**Weather forecast and Prediction System Using Time Series Analysis**

**Ali Farooqui, Adnan, Aveen and Javeria**

**Abstract:**

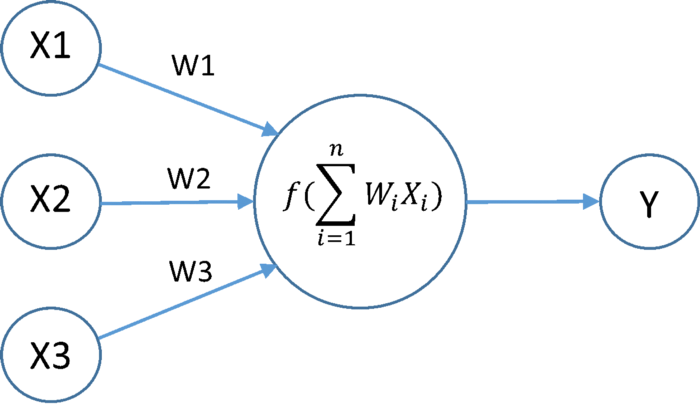
Weather forecasting is an important , vital and indispensable requirement in the outline of people. It is a technique used to predict weather conditions from science and technology through data. It is considered to be one of the most difficult issues around the globe. It controls many factors for the country’s economy and current affairs such as agriculture, air commute and e.t.c. As Big data is a game changer in every field in today's world it can also play a very vital role in weather forecasting as well. Different techniques have been suggested in the past for predicting and forecasting weather using big data or machine learning techniques. This paper provides a complete analysis and comparison of forecasting over time series analysis with techniques such as ARIMA, SARIMA and Moving Average.

**Keywords:**

Forecasting, Time series analysis, ARIMA. SARIMA, MOVING AVERAGE, Big Data.

**Introduction:**

Weather plays an important role in the World economy, people day to day routine and e.t.c. Recent events of natural disaster like Australia wildfire, 2020 Karachi’s flood, cyclones and hurricanes in US and e.t.c proves that unstable weather can create havoc and can prove to be a massive hazzard. To prevent such hazzard weather forecasting and prediction comes into play. Factors such as humidity, wind direction, apparent temperature and e.t.c play a very important role in prediction of weather. Initially it was done manually by people with expertise but now it is heavily reliant on data, science and mathematics . Weather forecasting is the scientific process of predicting the climate at desired time frame and locations. This kind of prediction involves machine learning algorithms based on mathematical models to provide prediction or a forecast of future weather by the data of past and present. Many Weather prediction models were developed in the recent past such as weather research and forecasting models which introduced regional or limited area models. WRF proved to be the most used model due to its sharp accuracy, rate and usability in different other domains. ANN (Artificial neural networks) can be used for this purpose due to its learning capabilities on previous knowledge and ever improving learning with the help of back propagation. This makes Ann very appealing in this matter. For this research paper we will be using time series technique to train our model with the one dimensional data. Time series forecasting networks such as Recurrent Neural Network and Temporal Convolution Networks are often used. A variant of Recurrent Neural Network LSTM (Long-Short Term Memory) was recently introduced and gained popularity due to its brilliant performance compared to its predecessor. Artificial Neural Network mimics the human brain. Deep Networks consist of Stacked neural layers consisting of multiple nodes per layer and each layer connected with weighted connection. Activation functions reside in the neurons which helps to process data along with weights along connection and bias per layer. Network learns with the process of back propagations in multiple iterations (epochs) until certain accuracy is achieved. Stochastic gradient descent allows faster convergence in deep neural networks.



This Paper provides a deep and detailed analysis of achieving weather forecasting through time series techniques. This paper is divided into multiple components such as related work, definitions and detailed explanation of techniques used prior, our dataset analysis, our analysis on the data set which will be based upon our working and implementation of time series analysis on python. Related work will provide explanation of the previous work done on the related field and how they proposed new techniques. Dataset analysis will provide the details of our dataset such as attributes name, datatype what factors impact weather. Techniques used prior description will give insight into what those techniques are and how they work. Our analysis will give insight of what we have achieved in our implementation and how we processed data and made predictions on the analysis. The analysis is done in the python language and the SciPy library is used to implement deep neural networks.

**Related Work**

In recent years, many people have worked on the field of weather forecasting using big data and have applied numerous techniques to find the desired results. There are multiple methods for weather forecasting and predictions using big data and numerous people have either proposed or worked on this field. In this particular section we discuss some of the major work we found during our research that supports our work as well.

Mehboob Alam and & Mohd Amjad in June 2019, primarily presented weather forecasting, counting basic progressions and altered methodologies, and then big data. They explained the number of data structures used for big data and methodologies for using Map Reduce or hadoop. They worked on big data established analytics methodology in cloud setting to forecast highest, lowest and average temperature.

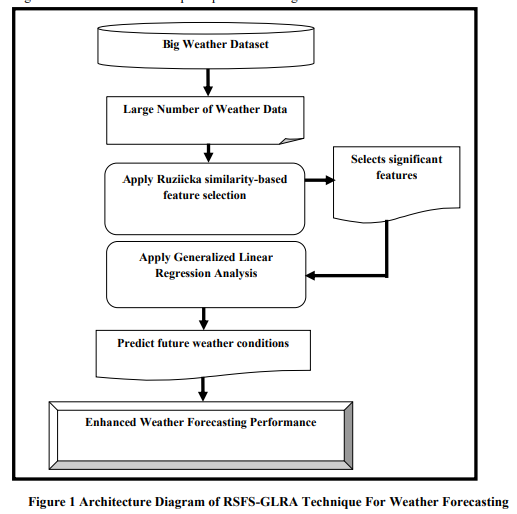
They used Hadoop with hive which uses map reduce within to progression and investigates the weather data. They predict the future weather highest, lowest and average temperature on the base of data composed from numerous weather sites. They used Hadoop to extract big data from NCDC through HDFS and then practice this data over Map Reduce.

The other technique included Deep Neural Networks, for working on big data techniques include Machine Learning, Deep Learning and Deep Neural Networks. Many researchers have worked on weather forecasting and prediction using algorithms entirely based on Deep Networks with different combinations of hyperparameters finding the optimal solution.

James N. K. Liu, Yanxing Hu, Yulin He, Pak Wai Chan and Lucas Lai shared in their paper “Deep Neural Network Modeling for Big Data Weather Forecasting” all about the challenges they experience working on big data and deep neural networks for weather forecasting. They engaged a new computational intelligence technology called stacked Auto-Encoder to mimic hourly weather data in 30 years. This technique can habitually pick up the structures and features from the enormous capacity of the dataset via layer-by-layer feature granulation, and the massive size of the dataset can clarify that the multi-layered profound model evades the over-fitting problem.

Another major technique includes different regression models such as Linear regression, Multiple Regression, Polynomial Regression and e.t.c. Apart from other techniques, Linear Regression is one of the most popular and usual techniques when it comes to Weather forecasting. It is also being used by many researchers to predict the weather forecasting system using big data.

Dr. R. Pushpalatha reasoned the RSFS-GLRA procedure achieves feature assortment and forecast development. Ruzicka Similarity-Based Feature Selection (RS-FS) procedure is accepted to select applicable features for carrying out weather forecasting. Afterwards feature assortment, Generalized Linear Regression Analysis based Weather Forecasting (GLRA-WF) procedure predicts future weather circumstances. Dr. R. Pushpalatha's proposed technique takes data from the server and feeds it to the system and performs feature selection along with prediction. It is a powerful tool that examines two separate variables on given historical data of weather and provides a relationship between them and provides a future. This evaluation provides higher accuracy with minimal time which makes it optimal. Architecture is explained in the below illustration.



Multi-class Weather Forecasting from Twitter Using Machine Learning Approaches by Kartika Purwandaria,∗ , Join W. C. Sigalinggingb, Tjeng Wawan Cenggoroa,c, Bens Pardameana,d in Sept 2020 gave a brief analysis of their experiment. They used weather data by Twitter which was acquired by text mining. They performed SVM, Naive Bayes and Logistic Regression methods. The experiment proved SVM to be superior among others with an accuracy level of 93% which further proved that SVM is suitable for Text Categorization. They also later used clustering algorithms as well.

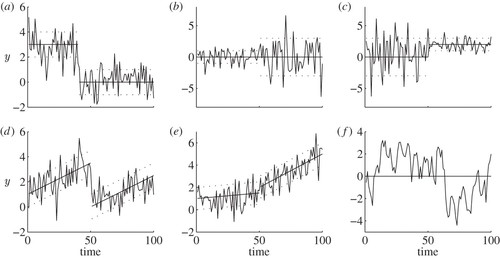


**Other Techniques Used Previously:**

During our research on topics like weather forecasting, we found aggregate researchers applying aggregate methods to do weather forecasting using big data. We hand-picked some and referenced the content of how efficiently they work.

**Change Point Analysis**

First technique was Change Point Analysis, it is progressively crucial trouble in a reach of applications, like to observe copy number variables. An individual approach to inferring the number and position of the change points is to present a model for the data inside a segment, and then exploit a penalized likelihood utility. This maximization can frequently be done precisely using dynamic scheduling, but the consequent algorithm has a procedure cost that is quadratic, or even cubic, in the number of data points.



**C5 Hybrid with K-Means Clustering**

Second technique was C5 Hybrid, K-Means Clustering, the weather forecasting model exercising the C5.0 algorithmic program with K-means clustering. The C5.0 is one of the amended decision tree classifiers, and the decision tree is an essential alternative for forecasting and prediction. The algorithm for clustering the K-means is used to cartel fraternal data put together.

**Hybrid Machine Learning**

Third technique was Hybrid Machine Learning. A vigorous hybrid technique is being used to forecast rainfall by the union of Particle Swarm Optimization (PSO) and Multi-Layer Perceptron (MLP) which are favorite kinds utilized in Feed Forward Neural Network (FFNN). The aim of using PSO with MLP is not just to forecast the rainfall but, to modify the execution of the system; this was evidenced by examination with several Back Propagation (BP) an algorithm such as Levenberg-Marquardt (LM) through with consequence of Root Mean Square Error (RMSE). RMSE for MLP supported PSO is 0.14 whereas RMSE for MLP supported LM is 0.18.

**Multi-Class Machine Learning using SVM AND LR**

The Fourth technique was Multi-Class Machine Learning using SVM AND LR. Different classification and regression models are used in order to make predictions. Svm and Linear regression standout. Linear Regression makes predictions on the basis of numerical data such as temperature and humidity and predicts numerical data temperature. While SVM predicts what kind of weather it is, be it binary or multiclass classification.

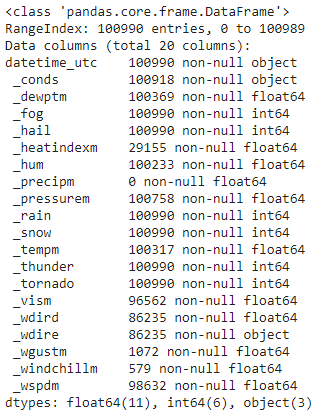
**Predicting Extreme Weather With Supercomputers and Simulations**

Massive data is fed to supercomputers to make simulations of future weather patterns, a process which is described as modelling physics from mathematics. Offices in Uk and NOAA (US) use massive and powerful machines to make micro-models of temperature, air pressure, humidity and wind speed, which are extrapolated to predict future weather events. NOAA collects 20TB of data every day.

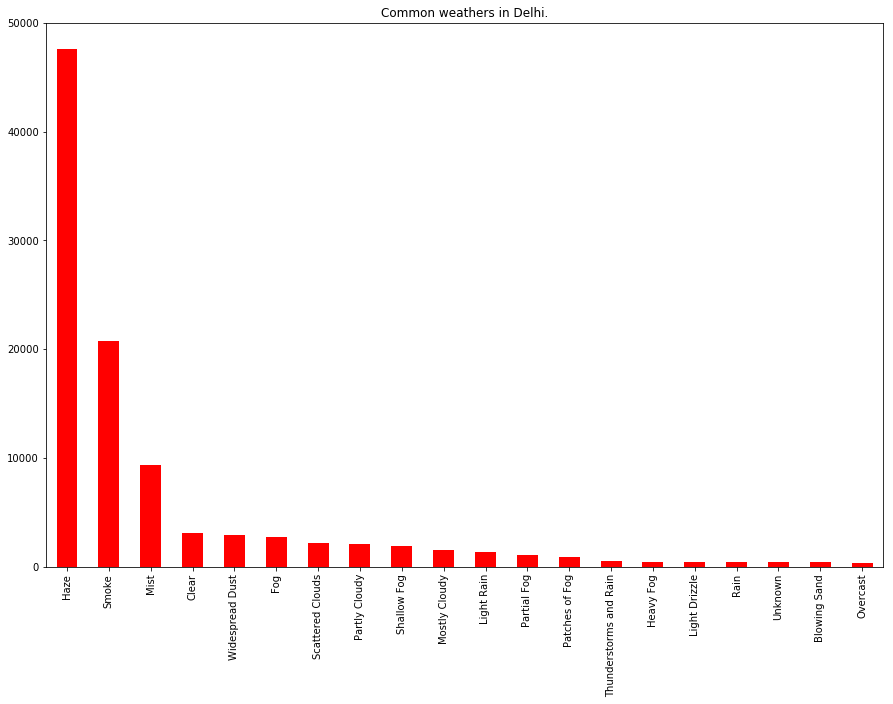
**Our DataSet:**

Here we have taken Dehli weather data which has the attributes such as date Time , weather conditions and other several attributes such as fog, rain, snow and condition in detail. We have checked columns and rows details of the dataset and the data types of each column. Also, we have filtered out the null values from our dataset. We have found nearly 10 million null data. The missing or Nan values are imputed or they are eliminated along with their data.

The following picture enlists the attributes of dataset along with their data types



We have filtered the most frequent category of **\_conds** which provides us with weather conditionsin Categorical form. We found out that the most common weather condition in Delhi was Haze, after that Smoke and then Mist. The least common weather condition in Delhi was Heavy Thunderstorms with Hail which was just 1.



Along with Light Freezing rain and Light Hail showers which were also 1 in count. Then we have put that data in a chart which shows more than 40000 times the weather of Delhi was Haze after those 20000 times the weather of Delhi was Smoke and then the count directly drops to 10000 that is Mist. And the list goes on.

We have also converted the Date Time field which converts the data into Date Time data frame for our own convenience so we can later process better .

The dataset consist of all the weather conditions and these datasets are gathered from automatic systems (big data). To process this kind of huge amount of data we need high technological hardware.

We have used forecasting techniques to process our data and make the most out of it. Furthermore we have used jupyter Anaconda along with scipy libraries on a 1080 Nvidia Graphic card which ensure smooth operation without any hardware bottleneck.

Although big data sources are available to do these things, the aim of our work is to review the approaches which are used for data forecasting in big data analytics and provide an analysis over them.

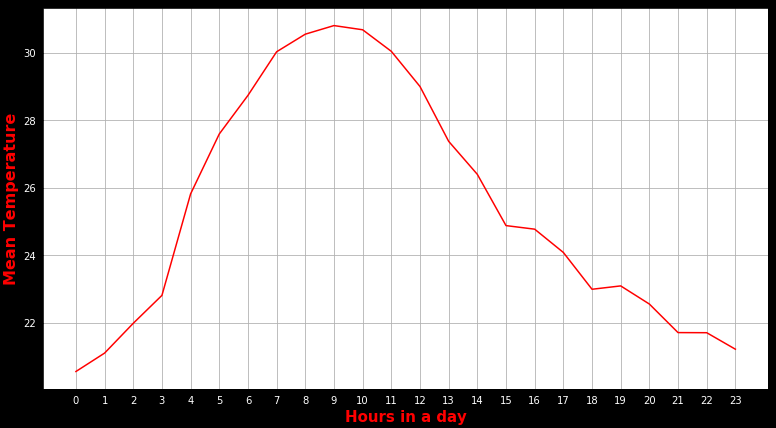
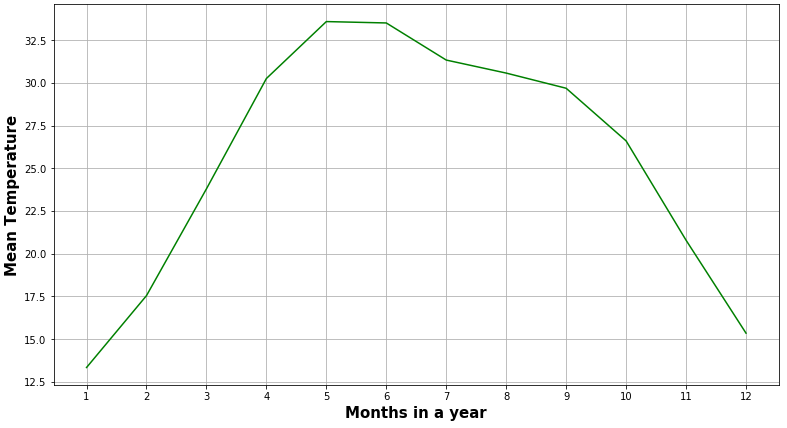
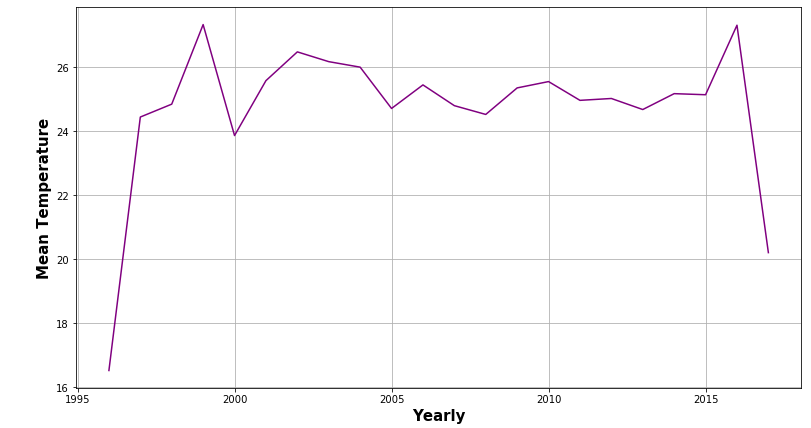
**Our Work:**

We have set the DateTime parameter to index. Now we have “datetime\_utc” as the index of our data. Now we have calculated the daily, monthly, hourly and yearly mean. Now by using hourly\_mean we will plot the graph of mean temperature on daily basis.

The graph shows that the temperature increases from 7:00 and again starts dropping at 13:00.

Again we are using monthly\_mean and after plotting the graph we came to a conclusion that April to September the temperature is highest as compared to other months.

By plotting the graph of “yearly\_mean” we can see that there is a sharp increase in temperature from 1990 and have stabalized after 2002. Following are the graphical representations.



We have also done the cleaning of data and also filled missing values. Also we have plotted the actual vs predicted data.And through the graphical representation we can see that actual data and predicted data was so close, almost identical.

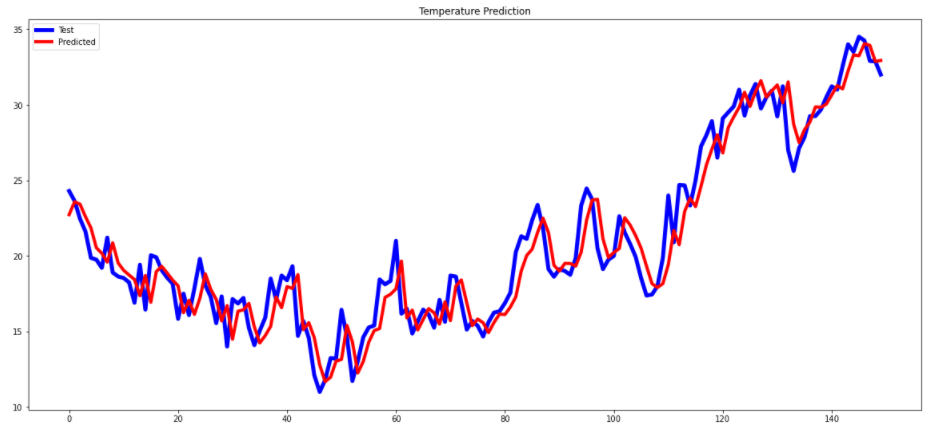
**Disadvantage of MA**

* We have used the past temperature, and we will be predicting future temperatures. This is so Naive of MA models.
* We are also seeing global warming is increasing the temperature. And in future this will increase more then we can also face change in the predicted values . To overcome this we will consider those variables which were neglected before.

RMSE value is found to be 3.6 which is pretty good.

Now we will do the forecasting using the LSTM Method. Now we will use the same dataset and we'll split the data into 100940 training records and 50 testing records. Then we'll evaluate the performance of the model.

The following graph represents the actual and predicted values using the SRIMAX model.



**Conclusion:**

We can conclude that we have compared two models. LSTM and SRIMAX. And we have found that LSTM’s RMSE is comparatively higher than theSRIMAX’s RMSE. Hence the SRIMAX is performing far better.We have obtained the RMSE value of the SARIMAX model. Which is 3.69. However Neural Networks always provide room for model improvement,thus by using valid activation functions and neurons this can be improved.

**References:**

1. **Mahboob Alam (2019) Weather forecasting using parallel and distributed analytics approaches on big data clouds** [**https://www.tandfonline.com/doi/abs/10.1080/09720510.2019.1609559**](https://www.tandfonline.com/doi/abs/10.1080/09720510.2019.1609559)
2. **James N. K. Liu Yanxing HuYulin HePak Wai ChanLucas Lai (2014) Deep Neural Network Modeling for Big Data Weather Forecasting**
3. [**https://link.springer.com/chapter/10.1007/978-3-319-08254-7\_19**](https://link.springer.com/chapter/10.1007/978-3-319-08254-7_19)
4. **Mani Mehrotra1 , Nakul Joshi (2017) Anomaly Detection in Temporal data Using K Means Clustering with C5.0** [**https://www.theijes.com/papers/vol6-issue5/L0605017781.pdf**](https://www.theijes.com/papers/vol6-issue5/L0605017781.pdf)
5. **Dr.R.Pushpalatha (2021) Ruziicka Similarity Feature Selection Based Generalized Linear Regression Analysis For Weather Forecasting Using Big Data**
6. **Robert Northcott() Big data and prediction: Four case studies**
7. **Pradeep Hewage1, Marcello Trovati1, Ella Pereira1, Ardhendu Behera (2020) Deep learning‑based effective fine‑grained weather forecasting model**
8. **Dr. S. Santhosh Baboo and I.Kadar Shereef () An Efficient Weather Forecasting System using Artificial Neural Network**
9. **Kartika Purwandaria , Join W. C. Sigalinggingb, Tjeng Wawan Cenggoroa, Bens Pardameana (2020) Multi-class Weather Forecasting from Twitter Using Machine Learning Aprroaches**
10. **Marzieh Fathi1, Mostafa Haghi Kashani, Seyed Mahdi Jameii, Ebrahim Mahdipour (2020) Big Data Analytics in Weather Forecasting: A Systematic Review**
11. **Munmun Biswas, Tanni Dhoom, Sayantani Barua (2018) Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data**
12. **Hatem Abdul-Kader, Mustafa.Abd-El salam, and Mona Mohamed (2020) Hybrid Machine Learning Model for Rainfall Forecasting**
13. **Stephan Rasp , Peter D. Dueben , Sebastian Scher , Jonathan A. Weyn , Soukayna Mouatadid , and Nils Thuerey (2020) WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting**
14. **David Kreuzer , Michael Munz, Stephan Schlüter (2020) Short-term temperature forecasts using a convolutional neural network — An application to different weather stations in Germany**
15. **Jie Lu, Tianrui Li (2020) Deep Uncertainty Quantification: A Machine Learning Approach for Weather Forecasting**
16. **Lans P. Rothfusz, Russell Schneider, David Novak, Kimberly Klockow-McClain, Alan E. Gerard, Chris Karstens, Gregory J. Stumpf, and Travis M. Smit (2021) A Proposed Next-Generation Paradigm for High-Impact Weather Forecasting**
17. **Hwan-Jin Song and Soonyoung Roh (2021) Improved Weather Forecasting Using Neural Network Emulation for Radiation Parameterization**
18. **Afan Galih Salman , Yaya Heryadi , Edi Abdurahman , Wayan Suparta (2018) Weather Forecasting Using Merged Long Short-term Memory Model**
19. **Elizabeth M. Argyle, Jonathan J. Gourley, Zachary L. Flamig, Tracy Hansen, and Kevin Manross (2017) TOWARD A USER-CENTERED DESIGN OF A WEATHER FORECASTING DECISION SUPPORT TOOL**