CEREAL RATING PREDICTION USING NEURAL NETWORK:

A PROJECT REPORT

for

SOFT COMPUTING (SWE1011)

in

M. TECH SOFTWARE ENGINEERING

by

YALAMANCHILI JOSMITHA– 17MIS0050

SIGE SWETHA- 17MIS0084

CHENJI IPSHITHA ROYAL– 17MIS0232

FALL SEMESTER, 2019

Under the Guidance of

Prof. AGILANDEESWARI.L

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Information Technology and Engineering

NOV, 2019

CERTIFICATE

This is to certify that the project report entitled **“CEREAL RATING PREDICTION USING NEURAL NETWORK**

” submitted by **YALAMANCHILI JOSHMITHA (17MIS0050), SIGE SWETHA(17MIS0084) AND CHENJI IPSHITHA ROYAL (17MIS0232)** to Vellore Institute of Technology University, Vellore in partial fulfillment of the requirement for the award of the course **Soft Computing (SWE1011)** is a record of bonafide work carried out by them under my guidance.

Prof. Agilandeeswari L
GUIDE
Asso. Professor, SITE

CEREAL RATING PREDICTION USING NEURAL NETWORK

JOSHMITHA, SWETHA, IPSHITHA

Department of Information Technology, VIT University, Vellore, Tamil Nadu, India

ABSTRACT:

In our project we apply RBF method for the algorithm to predict cereals rating. Prediction is the method of making statements about certain event whose actual results have not been observed. It seems to be an easy process but is not. It requires a lot of analysis on current and past outcomes in order to give timely and accurate timely predicted results. Radial Basis Function (RBF) is a method proposed in machine learning for making predictions and forecasting. It has been used in various real time applications such as weather prediction, load forecasting, prediction about number of tourist and many such applications. Hence, we compare the accuracy of this RBF classification algorithm and the accuracy of the cereal's classification done by BP-MLP algorithm i.e. Back propagation for Multi-layer perceptron.

OBJECTIVE:

- To predict approximate value of cereal price.
- Before an investor invests in any cereal market, he needs to be aware how the stock market behaves.
- To identify factors effecting cereals share market.

KEYWORDS:

Artificial Neural Networks, Back propagation, Forward Propagation, Radial basis function network, Fuzzy set, Raw data

1. INTRODUCTION:

Neural network is an information-processing machine and can be viewed as analogous to human nervous system. Just like human nervous system, which is made up of interconnected neurons, a neural network is made up of interconnected information processing units. The information processing units do not work in a linear manner. In fact, neural network draws its strength from parallel processing of information, which allows it to deal with non-linearity. Neural network becomes handy to infer meaning and detect patterns from complex data sets. Neural network is considered as one of the most useful technique in the world of data analytics. However, it is complex and is often regarded as a black box, i.e. users view the input and output of a neural network but remain clueless about the knowledge generating process.

2. BACKGROUND:

RBF NEURAL NETWORKS:

Radial Basis Function Network (RBFN) is a neural network. The RBFN approach is more intuitive than the MLP. An RBFN performs classification by measuring the input's similarity to examples from the training set. Each RBFN neuron stores a "prototype", which is just one of the examples from the training set. When we want to classify a new input, each neuron computes the Euclidean distance between the input and its prototype. If the input more closely resembles the class A prototypes than the class B prototypes, it is classified as class A. Steps involved in this algorithm:

Choosing an appropriate value for the center.

There are various methods which can be applied for the same, such as:

- Randomly Selecting from the data set
- K means algorithm
- OLS Algorithm

Deciding the activation function to be implemented in hidden layer (Linear, Gaussian)

Deciding the bias/spread value

Adjusting the output weights (Gradient Descent, Least Square etc.) where there is a need for normalization to be done as well.

This can be done by using

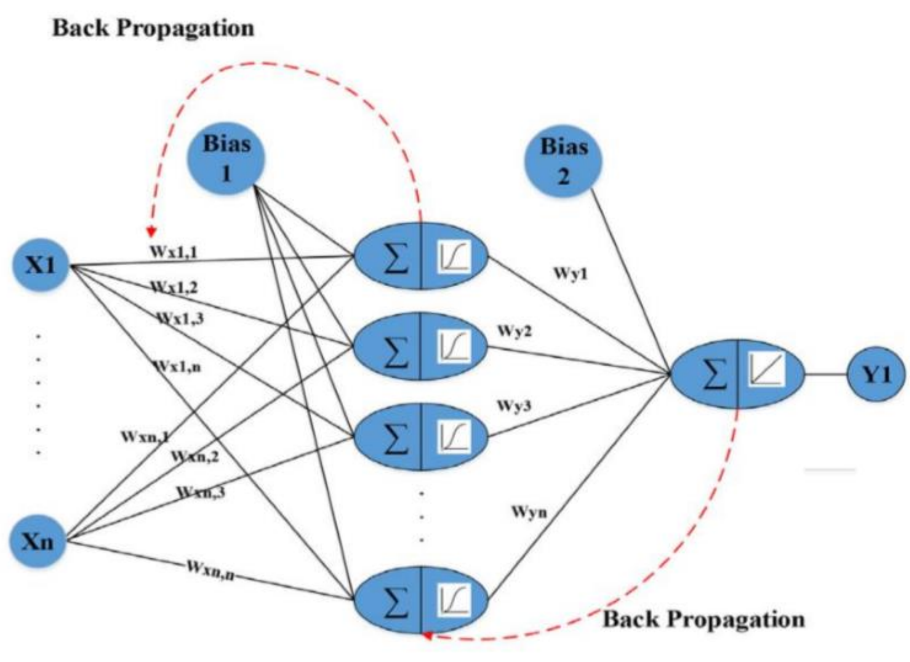
$$\bar{y} = \frac{Y - Ymin}{Ymax - Ymin}$$

Where Y is the actual value of the data sample, Ymax takes the value larger than the foresting year, Ymin takes the value that is minimum from the sample of data.

BACK PROPAGATION ALGORITHM:

Backpropagation is a method used in artificial neural networks to calculate a gradient that is needed in the calculation of the weights to be used in the network. Backpropagation is shorthand for "the backward propagation of errors," since an error is computed at the output and distributed backwards throughout the network's layers. It is a special case of a more general technique called automatic differentiation. In the context of learning, back propagation is commonly used by the gradient descent optimization algorithm to adjust the weight of neurons by calculating the gradient of the loss function. This technique is also sometimes called backward propagation of errors, because the error is calculated at the output and distributed back through the network layers.

ARCHITETURE:



The Backpropagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. The

weights that minimize the error function is then considered to be a solution to the learning problem.

3. LITERATURE SURVEY:

Authors &Year	Methodology or Techniques used	Advantages	Issues	Metrics used
Silvia Soledad Moreno Gutiérrez1 Alfredo Toriz Palacios Sócrates López Pérez3 Abraham Sánchez López YEAR- 2019	This study attempts to improve the prediction capacity of the stock price via an integrated prediction model based on kernel Principal Analysis (KPCA) is used KPCA is firstly introduced to reduce the feature dimensions Support Vector Machines for Regression(SVR) is used SVR is used to build a short-term investment decision system.	Results show that SVR by feature extraction using KPCA can achieve better generalization performance.	High of cost	Evaluation metrics

Mikhail Goykhman, Ali Teimouri YEAR-2018	The study of the simulated stock market framework defined by the driving sentiment processes. We focus on the market environment driven by the buy/sell trading sentiment process of the Markov chain type. We apply the methodology of the Hidden Markov Models and the Recurrent Neural Networks.	Hidden Markov Model allows to successfully reproduce the sentiment transition probabilities matrix.	Uncertainty	Accuracy
Asil Oztekina, Recep Kizilaslanb, Steven Freundc, Ali Iseri d YEAR-2019	This study develops a generic methodology to predict daily stock price movements. Used methods adaptive Neuro-fuzzy inference systems, artificial neural net-works, and support vector machines.	• Stock price forecasting can be effectively performed. • It achieves significantly better forecasting performance in terms of the accuracy rates.	Market fluctuations	Evaluation metrics

Yangseon Kim , Jae-Hwan Roh and Ha Young Kim YEAR- 2018	We propose an artificial-intelligence-based model for rice blast disease prediction. The predictive performance of the proposed LSTM model is evaluated by varying the input variables. Rice blast fungus prediction using the proposed LSTM model is variety-based and the deep learning model used in this study implements a data-driven approach.	• The utility of the LSTM models is expected to be high. • The findings and developed system will be helpful for the various countries in which rice is grown as a primary crop.	This increase in rice blast disease will impure the quality of the rice and increase the rate of disease in the world.	Precision

Hadi Karimi ,Søren Skovsen , Mads Dyrmann and Rasmus Nyholm Jørgensen YEAR-2018	In this study, a system for locating cereal plant stem emerging points (PSEPs) has been developed. The PSEPs of the cereal plants were marked manually and used to train a fully-convolution neural network. An accurateDistribution of the cereals will minimize competition among plants and provide proper conditions for nutrientsand light, leading to a higher yield	• Easy to use • Cost of usage is less	The developed seed evaluation systems, which can only track seeds until the delivery tube of seed drills, could not provide precise locations of crops in the fields.	Accuracy, Specificity

Erkam Guresen,Gulgun Kayakutlu , Tugrul U. Daim YEAR-2019	Analyzed in comparison to classical Multi-Layer Perceptron (MLP) model. The analysed models will be tested on NASDAQ	It helps you to invest wisely to make good profits.	The financial time series models expressed by financial theories have been the basis for forecasting a series of data	Financial metrics

	index data for nine months and the methods will be compared by using Mean Square Error (MSE) and Mean Absolute Deviation (MAD)		in the twentieth century. Yet, these theories are not directly applicable to predict the market values which have external impact.	

Jimenez Daniel, Perez-Uribe Andres, Satizábal Héctor Barreto Miguel , Van Damme Patrick , and Tomassini Marco YEAR-2017	Modelling agro ecological processes	Artificial neural network models showed better performance than traditional approaches.	we often refer to these models as “black boxes”. These models can perform quite well either on classification or regression tasks, but we do not know very well how they are using the input data to come-up	black-box linear parametric models

M. Safa, S. Samarasinghe , and M. Nejat Year 2015	Nonlinear statistical methods (NNs methods, Artificial networks, fuzzy Logics)	NNs can be successfully trained to describe the influence of direct and indirect factors on wheat production	farmers who kept more livestock usually had lower wheat yields than farmers who concentrated on crop production.	Specificity

Mads Dyrmann, Søren Skovsen, Morten Stigaard Laursen1, and Rasmus Nyholm Jørgensen1 YEAR-2018	We addressed the problem of detecting weed instances by using a modified version of the SSD(Solid State Drive) detector. This network has been trained to distinguish monocot weeds and dicot weeds from cereals. For detecting weeds, this study proposes a fully	Detecting weeds using a fully convolution neural network showed problems in handling both monocots and dicots simultaneously.	Weed detection have had trouble with detecting weed instances in cereal fields due to heavy leaf occlusion.	Accuracy, Sensitivity

	convolution neural network and deep learning.			

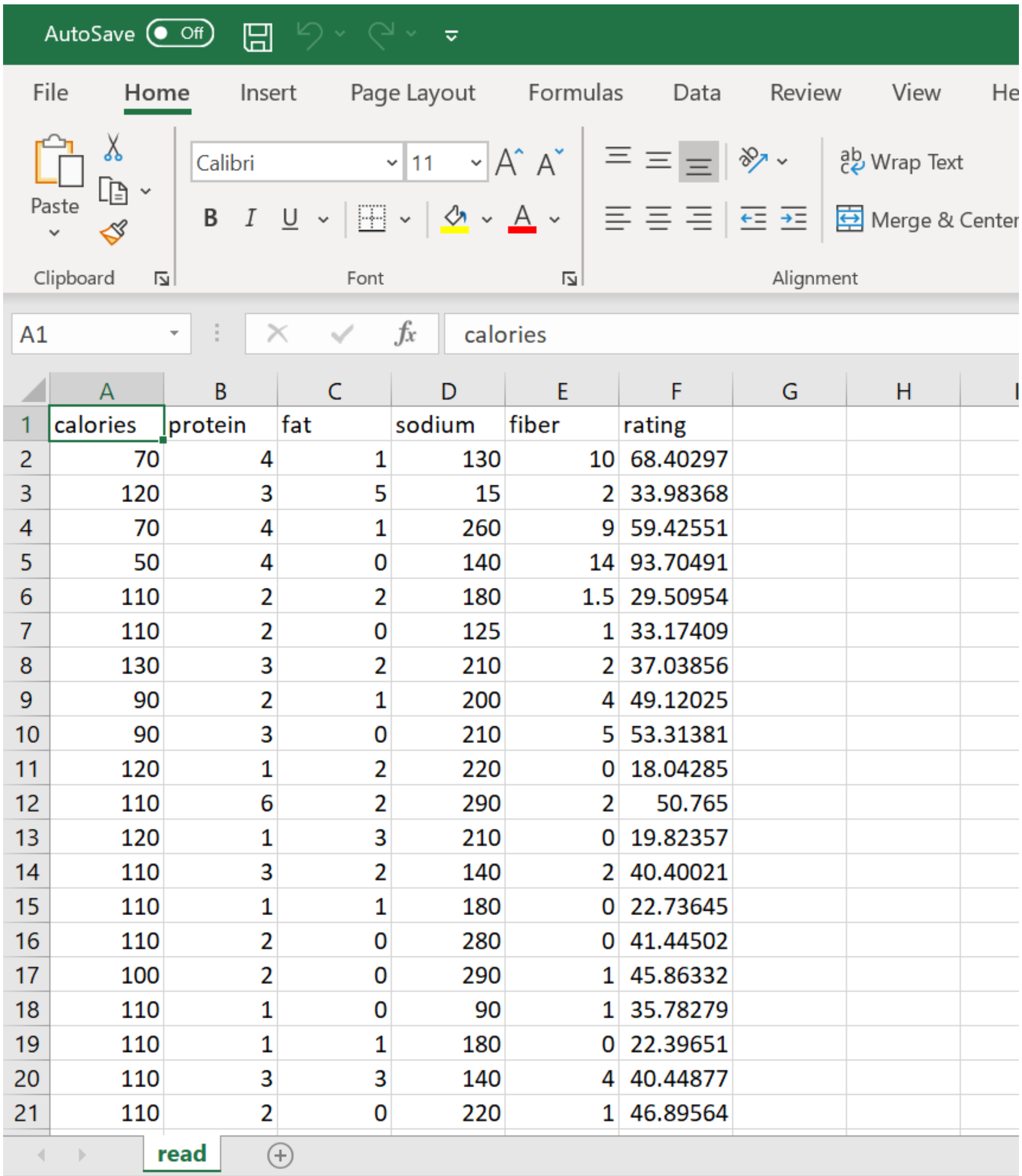
Takashi Kimoto and Kazuo Asakawa Morio Yoda and Masakazu Takeoka YEAR-2016	It is based on modular neural networks and these are being applied to a widely expanding range of applications in addition to the traditional areas such as pattern recognition and control.	This paper has discussed a prediction system that advises the timing for when to buy and sell stocks.	The current prediction system requires much simulation to determine moving average. Difficult in finding accurate values.	None

Zan Huanga, Hsinchun Chena , Chia-Jung Hsua , Wun-Hwa Chenb , Soushan Wuc YEAR-2017	This article introduces a relatively new machine learning technique, support vector machines (SVM), to the problem in attempt to provide a model with better explanatory power. We used back propagation neural network (BNN). Artificial Intelligence (AI) methods achieved better performance than traditional statistical methods.	To evaluate the prediction performance, we followed the 10-fold cross-validation procedure, which has shown good performance in model selection. SVM method helps in detecting accurate values.	We could conclude that bond rates largely rely on a small list of financial variables to make rating decisions. However, it is generally difficult to interpret the relative importance of the variables in the models for either support vector machines or neural networks. In fact, this limitation has been a frequent complaint about neural networks in the literature	Bond Rating Prediction. prediction accuracy 75.5%

DATASET DESCRIPTION AND SAMPLE DATA:

- Here we used cereals dataset in which it contains the data on several variables of different brands.
- Here we used the dataset Kaggle and the link is https://www.kaggle.com
- Dataset link is https://www.kaggle.com/virgodata/cereal

SAMPLE DATA:



calories	protein	fat	sodium	fiber	rating
70	4	1	130	10	68.40297
120	3	5	15	2	33.98368
70	4	1	260	9	59.42551
50	4	0	140	14	93.70491
110	2	2	180	1.5	29.50954
110	2	0	125	1	33.17409
130	3	2	210	2	37.03856
90	2	1	200	4	49.12025
90	3	0	210	5	53.31381
120	1	2	220	0	18.04285
110	6	2	290	2	50.765
120	1	3	210	0	19.82357
110	3	2	140	2	40.40021
110	1	1	180	0	22.73645
110	2	0	280	0	41.44502
100	2	0	290	1	45.86332
110	1	0	90	1	35.78279
110	1	1	180	0	22.39651
110	3	3	140	4	40.44877
110	2	0	220	1	46.89564

ISSUES OF THE EXISITING:

- Cereals Rating Prediction is an important issue for financial investors to decide which cereals one should buy and sell.
- It can also be said that longer periods cannot be forecasted by simply technical analysis.
- Finding accuracy is difficult
- Some techniques are of high of cost which is not affordable.
- Market fluctuations.

METRICS NEED TO BE MEASURED:

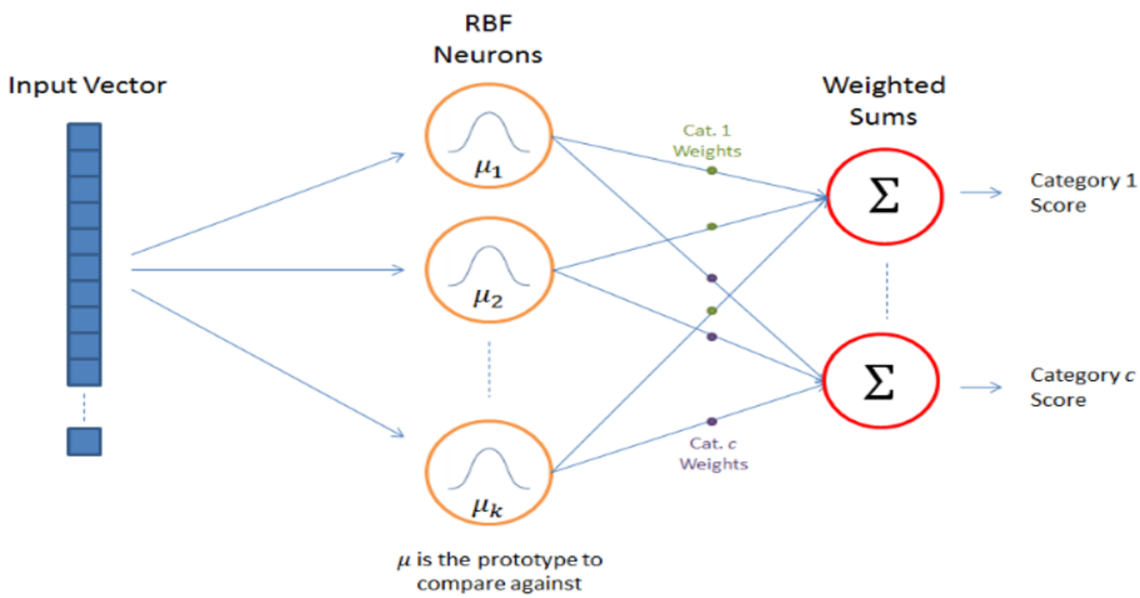
- Accuracy
- Precision
- In problems of cereal market prediction, simply comparing the accuracy, that the ration of correct predictions to total predictions is not enough.
- Investment metrics- In some case, there can some difficulty in choosing the crops.

4. PROPOSED ALGORITHM WITH FLOWCHART:

RBF NEURAL NETWORKS:

Radial Basis Function Network (RBFN) is a neural network. The RBFN approach is more intuitive than the MLP. An RBFN performs classification by measuring the input's similarity to examples from the training set. Each RBFN neuron stores a "prototype", which is just one of the examples from the training set. When we want to classify a new input, each neuron computes the Euclidean distance between the input and its prototype. If the input more closely resembles the class A prototypes than the class B prototypes, it is classified as class A.

RBF NETWORK ARCHITECTURE:



FORWARD PROPAGATION:

- We first initialize some random value to 'W' and propagate forward.
- We will calculate the net input value and output of the hidden layer, using the randomly selected weight value between input and hidden layer as shown below:

net h1 = w1*i1 + w2*i2 + b1*1

out h1 = $1/1+e^{-net\ h1}$

- Hence the final output from the output layer is found and

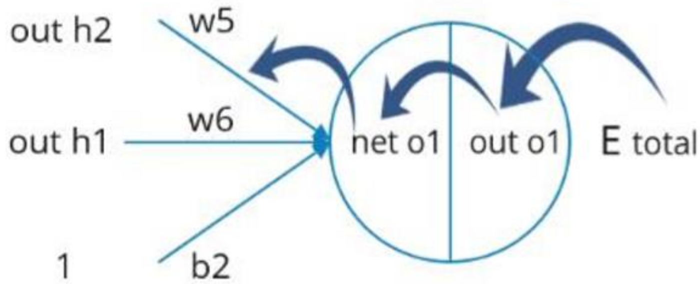
$E\ o1 = \Sigma 1/2(target\ - output)^2$

- If we notice that there is some error. To reduce that error, we propagate backwards and increase the value of 'W'.

BACKWARD PROPAGATION:

- Now, we will propagate backwards. This way we will try to reduce the error by changing the values of weights and biases.

$$\frac{\delta Etotal}{\delta w5} = \frac{\delta Etotal}{\delta out\ o1} * \frac{\delta out\ o1}{\delta net\ o1} * \frac{\delta net\ o1}{\delta w5}$$



The local gradient of output layer Y2 and local gradient of hidden layer Y1 is found using this gradient descent- based learning rule as follows:

Y2 = 2*dlogsig(n2,a2)*e;

Y1 = diag(dtansig(n1,a1),0)*w2'*Y2;

And then Weights Update is done,

$$w5^{+} = w5 - n\frac{\delta Etotal}{\delta w5}$$

- Then we feed forward and find the error. Even after if we notice that the error is increased. Finally, we come to know that, we can’t increase the ‘W’ value.
- So, we again propagate backwards, and we decrease ‘W’ value.
- Now, if we can notice that the error is reduced, that weight value is a solution to the learning problem.

So, we are trying to get the value of weight such that the error becomes minimum.

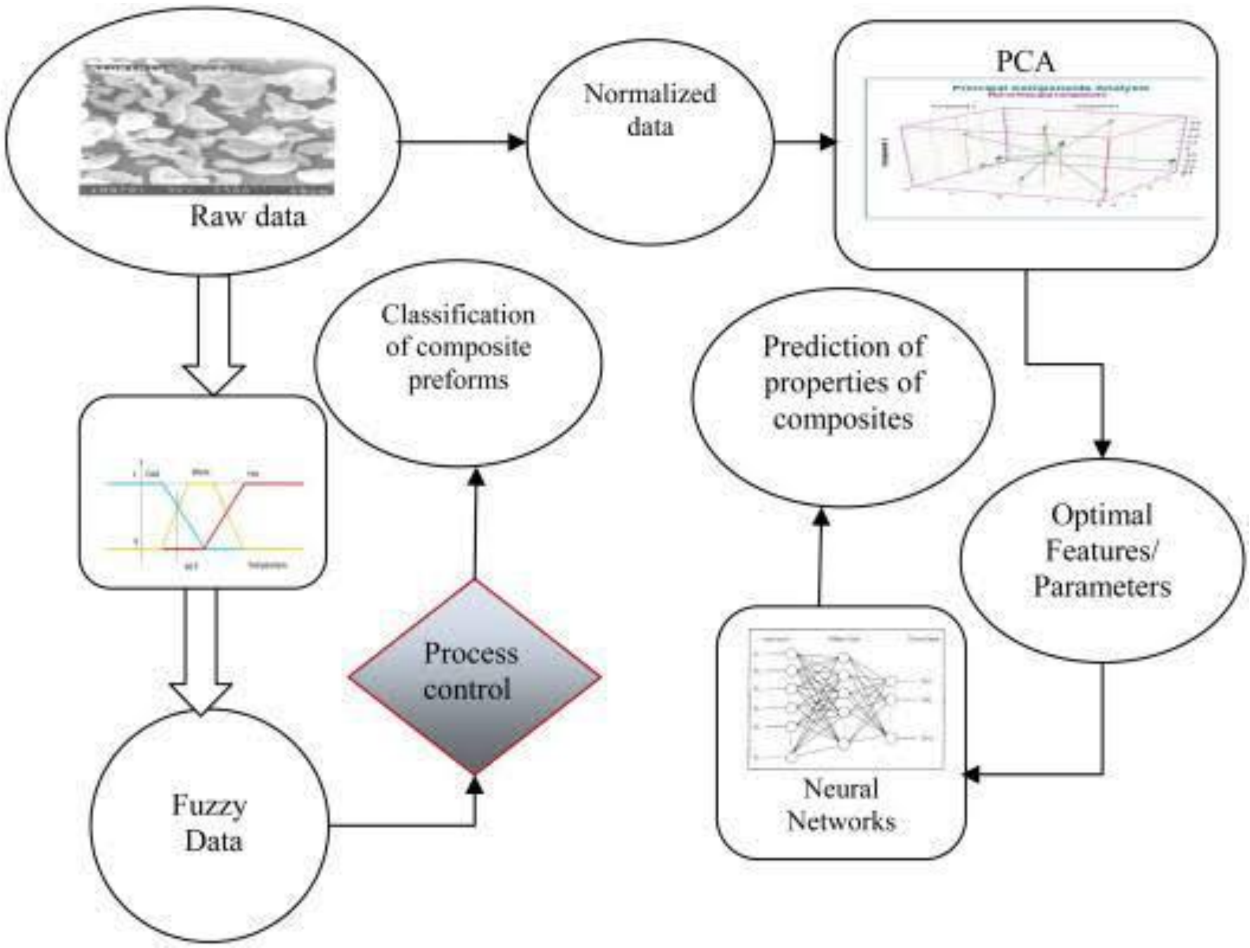
Basically, we need to figure out whether we need to increase or decrease the weight value.

Once we know that, we keep on updating the weight value in that direction until error becomes minimum.

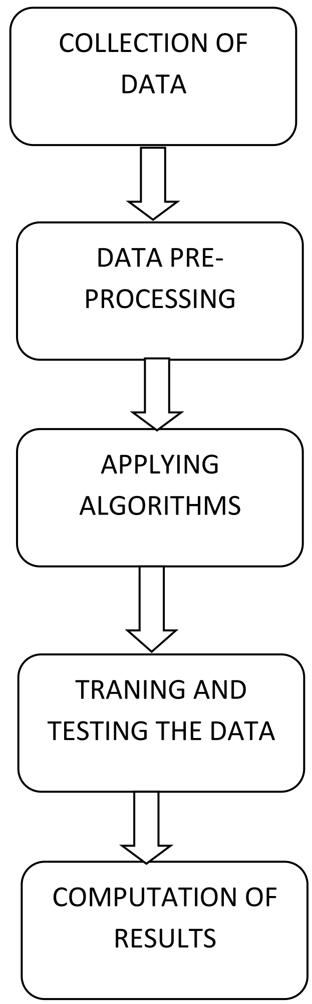
You might reach a point, where if you further update the weight, the error will increase.

At that time, you need to stop, and that is your final weight value. We need to reach the 'Global Loss Minimum' in the squared error.

ARCHITECTURE:



FLOWCHART:



CODE:

```
cereal <- read.csv(file.choose())

head(cereal)

nrow(cereal)

#constructing a neural network for cereal rating


#creating index

samplesize = 0.60 * nrow(cereal)#splitting the cereal data set to extract 60% of the data
```

```
set.seed(80)

index = sample( seq_len ( nrow ( cereal ) ), size = samplesize )

index

#creating training and test set with the index

train = cereal[index,]
test = cereal[-index,]

#scale the data for analysis

max = apply(cereal , 2 , max)
min = apply(cereal, 2 , min)
max
min

scaled = as.data.frame(scale(cereal, center = min, scale = max - min))

library(neuralnet)

trainNN = scaled[index , ]
testNN = scaled[-index , ]
```

set.seed(20)

NN = neuralnet(rating ~ calories + protein + fat + sodium + fiber, trainNN, hidden = 3 , linear.output = T)

plot(NN)

predict_testNN = compute(NN, testNN[,c(1:5)])

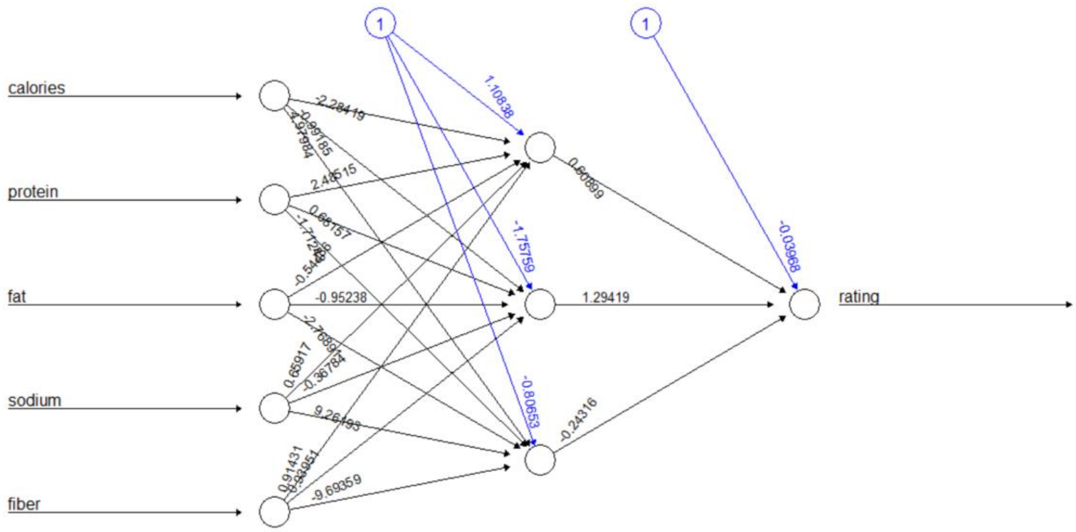
predict_testNN = predict_testNN$net.result * (max(cereal$rating) - min(cereal$rating)) + min(cereal$rating)

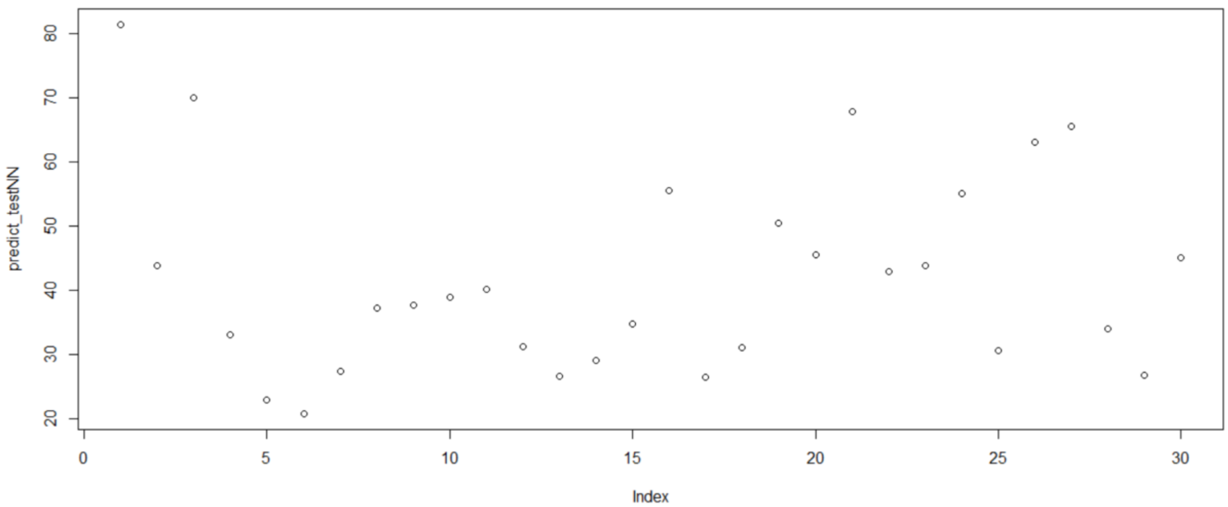
plot(predict_testNN)

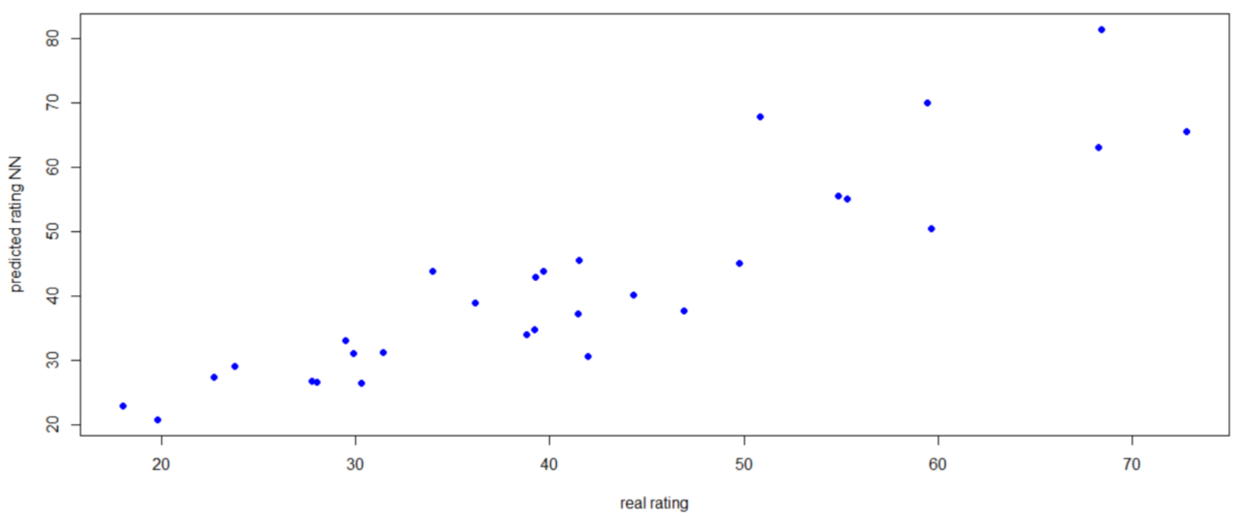
plot(test$rating, predict_testNN, col='blue', pch=16, ylab = "predicted rating NN", xlab = "real rating")

abline()

5. EXPEIMENTAL RESULT:




predict_testNN
Index
20
30
40
50
60
70
80
0
5
10
15
20
25
30


predicted rating NN
real rating
20
30
40
50
60
70
80
20
30
40
50
60
70

```
> cereal <- read.csv(file.choose())
> head(cereal)
  calories protein fat sodium fiber   rating
1       70       4   1    130  10.0 68.40297
2      120       3   5     15   2.0 33.98368
3       70       4   1    260   9.0 59.42551
4       50       4   0    140  14.0 93.70491
5      110       2   2    180   1.5 29.50954
6      110       2   0    125   1.0 33.17409
> nrow(cereal)
[1] 75
```

```
> index = sample( seq_len ( nrow ( cereal ) ), size = samplesize )
> index
 [1] 11 43 31 65 50 36  8  4 13 35 54 16 27  9 25 40 38  7 74 53 33 70 61 34 44
[26] 51 32  6 48 23 18 55 41 75 46 63 71 69 17 67 19 29 37 26 66
```

```
> max = apply(cereal , 2 , max)
> min = apply(cereal, 2 , min)
> max
 calories   protein       fat    sodium     fiber    rating
160.00000   6.00000   5.00000 320.00000  14.00000  93.70491
> min
calories  protein      fat   sodium    fiber   rating
50.00000  1.00000  0.00000  0.00000  0.00000 18.04285
```

6. COMPARITIVE STUDY:

Neural Networks (NN) have been the forefront of growth in recent years due to their variety, the opportunities they provide and most importantly their dynamic nature. A control system for catering robots for path planning is proposed with the help of neural networks as a comparative study. Various parameters such as training time, performance of the network, forecasted distance are considered after iterating to obtain the optimal dataset.

7. CONCLUSION AND FUTURE WORK:

In our project we have taken the cereals dataset in which contains the data on several variables of different brands of cereal. We have compared RBF algorithm and Back propagation for multi-layer perceptron to find out which is more accurate. In our proposed solution, RBF neural network is the best approach for cereals classification as it is more accurate than BP-MLP algorithm. No manual feature extraction stages are needed. This network architecture provides the significant classification accuracy. The mean square error for RBF is lesser when compared to BP-MLP. Hence RBF is more accurate than BP-MLP.

8. REFERENCES:

1. Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M. (1990, June). Stock market prediction system with modular neural networks. In *1990 IJCNN international joint conference on neural networks* (pp. 1-6). IEEE.
2. Klem, K., Vanova, M., Hajslova, J., Lancová, K., & Sehnalová, M. (2007). A neural network model for prediction of deoxynivalenol content in wheat grain based on weather data and preceding crop. *Plant Soil and Environment*, *53*(10), 421.
3. Daniel, J., Andrés, P. U., Héctor, S., Miguel, B., & Marco, T. (2008). A survey of artificial neural network-based modeling in agroecology. In *Soft Computing applications in industry* (pp. 247-269). Springer, Berlin, Heidelberg.
4. Samarasinghe, S., Safa, M., & Nejat, M. (2018). Prediction of Wheat Production Using Artificial Neu ral Networks and Investigating Indirect Factors Affecti ng It: Case Study in Canterbury Province, New Zealand.
5. Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, *38*(8), 10389-10397.
6. Chen, M. Y., Chen, C. C., & Liu, J. Y. (2013, June). Credit rating analysis with support vector machines and artificial bee colony algorithm. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 528-534). Springer, Berlin, Heidelberg.
7. Dyrmann, M., Skovsen, S., Laursen, M. S., & Jørgensen, R. N. (2018, June). Using a fully convolutional neural network for detecting locations of weeds in images from cereal fields. In *International Conference on Precision Agriculture*. International Society of Precision Agriculture.
8. Gutierrez, S. S. M., Palacios, A. T., Pérez, S. L., & López, A. S. (2018). Strategies and technologies of cereal sector in the face of climatic variability. A literatura review. *Strategy, Technology & Society*, *6*(1).
9. Nahil, A., & Lyhyaoui, A. (2018). Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange. *Procedia Computer Science*, *127*, 161-169.

10. Goykhman, M., & Teimouri, A. (2018). Machine learning in sentiment reconstruction of the simulated stock market. *Physica A: Statistical Mechanics and its Applications*, *492*, 1729-1740.
11. Kizilaslan, R., Freund, S., & Iseri, A. (2016). A data analytic approach to forecasting daily stock returns in an emerging marketAuthor-Name: Oztekin, Asil. *European Journal of Operational Research*, *253*(3), 697-710.
12. Oztekin, A., Kizilaslan, R., Freund, S., & Iseri, A. (2016). A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal of Operational Research*, *253*(3), 697-710.
13. Kim, Y., Roh, J. H., & Kim, H. Y. (2018). Early forecasting of Rice blast disease using long short-term memory recurrent neural networks. *Sustainability*, *10*(1), 34.
14. Minh, D. L., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *IEEE Access*, *6*, 55392-55404.
15. Hayashi, K., Llorca, L., Rustini, S., Setyanto, P., & Zaini, Z. (2018). Reducing vulnerability of rainfed agriculture through seasonal climate predictions: A case study on the rainfed rice production in Southeast Asia. *Agricultural systems*, *162*, 66-76.
16. HUANG, M., XU, Y. H., & WANG, H. Q. (2019). Field identification of morphological and physiological traits in two special mutants with strong tolerance and high sensitivity to drought stress in upland rice (Oryza sativa L.). *Journal of Integrative Agriculture*, *18*(5), 970-981.
17. Karimi, H., Skovsen, S., Dyrmann, M., & Nyholm Jørgensen, R. (2018). A Novel Locating System for Cereal Plant Stem Emerging Points’ Detection Using a Convolutional Neural Network. *Sensors*, *18*(5), 1611.