A Review of Weather Forecasting Models-Based on Data Mining and Artificial Neural Networks

G.Vamsi Krishna

Computer science and engineering,GITAM University,Visakhapatnam,India

Abstract: A large amount of the world's population relies mostly on the monsoons. These monsoons have a great impact on the livelihood of the Indian families where cultivation is a major source of livelihood. Many models have been developed and utilized for prediction, forecasting of the weather data based on techniques such as data mining, time series analysis, neural networks. In this paper a brief review of the works carried out in this area in the recent past are presented.

Index terms: Time-series, Data mining, Artificial Neural Networks, weather forecasting, prediction.

1. Introduction:

Weather forecasting is one of the challenging problems which are addressed by the meteorology departments all over the globe. In this procedure a day-to-day analysis of the weather is to be predicted and these forecasting should be communicated for the end users for taking decisions. It is a most challenging issue since the decisions that are taken are mostly with uncertainty. Researchers in this domain have categorized these forecasting strategies into a 2-fold, based on numeric modeling and based scientific processing. Among these categorizations, the predictions of the rainfalls are mostly subjected to the numerical methods of analysis. Since the data changes dynamically, in particular with rainfall/weather, results into uncertainties in most of the cases. This is due to the choice of the initial conditions which are mainly used for weather predictions and which are viable of changing. Therefore to update these methods statistical techniques have been used at a high pace. Among these models time-series models are mostly considered to understand and interpret the patterns of the rainfalls. However these statistical methods suffer from disadvantages like existing of non linear relationships between the monsoon data and the prediction data and secondly for evaluating the analysis, these statistical models needs to identify the predictions more effectively. Therefore identifying these parameters has become a challenging task. Many researchers have highlighted that most of the changes in the climatic conditions are mostly due to the global weather changes. Therefore if these weather changes are identified, effective prediction techniques can be planned. Data mining techniques play a vital role in the prediction of global parameters with these assumptions ,the knowledge discovery is carried out by integrating the time series analyzing together with data mining techniques. Artificial neural networks also helps in effective forecasting as it can handle the non-linearity cases arose from the statistical models. This paper presents a brief review of literature carried out in the area of data mining and artificial neural networks for the betterment of research works in this area.

2. Data mining techniques:

Data mining can be considered as effective techniques for identifying the patterns hidden in the data. Data mining can be considered as a general case in every knowledge discovery models. There are several models available in the literature based on classification and clustering. The main contribution in this paper is to highlight the various data mining techniques available for usage in this area. The rest of the paper is organized as follows section-3.1 the brief review of classification models are highlighted, section-3.2 highlights about the clustering models and models based on artificial neural networks are highlighted in section-3.3. The conclusions derived thereoff are presented in section-4.

3. Literature on classification techniques:

3.1 works releated to classification techniques

Classification is considered as grouping the data based on particular class label. This classification helps to predict the rainfall more accurately. Lots of works are showcased in the literature based on classification models. This section of the paper deals with brief review of classification models available for effective analysis of weather data based on classification.

Valmik.B et al. proposed a model for predicting the weather data based on classification technique and considered several attributes such as wind pressure, humidity, vapor, wind speed and cynical results obtained. The results showcased good accuracy by correlating the above parameters.

F.Dell presented a more general approach for identifying the varying the wind speed. The metrological data is divided into 2 groups and performed the classification on these groups to identify the hurricanes and non hurricanes.

Prediction of data using PCA is proposed by Evangelos.T et al. the works are carried out ERA-40 data set and the results are showcased using high relevance accuracy. Meteorology's are presented using image acquisition , wind pressure, temperature and future extraction models.

Improved Nave Bayesian classification is considered by James-N.K.Liu by varying the weather condition data. The Bayesian classification is used for identifying the weather prediction. The algorithm is compared with different models based on genetic algorithms and results showed that the considerable amount of accuracy is obtained.

Works on k-nearest neighbor classification are utilized to estimate the wind speed and using the Manhattan and miniskowski distance the speed of wind is estimated. The authors Menhet.w et al. have also analyzed these works using three parameters wind direction, air temperature and atmospheric pressure.

The fluctuating wind series are modeled by P.J Trombe et al. using the time series analysis together with Markovian switching model and there by a methodology is carried out for estimating the weather changes in the sea using radars. The classification is used to find the relationship between precipitation and wind attributes.

A renewable energy source optimization model is proposed by M.Iqbal et al. for classifying the wind sources using optimization classification algorithm. The works are further extended by Vibav.M by integrating clustering and classification by which wind speed evaporation is calculated and minimum and maximum temperatures can be comforty obtainable. k-means algorithm is utilized for clustering the data.

3.2 works related to clustering techniques:

Clustering is considered as grouping the homogenous data from the heterogeneous group of data available in the networks. Many clustering models are available in the literature for effective analysis of weather data. Among these models, in the literature works are highlighted using both partition clustering and aggloromative clustering techniques. This section of the paper highlights a brief review of the works carried out in the area of weather prediction in the recent past.

TV Rajinikanth et al. has proposed a methodology for analyzing the metrological data that is very much suited in particular to the Indian weather environments using data mining techniques. The related data is grouped basing on k-means algorithm,a decision tree is used for predicting the observations. J.48 classification technique is used for the classification purpose. The experimentations are considered based on time series analysis on the data available during 1955 to 1965.

Kavita.P et al. have presented a model for interpreting the weather using k-means clustering an attempt is made to analyze the temperature at atmospheric pressures with 400hpa, 500hpa,700hpa. The experimentation are also followed by varying the humidity and for presenting the model the weather data set of Bangladesh is considered.

Badhiye S.S. et al. presented an approach to analyze and predict data of temperature, humidity values for future using clustering technique. The outlier analysis is considered to detect the outliers with respect to the data and clustering analysis is used to partitioning the data based on the similarity of the objects. K-nearest neighbor algorithm is used to predict the values of temperature and humidity parameters of climate. To find the distances between datasets in the K-NN algorithm absolute distance and Euclidean distance measuring are used.

 Zahoor Jan M. Abrar et al. developed a system for seasonal to inter-annual climate prediction by using data mining K-NN technique. The main purpose is how to use a data mining technique K-NN and how to develop a system that uses numeric historical data to forecast the climate of specific region, city or country months in advance. Sea surface temperature is considered as a main factor and regression tree technique is used to find the predication of climate. Euclidean distance metric is used to find the distance between the samples. The data set for the system was collected from the Pakisthan meteorological department and the Capital and National Climate Data (NCDC) which consists of 10 years of historic data and stored in MS ACCESS format.

 Sarah N.Kohail and Alaa M.El-Halees presented a model for analyzing meteorological data using data mining techniques. The authors tried to extract useful knowledge from weather daily historical data collected at Gaza strip city during the duration of 9 years period (1977-1985). Daily average relative humidity, average temperature, wind speed, wind direction, the time of the highest wind speed and rainfall are the parameters considered. Time series analysis is used to identifying temperature and linear interpolation method is used to fill missing values. Outliers are removed by using outlier analysis. The authors use cluster analysis to partitioning data by using K-means algorithm at K=4. ANN and least median squares linear regression method are the prediction methods. Classification and association rules are also used .

 MeghaliA.Kalyankar and S.J.Alaspurkar presented a data mining technique to analyze the meteorological data. The weather data is extracted to find hidden patterns inside the large dataset by using clustering technique i.e., K-means partitioning algorithm. The authors used 10 channel midi-data logger system that provide weather data

in form of excel sheets and can automatically collect data on a 24-hour basis. Data preprocessing technique is used applied to the weather data which consists of various parameters as temperature, humidity, rain, wind speed etc. data preprocessing involves Data cleaning, Data transformation is used to reduce noisy and to fill the missing values. Data transformation is a phase in which the selected data is transformed into appropriate forms.

Sanjay Chakraborty et al. presented a methodology for weather forecasting using incremental K-means clustering. The weather forecasting is done based on the incremental air pollution data base of west Bengal in the years of 2009 and 2010. In this approach, the authors proposed a generic methodology for time series forecasting through clustering. The air pollution data is collected from "west Bengal air pollution control board." This data base consists of 4 air pollution elements or attributes. They are carbon dioxide (co2), Respirable Particulate Matter (RPM), sulphur dioxide (so2) and Oxides of Nitrogen (No x). K-means is applied for initial data and incremental K-means is applied to the new coming data by using Manhattan metric measure.

Marwa F.Al-Roby and Alaa M.El-Halees presented an data mining technique for wind speed analysis. The data recorded between 2004 to November 2006 daily historical data by meteorological station of Gaza is considered. After preprocessing the data, they applied data mining techniques, association rules, classification, cluster, outlier analysis are used. The observed data of wind speed contain 4 years from January 2003 to November 2006 daily historical data. To normalize the data, Normalization is applied.

Abay Kumar et al. presented a model using K-means clustering algorithm. The authors have used Probability Density Function Algorithm to generate numerical results in the k-means clustering for weather –related predictions. In this paper, they constructed a model for predicting the probability of play class for YES and NO category through k-means clustering. The original data is collected from Quinlan in 1986. They considered five attributes namely outlook, temperature, humidity, windy and play. The attribute outlook has three possibilities sunny, overcast and rainy. The attribute windy can be either true or false. The outcome class play is either yes or no. in this the authors obtained the results 77.36% of overall accuracy, 85.635 of precision, 91.63% of recall.

Zohreh Nazeri and Jainping Zhang presented a methodology to analyze severe weather impacts on National Airspace System (NAS) performance by using data mining applications. The frame work of this approach consists of three phases: data preparation, feature extraction and data mining. For this experiment, the authors used three sources: Airline Service Quality Performance (ASQP), Enhanced Traffic Management System (ETMS) and National Convective Weather Forecast (NCWF) supplied by National Center for Atmosphere Research from april to September in 2000 to represent the severe weather season. 152 days of complete weather data were used. The feature extraction phase consists of four steps in this work: image segmentation, weather feature extraction, air traffic feature extraction and representation conversion. The authors used segmentation algorithm to identify the severe weather areas. They considered correlation analysis to calculate the eight weather and air traffic features with the performance. They applied clustering to group the days with similar weather impacts on NAS performance. K-means clustering algorithm was also applied. The authors used classification method to partition the performance measure into three classes: bad, medium and good. They applied decision tree algorithm C5.0 to learn rules for these classes. They achieve 86% classification accuracy using 100-fold cross validation on the two-class problem and 77% on the three-class problem.

M.Mayilvahanan and M.Sabitha presented an approach for estimating the availability of sunshine using data mining techniques. This paper deals with estimating the temperature values of four cities in Tamil Nadu, South India namely Chennai, Coimbatore, Madurai and Kanyakumari. The dataset was collected from the India Meteorological Department portal. The dataset consists of monthly mean maximum, minimum temperature and total rainfall based upon 1901-2000 data of various cities in India. The attributes present in the data are station name, month, period(years), number of years, maximum mean temperature in degrees Celsius, minimum mean temperature Celsius and mean rainfall in millimeters. Data preprocessing was applied to the data to clean and improve the quality of data. K-Means clustering algorithm and Expectation Maximization algorithms were used to compare the availability of sunshine in the cities. The authors claimed that the maximum amount of sunshine is recorded in the city of Chennai when compared to the other cities.

Stefnao Serafin et al presented an application of cluster analyzes technique to the verification of quantitative precipitation forecasts. Clustering analyzes is adopted as an objective method to create groups of rain gauges displaying interrelated measurements. In this paper, most of the events are considered with an upper level throw and a surface level low approaching northern Italy from the west. The authors detected the cluster by using the average linkage algorithm. The results of the verification of the precipitation forecasts are highly affected by distribution of rain gauges.

J.Mgutierrez et al. presented an application of clustering methods for statistical down scaling in short range weather forecasts. In this paper, the clustering technique compared with the standard nearest neighbors analogue methods. Some validation results daily precipitation and maximum wind speed operative downscaling (lead time 1 to 5 days) on a network of hundred stations in the iberian peninsula for the period 1998 to 1999 are collected as data. In this paper, self-Organizing maps (SOMS) to define weather classes and study the associated

distributions of local precipitation. The author claimed the outputs of an operative circulation model on different local area or large-scale greets are considered to characterize the atmospheric circulation patterns.

Vipul Kedia et al presented a methodology for time series forecasting through clustering. In this paper, the author used subset of the data set to build up the system model by compressing the information through clustering and coming up with inherent patterns for data. The pattern was represented as the curves such that any time series from the given set is expected to follow. The authors used linear regression method by matching to the closest pattern to each time series that has to be predicted. In this paper, the authors classified the research paper in to two partitions these are dead or alive papers.

N.Raja shekar, T.V.Rajinikanth presented an weather analyzers of Guntur district of Andhra region using SVM data mining techniques. The authors developed a machine learning algorithm that is hybrid SVM (support vector machine) model. The model is used for effective weather prediction by analyzing the given weather data and to recognize the patterns exiting it. SVM comes under the set of supervised learning methods for classification and regression.in this paper, Guntur district weather data sets were considered for analyzers using the hybrid SVM data mining techniques. The authors applied K-means clustering technique over the clustered data set they considered the data set for about 102 years namely, monthly mean for each year average temperature (1901- 2002). The authors yielded good result in predicting the weather that the existing machine learning programming techniques

 Ankita Singh et al. presented an efficient clustering methods for atmospheric conditions prediction using art algorithm. The aim of this research is to develop artificial neural networks based on clustering method for ambient atmospheric conditions prediction in an Indian city. In this paper they presented a clustering method that classifies cities based on atmospheric conditions like temperature, pressure, wind speed and humidity. The data represents the month wise atmospheric conditions under two parameters namely temperature and pressure. And collected from the ten different cities namely Delhi, Kolkata, Bhopal, Mumbai, Jaipur, Amritsar, Cochin, Luck now, Bhubaneshwar, Guwahati.

Ngugen Dinh Hoa et al. presented a methodology for weather now casting from satellite image sequences using picture fuzzy clustering and spatiotemporal. Weather now casting comprises the detail description of the current weather along with forecasts obtained by the extrapolation for every short range period of zero to six hours ahead. In this paper, firstly the satellite image pixels are partitioned into clusters by using fuzzy clustering-a fuzzy clustering method based on the theory of picture fuzzy sets secondly, the fast fourier transform method is used to filter out non predictable scales leading to the increasing time ranges of predictability. Finally the spatiotemporal regression method is used to forecast the predicted sequences of images the authors claimed the better results than the relevant once in weather now casting.

 S.Gokila et al. presented an approach based on clustering and classification in support of climatology to mine weather data. The aim of this paper is to provide a review report on various data mining techniques applied on weather data set in support of weather prediction and climate analyzers. In this paper, clustering techniques applied on the climate data helps to produce similar patterns of climate with the consideration of spatial nature. The classification techniques are used to relate the attributes of weather data to predict the future climate.

Aastha Sharma et al. presented a semi supervised technique for weather condition prediction using DBSCAN and KNN. In this paper, certain atmospheric parameters are taken for four years on a day wise basis in a certain city the weather attributes snow, rain, fog, are considered by authors in this paper, they proposed a new approach that combines data mining technologies like classification and clustering. The semi supervised clustering algorithm to identify clusters in large spatial daya sets. Areter clusters are formed they are given for training to KNN. K-nearest neighbor rule (KNN) has been one of the most known supervised learning an algorithm in pattern classification. In this paper, they considered overall accuracy, class wise accuracy, class wise precision, class wise recall, class wise f-measure as elevation parameters in performance analyzers they considered three class fog, rain and snow for the following conditions accuracy, precision, recall.

 HANS HENRICK BENZON and THOMAS BOVITH et al. presented a methodology for simulation and prediction weather radar cluster using a wave propagator on high resolution NWP data. Cluster is a common problem caused by nonstandard wave propagation in weather radar application. In this paper, the proposed method uses a wave propagator to identify areas. The wave propagator uses a three dimensional refractivity field like temperature, humidity and pressure. These methods are obtained from high resolution. The wave propagator is based on the parabolic equation. The authors obtained the results by removing of cluster echoes while preserving precipitation echoes in this paper.

DANIEL ALEJANDRA VILLA et al. presented a methodology for forecast and tracking the evolution of cloud clusters for TRACC using satellite infrared imagery. The main feature of this system of the following: 1. A cloud cluster detection method based on a threshold temperature (235K) 2. A tracking technique based on MCS (MESOSCALE convective systems) overlapping areas in successive images and 3.A forecast model based on the evolution of each particular MCS in previous steps. Statistical information about MCS evolution during the

wet season atmospheric MESOSCALE campaign (WETAMC) of the large scale atmospheric experimental in Amazonia (LBA) was used to obtain area expansion mean rates for different MCS according to their lifetime duration. This now casting tool was n applied to evaluate MCS displacement and size evolution over the DEL PLATA basin in South America up to 120 MIN with 30 MIN intervals. The obtained results are quite reasonable by this model.

3.3 Works Related To Artificial Neural Networks:

Artificial nueral networks is a system inspired by the biological nerve system. It consists set of interconnected neurons. The neurons are connected by means of links and each link is associated with there wait. In ANN system consists of a 3-layered architecture, input layer, output layer and hidden layer. The ANN is used to understand the patterns for which pre assignment is not available.

Godfrey C.Onwubolu et al.have presented a self organizing data mining methodology for weather forecasting. In this paper, the authors employed the enhanced Group Method of Data Handling [e-GMDH] for the self-organizing data mining approach. The authors uses artificial neural networks for self-organizing. Classification, clustering, modeling, time series forecasting, sequential patterns were the data mining functions used in it. In this paper, the weather data was acquired at the school of engineering physics, University of the South Pacific,Fiji. This data includes daily temperature and pressure observed from 2000 to 2007. They analyzed 13years of rainfall data in Suva which is capital of Fiji. The authors compared the performance with other variants of GMDH. Those are polynomial neural network (PNN) and enhanced versions are e-PNN and e-GDMH. The statistical data of minimum, maximum, standard deviations is 27.3mm, 645.6mm and 135.5mm respectively..

Gurubrinder Kaur presented a survey based on the meteorological data mining techniques. In this paper the authors used data mining techniques such as decision trees, fuzzy logic, artificial neural networks, rule based techniques and some other techniques for the climate prediction. In this paper, the authors found that neural network based algorithms are performed well when compared to other data mining techniques. In order to improve the presentation of neural network algorithms, other statistical based feature selection techniques and fuzzy techniques can be incorporated to achieve better predictability.

SS. Kotsiantis Et al.presented a procedure for estimating minimum, maximum and average daily temperature values using data mining techniques. The authors conducted experiments with regression algorithms using temperature data from the city of Patras in Greece. In this paper the data is collected from meteorological station of the Laboratory of Energy and Environmental Physics of the Department of Physics of University of Patras during the period 2002-2005. Classical linear least-squares regression, model trees and MS rules algorithms are used to predict the data in this paper. The authors used the K-nearest neighbor learning algorithm for grouping the data and to estimate the values of the weights of a neural network, back propagation algorithm is used in the artificial neural networks. They presented the regression method in predicting minimum daily temperature values as input.

Folorunsho Olaiya presented an application in weather prediction and climate changes using data mining techniques. In this paper, the authors investigate the uses of data mining techniques in forecasting temperature, rainfall, evaporation, wind speed and the data was collected between 2000 and 2009 from the city of Ibadan, Nigeria. In this work, both artificial neural networks and decision trees were used to meteorological data and to fill the missing values, data cleaning techniques are used. The authors used evaluation metrics such as co-relation co-efficient, mean squared error, the mean-squared error to select the algorithms, parameters and to predict the values of weather forecasting.

 Arti R. Naik and S.K. Pathan presented an application of artificial neural networks in weather classification and prediction in the forecasting, predicted values are obtained simply by observation. After the numerical methods were developed, the images from satellite were used to retrieve data. The weather data is updated for every 24 hours and records the values till next 5 days. In this paper weather forecasting is done by using Back-propagation Feed forward Neural Network. The authors collected the data by using wireless sensors like anemometer sensor, thermo hydro sensor during 2012. The noise in the data is removed in pre-processing stage and in the back propagation algorithm the error is back-propagated and respectively the weights are adjusted so as to reduce the error.

GauravJ.Sawale and sunilR.Gupta presented a methodology for forecasting weather using Artificial Neural Network in data mining. In this paper a neural-network based algorithm is used to predict the atmosphere for a future time. The authors used Back Propagation Neural (BPN) Network for initial modeling. The results obtained by BPN model are fed to a hop field network. The performance of the proposed ANN-based method tested on 8 years weather dataset comprising 15000 records containing attributes like temperature, humidity and wind speed. The main focus of this paper is based on predictive data mining by which the can extract interesting patterns or knowledge from huge amount of meteorological data. The authors use Data cleaning to fill the missing values and remove noisy data. Relevance analysis can be used to detect attributes that do not contribute

to the classification or prediction task. In this paper, Normalization is applied to the data scaling all values the performance of the Back Propagation Network and Hope field Network Model was satisfactory as there were not sustain number of errors.

Mehrnoosh Torabi and Sattar Hashemi presented an approach for short-term energy consumption to forecast weather by using data mining. This approach used to predict ranging from hourly electric energy usage to weather data in a predefined period of time. The authors used Neural Networks and Support Vector Machine (SVM) methods to extract energy consumption patterns. The authors used three different kinds of data. Those are Hourly electric energy usage for each day, Zero power consumption data and weather data for each hour. In data cleaning process missing values missing values of electric energy consumption is calculated as the hourly average value and for everyday Load patterns are described through charts. ANN has capability of identifying patterns in unsupervised clustering. SVM is based on statistical learning theory and it improves performance.

Oleg V. Divyankov et al.presented a method to forecast weather by using Artificial Neural Networks. The authors proposed a special technique to construct the learning set of images to forecast the weather events, which represents the actual data. Forecasting of the forthcoming weather events is based on the recognition of known data image using artificial neural networks and metric methods. Learning images set and recognition technique are used. 30 year database was considered for the testing of true forecast probability.

Naveed Answer et al proposed a methodology for measuring weather prediction accuracy using sugeno based adaptive Neuro Fuzzy Inference System, Grid partitioning and Guassmf. In this paper, the authors solved weather event puzzle for known industrial city Pakistan, Sialkot by implementing a fuzzy rule based system using sugeno Fuzzy Inference during the period of 7 years from January 2003 to December 2010. In this paper, two separate experimental settings have used. To develop a Fuzzy Inference System, the first experimental dataset consists of 2100 instances with 14 inputs and 5 weather events. The second dataset also consists of 2100 instances but 6 input parameters. Adaptive Neuro Fuzzy Inference System (ANFIS) model was used for fuzzy inference system in this paper. In this work, the training data consists of 70% of whole dataset was loaded in ANFIS editor.

Soumadip Ghosh et al presented a methodology for weather Data mining using artificial neural network. The main focus of this paper is based on predictive data mining by which they have extracted interesting patterns or knowledge from huge amount of meteorological data. The authors have used back propagation neural BPN network for initial modeling. The results obtained by BPN model are fed to a Hopfield Network. The performance of this approach is tested on three years weather dataset comprises 15000 records containing attributes like temperature, humidity and wind speed. Data cleaning, relevance analysis, data transformation and reduction are the preprocessing steps applied to the data to improve accuracy and efficiency.

L.A1-Matarneh et al.developed a temperature based weather forecasting model using neural networks and fuzzy logic the authors proposed computer based models for weather forecasting based on the temperature to predict the daily temperature using two techniques, artificial neural networks and fuzzy logic the main proposed from the study is to develop different weather forecasting models based on two techniques over different regions. The models had established on two different regions, amman airport and taipeil china. In this paper, back propogation algorithm was used in artificial neural networks technique they evaluate the efficiency of weather forecasting models using two measures, variance accounted for (VAF), and mean absolute error (MAE).

.EIMAN THAIMA A1 SHAMMARI et al.presented an adaptive neuro fuzzy approach on the estimation of wind turbine wake effect in this paper, the adaptive neuro fuzzy inference system (ANFIS) is designed and adopted to estimate the wind form wake effect in the wind form according to wind turbine position and distances between turbines in the wind form and router radius.

DIMITRI P.SOLOMATINE and KHADA N.DULAL (2003) presented a model for model trees as an alternative to neural networks in rainfall. In this paper The authors presented the comparative performance of two data driven modelling technique namely artificial neural network (ANNS) and modelling trees (MY`s) in rainfall runoff transformation. The applicability of these techniques is studied by predicting runoff one, three and six hours ahead for a European catchment. The result shows that both ANNs and MTs produce excellent results for 1-h ahead prediction acceptable results for 3-h ahead prediction and conditionally acceptable result for 6-h ahead prediction of runoff but the result of the ANN is slightly better than the MT for higher lead times. The advantage of the MT is that the result is more understandable.

M.A.MOHANDES et al, presented a model based on support vector machines for wind speed prediction. In this paper the authors introduced the neural network algorithm support vector machines (SVM) to wind speed prediction and compares their performance with their multi layer perception (MLP). Neural network regression method was considered in the SVM regression method is to map the data in a high dimensional feature space non linear mapping and to perform a linear regression in this feature space. The authors collected wind speed data cover a period of 12 years between 1970 and 1982. The data is divided into 3 parts training data, validation

data testing data. In this paper the performance is measured by using mean square error (MSE) the authors obtained the results as that SVM out performs MLP for all systems with orders ranging between 1 and 11. The lowest MSE on testing data for the MLP is 0.0090 while it is 0.0078 for the SVM data of order 11.

 G PETER ZHANG presented a time series forecasting model using a hybrid ARIMA and neural network. In this paper both auto regressive integrated moving average (ARIMA) and artificial neural network ANN used in linear and nonlinear modeling. With ANN`S, the non linear model form as well as the order of the model must be estimated from the data. In this paper, the proposed methodology consists of two steps. In the first step, an ARIMA model is used to analyze the linear part of the problem. In the second step, a neural network model was developed to model the residuals from the ARIMA model three well known data sets the wolf`s sunspot data, the Canadian lynx data and the British pound/ us dollar exchange rate data are used in this paper to demonstrate the hybrid method.

MOHAMMED A.MOHANDES et al .presented a neural network approach for wind speed prediction. In this paper they introduced neural network technique for wind speed prediction and compared its performance with an auto regressive model. In this method, auto correlation coefficients are used to compute the values. Back propagation algorithm was used in this paper using a generalized least-mean square algorithm the back propagation algorithm minimizes the mean square difference between the real network output and their desired output. Root mean square error (RMSE) was used to test the data the results obtained by the neural network model is better than the AR model.

4.   Conclusion:

This study concentrates on the various methodologies available in the area of weather forecasting. This paper presents a review on the brief study of models available based on data mining techniques and artificial neural networks.

References:

[1].   Valmik B Nikam and B.B. Meshram, "Modeling Rainfall Prediction using DataMining Method", Fifth International Conference on Computational Intelligence, Modeling and Simulation, Issue No: 2166-8531, PP:132-136, 2013.

[2].   F.Dell Acqua and P.Gamba, "A simple Model Approach to the problem of meteorological object Tracking", Volume: 1, Issue No: 7803-6359, PP: 2152-2154, Italy, 2000.

[3].   Evangelos Tsagalidis and Georgios Evangelidis, "The effect of Training Set Selection in Meteorological Datamining", 14th Panhellenic Conference on Informatics, Volume: 3, Issue No: 7695-4172, PP: 61-65, 2010.

[4].   James N.K.Liu, Bavy N.L.Li, and Tharam S.Dillon, "An Improved Naïve Bayesian Classifier Technique Coupled with a Novel Input Solution Method", IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews, Vol.31, No.2, PP:249-256, May 2001.

[5].   MEHMET YESILLBUDAK, SEREF SAGIROGLU and ILHAMI COLAK, "a new approach to very short term wind speed prediction using K-nearest neighbor classification", energy conversion and management, volume- 69, PP: 77-86, 2013

[6].   PIERRE JULIEN TROMBE and PIERRE PINSON, "automatic classification of offshore wind regimes with weather radar observations", IEEE journal of selected topics in applied earth observations and remote sensing, volume: 07, No: 1, PP: 116-125, January 2014

[7].   M.IQBAL, M.AZAM, M.NAEEM, A.S.KHWAJA and ANPALAGAN, "optimization classification algorithms and tools for renewable energy: a review", renewable and sustainable energy reviews, volume: 39, PP: 640-654, 2014

[8].   T.V. Rajini Kanth, v.v.s.s.s. Balaram, N.Rajashekar, "Analysis of Data Sets using Data Mining Techniques", Computer science and technology, DOI: 10.5121/esit.2014.4510, PP- No 89-94, 2014.

[9].   Kavita Pabreja , Research scholar, " Clustering Technique to Interpret Numerical Weather Prediction Output Products for Forecast of Cloudburst", International Journal of Computer Science and Information Technologies (IJCSIT), vol-3(1), ISSN:0975-9646, PP. 2996-2999, 2012.

[10].   Badhiye S.S, Dr.Chatur P.N, Wakode B.V, "Temperature and Humidity Data Analysis for Future Value Prediction using clustering Technique: An Approach", International Journal of Emerging Technology and Advanced Engineering, ISSN:2250-2459, volume-2, Issue-1, PP- 88-91, January 2012.

[11].   Zahoor Jan, M.Abrar, Shariq Bashir, Anwar M.Mirza, "Seasonal to Inter-Annual Climate Prediction using Data Mining KNN Technique", IMTIC 2008, CCIS 20,PP:40-51, 2008.

[12].   Sarah N.Kohail, Alaa M.EL-Halees, "Implementation of Data Mining Techniques for Mateorological Data Analysis", International Journal of Information and Communication Technology Research, ISSN:2223-4985, Volume-1, No:3, PP:96-100, July 2011.

[13].   Meghali A.Kalyankar, Prof.S.J. Alaspurkar, "Data Mining Technique to Analyse the Mateorological Data", International Journal of Advanced Research in Computer scince and Software Engineering, ISSN:2277-128X, Volume-3,Issue:2,PP:114-118, February 2013.

[14].   Sanjay Chakraborty, Prof.N.K. Nagwani, Lopamudra dey, "Weather Forecasting using Incremental K-means Clustering", International Conference in High Performance Architechture and Grid Computing, Vol.169, Part-2, PP:338-341,2011.

[15].   Marwa F.AI-Roby, Alaa M.El-Halees, Data Mining Techniques for Wind Speed Analysis",Journal of Computer Engineering, ISSN:2010-1619, Vol-2 ,No:1, PP:1-5, 2011.

[16].   Abhay Kumar, Ramnish Sinha, Daya Shankar Verma, Vandhana Bhattacherjee, Satendra Singh (2012), "Modeling using K-Means Clustering Algorithm", First International Conference on Recent Advances in Information Technology, Vol:4, Issue-1, Issue No: 4577-0697, PP:1-5.

[17].   Zohreh Nazeri and Jianping Zhang " Mining Aviation Data to Understand Impacts of Severe Weather on Airspace Syatem Performance", Proceedings of the International Conference on Information Technology: Coding and Computing[ITCC'OL], Volume-1, ISSN: 7695-1506, PP: 1-6, 2002.

[18].   M.Mayilvahanan and M. Sabitha, "Estimating the Availability of Sunshine using Data Mining Techniques", 2013 International Conference on Computer Communication and Informatics (ICCCI-2013), PP:1-4,2013

[19].   J.M.GUTIERREZ and A.S.COFINO, R.CANO, M.A.RODRIGUEZ, "clustering methods for statistical downscaling in short range weather forecasts", American meteorological society, volume: 132, PP:2169-2183, September 2004

[20].   VIPUL KEDIA, VAMSIDHAR THUMMALA and KAMALAKAR KARLAPALEM "time series forecasting through clustering- a case study", PP1-9.

[21].   N.RAJASHEKAR and T.V.RAJINIKANTH, "weather analysis of Guntur district of Andhra region hybrid SVM data mining techniques", international journal of engineering and advanced technology (IJEAT), ISSN: 2449-8958, Volume-3, issue-4, PP: 133-136, April 2014

[22].   ANKITH SINGH, Dr. BHUPESH GOUR, ANSHUL KHANDELWAL and HARSHA LACKHWANI, "an efficient clustering method for atmospheric conditions prediction using art algorithm", international journal of advan ced research in computer engineering and technology, volume-1, issue-1, PP: 12-17, march 2012.

[23].   NGUYEN DINH HOA, PHAM HUY THONG, LE HOANG SON, "weather now casting from satellite image sequences using picture fuzzy clustering and spatio-temporal regression", international symposium on geo informatics for spatial infrastructure development for earth and allied sciences, PP: 1-6, 2014.

[24].   S.GOKILA, K.ANAND KUMAR, A.BHARATHI, "clustering and classification in support of climatology to mine weather data-a review", international conference computing and intelligence systems, ISSN: 2278-2397, volume: 04, pages: 1336-1340, march 2015.

[25].   AASTHA SHARMA, SETU CHATURVEDI and BHUPESH GOUR, "a semi-supervised technique for weather condition prediction using DBSCAN and KNN", international journal of computer applications (0975-8887), volume 95, No: 10, PP-21-26, June 2014

[26].   HANS-HENRIK BENZON and THOMAS BOVITH, "simulation and prediction of weather radar clutter using a wave propagator on high resolution NWP data", IEEE transactions an antennas and propagation, volume: 56, No: 12, PP: 3885-3890, December 2008

[27].   DANIEL ALEZANDRO VILA, LUIZ AUGUSTO TOLEDO MACHADO, HENRI LAURENT and INES VELASEO, "fore cast and tracking the evolution of cloud clusters (for TRACC) using satellite infrared imagery: methodology and validation", weather and forecasting, VOLUME: 23, PP: 233-245, April 2005

[28].   Godfrey C.Onwubolu, Petr Buryan, Sitarama Garimella, Visagaperuman Ramachandran, Viti Buadromo, Ajith Abraham, "Self-Organizing Data Mining for Weather Forecasting", IADIS European Conference Data Mining 2007, ISBN:978-972-8924-40-9,pp.81-88, 2007.

[29].   Gurbrinder Kaur, "Meteorological Data Mining Techniques: A Survey", International Journal of Emerging Technology and Advanced Engineering, Volume-2, ISSN-2250-2459, Issue-8, PP-325-327, Auguest 2012.

[30].   S.Kotsiants, A.Kostoulas, S.Lykoudis, A.Argirou, K.Menagias, "Using Data Mining Techniques for Estimating Minimum, Maximum and Average Daily Temperature values", International Journal of Computer , Information, Systems and Control Engineering, Vol-1,

[31].   Folorunsho olaiya, Adesesan Barnabas Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Stuies", I.J. Information Engineering and Electronic Business, DOI:10.5815/ijieeb.2012.01.07, pp-51-59, February 2012.

[32].   Arti R.Naik, Prof.S.K. Pathan, "Weather Classification and Forecasting using Back Propagation Feed Forward Neural Network", lnternational Journal of Scientific and Research Publications,ISSN:2250-3153, volume-2, Issue-12, PP:1-3, December 2012.No:2, PP-382-386, 2007.

[33].   Gaurav J.Sawale, Dr.Sunil R.Gupta,"Use of Artificial Neural Network in Data Mining for Weather Forecasting", International Journal of Computer Science and Applications, ISSN:0974-1011, vol.6, No.2, PP:383-387,April 2013.

[34].   Mehrnoosh Torabi and Sattar Hashemi, "A DataMining Paradigm to Forecast Weather Sensitive Short-Term Energy Consumption", The 16th CSI International Symposium on Artificial Intelligence and Signal Processing, Volume: 4, Issue No: 4673-1479, PP: 579-584, 2012.

[35].   Oleg V.Diyvankov, Vladimir A.lykov and Serge A.Terekhoff (1992), "Artificial Neural Networks in Weather Forecasting", PP: 829-835, February 1st, 1992(in Russia).

[36].   Naveed Answer, Aneela Abbas, Aneela Mazhar and Syed Hassan, "Measuring Weather Prediction Accuracy using Sugeno based Adaptive Neuro Fuzzy Inference System, Grid Portioning and Guassmf", PP:214-219.

[37].   Soumadip Ghosh, Amitava Nag, Debashish Biswas and Jyoti Prakash Singh, "Weather Data Mining using Artificial Neural Network", PP:192-195, 2011.

[38].   L.A1-matarnch, A.SHETA, S.BANI-AHMAD, J.ALSHAER AND LA1-OQUILY,"development of temperature based forecasting models using neural networks and fuzzy logic", international journal of multimedia and ubiquitous engineering, VOL: 9, No:12, PP:343-366,2014.

[39].   EIMAN TAMAH A1-SHAMMARI, MOHSEN AMIRMOJAHEDI, SAHABODDIN SHAMSHIRBAND, DALIBOR PETKOVIC NENAD T.PAVLOVIC, HOSSEIN BONAKDARI, "estimation of wind turbine wake effect by adaptive neuro fuzzy approach", flow measurement and instrumentation, volume: 45, PP: 1-6, 2015

[40].   DIMITRI P.SOLOMATINE and KHADAN N.DULAL, "model trees as an alternative to neural networks in rainfall runoff modelling", hydrological sciences journal, volume- 48, No- 3, PP: 399-411, 2003

[41].   M.A.MOHANDES, T.O.HALAWANI, S.REHMAN, AHMED A.HUSSAINE, "support vector machines for wind speed prediction", renewable energy, VOL: 29, PP: 939-947, 2004

[42].   G.PETER ZHANG, "time series forecasting using a hybrid ARIMA and neural network model", neuro computing, VOL: 50, PP: 159-175, 2003

[43].   MOHAMMED A MOHANDER, SHAFIQUR REHMAN and TALAL O.HALAWANI, "a neural networks approach for wind speed prediction", renewable energy VOL: 13, PP: 345-354, 1998