

A Novel Approach for Weather Forecasting using Machine Learning Techniques

B.S.Panda¹, P.V.Lasyasri², D.Maneesha³, P.Goutham⁴, K.Suresh⁵, Ch.Pranavsankar⁶

¹Associate Professor ²⁻⁶Students

Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India.

Abstract - Prediction of weather in general is a complex process and challenging task. It requires different parameters to estimate the climate. Observing its job in consistently today perspective is significantly more testing. This clarifies why determining can't be anticipated with less complex methods. In the current occasions there are top quality satellite pictures to precisely foresee the estimate of the up and coming days; however the procedure is neither straightforward nor conservative. Here this module causes us to foresee the climate utilizing the past information and dissect it with a decent pace of exactness and ends up being a straightforward one. The module includes the utilization of ideas identified with man-made brainpower and Machine Learning tools.

Among the various tools, we have chosen linear regression technique and deep neural networks technique which is part of deep learning. One thing which is to be done mandatorily by the user is to update the previous day's weather parameters or else the module fails to apply linear regression to predict, as every tool dealing with machine learning involves the constant renewal of past data. An intelligent prediction based on the available data is accomplished using machine learning techniques. The analysis and prediction is based on linear regression and Deep neural networks which predicts the next day's weather with good accuracy. An accuracy of more than 90% is obtained, based on the dataset. Recent studies have reflected that machine learning techniques achieved better performance than traditional statistical methods. Machine learning, a branch of artificial intelligence has been proved to be a robust method in predicting and analyzing a given data set.

Index Terms: Weather forecasting. Data mining. Machine learning. Linear regression. Deep neural networks.

I. INTRODUCTION

Weather expectation when all is said in done, is an enormous procedure and confounded assignment. It requires different parameters to figure the climate. Observing and foreseeing weather helps in different fields like farming, travel, contamination dispersal, correspondence, catastrophe the executives, and so on. Lately, masters and analysts lead ceaseless research in the meteorological zone and subsequently an enormous number of meteorological information are gathered. Human sum up and secure immense of meteorological estimating information through contemplating and summing up understanding from these archives which contains information with respect to the climate atmosphere and opposition of horrendous climate. Simultaneously the PC innovation grows combatively. Every one of these variables adds to the educational development of meteorological reason.

The feasible meteorological information and the kind of meteorological information are expanding argumentatively. Meteorological sounding information are chiefly assembled from surface meteorological perception stations and elevated perception stations, presently there are progressively climate stations, the day by day meteorological perception information which can be watched, obtained and handled are developing at an exponential speed. Presently these meteorological information are essentially put away in a message document or a database, and are basically utilized for information examination and the arrangement of climate gauging, debacle anticipating and other data, giving choice help to different departments.

II. RELATED WORKS

The authors in [1] managed the forecast of environmental temperature utilizing Bolster vector machine. This assisted with comprehension about imperfections of SVM. The forecast interim utilizing hydrological information which helped us to think about the vulnerability was talked about in [2]. The creators in [3], anticipated the measure of sunlight based vitality produced utilizing climate conjecture gave a case of how to utilize the figure in day by day life. Forecast of the most extreme temperature utilizing bolster vector machine helped us in the expectation temperature process was talked about in [4]. In [5], the creators gave an instinct of various pieces utilized in help vector machine. Anticipating utilizing fake neural system gave on the utilization of gauging utilizing ANN [6].

The below table-1 tabulates the various Machine learning techniques and ANN [7] methods used for weather prediction with different set of weather parameters.

Table1. Various Machine learning techniques

Author Name	Prediction	Technique	Parameters
Abhishek Saxena	Weather forecasting	Artificial Neural Networks	Maximum and Minimum, Temperature, Rainfall and Wind stream and Direction [8].
Rupa	Rainfall	Data Mining, Neural Networks and Fuzzy Logic	Surface pressure, Heat, Pressure, Breeze [9].
Pallavi	Rainfall	AI, ANN, Fuzzy Logic	Relative humidity, Wind direction, Temperature, Total cloud cover [10].
Pinky saikia Dutta	Rainfall	Data Mining	Maximum and Minimum Temperature, Mean level pressure, Wind speed and Rainfall [11].

III. LITERATURE SURVEY

A. Data mining (DM):

DM is another data innovation as the improvement of database innovation and man-made consciousness innovation. Data mining technique [12] is the way toward separating the concealed data and information that individuals don't know ahead of time yet conceivably helpful. The point is to find obscure connections and sum up information in the inventive method of comprehension by the information proprietor and esteem, and anticipate conceivable future conduct, in order to offer more grounded help for dynamic. As indicated by the various types of the primary information structure, information mining can be partitioned into three classes when all is said in done that are information mining, Web information mining and content information mining. The information digging is for organized information, for example, SQL, Server, Prophet, Informix and other information or information distribution center. At present, the accompanying programming DB2 Shrewd Digger for Information SAS Undertaking Excavator of IBM can be utilized.

B. Web mining (WM):

The WM is a customary database or information warehousing, and Web information mining is the different Web information including Site pages, structure between pages, the client access to data, and business exchange data, which is to find valuable information to assist individuals with separating information from the Internet, improving webpage plan and online business to better or improve administrations. Web information mining can be separated into Web content mining, Web use mining and Web structure mining.

C. Text data mining (TDM):

At the point when the objects of information mining is made completely out of content kind, the procedure of mechanized data handling and examination gigantic content data is called content information mining, which is obstructing information mining calculations and data recovery calculations. It incorporates highlight extraction, content synopsis, content order and bunching, idea tasks and exploratory information investigation. The strategies of content information mining contains the vector of word recurrence, the word string portrayal, Bayesian classifier, Pack of word, content bunching calculation dependent on the idea and characterization calculation of K - the closest neighbor. In the zone of data the executives, it is the best way to accomplish improvement of information recovery and information the board, as long as we incorporate information mining innovation with man-made reasoning innovation, access to client information, writing and different sorts of information.

E. Statistical analysis (SA):

Statistical analysis is to extract unknown mathematical model from samples analysis. . In information mining, it frequently includes a specific factual methodology, judgment theory and mistakes control. Climate determining depends on Numerical Climate Expectations (NWP) that catches the condition of the air and reenacts its development dependent on physical and compound models. Worldwide NWP models typically give countless parameters speaking to various physical factors in reality. In view of the absence of spatial and transient goals, these fields should be deciphered by profoundly qualified work force to create gauges for a particular area. This is still today a human based procedure, which depends on explicitly prepared and experienced experts to decipher displayed and watched information NWP factors characterize the condition of the environment and its progressions through existence. NWPs characterize a profoundly organized dataset in which the connections between its factors are characterized by material science conditions, for example, preservation of mass, force and vitality.

IV. EXISTING SYSTEM

Time series information mining is to remove data and information from a great deal of time arrangement information, these data and information are not known ahead of time for individuals yet they are conceivably helpful and time-related, and for present moment, medium-term or long haul conjectures, directing individuals' conduct, for example, society, economy, military and life. Truth be told, practically all information in the meteorological field are time arrangement information, and future meteorological information can be anticipated better by methods for the time arrangement mining calculation. The time arrangement mining calculation received in the article is ARIMA time arrangement mining calculation. The complete name of the ARIMA (p,d,q) model is distinction auto-relapse moving normal model.

Forecast of the most extreme temperature utilizing bolster vector machine helped us in the expectation of temperature process was additionally examined. Further, the creators gave an instinct of various pieces utilized in help vector machine. Determining utilizing fake neural system gave on the utilization of estimating utilizing ANN.

V. METHODOLOGY

Weather cannot be predicted with good percentage of accuracy. It is an art to forecast weather with very low deviations and making it to fetch good results. But weather forecasting tends to deviate more and has moderate accuracy. Weather forecasting is based on Numerical Weather Predictions (NWP) that capture the state of the atmosphere and simulate its evolution based on physical and chemical models. Global NWP models normally provide a large number of parameters representing different physical variables in space and time. Because of the lack of spatial and temporal resolution, these fields need to be interpreted by highly qualified personnel to produce forecasts for any specific region. This is still today a human based process, which relies on specifically trained and experienced professionals to interpret modeled and observed data.

To build up an intelligent weather predicting module since this has become a vital instrument. These instruments consider measures, for example, most extreme temperature, least temperature and precipitation for an examined time of days and are investigated. A canny expectation dependent on the accessible information is practiced utilizing AI strategies. The investigation and expectation depends on straight relapse and DNN regressor which predicts the following day's climate with acceptable exactness. The module assumes an essential job in agrarian, modern and calculated fields where the climate gauge is a significant measure.

VI. PROPOSED SYSTEM

Weather forecasting is rather a statistical measure than a binary decision. We intend to develop an intelligent weather predicting module since this has become a necessary tool. This tool considers measures such as maximum temperature, minimum temperature and rainfall for a sampled period of days and is analyzed. The analysis and prediction is based on linear regression and DNN regressor which predicts the next day's weather with good accuracy. An accuracy of more than 90% is obtained, based on the dataset. Recent studies have reflected that machine learning techniques achieved better performance than traditional statistical methods. Machine learning, a branch of artificial intelligence has been proved to be a robust method in predicting and analyzing a given data set. The DNN regressor shows more accuracy when compared with linear regression algorithm (Fig.1)

Data is extracted from the website known as <http://darksy.net> and 'JaipurFinalCleanData.csv' file (Fig.2) is fetched with the help of geographic location jaipur. The file name of .csv is jaipurFinalCleanData.csv. So, this csv file is passed as data set into algorithms and operations like training, testing, and validations are performed and weather prediction of whole month is predicted with good accuracy.

To simulate the results the python program is used and pandas library is used to extract the data frame.

```
df = pd.read_csv('JaipurFinalCleanData.csv').set_index('date')
# execute the describe() function and transpose the output so that it doesn't overflow the width of the screen
print(df.describe().T)
# execute the info() function
print(df.info())
# First drop the maxtempm and mintempm from the dataframe
df = df.drop(['mintempm', 'maxtempm'], axis=1)
```

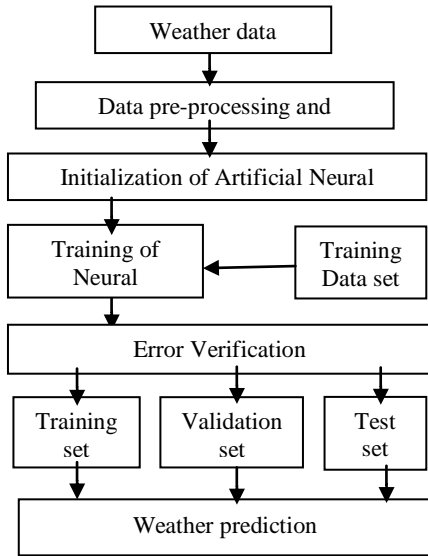


Fig.1. Proposed System

Date	temperature	humidity	wind	pressure	visibility	clouds	uv	precipitation	weather
2017-01-01	20	65	10	1013	10	0	1	0	Clear
2017-01-02	22	68	12	1015	12	0	1	0	Clear
2017-01-03	25	70	15	1018	15	0	1	0	Clear
2017-01-04	28	72	18	1020	18	0	1	0	Clear
2017-01-05	30	75	20	1022	20	0	1	0	Clear
2017-01-06	32	78	22	1024	22	0	1	0	Clear
2017-01-07	34	80	25	1026	25	0	1	0	Clear
2017-01-08	35	82	28	1028	28	0	1	0	Clear
2017-01-09	33	80	25	1026	25	0	1	0	Clear
2017-01-10	30	78	22	1024	22	0	1	0	Clear
2017-01-11	28	75	20	1022	20	0	1	0	Clear
2017-01-12	25	72	18	1020	18	0	1	0	Clear
2017-01-13	22	68	15	1018	15	0	1	0	Clear
2017-01-14	20	65	12	1015	12	0	1	0	Clear
2017-01-15	18	62	10	1013	10	0	1	0	Clear
2017-01-16	15	60	8	1010	8	0	1	0	Clear
2017-01-17	12	58	6	1008	6	0	1	0	Clear
2017-01-18	10	55	5	1005	5	0	1	0	Clear
2017-01-19	8	52	4	1002	4	0	1	0	Clear
2017-01-20	5	50	3	1000	3	0	1	0	Clear
2017-01-21	3	48	2	998	2	0	1	0	Clear
2017-01-22	2	45	2	995	2	0	1	0	Clear
2017-01-23	1	42	1	992	1	0	1	0	Clear
2017-01-24	0	40	1	990	0	0	1	0	Clear
2017-01-25	-1	38	1	988	-1	0	1	0	Clear
2017-01-26	-2	35	1	985	-2	0	1	0	Clear
2017-01-27	-3	32	1	982	-3	0	1	0	Clear
2017-01-28	-4	30	1	980	-4	0	1	0	Clear
2017-01-29	-5	28	1	978	-5	0	1	0	Clear
2017-01-30	-6	25	1	975	-6	0	1	0	Clear
2017-01-31	-7	22	1	972	-7	0	1	0	Clear
2017-02-01	-8	20	1	970	-8	0	1	0	Clear
2017-02-02	-9	18	1	968	-9	0	1	0	Clear
2017-02-03	-10	15	1	965	-10	0	1	0	Clear
2017-02-04	-11	12	1	962	-11	0	1	0	Clear
2017-02-05	-12	10	1	960	-12	0	1	0	Clear
2017-02-06	-13	8	1	958	-13	0	1	0	Clear
2017-02-07	-14	5	1	955	-14	0	1	0	Clear
2017-02-08	-15	3	1	952	-15	0	1	0	Clear
2017-02-09	-16	2	1	950	-16	0	1	0	Clear
2017-02-10	-17	0	1	948	-17	0	1	0	Clear
2017-02-11	-18	-2	1	945	-18	0	1	0	Clear
2017-02-12	-19	-4	1	942	-19	0	1	0	Clear
2017-02-13	-20	-6	1	940	-20	0	1	0	Clear
2017-02-14	-21	-8	1	938	-21	0	1	0	Clear
2017-02-15	-22	-10	1	935	-22	0	1	0	Clear
2017-02-16	-23	-12	1	932	-23	0	1	0	Clear
2017-02-17	-24	-14	1	930	-24	0	1	0	Clear
2017-02-18	-25	-16	1	928	-25	0	1	0	Clear
2017-02-19	-26	-18	1	925	-26	0	1	0	Clear
2017-02-20	-27	-20	1	922	-27	0	1	0	Clear
2017-02-21	-28	-22	1	920	-28	0	1	0	Clear
2017-02-22	-29	-24	1	918	-29	0	1	0	Clear
2017-02-23	-30	-26	1	915	-30	0	1	0	Clear
2017-02-24	-31	-28	1	912	-31	0	1	0	Clear
2017-02-25	-32	-30	1	910	-32	0	1	0	Clear
2017-02-26	-33	-32	1	908	-33	0	1	0	Clear
2017-02-27	-34	-34	1	905	-34	0	1	0	Clear
2017-02-28	-35	-36	1	902	-35	0	1	0	Clear
2017-02-29	-36	-38	1	900	-36	0	1	0	Clear

Fig.2. Sample data set

VII. TESTING & ERROR DETECTION

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

The training set error and cross validation error are calculated in order to understand the errors present in the prediction. The error normally be because of overfitting (high variance) or underfitting (high bias) of curves. A 20% of data is taken for the cross validation set and another 20% of data is taken for the test set. Remaining 60% of data is evaluated for the training set.

Mean absolute error:

In statistics, mean absolute error (MAE) is a measure of errors between paired observations expressing the same phenomenon. Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement. MAE is calculated as:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n} \quad (1)$$

Median absolute error:

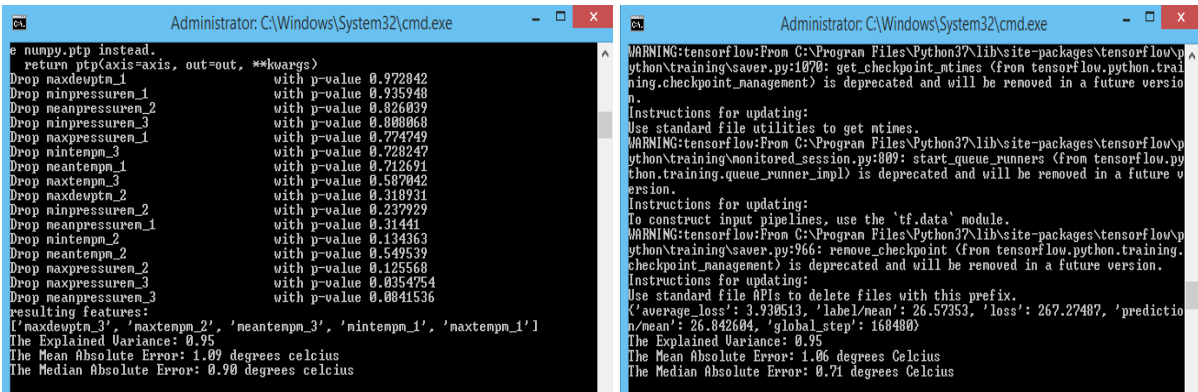
In statistics median absolute error is particularly interesting because it is robust to outliers. The loss is calculated by taking the median of all absolute differences between the target and the prediction. If y^{\wedge} is the predicted value of the ith sample and y_i is the corresponding true value, the median absolute error estimated over n samples is defined as follows:

$$MAE(y, y^{\wedge}) = \text{median}(|y_1 - y_1^{\wedge}|, \dots, |y_n - y_n^{\wedge}|) \quad (2)$$

VIII. RESULTS

The prediction for the next day's weather with good accuracy can be done with the help of integrated weather predicting module which takes parameter like temperature, pressure, humidity, and various other parameters in the form of data set. The data set is passed to the algorithms like Deep neural networks (DNN) and linear regressor (LR) and some operations like training, validation, testing are processed and ultimately predicts every days temperature by showing the variance means difference between each day temperature and also predicts mean temperature for whole month. It also shows 90% accuracy by displaying mean absolute error and median absolute error rate. So, complex task like weather prediction can be done with the help of Deep Neural Networks regressor and linear regressor techniques. Deep neural networks regressor will provide more accurate results than linear regression algorithm.

When jaipurFinalCleanData.csv file is passed as data set into algorithms and operations like training, testing, and validations are performed and weather prediction of whole month is predicted with good accuracy (i.e based on mean and median absolute error rate).

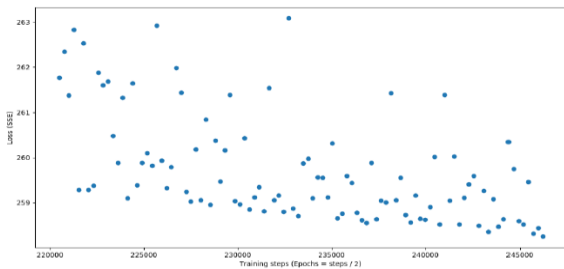


```

Administrator: C:\Windows\System32\cmd.exe
python numpyp.py instead.
return npf(axis=axis, out=out, **kwargs)
Drop maxdeuptn_1 with p-value 0.972842
Drop minpressure_1 with p-value 0.935948
Drop meanpressure_2 with p-value 0.826039
Drop minpressure_3 with p-value 0.888068
Drop maxpressure_1 with p-value 0.774749
Drop nintemp_3 with p-value 0.728247
Drop neantemp_1 with p-value 0.712691
Drop naxtemp_3 with p-value 0.587842
Drop maxdeuptn_2 with p-value 0.318931
Drop minpressure_2 with p-value 0.237929
Drop meanpressure_1 with p-value 0.314441
Drop nintemp_2 with p-value 0.134363
Drop neantemp_2 with p-value 0.549539
Drop maxpressure_2 with p-value 0.125568
Drop maxpressure_3 with p-value 0.0354754
Drop meanpressure_3 with p-value 0.0841536
resulting features:
['maxdeuptn_3', 'naxtemp_2', 'neantemp_3', 'nintemp_1', 'naxtemp_1']
The Explained Variance: 0.95
The Mean Absolute Error: 1.09 degrees celcius
The Median Absolute Error: 0.90 degrees celcius

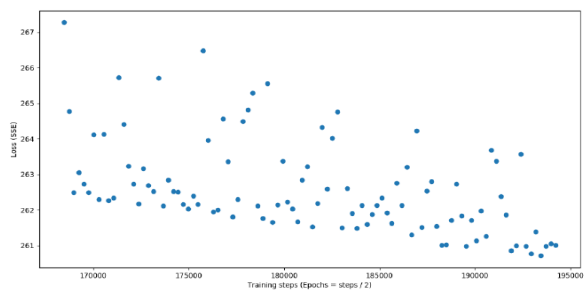
Administrator: C:\Windows\System32\cmd.exe
WARNING:tensorflow:From C:\Program Files\Python37\Lib\site-packages\tensorflow\python\taining\saver.py:1070: get_checkpoint_names (from tensorflow.python.training.checkpoint_management) is deprecated and will be removed in a future version.
Instructions for updating:
Use standard file utilities to get names.
WARNING:tensorflow:From C:\Program Files\Python37\Lib\site-packages\tensorflow\python\taining\monitored_session.py:809: start_queue_runners (from tensorflow.python.training.queue_runner_impl) is deprecated and will be removed in a future version.
Instructions for updating:
To construct input pipelines, use the 'tf.data' module.
WARNING:tensorflow:From C:\Program Files\Python37\Lib\site-packages\tensorflow\python\taining\saver.py:966: remove_checkpoint (from tensorflow.python.training.checkpoint_management) is deprecated and will be removed in a future version.
Instructions for updating:
Use standard file APIs to delete files with this prefix.
{'average_loss': 3.930513, 'label/mean': 26.57353, 'loss': 267.27487, 'prediction/mean': 26.842604, 'global_step': 168480}
The Explained Variance: 0.95
The Mean Absolute Error: 1.06 degrees Celcius
The Median Absolute Error: 0.71 degrees Celcius
  
```

Observation 1.Linear regression



Scatter plot 1. LR

Observation 2. Deep Neural Networks regressor



Scatter plot 2. DNN

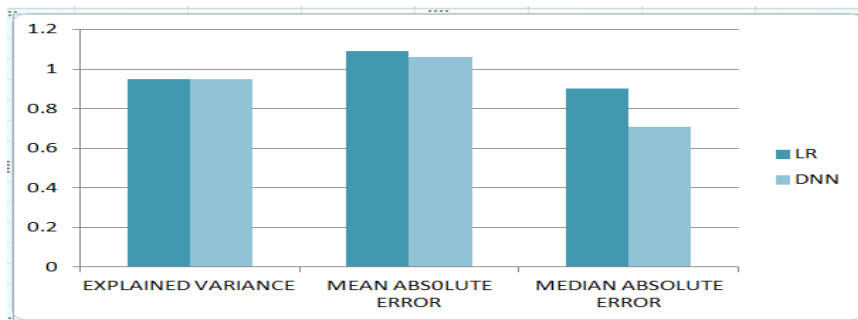


Fig3. Comparison between LR and DNN

Here, clearly seethat the error rate of mean and median of linear regression is more when compared to DNN regressor. So, DNN regressor algorithm is more accurate when compared with linear regression as it has less error rate.

IX. CONCLUSION AND FUTURE SCOPE

Weather conditions are very important to be predicted to save the life of people .And that is a challenging problem. Machine Learning techniques can be used to predict the extreme weather incidents. This paper talks about the use of machine learning algorithms to filter and visualize the extreme weather events. In this, the decision tree classification algorithm was used to and rules for classifying weather parameters such as maximum temperature, minimum temperature, evaporation, humidity and wind speed in terms of the month and year. The results show how these parameters have an influence on the weather observed in this period over the study.

The outcomes show that above 90% of extraordinary and non outrageous occasions are accurately distinguished by the proposed technique. In future research works neuro-fuzzy models will be utilized for climate forecast process. This work is critical to climatic change considers in light of the fact that the variety in climate conditions as far as temperature, precipitation and wind velocities can be examined utilizing these strategies. The outcomes can likewise be confirmed with multi arrangement utilizing straight relapse and with Counterfeit Neural System. Yet, the weakness of utilizing Counterfeit Neural Systems and multi arrangement is that, they

give us a yield of the day and not the closest worth likelihood of how the day will be. The SVM can likewise be utilized in foreseeing the information and works best when there are countless highlights and orders present however excess highlights must be maintained a strategic distance from. The main exertion which is to be taken by the client is to refresh the informational index so as to show exact outcomes.

REFERENCES

- [1] Y.Radhika., M.Shashi. Atmospheric Temperature Prediction using Support Vector Machines “Vol. 1, No. 1, April 2009.
- [2] Durga L Shrestha. Dimitri P. Solomatine, Machine learning approaches for estimation of prediction interval for the model output Neural Networks 19, 225-235, 2006.
- [3] Navin Sharma.,Pranshu Sharma., David Irwin., Prashant Shenoy.: Predicting Solar Generation from Weather Forecasts Using Machine Learning, IEEE Smartgridcomm.
- [4] Paniagua-Tineoa, S. Salcedo-Sanz a., C. Casanova-Mateoc.,E.G. Ortiz-Garcíaa.,M.A. Cony b.,E. Hernández-Martínc.: Prediction of daily maximum temperature using a support vector regression algorithm. Renewable Energy 36, 2011.
- [5] N. Cristianini., J. Shawe-Taylor.:An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000.
- [6] Kabir Rasouli , William W. Hsieh , Alex J. Cannon Daily, stream flow forecasting by machine learning methods and other climate inputs .Journal of Hydrology414-415,pp.285-293, 2011.
- [7] Dr. S. Santhosh Baboo, I.Kadar Shereef, An Efficient Weather Forecasting System using Artificial Neural Network. IJESD Vol.1 No.4 October 2013.
- [8] Saxena, Abhishek, Neeta Verma and Krishna Chandra Tripathi. “A Review Study of Weather Forecasting Using Artificial Neural Network Approach.” International journal of engineering research and technology volume.2, 2013.
- [9] Pinky Saikia Dutta, Hitesh Tabbilder “Prediction of Rainfall using Datamining Technique over Assam”, in Indian Journal of Computer Science and Engineering (IJCS), ISSN : 0976-5166 Vol. 5 No.2, pp. 85-90, May 2014.
- [10] Rupa, Shilpa Jain, “A Review Study of Rainfall Prediction using Neuro-fuzzy Inference System” in International Journal of Engineering Sciences & Research Technology (IJESRT), 4.(8): ISSN: 2277-9655 pp 918-922, August, 2015.
- [11] R. Samya, R. Rathipriya. “Predictive Analysis for Weather Prediction using Data Mining with ANN: A Study”, in International Journal of Computational Intelligence and Informatics, Vol. 6: No. 2, ISSN: 2349-6363 `149-153, September 2016.
- [12] B. S. Panda, Ashok Misra, S. S. Gantayat “Methods and Concepts of Data Mining Techniques to Impute Missing Data Information” in: Far East Journal of Electronics and Communications (FEJEC), ISSN: 0973-7006, 2019.