International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 11 | Nov 2019 www.irjet.net p-ISSN: 2395-0072

Weather Prediction for Tourism Application using ARIMA

Abhijit Kocharekar[1], Bharat V. Nemade[2], Chetan G. Patil[3], Durgesh D. Sapkale[4], Prof. Sagar G. Salunke[5]

[1,2,3,4]*Computer Engineering Department, Pimpri* Chinchwad *College of Engineering, Pune, India*
[5]*Project Guide, Computer Engineering Department, Pimpri Chinchwad College of Engineering, Pune, India*

Abstract - *In many areas, accurate projections of future occurrences are crucial, one of which is the tourism industry. 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 could assist both the economy and tourists.*

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 advent of information and computer technology have 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, ARIMA model is used for predicting time series. ARIMA is an acronym representing the AutoRegressive Integrated Moving Average. It is a model class that captures in time series data a suite of distinct normal temporal stuctures.

***Key Words*: Tourism Industry, Weather Forecasting, Time Series Analysis, ARIMA**

1. INTRODUCTION

Climate and weather are essential considerations in the decision making for visitors and also affect the effective operation of tourism enterprises. Tourist industry is a contributing sector to the global economy. Indeed, the economies of some nations derive most of their income from tourism. The rise 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. The World Travel and Tourism Council calculated that tourism generated 16.91 lakh crore or 9.2% of India's GDP in 2018 and provided 42.673 million employment, 8.1% of total employment. The industry is forecast to expand by 2028 (9.9 % of GDP) at an annual rate of 6.9% to 32.05 lakh crore. All tourist destinations are climate sensitive and climate has a main impact on travel planning and travel experience. 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. 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. Hence proper measures can be taken by the government and holiday planners can effectively plan the tours. 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 thus have received important 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, among others, are learning weather representation using a huge weather dataset quantity. Analysis of various information mining procedures is carried out for this purpose. 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.

2. RELATED WORKS

Related works included many distinct and exciting weather forecasting methods. While much of the present prediction technology includes physics-based simulations and differential equations, many fresh methods from artificial intelligence primarily used machine learning methods, mostly neural networks, while some used probabilistic models such as Bayesian networks. Two of the three papers on weather prediction machine learning we analyzed used neural networks while one used support vector machines. Neural networks, unlike the linear regression and functional

regression models we used, seem to be the common machine learning model option for weather forecasting due to the capacity to capture the non-linear dependencies of previous weather trends and future weather conditions. This gives the benefit of not assuming that all characteristics are easy linear dependencies over our models. Of the two solutions to neural networks, one used a hybrid model that used neural networks to model physics behind weather forecasting while the other[1] used learning more directly to predict weather conditions. Similarly, the strategy using support vector machines[2] also directly implemented the classifier for weather prediction but was more restricted in scope than approaches to the neural network.

Over the last decade, countless important attempts have been documented with effective outcomes to fix weather forecasting issues using statistical modeling, including machine learning systems[3][4].Different methods have been used in weather prediction systems such as neural network-based algorithms using Back Propagation Neural Network (BPN) and Hopfield Network[5], predictive analysis in Apache Hadoop Framework utilizing Naive Bayes Algorithm [7], Artificial Neural Network and Decision Tree Algorithms[6],Recurrence Neural Network (RNN), Conditional Restricted Boltzmann Machine (CRBM) and Convolutional Network (CN) models[8].

A few studies concentrate on evaluating information from social media, such as check-ins and remarks published by visitors, to infer the density of visitation over time. For example, in (Spencer A. Wood and Lacayo 2013)[9], the writers use the locations of photos in Flickr, a popular website for picture hosting, to estimate the number of visitors to some recreational sites around the globe. They study the link between the empirical estimates of mean annual user-days visitors and photo-derived ones. This is best described by a polynomial function with R2 = 0.386 and that categorizing the recreational parts into more specific profiles could improve correlations. However, they do not address predictions. In (Fisichelli et al. 2015)[10], the writers use a third-order polynomial temperature model to evaluate the climate and visitation information for U.S. national parks and claim that it explains 69 percent of the variability in historical visitation patterns. We were able to attain much greater levels of precision by exploiting a richer function set, including social media information. We demonstrate that even a simple linear regression model can achieve an precision of over 80 percent by exploiting social media as well as climate information, while a more robust algorithm, Support Vector Regression, generated a superior 94 percent precision outcome. The writers use Wikipedia usage trends in more latest work (Khadivi and Ramakrishnan 2016)[11] to predict Hawaii's travel demand. However, they only report the precision of their forecast outcomes by using an auto-regressive exogenous model where the external variable is a trend time series of Wikipedia usage. RMSE is an accuracy metric to compare forecast mistakes of distinct models for specific information and not between datasets as it is scale dependent (Hyndman and Koehler 2006)[12]. Although this work contains interesting statements and outcomes, there is no comparison of predictive models with other baselines or similar evaluation of the outcomes.

3. PROPOSED METHOD

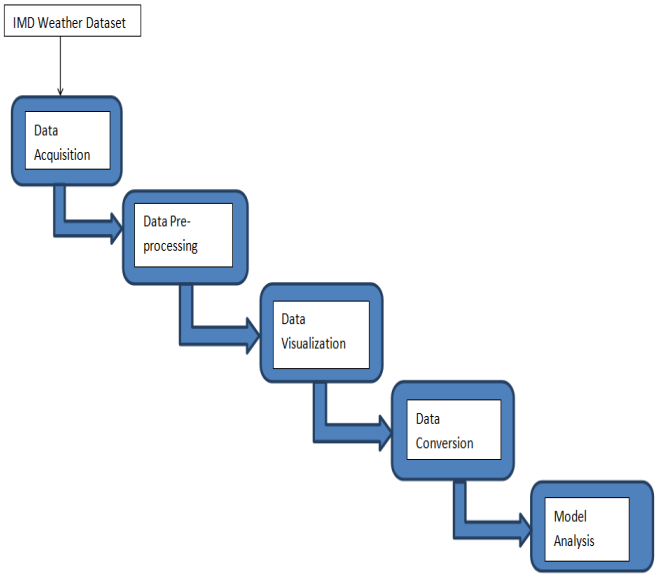

Fig -1: Workflow of the proposed system

Data Acquisition:

The weather data set supplied by the Indian Meteorological Department has been taken into account, consisting of various features such as air temperature, atmospheric pressure, relative humidity, average wind speed, maximum air temperature, complete cloud cover, horizontal visibility, minimum dew point air temperature. The dataset contains a total of eight records for one day. Each of these is held after every 3 hours of day. The dataset includes a total of 23,000 records.

Data Preprocessing:

As part of pre-processing, the missing values must be removed from the dataset to ensure that the results generated are more accurate. Since most data is available in numerical format, it is possible to calculate the mean, median or mode of the features and replace it with missing values. This is an approximation that can add variance to the dataset.Replacing the three approximations above is a statistical approach to the handling of missing values.After there are eight readings for each day, it must be normalised in such a way that for the particular day only one reading is obtained.

Data visualization:

Data visualization provides additional interpretation possibilities.Visual presentation for individuals is often more readable than any other presentation of dataWeather data in itself is massive. Due to the fact that the data is in enormous amounts, data visualization helps to understand the dataset. It helps to see patterns, trends and correlation in the attribute values that may go unnoticed. It also helps to identify climate data patterns and their variations throughout the year

Data Conversion

Weather data may be in non-stationary time series with statistical properties that change over time. It is necessary to ensure that these statistical properties are constant before starting any predictive modelling. A stationary time series is one whose statistical properties are relatively constant over time, such as mean, variance and auto correlation.

Conversion to stationary time series using:

a. Differencing: In differencing, the difference between the consecutive terms is calculated. Differencing is typically performed to get rid of the varying mean. Making a Stationary time series through differencing is an important part of the process of fitting an ARIMA model. Mathematically, differencing can be written as:

yt'=yt – y(t-1) ,where yt is the value at a time t.

b. Transformation: To stabilize a series' non-constant variance, transformations are used. Common methods of transformation include power transformation, square root transformation and log transformation.

Model analysis

Weather forecasting is the system's main objective. ARIMA model is the most suitable statistical method for this purpose. ARIMA model: ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average. An ARIMA model is a class of statistical models for time series data analysis and forecasting. The acronym is descriptive,

a. AR: Autoregression.A model that uses an observation's dependent relationship with a number of lagged observations

b. I: Integrated.The use of differencing of raw observationsin order to make the time series stationary.

c. MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations. Each of these components is specified explicitly as a parameter in the model.

4. ALGORITHM

A time series is a set of observations on the values taken at different times by a variable.

Such information must be gathered daily (e.g. weather), weekly (e.g. monthly supply) 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.

Stationary Series:

There are three basic criterion for a series to be classified as stationary series:

1. The mean of the series should not be a function of time rather should be a constant. The image below has the left hand graph satisfying the condition whereas the graph in red has a time dependent mean.

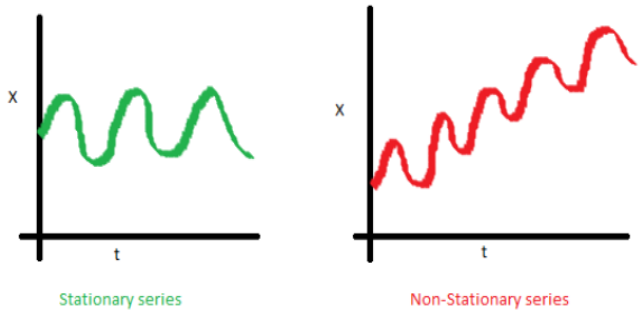

Fig -2: Time independent mean vs Time independent mean

2. The variance of the series should not a be a function of time. This property is known as homoscedasticity.

Following graph depicts what is and what is not a stationary series.

The varying spread of distribution in the right hand graph, indicates it is non-stationary.

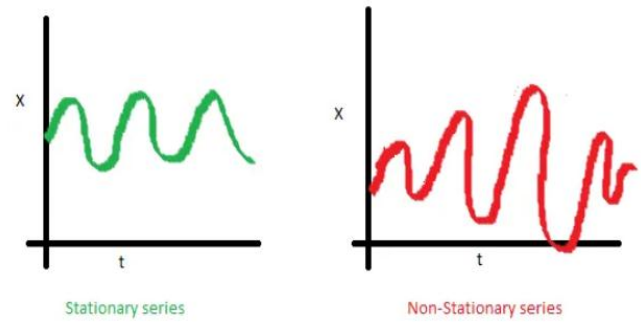

Fig -3: Spread of Distribution in Stationary and Non-Stationary series

3. The covariance of the i th term and the (i + m) th term should not be a function of time.

In the following graph, you will notice the spread becomes closer as the time increases.

Hence, the covariance is not constant with time for the 'red series'.

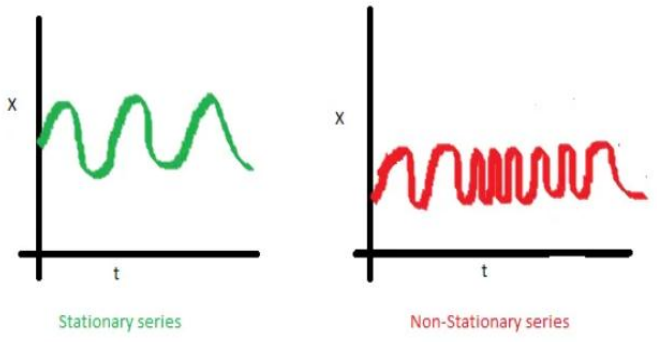

Fig -4: Covariance in Stationary and Non- Stationary series

Test for Stationarity of Series:

Plotting Rolling Statistics:

In this method ,the moving average or moving variance and see if it varies with time.

By moving average or variance ,it means that at any instant 't', we'll take the average or variance of the last year, i.e. last 12 months.

But again Rolling Statistics is more of a visual technique.

Dickey-Fuller Test:

This is one of the statistical tests for checking stationarity.

Here the null hypothesis is that the Time Series is non-stationary.

The test results comprise of a Test Statistic and some Critical Values for difference confidence levels.

If the 'Test Statistic' is less than the 'Critical Value', we can reject the null hypothesis and say that the series is stationary.

ARIMA MODEL

ARIMA stands for Auto-Regressive Integrated Moving Averages.

The ARIMA forecasting for a stationary time series is nothing but a linear (like a linear regression) equation.

Number of AR (Auto-Regressive) terms (p): AR terms are just lags of dependent variable. For instance if p is 5, the predictors for x(t) will be x(t-1)….x(t-5).

Number of MA (Moving Average) terms (q): MA terms are lagged forecast errors in prediction equation. For instance if q is 5, the predictors for x(t) will be e(t-1)….e(t-5) where e(i) is the difference between the moving average at ith instant and actual value.

Number of Differences (d): These are the number of non-seasonal differences, i.e. in most case we took the first order difference. For d=2,it means that the differences has been calculated 2 times. By visualizing these newly created series, we can identify ideal transformation for our use-case.

5. CONCLUSION

The primary goal of this project is that, the tourist should be able to plan his holidays/trips based on the predictions generated by the proposed system .This System should be able to predict 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 particularly, ARIMA model can deliver higher accuracies for the prediction. 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, 2004. Proceedings of 2004 International Conference on. Vol. 7. IEEE, 2004.

[2] Radhika, Y., and M. Shashi. "Atmospheric temperature prediction using support vector machines."International Journal of Computer Theory and Engineering1.1 (2009):55.

[3] A. G. Salman, B. Kanigoro and Y. Heryadi, "Weather forecasting using deep learning techniques," 2015 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," 2011 IEEE Recent Advances in

Intelligent Computational Systems, Trivandrum, 2011, pp. 192-195.

[6] Wang, ZhanJie & 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. Jobin Thomas, Mr. Pintukumar Yadav, Ms. Shakila Shaikh, "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, B. Kanigoro and Y. Heryadi, "Weather forecasting using deep learning techniques," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2015, pp.281-285.

[9] Spencer A. Wood, Anne D. Guerry, J. M. S., and Lacayo, M. 2013. Using social media to quantify nature-based tourism and recreation. Scientific Report 3.

[10] Fisichelli, N. A.; Schuurman, G. W.; Monahan, W. B.; and Ziesler, P. S. 2015. Protected area tourism in a changing climate: Will visitation at us national parks warm up or overheat? PLoS ONE 10(6).

[11] Khadivi, P., and Ramakrishnan, N. 2016. Wikipedia in the tourism industry: Forecasting demand and modeling usage behavior. In ICWSM, 4016–4021.

[12] Hyndman, R. J., and Koehler, A. B. 2006. Another look at measures o forecast accuracy. Int. J. Forecasting 22(4):679–688.

[13] Stern, H. (2008), The accuracy of weather forecasts for Melbourne, Australia. Met. Apps, 15: 65?71. doi:10.1002/met.67

[14] Abramson, Bruce, et al. "Hailfinder: A Bayesian system for forecasting severe weather."International Journal of Forecasting12.1 (1996): 57-71.

[15] Paras, S. Mathur, A. Kumar and M. Chandra (2007), "A Feature Based Neural Network Model for Weather Forecasting", World Academy of Science, Engineering and Technology.

[16] Afan Galih Salman ; Bayu Kanigoro; Yaya Heryadi Weather Forecasting using Deep Learning Techniques

[17] Munmun Biswas, Tanni Dhoom, Sayantanu Barua. "Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data".

[18] Mehrnoosh Torabi, Sattar Hashemi, "A Data Mining Paradigm to Forecast Weather", The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012),IEEE, pp 579-584