Journal of Green Engineering (JGE)

Volume-10, Issue-3, March 2020

Analysis of Timeseries Forecasting Models Using TamilNadu Environmental Weather Data

[1]S.Bangaru Kamatchi and [2]R.Parvathi

[1]*Research scholar, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India. Email:bangaru.kamatchis2014@vit.ac.in*
[2]*Professor,School of Computing Science and Engineering, Vellore Institute of Technology, Chennai,India. E-mail:parvathi.r@vit.ac.in*

Abstract

The prediction of yield and crop recommendation in the field of agriculture depends on the agro climatic conditions, which determines the success ratio of the crops. The minor changes in climatic condition will also affect the crop growth, production and yield. Analyzing the different climatic patterns over the years and predicting the future agro climatic condition is still a difficult task. Many machine learning prediction of climatic conditions are used over the decades, here to get better optimal results the concept of time series forecasting is taken for validating and predicting the data. the paper is dealt with the different attributes of agro climatic conditions like vapor pressure, precipitation ,maximum temperature ,minimum temperature average temperature are analyzed by different time series models and its accuracy are compared to find the optimal model for forecasting.

Keywords: ARIMA, accuracy, double exponential smoothing, Exponential smoothing, Root mean square error

Journal of Green Engineering, Vol. 10_3, 1208-1217.

1 Introduction

Weather forecasting is one of the most interesting applications in the field of data science. To predict the atmospherically conditions and changes in each location it needs tremendous computation due to varying nature of the atmosphere. The probability of error in forecasting is usually a big challenge because of this varying nature. The weather forecasting is useful in number of applications like agriculture, detecting natural calamities etc.[1][2].In the field of agriculture weather plays the major role in bringing up the vegetation. Alerting farmers about the variation in atmospheric condition will make them knowledgeable at the time of vegetation and maintenance of the crop. In this paper the prediction of climatic condition is done for agriculture. The agriculture zones climatic conditions are collected and prediction of climatic condition is done for each zone to help agriculture production.

Climate is one of the unpredictable events that affect anything and everything in the world. The major changes will cause the natural calamities will affect the people directly but the victim of minor changes in climatic conditions are the crop. The crop's each and every stage is determined by the amount of heat it gets, the water it needs, humidity level in the crop and many more. The pattern of weather occurring over number of decades does not show any steep changes but the seasonality change over them is noticeable. The temperature increase is well known. Over the years the occurrence of minimum temperature in a month has increased. It is well known that the temperature is going to increase in the upcoming years in order to predict that many machine learning algorithms are used. Each and every model produces different kind of results with different level of accuracy.

Early endeavors to consider time arrangement, especially in the nineteenth century, were for the most part portrayed by the possibility of a deterministic world. It was the significant commitment[3]that propelled the concept of stochasticity in time arrangement by hypothesizing that each time arrangement could be regarded as recognizing a stochastic process. In light of this straightforward thought, various time arrangement techniques have been created from that point forward. Wold's decay hypothesis prompted the plan and arrangement of the direct estimating issue[4] A quarter century prior, exponential smoothing strategies were regularly viewed as an assortment of specially appointed methods for extrapolating different kinds of univariate time arrangement. Albeit exponential smoothing techniques were generally utilized in business and industry, they had gotten little consideration from analysts and didn't have a well-created measurable

establishment.[5][6] The Exponential smoothing methods which was introduced in the late years of 1985.The number of smoothing methods used this as foundation in the prediction using time series model. [7] provided a detailed view on how an exponential smoothing method is used fin classification and analyze the trend pattern present in the data. Many machine learning algorithm like support vector machine [8] is used for financial time series forecasting in addition to that adaptive parameters are also used in [9]to make the better forecasting. The ARIMA model in [10]a hybrid approach is introduced along with Ann which outperforms the other one. The nonlinear data forecasting which can be handles by another hybrid model in [11].seasonal arima model is also introduced in [12] for handling seasonal component This paper brought together a lot of existing time series methods on handling weather related attributes which paves the way for understanding the model and prediction of data for this kind of weather The role of machine learning and methods are also used in radiology.[13]for daily weather prediction [14]and also in weather forecasting of various regions[15]

2 Materials and Methods

The climatic data are collected from the indiawaterportal website on which it has general monthly mean of different attributes from the year 1901 to 2012. The month wise data for all years are considered for prediction. Attributes like minimum and maximum temperature and vapor pressure has been used since they play key role in plant growth and production. The relative humidity which is one of the most important factor for plant growth hence the relative humidity is calculated from the vapor pressure data using saturated vapor pressure (v_s) and the actual vapor pressure (v)

$$v = 6.11*10(7.5*Td/237.3+Td) \quad (1)$$

$$v_s = 6.11*10(7.5*T/237.3+T) \quad (2)$$

2.1 Time Series Forecasting

In time series any particular metric which is recorded over regular time interval as a sequence .It uses number of methods for both univariate and multivariate forecasting. For understanding the distribution and to analyze the data suitable model has to be selected. There is list of different models available in Time series forecasting .In analyzing the agro climatic parameters 4 different models of time series forecasting has considered and their results are compared to show the better model for this kind of data.

There are twelve different attributes present in the dataset and only concern methods taken here are the univariate methods. Each attribute is forecasted individually, to show it as an example minimum temperature is forecasted and its distribution over time is shown in Fig.1 Time series which has a specific behavior over time and there is a very excessive probability that will observe the same in future .In order to prove that the data points of time series should be stationary. The term stationary here refers that it must have the constant mean, standard deviation and the time independent auto covariance.

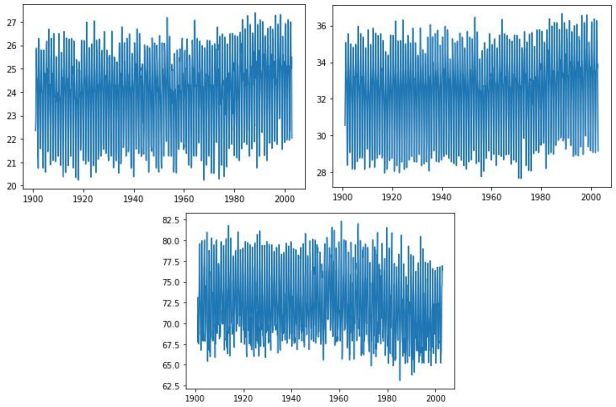

Fig.1 .Minimum Temperature,Maximum Temperature &Relative Humidity

To check whether the data points are stationary two popular test is carried out one is rolling statistics shown in figure 2 and the other one is dicky fullers test. Here both test has been carried out to check the stationarity. For the rolling statistics rolling mean and rolling standard deviation is calculated which is shown in figure for that the window size is set to 12 so that for every 12 value the mean and standard deviation gets updated and in the Dicky-Fuller test statistics and critical value for 1% 5% and 10% is calculated shown in Table1. If the t value is less than the critical value the data points are assumed to be stationary and if not further differencing has to be done.

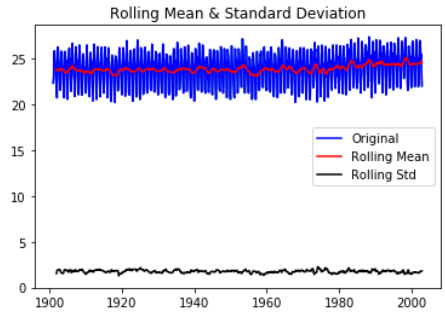

Fig.2. Rolling statistics

The below table1 shows that the t value is less than the critical value hence no need of further differencing and the points are stationary. Once the data points are found to be stationary three important components of time series forecasting has to be generated they are trend, seasonal and residual. The trend component which shows any increasing or decreasing pattern over the time, the seasonal component shows the repeated pattern over a time of period and the residual is the error, these three components for minimum temperature is shown in the Fig.2.

Table1: Results of Dickey-Fuller Test

Test Statistic	-2.557580
p-value	0.102085
#Lags Used	23.000000
Number of Observations Used	1200.000000
Critical Value (1%)	-3.435811
Critical Value (5%)	-2.863952
Critical Value (10%)	-2.568054

From the Fig.3 it is shown that there is no trend for this particular data points but seasonal component is present depending on that time series model has to be built. Here 4 different models are taken for consideration.

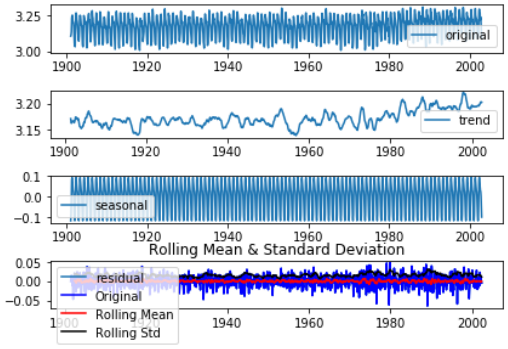

Fig.3. Original Trend, seasonal Residual Component

2.2 ARIMA Model

Autoregressive integrated moving average model which is one of the well-known and most effective models of time series data. It works on its own lagged data and errors on that lag to forecast the future values. The univariate time series which uses only the previous values to predict the future data .

$$Y_t=\alpha+\beta_1 Y_{t-1}+ \beta_2 Y_{t-2}+.+ \beta_p Y_{t-p}\varepsilon_t+\Phi_1\varepsilon_{t-1}+ \Phi_2\varepsilon_{t-2}+\ldots+ \Phi_q\varepsilon_{t-q} \quad (3)$$

Where Y_t is the differenced series .The "predictors" on the right hand side include both lagged values of Y_t and lagged errors.

p= order of autoregressive part
d=degree of first differencing involved
q=order of moving average part

Selecting the value for p & q is quite difficult, to find out p Autocorrelation plot is plotted and for q Partial autocorrelation plot are drawn which is shown in the Fig.4 .here in ACF the curve first crosses the horizontal line somewhere between 1 to 3 and in PACF between 1 to 2 hence the p and q values are taken as ARIMA (2,1,2).

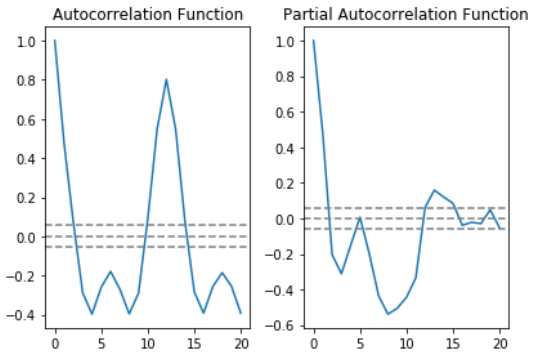

Fig.4 ACF and PACF

2.3 Simple Exponential Smoothing

In simple exponential methods the forecast is done by calculating the weighted average in which more weights are given to the most recent one and minimum weight for the data in the past. A is the smoothing parameter. The one-advance ahead conjecture for time weighted normal of the entirety of the perceptions in the arrangement of y The rate at which the loads decline is constrained by the parameter αα. For any αα somewhere in the range of 0 and 1, the loads connected to the perceptions decline exponentially as we return in time, thus the name "exponential smoothing". On the off chance that αα is little (i.e., near 0), more weight is given to perceptions from the more inaccessible past. In the event that αα is huge (i.e., near 1), more weight is given to the later perceptions.

2.4 Double Exponential Smoothing

In univariate time series to explicitly support trend an extension in exponential smoothing method is done and named as Double Exponential .In order to do this a parameter is added along with the alpha parameter to control the decay of the influence of the change in trend called *beta* (*b*). The approach supports trends that alternate in unique ways: an additive and a multiplicative, relying on whether or not the style is linear or exponential respectively. Double Exponential Smoothing with an additive trend is classically referred to as Holt's linear trend model, named for the developer of the approach Charles Holt. For longer range (multi-step) forecasts, the style may proceed on unrealistically. As such, it can be beneficial to dampen the vogue over time. Dampening ability lowers the measurement of the style over future time steps down to a straight line (no trend).

2.5 Holt-Winter Multiplicative Model

Holt-Winter is utilized for exponential smoothing to make transient conjectures by utilizing added substance" or "multiplicative" model with expanding or diminishing pattern and regularity, Smoothing is estimated by beta and gamma parameters in Holt's model. • If the beta parameter is set to FALSE, the capacity performs exponential smoothing • Gamma parameter is utilized for the regular segment. On the off chance that the gamma parameter is set to FALSE, a non-regular model is fitted. The gamma and beta qualities are set somewhere in the range of 0 and 1, the qualities are near 0 (zero) determines that weight is put on latest perceptions when building gauge of future qualities.

3 Results and Discussion

The different time series model which are validated using the metrics root mean square value which are shown in the table 2 below. Among all that the exponential smoothing model which shows the minimum rmse value compared to other models. In SES, Holt winter and ARIMA model no seasonal variation is predicted. The exponential model which shows the seasonal variation

Table 2. Comparison of error value of all the models

Index	Model	Rmse
0	SES	2.945668
1	ES	0.627927
2	HOLT'S	3.786478
3	ARIMA	3.65543

As an example the above models are built to forecast the minimum temperature similarly other climatic parameters are forecasted. On forecasting all the other parameters ES model performs better with the minimal error which is shown in Table2.

4 Conclusion

The agro climatic variables affect the crop growth and yield of the crop throughout the year. By predicting these kind of variables helps the farmers in making necessary precautions of what must be done with the crop if there is any irregularities on these parameters .The time series forecasting methods

which has number of models among all that four different kind of univariate methods ARIMA, Simple exponential smoothing, Double exponential smoothing and multiplicative holt's model are used to find the optimum model for predicting the agro climatic variables in future for increasing the yield in agriculture crops. The metrics which is used for comparison is the RMSE value. The ELT model which produce the minimal error of 0.62 and handle the seasonal component in the data points.

References

[1] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," vol. 22, no. 44, pp. 443–473, 2006.

[2] V. M. Krasnopolsky and M. S. Fox-Rabinovitz, "Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction," Neural Networks, vol. 19, no. 2, pp. 122–134, 2006.

[3] R. Society, "On a Method of Investigating Periodicities in Disturbed Series , with Special Reference to Wolfer ' s Sunspot Numbers Author (s): G . Udny Yule Source: Philosophical Transactions of the Royal Society of London . Series A , Containing Papers of a Mathem," vol. 226, no. 1927, pp. 267–298, 2017.

[4] Y. Radhika and M. Shashi, "Atmospheric Temperature Prediction using Support Vector Machines," Int. J. Comput. Theory Eng., vol. 1, no. 1, pp. 55–58, 2009, doi: 10.7763/ijcte.2009.v1.9.

[5] I. Maqsood, M. R. Khan, and A. Abraham, "An ensemble of neural networks for weather forecasting," Neural Comput. Appl., vol. 13, no. 2, pp. 112–122, 2004.

[6] K. Abhishek, M. P. Singh, S. Ghosh, and A. Anand, "Weather Forecasting Model using Artificial Neural Network," Procedia Technol., vol. 4, pp. 311–318, 2012.

[7] E. S. Gardner, "Exponential Smoothing: The State of the Art," vol. 4, no. August 1984, pp. 1–28, 1985.

[8] K. J. Kim, "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, no. 1–2, pp. 307–319, 2003.

[9] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," IEEE Trans. Neural Networks, vol. 14, no. 6, pp. 1506–1518, 2003.

[10] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," Appl. Soft Comput. J., vol. 11, no. 2, pp. 2664–2675, 2011.

[11] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," Appl. Math. Lett., vol. 22, no. 9, pp.

1467–1470, 2009.
[12]J. C. B. Gamboa, "Deep Learning for Time-Series Analysis," 2017.
[13]P. Radha and B. Meena Preethi, "Machine learning approaches for disease prediction from radiology and pathology reports," Journal of Green Engineering, vol. 9, no. 2, pp. 149–166, 2019.
[14]M. Narvekar and P. Fargose, "Daily Weather Forecasting using Artificial Neural Network," Int. J. Comput. Appl., vol. 121, no. 22, pp. 9–13, 2015.
[15]A. Tealab, H. Hefny, and A. Badr, "Forecasting of nonlinear time series using ANN," Futur. Comput. Informatics J., vol. 2, no. 1, pp. 39–47, 2017.

Biographies

Bangaru Kamatchi S research scholar at Vellore Institute of Technology, Chennai. She has completed her Masters in Computer Science Engineering from Anna University, Chennai. She has a teaching experience of 1.5 years.

Parvathi R completed her doctoral degree from Anna University, Chennai, India, by contributing her ideas to the field of spatial data mining. She has a teaching experience of over 20 years in the field of computer applications. Her research interests include Data Mining, Recommendation Systems and Social Network Analysis. She has authored articles in big data analytics for renowned publications.