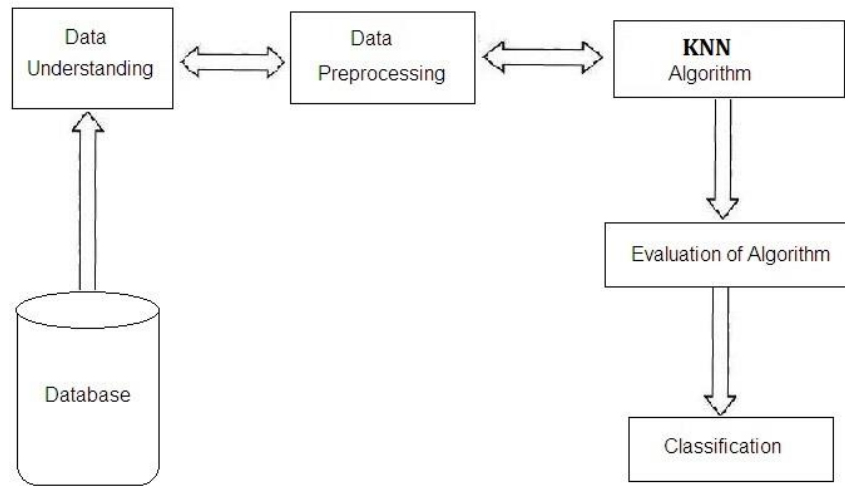


Fig 1. System Architecture of the Proposed System

## Methodology

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

We gather data on stress in this stage of the employee stress prediction process. Information that has been gathered from a variety of sources and includes variables like gender, age, financial, family, working hours, learning method, health, partiality fix, colleague issues, pressure, regular interaction, etc.

### Step 2: Data Preparation

Only significant data was collected after analysis of the stress data. Data that is needed for processing is extracted and segmented in accordance with the specifications. Because complete data is not needed for processing and if we entered all data, processing would take too long, necessary data extraction is performed.

### Step 3: Specify Constraints

The stress parameters that are used to predict stress levels are fetched. Gender, age, financial problems, family problems, working hours, learning methods, health problems, partiality problems, colleague problems, pressure problems, regular problems, interaction problems, etc.

### Step 4: ML Algorithms

#### ▪ Supervised Learning

A method of machine learning called supervised learning is based on training data that contains predicted responses.

#### ▪ The KNN Algorithm and Naive Bayes, as well as the Decision Tree

1. a strong classifier.
2. It performs well with fewer parameters as well as many parameters.
3. effective with both small and big data sets.
4. more precise outcomes.

### Step 5: Stress Prediction.

The system uses a machine learning algorithm to predict the stress levels of working employees based on the factors. KNN algorithm and Naive Bayes algorithm are the two algorithms we employ to predict stress.

### Step 6: Obtaining Results

Here, we separate the training datasets into training and testing datasets to determine the algorithm's correctness. 90% of the datasets are regarded as training datasets, and 10% as testing datasets.

### Step 7: Visual Representation

The GUI displays the output to the user.

Users:

- ▶ **Administrator** - the one who maintains the whole application. the one who has full authority.
- ▶ **Employees** - working employees are service receivers and the one who can post any queries to admin.

### Admin Modules

1. **Login Module** - here admin gets login to the application by inputting admin id and password.
2. **Add Employees** - here admin will add all employees of different departments.
3. **Set Id and Password** - here admin can set unique id and password for the individual working employee.
4. **View Employees** - here admin can view all the employees that are present in the organization along with their details.
5. **Manage Dataset** - here admin can view the training dataset used in the project which is imported from an excel file.
6. **Stress Prediction Module** - this is the core module where system finds the stress problems faced by working employees.
  - Here we use data science technique called as "supervised learning" KNN algorithm. (YES OR NO PREDICTION)
  - Admin can predict output for the testing dataset stored in an excel sheet in order to analyze and evaluate the accuracy and performance of the system.
  - Result Analysis (accuracy and efficiency) for the testing dataset imported from excel sheet.
7. **Solution Module** – Admin can suggest and manage solutions for each employee in the organization based on the different stress level groups.
  - The administrator can manage all the solutions for each stress level group.
8. **Queries** - admin can view the employee queries and send reply. Queries module also has pending and answered queries modules, to view both respective queries.

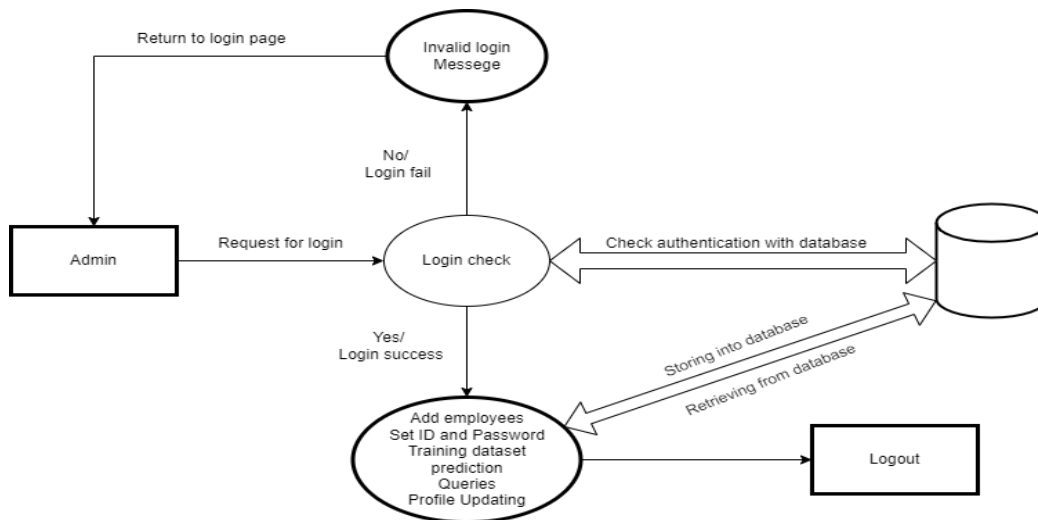


Fig. 2 Admin module dataflow Diagram

### Employee Modules

1. **Login Module** - Users can login by using specific employee id and password.
2. **Employee Stress Module** – user will fill in all the parameters required for stress level prediction.
3. **Stress Prediction Module** - this is the core module where system finds the stress problems faced by working employee. Here we use data science technique called as "supervised learning" KNN algorithm.
4. **Solution Module** - here system will find some solutions defined by admin for the stress problems of working employees.
5. **Queries Module** – It is same as in the admin module only change is Post queries module.
  - **Post Queries** – User can post queries to admin if any and view responses.

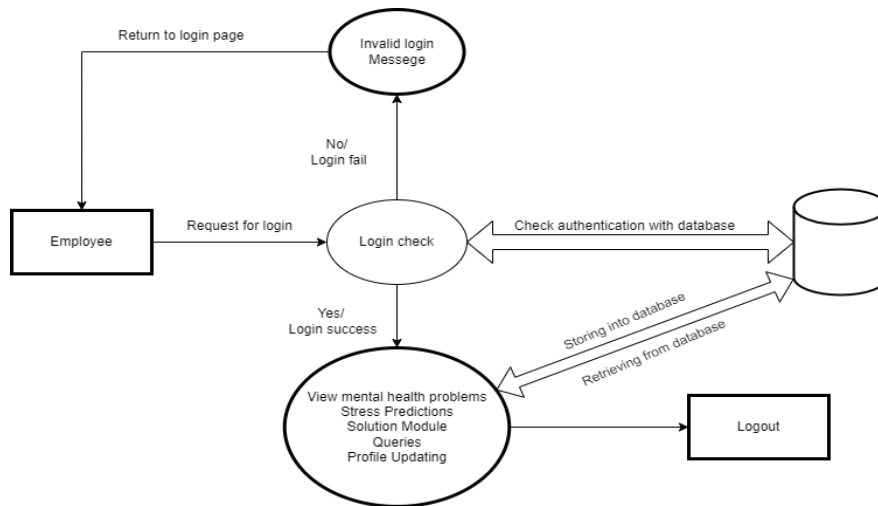


Fig.3 Employee module dataflow Diagram

**Project Outcome**

**Input:** System uses many parameters such as gender, age, family history, provided health benefits, share about illness, tech company, tech role, acquiring leave etc. and Old data-sets for processing.

**Output:** classifies the employees into Stress and Stress Free.

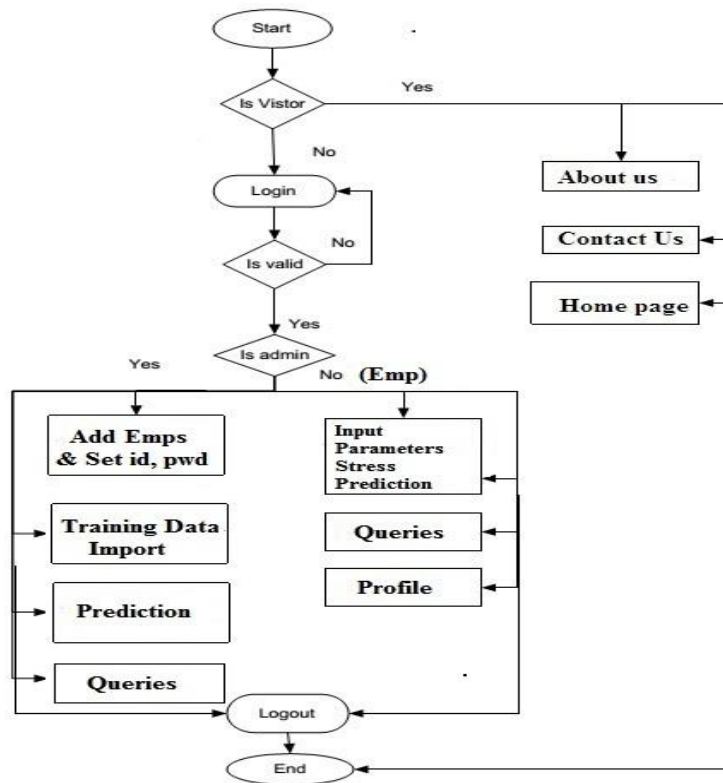


Fig. 5 System flow chart

**Conclusion:**

When evaluating whether a person may experience mental health problems, factors like gender, family history of disease, and whether an organisation offers mental health benefits to its staff were more significant than other considerations. According to the results of our study, even those who did not have tech-related jobs were marginally more likely to experience stress than those who did. Corporations can effectively use this data to develop better HR policies for their employees.

---

**Reference**

---

- [1]. [1] Basu, S., and R. Bhattacharyya (2018). India Inc. is working to reduce the mounting stress among employees. drawn from "The Economic Times."
- [2]. [2] Dataset from Kaggle's 2017 OSMI Mental Health in Tech Survey.
- [3]. [3] J. Van den Broeck, S. A. Cunningham, R. Eeckels, & K. Herbst (2005). Data cleaning entails identifying, diagnosing, and correcting data irregularities. 2(10), e267 in PLoS medicine.
- [4]. [4] Predictive analysis utilising categorization techniques in the healthcare industry. Journal of Computing and Linguistics International, ISSN: 2456-8848, Vol. I, Issue. I, June-2017.
- [5]. [5] Tomar, D., and S. Agarwal (2013). a review of data mining techniques used in healthcare. 241-266 in International Journal of Bioscience and Biotechnology, volume 5(5).
- [6]. [6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, & J. Vanderplus.