Report

Abstract

Increasing amount of workload and changing lifestyles contribute to high stress level among employees.Employees working in IT industry are prone to develop a lot of health problems due to continuous physical and mental stress of their work. Diseases are either induced, sustained or exacerbated by stress. A lot of minute factors contribute towards this stressful nature like gender,employee status,income,etc.

Here the data from "IBM " was taken into consideration.This project aims towards determing the factors that strongly contribute towards stress levels by analyzing using various Machine Learning techniques like decision tree,regression,etc after data preprocessing. Using deep learning techniques like CNN (Convoluted NeuralNetworks) and verify how the model performs for the given dataset is also the goal ,this project focuses on implementing various techniques and analyzing them also gives an insight of the accuracy provided by these algorithms as well as some hidden attributes contributing to the same.This can be useful for the industry as well the employess to analyze thereby reducing stress levels and provide better environment in the workplace.

Literature Survey

In the paper referred, they have worked on OSMI Mental health record. Moreover, they have used algorithms like logistic regression, Decision Tree, Random forest, Bagging and boosting. They have compared accuracy using all these methods. They achieved highest accuracy using random forest. The authors suggested for future

scope to use neural networks in order to improve the accuracy of the model.

Statement of Proposal

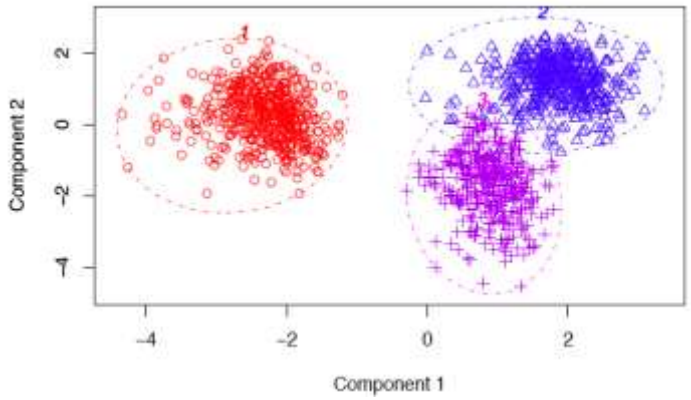
To determine the level of stress in working employees by using supervised and unsupervised algorithms in machine learning and compare accuracy between SVM and Neural Networks.

Results on sample set of data

Algorithms and Techniques

1. K means clustering algorithm

This algorithm is used in our project for clustering. Using this algorithm we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as euclidean-based distance or correlation-based distance. We have two clusters in our project



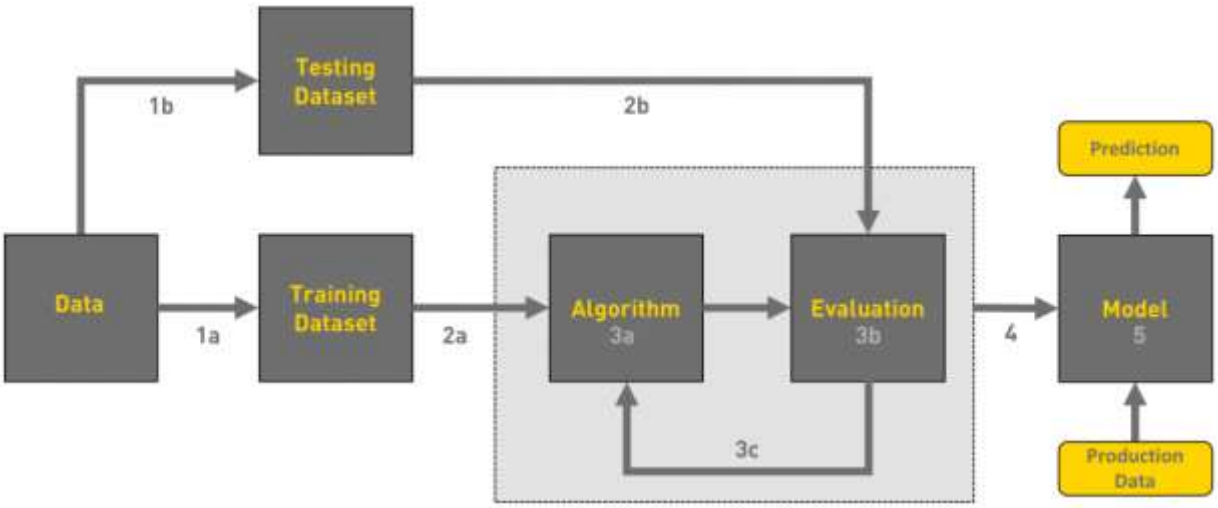
2. SVM

A support vector machine (**SVM**) is a supervised machine learning model that uses classification algorithms for two-group classification problems. In this project SVM is used for classifying into two groups whether the employee is stressed or not. Using SVM we have achieved an accuracy of 92 percent.

3. Neural networks

Neural networks is not yet implemented in this project. But the next plan is to implement neural networks and compare accuracy with respect to SVM. Neural networks will be helpful in increasing the accuracy of the model.

COMPLETE SCHEMA


Testing Dataset
1b
2b
Prediction
Data
Training Dataset
1a
2a
Algorithm
3a
Evaluation
3b
4
Model
5
3c
Production Data