A STUDY ON STATISTICAL ANALYSIS AND VALIDATION OF STUDENT STRESS LEVEL USING ANN-CASE STUDY

E.Manokaran[1] , G.Muthu guru rathinesh [2], S.Senthilvel[3] , R.Muruganandham[4] , M.Vijoe [5] & K.Ravichandran[6]

Abstract: *In current corporate scenario, the performance of human beings has become more challenging due to several aspects like work environment, peers competition, organizational conflicts, working stress, project deadlines, etc, . Among all factors stress plays a major role in motivating or de-motivating an individual. Our paper has tried to peruse the origin, nature and effects of stress for budding professionals like engineering students. The paper has categorized the factors responsible for causing the stress to the professional students of Information Technology. Optimization is an active and fast growing research area and has a great impact on the real world .Optimization technique is one of the popular method to find a solution for complex decision-making problem and also finding ranking performance under the given constraints, by maximizing desired factors and minimizing undesired ones. In this Paper Optimization technique ANN tool has been used to identify the stressed professionals. The results have been validated using Statistical analysis.*

KeyWords: 1. Multi-criteria Decision Making (MCDM) 2. Artificial Neural Network (ANN) 3. Optimization Technique. 4. Statistical Analysis 5. Feed Forward algorithm.

[1] P.G Student , Thiagarajar College of Engineering, Madurai, TamilNadu, India , manoips@rediffmail.com
[2] U.G Student , Thiagarajar College of Engineering, Madurai, TamilNadu,India gsakthishankarr@gmail.com
[3] U.G Student , Thiagarajar College of Engineering, Madurai, TamilNadu, India, senthilveltce@gmail.com
[4] Faculty in Mechanical Engineering, Thiagarajar College of Engineering, Madurai, TamilNadu, India, rmmech@tce.edu
[5] U.G Student , Thiagarajar College of Engineering, Madurai, TamilNadu, India, Josephvijoe@gmail.com
[6] Professor in MKU, Madurai Kamaraj University, TamilNadu,India, ravimba_mku@yahoo.co.in

1.0 Introduction

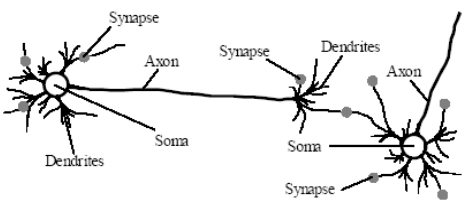
Optimization is an active and fast growing research area and has a great impact on the real world. Despite of the enormous amount of work that has been conducted both theoretically and empirically and the huge success that has been achieved in different aspects, it is still an ongoing and long-term task to develop competent techniques, which could effectively solve large-scale optimization problems. Optimization technique is one of the popular method to find a solution for complex decision-making problem and also finding ranking performance under the given constraints, by maximizing desired factors and minimizing undesired ones. In computer simulation (modelling) of business problems, optimization is achieved usually by using linear programming techniques of operations research .In this paper, ANN methodological tool have been used to validate the stress causing factors and to rank according to the scores and also compare the results using statistical analysis.

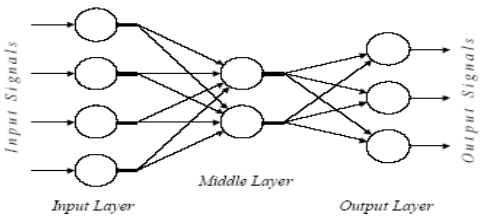
2.0 PROBLEM IDENTIFICATION:

Various optimization techniques are used to solve real time problems. In this ANN tool has been used to rank the stress causing factors according to the results and also compare the results with statistical analysis to determine the most significant factor effect.

3.0 ANN:

An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well. And for the validation process ANN is followed, the human brain provides proof of the existence of massive neural networks that can succeed at those cognitive, perceptual, and control tasks in which humans are successful.





4.0 STATISTICAL ANALYSIS:

Statistics are a tool, not an aim. Simple inspection of data, without statistical treatment, by an experienced and dedicated analyst may be just as useful as statistical figures on the desk of the disinterested. The value of statistics lies with organizing and simplifying data, to permit some objective estimate showing that an analysis is under control or that a change has occurred. Equally important is that the results of these statistical procedures are recorded and can be retrieved.

Statistical approach of analyzing the student Stress data:

Basic Statistics Parameters for Analysis

1. Mean
2. Standard deviation
3. Relative standard deviation. Coefficient of variation
4. Confidence limits of a measurement
5. Propagation of errors.

5.0 DATA COLLECTION & METHODOLOGY:

Methodology:

- Feed Forward Algorithm:

  A feed-forward network has a layered structure. Each layer consists of units which receive their input from units from a layer directly below and send their output to units in a layer directly above the unit. There are no connections within a layer.

Data Collection

For the analysis of students stress level the data has been collected among various students by conducting pilot study, questionnaire technique. Based on the data's most significant factor effect can be determined using ANN.

From the ranking results I have taken IT –Dept to determine the significant factor for causing stress. Hence factor influence to causing stress has shown below.

Rank	Student no	Fat%	BMI	System time	Physical activity in min per day
1	42	23	23	8	0
2	3	18	22	7	10
3	4	23	21	8	0
4	1	22	23	8	0
5	22	23	24	7	0
6	38	21.5	28	8	0
7	32	16	26	4	0
8	17	14	21	8	15
9	8	23	22	8	10
10	35	22	21	3	0
11	11	32	22	4	0
12	10	22	27	4	0
13	31	24	26	3	0
14	23	23	21	8	20
15	34	22	27	8	15
16	7	22.5	26	7	0
17	20	24	21	8	15
18	27	23	23	3	10
19	33	18	22	8	10
20	9	15	22	7	0
21	39	23	23	8	0
22	6	35	21	3	0
23	30	17	20	7	0

24	14	16	28	3	0
25	12	21.5	23	3	0
26	24	22	27	8	20
27	36	15	21	8	15

6.0 Analysis and Results:

Based on the above factors using Back propagation algorithm find weightage values for each factor and take summation value from each factor to determine the most significant factor.

Sample calculation:

Old weight:

W1=1, w2=0.869, w3=0.232, w4=0.484**.**

P=(35*1)+(23*0.869)+(8*0.232)+(0*0.484)

=56.8 +1= 57.8 = Ans *0 .1=5.78.

New weight:

W new= w old +P

W1 new = 6.78, w2 new= 6.649, w3 new= 6.012, w4 new= 6.26.

Rank	Student no	Stress value %	Fat%	BMI in kg/m2	System time spends in hours per day	Physical activity in min per day
1	42	83.56	23.9	15.4	5.0	0.2
2	3	82.1	23.2	14.8	4.4	0.2
3	4	81.2	22.5	14.1	5.0	6.4
4	1	81.1	21.8	15.4	5.0	0.2
5	22	80.2	21.2	14.1	5.0	9.5
6	38	79.8	21.2	13.4	4.4	0.2
7	32	77.1	19.1	14.8	5.0	0.2
8	17	76.2	18.5	15.4	2.6	0.2
9	8	75.4	18.5	14.1	5.0	9.5
10	35	73.4	19.1	14.8	5.0	6.4

11	11	73.2	19.1	14.1	2.0	0.2
12	10	70.4	19.8	14.8	2.6	0.2
13	31	69.5	17.1	18.1	2.6	0.2
14	23	67.1	16.4	17.4	2.0	0.2
15	34	65.4	17.8	14.1	5.0	12.7
16	7	64.72	16.4	16.1	5.0	9.5
17	20	64.71	15.7	15.4	4.4	0.2
18	27	63.8	15.1	14.1	5.0	9.5
19	33	62.3	17.8	15.4	2.0	6.4
20	9	62.21	14.4	13.4	5.0	6.4
21	39	62.2	13.6	14.8	4.4	0.2
22	6	61.4	15.1	15.4	5.0	0.2
23	30	61.1	13.7	14.1	2.0	0.2
24	14	60.7	15.7	13.4	4.4	0.2
25	12	59.8	13.7	15.4	2.0	0.2
26	24	59.8	13.7	15.4	2.0	0.2
27	36	59.04	15.1	14.1	5.0	12.7
28	19	58.92	16.4	14.8	2.0	12.7
29	18	58.91	13.7	15.4	2.0	112.8
30	28	58.18	13.0	13.4	2.0	18.9

From the table take summation value for each factor,

FAT=764 .Hence the factor FAT has more effect.

7.0 Comparison of Results:

Results can be compared using statistical analysis.In statistical analysis correlation value can be determined between students stress level and factors to analysis relation between them and also find the Normal distribution to predict the most significant factor.

(i) Correlation:

The degree of relationship between the variables is said to be correlation. Here correlation is taken between factor and stress value to predict which factor has more correlated.

$R(x,y) = cov(x,y)/\sigma x * \sigma y.$

$Cov(x,y) = 1/n \sum xy - \bar{x}\,\bar{y}.$

$\sigma x2 = 1/n \sum x2 - \bar{x}2.$

$\sigma y2 = 1/n \sum y2 - \bar{y}2.$

By using above formula calculate correlation value for each factor,

FAT = 0.93

BMI = -0.04

TIME SPENDS IN SYSTEM = 0.41

PHYSICAL ACTIVITY = -0.30.

Based on correlation results it seems factor FAT has more related to stress level.

(ii) Normal Distribution:

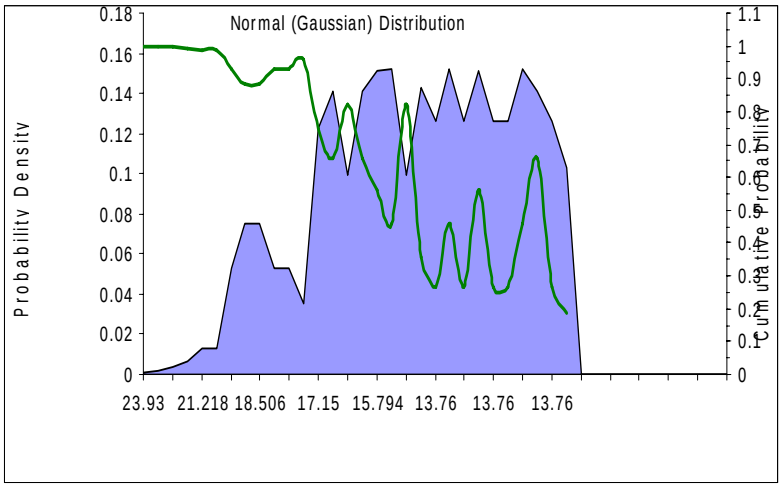
To confirm the correlation results normal distribution is plotted between factor and stress level. Take sample size of 30 students.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

Using above formula find value for f(x) and plot it,

x	***f*(*x*)**
23.93	0.000706
23.252	0.001605
22.574	0.00341
21.896	0.006768
21.218	0.012549
21.218	0.012549
19.184	0.053209
18.506	0.07517
18.506	0.07517
19.184	0.053209
19.184	0.053209
19.862	0.035188
17.15	0.122338
16.472	0.140936
17.828	0.099213
16.472	0.140936
15.794	0.151688
15.116	0.152527
17.828	0.099213
14.438	0.143288
13.76	0.125759
15.116	0.152527
13.76	0.125759

15.794	0.151688
13.76	0.125759
13.76	0.125759
15.116	0.152527
16.472	0.140936
13.76	0.125759
13.082	0.103119



From the graph it identify the factor FAT has more effect and so it reaches high value.

(iii) ANOVA:

Hence it confirmed the factor FAT has more effect on stress level. FAT has various ranges on human body. To determine which range has more adversely affected find ANOVA and compare with residuals.

NULL HYPOTHESIS : Factor FAT has no significant effect on stress level.

ALTERNATE HYPOTHESIS: FAT has significant effect on stress level.

FAT	STRESS						TOTAL **Yi.**	AVG
15	62.2	61.1	59.8	59.0	58.9	58.1	359.1	59.8
20	64.71	63.8	62.2	61.4	60.7	59.04	371.8	61.9
25	69.5	67.1	65.4	64.7	62.2	58.9	387.8	64.6
30	77.1	76.2	73.4	70.4	75.4	73.2	445.7	74.2
35	83.5	81.1	79.8	81.25	82.1	80.2	487.9	81.31
Yj.	357.01	349.3	340.6	336.7	339.3	329.4	**Y.**.=2052.3	
AVG	71.40	69.8	68.1	67.3	67.8	65.8		

1. SS treat = 1/n $\sum$ yi.2-y..2/N
2. SS total = yij2 - y..2/N
3. SS error = SS total – SS treat.

Using above formula find the values ,

Sum of squares	DOF	Mean squares	Fo
SS treat	4	496.72	87.9
SS total	29		
SS error	25	5.65	

Compare with table value at $\alpha = 0.05$, $F_{4,25}=2.75$

$F_0 > F\alpha$

We reject the H_0

(iv) Residuals:

Find the range which has more effect for causing stress.

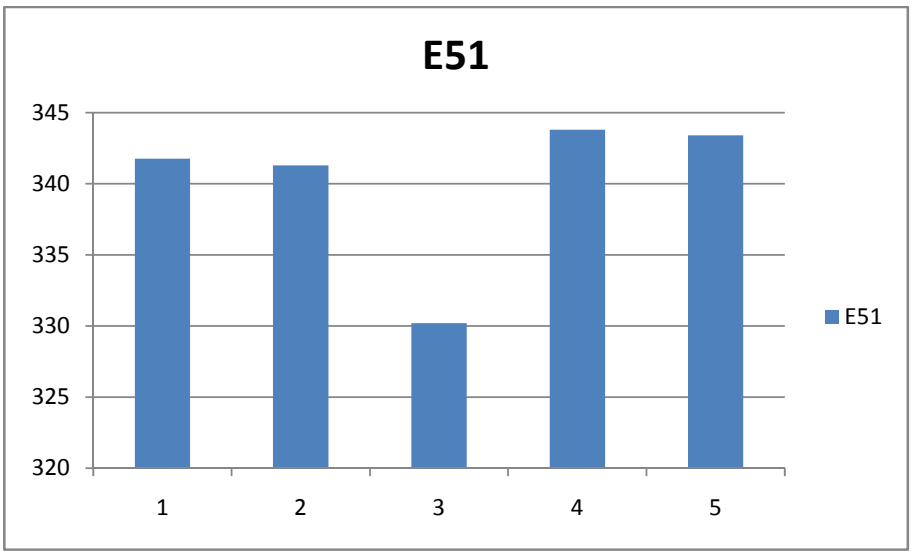
Eij = yij - $\bar{y}$ij.

$\bar{y}$ij = yi.+y.j-y..

Using this formula find the values and plot it.

FAT	E11	E21	E31	E41	E51	E61
15	341.46	341.96	342.36	342.36	341.76	342.96
20	341.8	342.4	339.9	342.5	341.3	341.7
25	343.9	343.1	343.1	343.2	330.2	338.9
30	341.9	342.6	341.5	339.3	343.8	343.6
35	341.2	340.4	340.8	343.1	343.4	343.5

From the above table the residuals E61 have more variation.



Hence finally it conclude if the FAT range is more than 30% in human body then the stress level also get increased.

8.0 Conclusion:

It is quite clear that selection of students with human factor involves a large number of considerations. The use of ANN method is observed to be quite capable and computationally easy to evaluate and select significant effect of stress from a given data. Thus the ANN result is validated and analysis using various statistical techniques and also determine the most significant factor for causing stress.

9.0 References:

[1] P. Ajith Abraham Oklahoma The Book "Artificial Neural Networks",State University, Stillwater, OK, USA

[2] Aizenberg I., Moraga C.: Multilayer feedforward neural network based on multi-valued neurons (MLMVN) and a backpropagation learning algorithm. *Soft Computing* **11** (2), 169-
183, 2007..

[3] Coryn A.L. Bailer-Jones Max-Planck-Institute"*An* introduction to artificial neural networks" f¨ur Astronomie, K¨onigstuhl 17, 69117 Heidelberg, Germany email: calj@mpia-hd.mpg.de Ranjan Gupta IUCAA, Post Bag 4, Ganeshkhind, Pune-411007, India email: rag@iucaa.ernet.in Harinder P. Singh Department of Physics, Sri Venkateswara College, Benito Juarez Road, Dhaula Kuan New Delhi-110021, India email: hps@ttdsvc.ernet.in.

[4] Claudio S. Moraga," Design of Neural Networks**",** European Centre for Soft Computing, 33600 Mieres, Asturias, Spain and Dept. Computer Science, University of Dortmund, 44221 Dortmund, Germany.

[5] Hwang, C.L., Yoon, K.," Multiple Attribute Decision Making Methods and Applications"**.**
Springer, Berlin, Heidelberg, 1981.

[6] Claudio S. Moraga," Design of Neural Networks**",** European Centre for Soft Computing, 33600 Mieres, Asturias, Spain and Dept. Computer Science, University of Dortmund, 44221 Dortmund, Germany.

[7] Laura P. Wobble "IMPACT OF STRESS FACTORS ON COLLEGE STUDENTS ACADEMIC PERFORMANCE" University of North Carolina at Charlotte.

[8] R. Venkata Rao, PhD, Department of Mechanical Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, "Decision Making in the Manufacturing Environment Using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods", Springer.

[9] "*The Week*", June 29, 2008.

.