International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-11, September 2020

Weather Prediction for Tourism Application using Time Series Algorithms

Abhijit Kocharekar, Bharat Nemade, Chetan Patil, Durgesh Sapkale, Sagar Salunke

Abstract*: Precise projections of future events are crucial in many areas, one of which is the tourism sector. Usually counter-trials and towns spend a enormous quantity of cash in planning and preparation to accommodate (and benefit) visitors. Precisely predicting the amount of visits in the days or months, that follow would benefit the economy and tourists both. Previous studies in this field investigate predictions for a nation as a whole rather than for fine-grained fields within a nation. Weather forecasting has drawn the attention of many scientists from distinct research communities due to its impact on human life globally. The developing deep learning methods coupled with the wide accessibility of huge weather observation data and the advancement of machine learning algorithms has motivated many scientists to investigate hidden hierarchical patterns for weather forecasting in large amounts of weather data over the previous century. To predict climate information accurately, heavy statistical algorithms are used on the big quantity of historical information. Time series Analysis enables us know the fundamental forces leading to a specific trend in time series data points and enables us to predict and monitor information points by fitting suitable models into them. In this study, Holt-Winter model is used for predicting time series. The forecasting algorithm for Holt-Winters enables users to construct a time series and then use that data to forecast interest areas. Exponential smoothing allocates weights and their respective values against past data to decrease exponentially, to decrease the weight value for older data.

***Keywords* : Tourism Industry, Weather Forecasting, Time Series Analysis, Holt-Winter algorithm.**

I. INTRODUCTION

Climate and weather are essential considerations in tourism decision making and also affect the effective system of tourism businesses. Tourist industry is a contributing sector to the world economy.

Revised Manuscript Received on September 30, 2020.
***** Correspondence Author
Abhijit Kocharekar*, Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: abhijitkocharekar8@gmail.com
Bharat Nemade, Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: bharatnemade21@gmail.com
Chetan Patil, Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: patilchetan2498@gmail.com
Durgesh Sapkale, Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: d2s2802@gmail.com
Prof. Sagar Salunke, Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: sagargsalunke@gmail.com

Indeed, the economies of some nations derive most of their income from tourism industry. The upheaval in individual revenue and the promotion of their attractions by distinct nations led the sector to evolve.

For the economy of the country, tourism in India is essential and is increasing quickly. Many kinds of tourism depend on the weather and, by extension, depend on the climate. Therefore, sooner or later, climate change is probable to impact your business area. Climate change can, for instance, decrease snow cover, boost and prolong heat waves, or change annual rainfall patterns. Majorly, all tourist destinations are climate sensitive, climate has a major impact on travel planning and travel experience. Risk identification can be accomplished by studying this climate change and its effect on the tourism industry. Proper tourism management and tour planning can be efficiently done by evaluating these risk variables. The government can therefore take proper measures and the holiday planners can plan the tours effectively. Weather forecasting is an appealing research topic with extensive potential applications ranging from flight navigation to farming and tourism. Also other thrust areas where weather forecasting can be proved to be essential include Air Traffic Control (ATC), Voyage planning, Military applications, Transport industry etc. Weather forecasting can also have a significant effect on various sports. Intelligent systems based on machine learning algorithms have the ability to learn from previous knowledge or historical information and have thus gained considerable recognition in the Computer Science Community. Weather Prediction and Forecasting is an application of science, research and technology to predict the climate for a specified place and specified instance of time using machine learning algorithms. The weather forecasting problems involve learning how to interpret the weather representations using an immense amount of weather data. For this purpose, analysis of various information using mining procedures is to be carried out. Data mining methods allow users to analyze, classify and condense the known associations from a broad range of sizes or angles. Classification, learning and prediction are some basic terms linked to data mining.

II. RELATED WORKS

Related research included many distinct and exciting methods of weather forecasting. While much of the present prediction technology includes physics-based simulations, many advanced methods of artificial intelligence primarily used machine learning methods, majorly neural networks, though other methods operate upon Bayesian networks based on probabilistic models.

Neural networks unlike the other regression techniques are seen as the usual machine learning model alternative for weather prediction due to the capacity to acquire all the non-linear dependence on pre-existing weather trends and future weather.

This gives the benefit of not assuming that all characteristics are easy linear dependencies over our models. One of the research suggested hybrid system which utilizes NN for weather prediction[1].The other studies have used direct learning which predicts weather parameters such as, SVM models used by some of the researches[2] which directly enacted the classifier. Over the last decade, countless important attempts have been documented with effective outcomes and using statistical modeling to fix weather prediction problems, including machine learning approach[3][4].Different approaches have been employing the weather forecasting systems such as neural network-based algorithms utilising the hopfield networks and BPN[5], Predictive analysis incorporating Naive Bayes Algorithm in Apache Hadoop Framework [6], Decision Tree Algorithms and Artificial Neural Network[7],RNN and CNN algorithms[8]. There are some studies to assess the holt-winters forecasting algorithm. This algorithm is a model of time dependent characteristics. It is based on time series forecasting method. One of the studies has done the comparison of Holt-Winters and adaptive regression in terms of their performance[9]. It shows there are many upsides to the method. For a wide school of data types it is simple to understand and to implement. The study revealed that this model's prediction was moderately accurate than the results of regression. Some researches have used the holt-winters exponential smoothing for forecasting, one of it has done the accuracy based comparison of two holt-winters approaches[10], additive and multiplicative. It depicts that Holt - Winters multiplicative method forecasting accuracy estimates were better than the additive method. Additive methodology predicted monthly data on the basis of the precision measurements. One of the study paper we studied focuses on analyzing data from the seasonal time series using exponential smoothing methods from Holt-Winters[11].In this scenario, additive seasonal model and multiplicative seasonal model are two approaches discussed in this paper. The test results of five different time series were also obtained by the author. Mean absolute percentage error is the principal parameter that is selected to estimate the error.The Results obtained indicate that for all the series the multiplicative model gave better outcome than the additive model. The adaptive multiplicative model gave better performance for most of the series as compared with the non-adaptive multiplicative model.

III. PROPOSED METHOD

1. *Data Acquisition:*

The weather data set supplied by the Indian Meteorological Department has been taken into account, consisting of various attributes such as air temperature, atmospheric pressure, relative humidity, average wind speed, maximum air temperature, horizontal visibility,minimum dew point air temperature.The dataset contains a single daily record which are noted from January 1, 2000 to April 29, 2019.

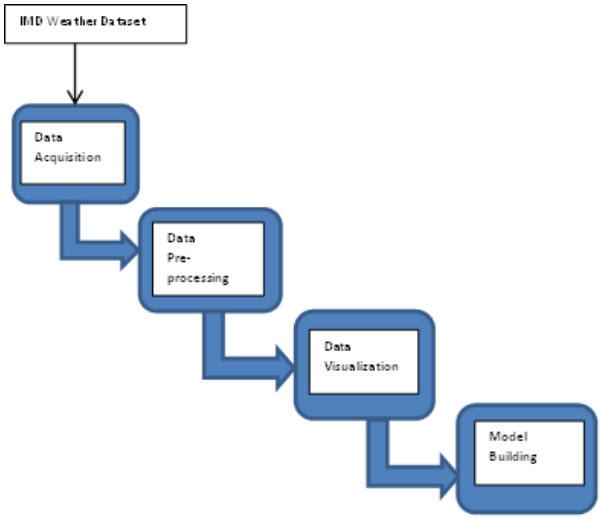

Fig. 1. Workflow of the proposed system

2. *Data Preprocessing:*

As part of pre-processing, the missing values must be handled in the dataset to ensure that the results generated are more accurate. Since most data is available in numerical format, and many a times the data values are missing over a particular range as well as discrete data values. Hence the missing values in this case are handled by forward filling method and the missing dates in the dataset are handled using "date_range" function in pandas library.

3. *Data visualization:*

Data visualization provides additional interpretation possi bilities.Visual presentation for individuals is often more read able than any other presentation of dataWeather data in itself is massive. Due to the fact that the data is in enormous amo unts, data visualization helps to understand the dataset. It hel ps to see patterns, trends and correlation in the attribute values that may go unnoticed. It also helps to identify climat e data patterns and their variations throughout the year.
These visualisations can be well understood by the graphical results we have obtained from the HP-Filter, Decomposition and Moving Averages and hence understand and analyse the dataset based on graphical visualisations.

4. *Hodrick-Prescott Filter for Data Smoothing*

HP filter is a method of smoothing data ,which eliminates variations in the short time frame. Data Smoothing helps to eliminate randomness in the data and helps to split the trend and cyclic characteristics of the data. The Elimination of the short term variations is useful to understand the long term variation in the time series graph. The lambda parameter for modelling the time series data used is 6.25,as the weather data is annual data. This filter helps determine variations in the long term time frame by slashing importance regarding short-term variations.

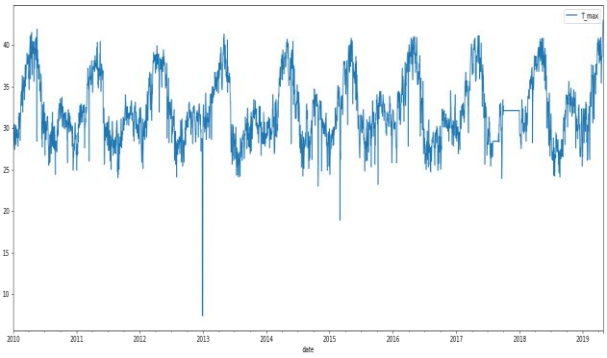

Fig. 2. T_max without using any filter

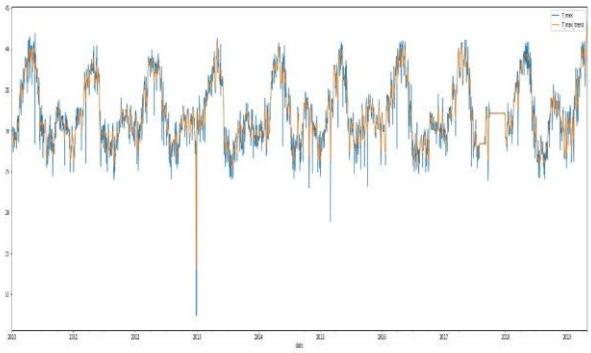

Fig. 3. T_max vs T_max_trend using HP-filter

5. ***Decomposition***

Time series decomposition provides a useful abstract model for time series analysis and individual component analysis. Trends, error and seasonality are the main factors of the TSA. As the time series seasonal factors are relatively constant in our use-case, we have used the decomposition additive model. A time series additive model is summation of Trends, Error and Seasonal elements of TSA model.

Additive Decomposition:

Time series = Seasonal + Trend + Random

Trend: The trend part helps to describe the low frequency variations in a time series, having filtered out the high and medium frequency fluctuations.

Seasonality: The Seasonality part helps to describe the behaviour that is common in every cycle of the time series graph. In case of weather data, in summer the temperature is high as compared to the temperature in winter season and this characteristic is constant for every year in the weather dataset.

Error :Error part is used to describe the characteristics that are left over after the seasonal and trend parts are eliminated.

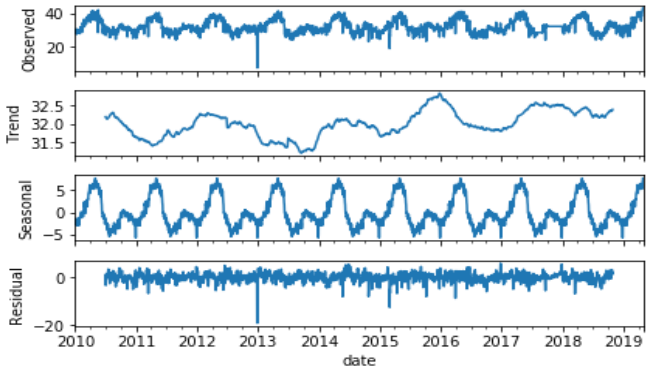

Fig. 4. Time Series Decomposition technique

6. ***Moving average***

Moving average(MA) calculation means building a new time series where the values are generated with help of calculating averages in the original time series.

Moving averages helps to understand how weather data set values have evolved in a given time frame.MA help to indicate the existing trends in the dataset .The main categories of moving average include methods such as simple(SMA) and exponential(EMA).

The SMA can be calculated by taking average considering the last n terms, in this case taking the last n data points from the weather dataset.

SMA = ((P1 + P2 +... + Pn) / n)

The EMA can be calculated by giving higher importance for the latest points in the dataset. The difference is EMA gives higher importance for the latest points while SMA gives the same importance for every point in dataset.

EMA can be expressed by the following formula,

$$EMA_{\text{Today}} = \left(\text{Value}_{\text{Today}} * \left(\frac{\text{Smoothing}}{1+\text{Days}}\right)\right) + EMA_{\text{Yesterday}} * \left(1 - \left(\frac{\text{Smoothing}}{1+\text{Days}}\right)\right)$$

Technique to calculate Exponential Weighted Moving Average is first we calculate the Simple Moving Averages .Then we generate the multiplier for EWMA by using the following equation:

[2 ÷ (selected time + 1)]

7. ***Model building:***

Weather forecasting is the system's main objective. Holt-Winter method is the most suitable method in this use case to predict weather for the next 15 days. In this case by analysing the effect of Exponential Smoothing techniques like SES,DES and TES, we find the basic parameters that are trend seasonality and error. The elements add together in an additive time series to form the time series. When you have an increasing trend, you will always see peaks and troughs about the same size in the graph. Hence, Holt-Winter additive method is used. Once the future 15 days forecasts are generated by the model, the most favourable date is selected in accordance to the destination and the weather conditions. For example, if a tourists intends to visit Juhu beach for vacations, the most favourable conditions are the days where the maximum temperature and Humidity is selected. Our system considers these parameters and returns the most favourable day for visiting the tourist destination to the end user.

IV. ALGORITHM

Times series provides a collection of observations assigned to a particular parameter at various time intervals. Such information must be gathered daily, weekly, monthly or annually (e.g. government budget) at periodic intervals. Time Series is used in statistics, finance, prediction of earthquake, forecasting of weather and many other applications.

The main use of Holt-Winters algorithm, is to predict a time based parameter as a dependent variable. This algorithm is used for analysis of data, hence the dependent variable is used to model future predictions established by previous data.
This model shows time dependent behaviour throughout the time series. This algorithm is a technique to represent factors such as a level , slope for specific time as well as a cyclic repetitive behaviour. This method incorporates exponential smoothing to embed many past values and forecast present and future values. In the time series the three components of actions such as seasonality, value and trend, thus this algorithm is termed as triple exponential smoothing.
This algorithm calculates actual and future parameters by measuring cumulative impact generated by these variables and assigns the parameters such as alpha, beta and gamma.
The 3 types of smoothing techniques and they are as follows,

1. Single Exponential Smoothing (SES) –

Without trend or seasonal pattern, this technique is optimal for data forecasting, where data thresholds can evolve over time. SES is the time series methodology that estimates the forecast for one-dimensional data lacking the above mentioned parameters. It needs a one-dimensional attribute, also termed as the smoothing component called alpha. The Alpha coefficient is assigned to a range of zero to one.

High ranges reflects latest historical outcomes, while in making a prediction, lower value suggests that more background is taken into consideration.

2. Double Exponential Smoothing (DES) –

This technique of smoothing is used to predict data where trends exist. Specifically, this approach is a modification in the SES to catch trends. DES is an extended form of SES that assigns aid to univariate time series trends. The beta parameter is for computing DES.DES with an additive pattern can be moulded into Holt Linear pattern model.

Trends associated with DES are,

Additive method is used to model one-dimensional patterns.

Multiplicative method is used to model exponentiality. The trend may continue unrealistically on longer range (multi-step) forecasts. As such, a dampening of the trend over time can be useful.

3. Triple Exponential Smoothing (TES) –

This technique is used with trend and or seasonality data for forecasting. TES is an extended form, which is specifically used to model seasonality into one-dimensional time series. The gamma parameter is used to model seasonality in TES. These 3 parameters can be modelled into additive or multiplicative procedure to frame the seasonality factor.
Trends associated with TES are:
Additive seasonality is used for TES with a linear seasonality.
Multiplicative seasonality is used for TES with an exponential seasonality.
Based on these terminologies, we have implemented the additive method, in Weather Prediction use-case.

4. Holt-Winters' additive method

The algorithm is modeled as follows,

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$
$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$
$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},$$

The three equations are for level, trend and seasonality components of time-series.
The level equation helps to derive the actual dependent term based on previous values. For example, the temperature will be derived based on the most recent data for temperature.
The trend equation helps to correlate how the data is varying over a period of time. Trend helps to determine whether the time series trajectory is going upwards, downwards or is remaining constant.
The Seasonality equation helps to determine the behavior that is common in every cycle of the time series graph. For example, the seasonality for temperature parameter in the summer season will be different as compared to that of winter season.
Consider maximum temperature as T_max that is an major attribute from our model. Hence by applying the above mentioned techniques in our use-case we have the following results,

1) SES and DES techniques over T_max from year 2010 to April 2019 ,

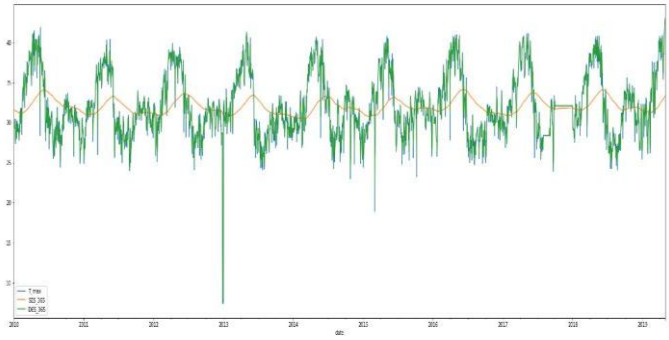

Fig. 5. T_max over SES and DES

2) DES and TES techniques over T_max from year 2010 to April 2019,

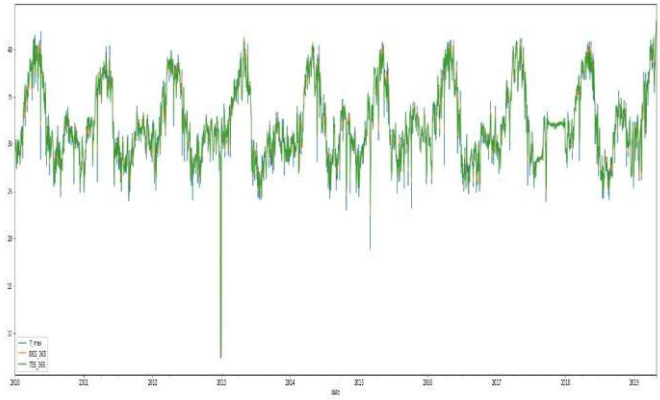

Fig. 6. T_max over DES and TES for 2010-2019

3) DES and TES upon T_max within last 365 days,

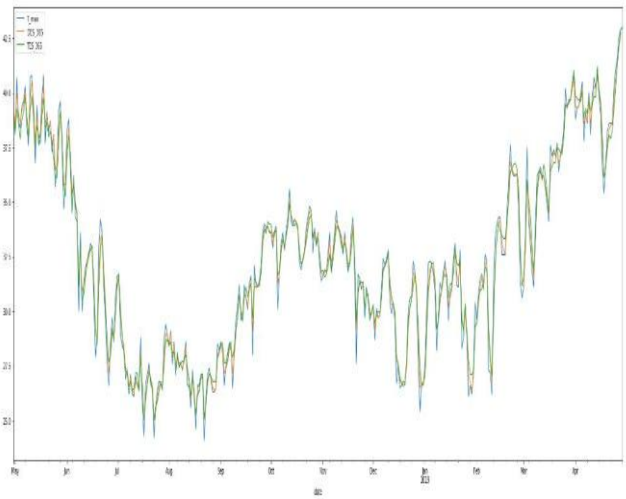

Fig. 7. T_max over SES and DES for last 365 days

Rather than the traditional moving averages(MA) methods that don't give an insight to the seasonality factor, Holt-Winters algorithm provides a detailed consideration of this factor. Apart from that the trend factor captured in this model is better than other exponential smoothing techniques. Another major advantage of using this technique is that the forecast error chart does not show trend.

V. RESULTS

Considering the scope of a specific area, we selected the weather data of Pune under consideration to demonstrate the below results. Similarly, this method is applied for all the other tourist places that are considered in our project. In view of the practical approach of this model to the Weather dataset, we reached the following results and they are as follows: i) The forecasts are done for the future two weeks. In this case, the mean absolute error incurred of parameter t_max (maximum temperature) is 1.543.The root mean square error is 1.918.

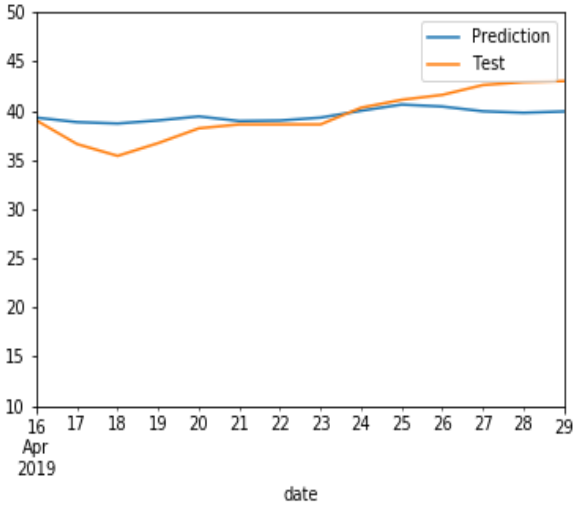

Fig. 8. Prediction vs Test results for Maximum Temperature attribute

ii) The forecasts are done for the future two weeks In this case, the mean absolute error incurred of parameter t_min (minimum temperature) is 2.341.The root mean square error is 2.923.

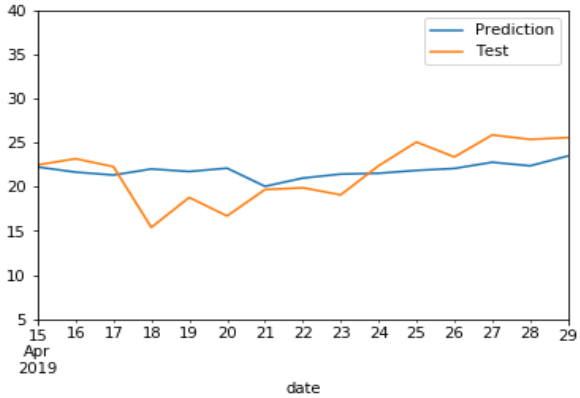

Fig. 9. Prediction vs Test results for Minimum Temperature attribute

iii) The forecasts are done for the future two weeks In this case, the mean absolute error incurred of parameter dpt (dew-point temperature) is 2.149.The root mean square error is 2.647.

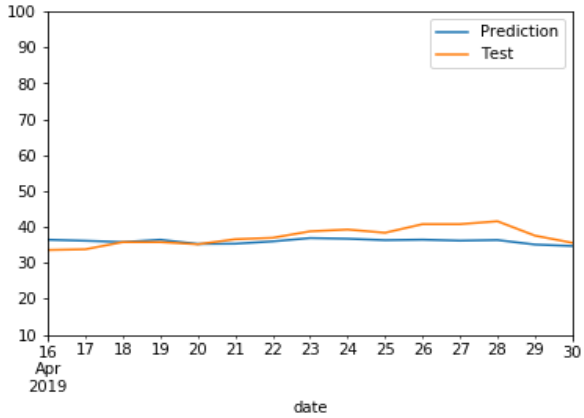

Fig. 10. Prediction vs Test results for Dew Point Temperature attribute

iv) The forecasts are done for the future 3 months.The objective is to predict the seasonal change in terms of humidity parameter. The mean squared error is 5.023..The root mean square error is 6.628.

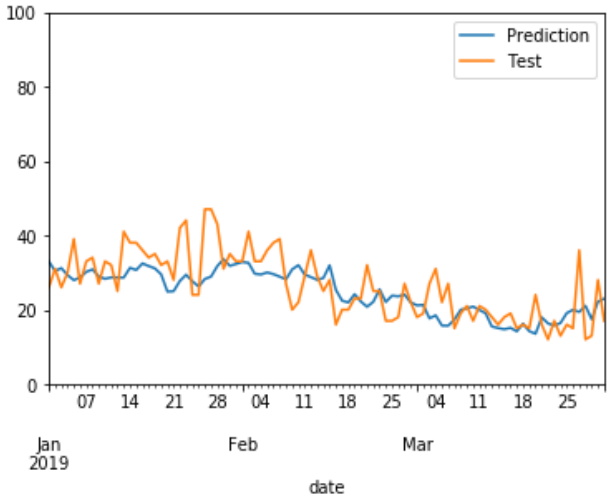

Fig. 11. Prediction vs Test results for Humidity attribute

From the results generated, the tourists can get a detailed forecast of the weather attributes that we have considered for the next 15 days .The following figures demonstrate the exact output of the respective attribute in a graphical format.

i) The following figure shows the maximum temperature results for the next 15 days,

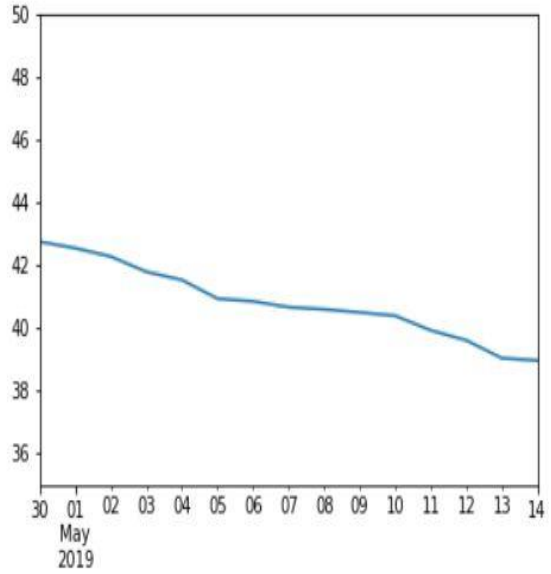

Fig. 12. Prediction for next 15 days maximum temperature

ii) The following figure shows the minimum temperature results for the next 15 days,

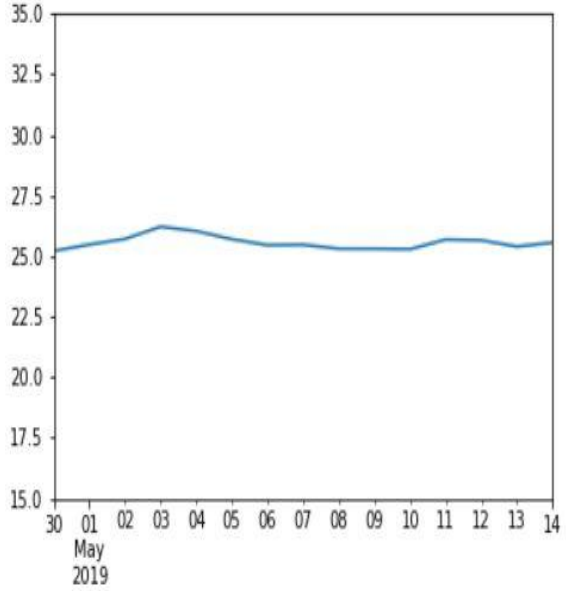

Fig. 13. Prediction for next 15 days minimum temperature

All the results are predicted for temperature of summer values of dataset. The papers from this results are studied have mostly used the Artificial Neural Network Techniques for forecasting.

As each paper has used the different datasets and different scenarios , it is not a correct measure to compare their results based on the accuracy. The table given below compares several algorithms along with their performance measures represented by the root mean squared error.

Table I. Comparative Study of Different Algorithms

ALGORITHMS	PERFORMANCE MEASURE (ROOT MEAN SQUARED ERROR)
MULTI-LAYERED PERCEPTRON APPROACH(MLPN)[14]	2.865
SUPPORT VECTOR MACHINE(SVM)[14]	2.749
RIDGE REGRESSION[15]	>4.0
RANDOM FOREST REGRESSOR(RFR) EXTRA-TREE REGRESSOR(EFR)[15]	≈3.0
AUTOREGRESSIVE INTEGRATED MOVING AVERAGE(ARIMA)[16]	1.6594
ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEMS (ANFIS)[16]	5.0808

VI. CONCLUSION

The primary resultant of this project is that, the tourist should be able to plan his holidays/trips based on the predictions generated by the proposed system. Such a system should be capable of forecasting the suitable weather forecast so that the tourist can have an a reliable application where he can simply enter the date and duration of his tour and hence validate it against the predictions and decisions generated by the System. Compared to traditional machine learning models such as Regression and Classification, Time series models particularly, Holt Winters model can deliver higher accuracies for the prediction. Data smoothing techniques by using Hodrick-Prescott filter improves the model accuracy by providing long-term trend. Since, this idea is not implemented as such by various Big Tourist Industry giants ; it can be a new development area to be explored which can definitely benefit the tourist by the predictions generated by our system.

REFERENCES

1. Lai, Loi Lei, et al. ”Intelligent weather forecast.” Machine Learning and Cybernetics. Proceedings of 2004 International Conference on. Vol. 7. IEEE, 2004.
2. M. Shashi and Radhika, Y. ”Atmospheric temperature prediction using support vector machines.” International Journal of Computer Theory and Engineering 1.1 (2009):55.
3. A. G. Salman,Y. Heryadi and B. Kanigoro "Weather forecasting using deep learning techniques," International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2015 pp.281-285.
4. Chen, S.-M., and J.-R. Hwang."Temperature prediction using fuzzy time series." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 30.2 (2000):263-275.
5. Ghosh et al., "Weather Data Mining using Artificial Neural Network " IEEE Recent Advances in Intelligent Computational Systems, Trivandrum, 2011, pp. 192-195.

6. Wang, ZhanJie and Mujib, A B M. (2017). "The Weather Forecast Using Data Mining Research Based on Cloud Computing". Journal of Physics, pp 1-6.
7. Mr. Sunil Navadia, Mr. Pintukumar Yadav, Ms. Shakila Shaikh, Mr.Jobin Thomas, "Weather Prediction:A novel approach for measuring and analyzing weather data", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), (I-SMAC 2017), IEEE, pp 414-417.
8. A. G. Salman ,Y. Heryadi and B. Kanigoro "Weather forecasting using deep learning techniques,"International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2015, pp.281-285.
9. Howard S. Burkom, Sean Patrick Murphy , Galit Shmueli." Automated time series forecasting for biosurveillance" Statist. Med. 2007; 26:4202–4218 DOI: 10.1002/sim.2835.
10. Md. Habibur Rahman1, Md. Moyazzem Hossain, Md. Tareq Ferdous Khan,Umma Salma,, "Revenue Forecasting using Holt–Winters Exponential Smoothing" ISSN: 2278-2273-online, ISSN: 2348-7909-print
11. Prajakta S. Kalekar(04329008) , Kanwal Rekhi School of Information Technology, "Time series Forecasting using Holt-Winters Exponential Smoothing " December 6, 2004
12. Imran Maqsood,Muhammad Riaz Khan, Ajith Abraham " An ensemble of neural networks for weather forecasting" Neural Comput & Applic (2004)13:112–122DOI 10.1007/s00521-004-0413-4
13. Dr. S. Santhosh Baboo and I.Kadar Shereef, " An Efficient Weather Forecasting System using Artificial Neural Network", International Journal of Environmental Science and Development, Vol. 1, ISSN: 2010-0264.
14. Y.Radhika and M.Shashi, "Atmospheric Temperature Prediction using Support Vector Machines", International Journal of Computer Theory and Engineering, Vol. 1, No. 1, April 2009 1793-8201
15. A H M Jakaria, Md Mosharaf Hossain, Mohammad Ashiqur Rahman," Smart Weather Forecasting Using Machine Learning:A Case Study in Tennessee", ACM Mid-Southeast Conference, At Gatlinburg, TN
16. Mahmudur Rahman , Rashedur M Rahman, A.H.M. Saiful Islam , Shah Yaser Maqnoon Nadvi " Comparative Study of ANFIS and ARIMA Model for Weather Forecasting in Dhaka", International Conference on Informatics, Electronics and Vision (ICIEV) 2013.

AUTHORS PROFILE

Abhijit Kocharekar, is pursuing Final Year Bachelor of Engineering in the Computer Engineering Department at Pimpri Chinchwad College of Engineering, Pune. He has been doing project work under the guidance of Prof.Sagar Salunke, Assistant Professor, Pimpri Chinchwad College of Engineering, Pune.

Bharat Nemade, is pursuing Final Year Bachelor of Engineering in the Computer Engineering Department at Pimpri Chinchwad College of Engineering, Pune. He has been doing project work under the guidance of Prof.Sagar Salunke, Assistant Professor, Pimpri Chinchwad College of Engineering, Pune.

Chetan Patil, is pursuing Final Year Bachelor of Engineering in the Computer Engineering Department at Pimpri Chinchwad College of Engineering, Pune. He has been doing project work under the guidance of Prof.Sagar Salunke, Assistant Professor, Pimpri Chinchwad College of Engineering, Pune.

Durgesh Sapkale, is pursuing Final Year Bachelor of Engineering in the Computer Engineering Department at Pimpri Chinchwad College of Engineering, Pune. He has been doing project work under the guidance of Prof.Sagar Salunke, Assistant Professor, Pimpri Chinchwad College of Engineering, Pune.

Prof.Sagar Salunke, is an Assistant Professor, Department of Computer Engineering at the Pimpri Chinchwad College of Engineering,Pune. His research interest includes Machine Learning and Deep Learning.