International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 6, Nov-Dec-2020, ISSN (Online): 2395-566X

The Application of Weather Forecast using Time Series Analysis

Vidhya Vasantrao Sambare, Prof. Arvind Jain
Dept. Of Computer Science & Engg.
Sagar Institute of Science Technology and Engineering (SISTEC-E), Bhopal

***Abstract*- Weather forecasting has been an important application in meteorology and one of the most scientifically and technologically challenging problem around the world. In my study, we have analyzed the use of data mining techniques in forecasting weather. This paper proposes a modern method to develop a service oriented architecture for the weather information systems which forecast weather using these data mining techniques. This can be carried out by using Artificial Neural Network and Decision tree Algorithms and meteorological data collected in Specific time. Algorithm has presented the best results to generate classification rules for the mean weather variables. The results showed that these data mining techniques can be enough for weather forecasting.**

***Keywords:*- Artificial neural network, prediction, Time series analysis, Mean Square error function, Support vector machines, Data Mining**

I.INTRODUCTION

Delay is one of the most remembered performance indicator of any transportation system. Notably, commercial aviation players understand delay as the period of time by which a right is late or postponed. Thus, delay may be represented by the differencebetween scheduled and real times of departure or arrival of a fight. Country regulator authorities have a multitude of indicators related to tolerance thresholds for fight delays. Indeed, fight delay is an important subject in the context of air transportationsystems. In 2013, 36% of fights delayed by more than five minutes in Europe, 31.1%of fights delayed by more than 15 minutes in the United States and in Brazil, 16.3% of fights were canceled or suffered delays greater than 30 minutes [1].

This indicateshow relevant this indicator is and how it affects no matter the scale of airline meshes. Flight delays have negative impacts, particularly economic, for passengers, airlinesand airports. Given the uncertainty of their occurrence, passengers usually plan totravel many hours before their appointments, increasing their trip costs, to ensuretheir arrival on time. On the other hand, airlines suffer penalties, fines and additionaloperation costs, such as crew and aircrafts retentions in airports [2]. Furthermore,from the sustainability point of view, delays may also cause environmental damage byincreasing fuel consumption and gas emissions [3,4].

CBR and fuzzy set theory each have their own unique welldemonstratedstrengths. So when both methods arecombined in one system, the system stands the chance ofinheriting the strengths of both methods. A recent increasein the number of such hybrid systems attests to theeffectiveness of combining CBR and fuzzy set theory (Pal et al. 2011).CBR is recommended to developers who are challengedto reduce the knowledge acquisition task, avoid repeatingmistakes made in the past, reason in domains that have notbeen fully understood or modeled, learn over time, reasonwith incomplete or imprecise data and concepts, provide ameans of explanation, and reflect human reasoning (Mainet al. 2012), and these are some of the challenges faced bydevelopers of weather forecasting systems.

1. Weather Prediction:

Weather prediction presents special challenges for CBR. Weather is continuous, data-intensive, multi dimensional, dynamic and chaotic. These properties make weather prediction a formidable proving ground for any CBR prediction system that depends on searching for similar sequences.

Fundamentally, there are only two methods to predictweather: the empirical approach and the dynamical approach **(Lorenz 2013).** The empirical approach is basedupon the occurrence of analogs (similar weather situations)and is often referred to by meteorologists as "analogforecasting." The empirical approach is useful forpredicting local-scale weather if recorded cases areplentiful (e.g., cloud ceiling height and visibility in a fewsquare kilometres around an airport).

The dynamicalapproach is based upon equations and forward simulationsof the atmosphere, and is often referred to as "computermodeling." Because of computer model grid coarseness,the dynamical approach, used by itself, is only useful formodeling large-scale weather phenomena (e.g., generalwind direction over a few thousand square kilometers). Inpractice, most weather

prediction systems use acombination of empirical and dynamical techniques.

2. Objective:

The Objectives of our project are pretty simple;
We wish to create a product that automates the work or as we explained the secondary functionalities of an Air Traffic Controller.

- Invent an automated system for an ATC.
- Automatically makes airport weather predictions.
- Helps assigning run way strips to aircrafts automatically as well, based on climate forecasts, and air traffic, for aircraft operations.
- Also, predict weather for oncoming days as well.
- Help decide the flight planand future flight operations and schedule.
- Making, Air travel safer and the airlines more profitable.

II. LITERATURE REVIEW

Beniwal, S. & Arora [1]; analyzed the departure and arrival data for ten majorairports in US and characterize the delay data in a statistical method.It revealed that the ground movement inefficiency contributed mostto surface delays and weather was the main cause of delay for airtraffic control. The departure and arrival delays were modeled withprobabilistic demand forecasting methods. Statistical approacheswere also applied in [5] with long-term trend and short-term pattern.The key component of this model is the estimation of thedelay propagation effect, where delay propagation was known asdelay built up from previous fights and effects on delays.

Rebollo and Balakrishnan [2]; proposed a new class of models using machinelearning techniques with air traffic network characteristicsto predict air traffic delays. The proposed models considered bothtemporal and spatial delay states as explanatory variables, and usedRandom Forest algorithms to predict departure delays 2-24 h in thefuture. Deep learning has also become a promising method for solvingdata analytics problems such as traffic prediction.

Kit Yan Chan (2016) [3]; selected the recurrent neural networks for the day-to-daydelay status prediction, which first trained the model for a reliabledelay status of a single day then the delay states of individual fightsof that day. Deep learning algorithms capture the sequential andtemporal relationships in the day-to-day delay data.

Narasimha Prasad, Prudhvi Kumar and Naidu MM [4]; presents a comparison of two sub sampling nonparametric methods for designing algorithms to forecast time series from the cumulative monthly rainfall. Both approaches are based on artificial feed-forward neural networks (ANNs).

Mark Ian Animas, Yung-Cheol Byun, Ma. Beth Concepcion and Bobby D. Gerardo [5]; proposed fuzzy inference system for monthly rainfall prediction in the northeast region of Thailand. The predicted show of the proposed model was compare to be conservative Box-Jenkins and artificial neural networks model. Accordingly, the experimental results show the modular FIS is superior another method to predict accurately. The predicted mechanism can be interpreted through fuzzy rules. Auto-regression, Seasonal auto regressive integrated moving average and ANN modular FIS provide better results. The experimental results give together accurate results and human-understandable prediction mechanism.

III. WEATHER DELAY PREDICTION TOOL

As stated earlier in the section on military planning, at present there is not tool to predict future NAS performance based on predicted weather. For efficient operation of the NAS,there is a need for the weather forecasting services and TFM products to estimate thereduction in capacity due to adverse weather. Weather forecast products are uncertain andthe uncertainty increases with lead-time. Useful applications of weather forecasts requireseither refinement, consultation, and application of the weather forecast to estimate air trafficcapacity or decision support tools that take forecasts and make predictions based on pastforecasts and those forecasts connections to NAS capacity [41]. This paper describes amethodology used to create one such decision support tool known as the Weather DelayPrediction Tool. With this tool, the user enters the TAF for a given day and airport and thetool provides AAR predictions which can be derived to estimate delay and Ground DelayProgram (GDP) time and duration.

1. Data Sources:

To test this theory, the following 8 airports were selected to collect data:

- Newark Liberty International(EWR)
- Chicago O'Hare International (ORD)
- Atlanta Hartsfield International Airport (ATL)
- Philadelphia International Airport (PHL)
- Reagan National Airport (DCA)
- Dulles International Airport (IAD)
- LaGuardia Airport (LGA)
- John F. Kennedy International Airport (JFK)

2. Data Mining In Weather Forecasting And Method:

Data Mining deals with what kind ofpatterns can bemined. Based on the kind of data to be mined, there are two kinds of functions involved in Data Mining such as Descriptive model and Predictive model. The Descriptivemodel identifies patterns or relationships in data and dealswith general properties of data in the database. The predictive

model is the process of finding a model whichdescribes the data classes or concepts, the purpose being tobe able to use this model to predict the class of objects whose class label is unknown[47].

Data mining techniques are mainly separated into two groups, viz. classification and clustering techniques. Classification techniques are designed for classifying unknown samples using information provided by a set ofclassified samples. This set is usually referred to as atraining set, because, it is generally used, to train theclassification technique i.e. how to perform itsclassification. If a training set is not available, there is no previous knowledge about the data to classify. [48]

Clustering technique is used to group the element that is particular area occupied by rainfall regions and the rainfall is predicted in a particular region. The different classification techniques for discovering knowledge are Rule Based Classifiers, Artificial Neural Network (ANN), Bayesian Networks (BN), Decision Tree (DT), Nearest Neighbour (NN), Rough Sets, Fuzzy Logic, Support Vector Machine (SVM), Genetic Algorithms.[49] The different clustering methods are Hierarchical Methods (HM), Partitioning Methods (PM), Density-based Methods (DBM), Grid-based Methods , Model-based Clustering Methods(MBCM) and Soft-computing Methods [fuzzy, neural network based], Squared Error—Based Clustering (Vector Quantization), network data and Clustering graph [50]

2.1 Decision trees:

It is commonly used in data mining to examine the data and to induce the tree and its rules that will be use to make predictions. A number of different algorithms may be used for building decision trees including CHAID (Chi-squared Automatic Interaction Detection), CART (Classification And Regression Trees), Quest, and C5.0 [20]. A decision tree is a tree in which each division node represents a choice between a number of alternatives, and every leaf node represents a decision. Depending on the algorithm, each node may have two or extra branches. For example, CART generate trees with only two branches at each node. Such a tree is called a binary tree. When more than two branches are allowed this is called a multiway tree [21].

2.2 Support vector machines (SVMs):

It is supervised learning methods used for classification [22,23, 23]. In their basic form, SVMs are used for classifying sets of samples into two displace classes, which are separated by a hyperplane defined in a suitable space. Note that, as consequence, a single SVM can only discriminate between two unlike classifications. However, as we will discuss later, there are strategies that allow one to extend SVMs for classification problems with more than two classes [23, 24].

The hyperplane used for separating the two classes can be defined on the basis of the information contained in a training set. showing a simplified functional flow diagram of the forecast algorithm. There are three main threads: the echo tops forecast shown in orange, the precipitation shown in tan, and the precipitation phase shown in blue. The image data are in 1 km resolution, with five-minute update rates and zero-to-two-hour forecast loops. NWP stands for Numerical Weather Prediction.

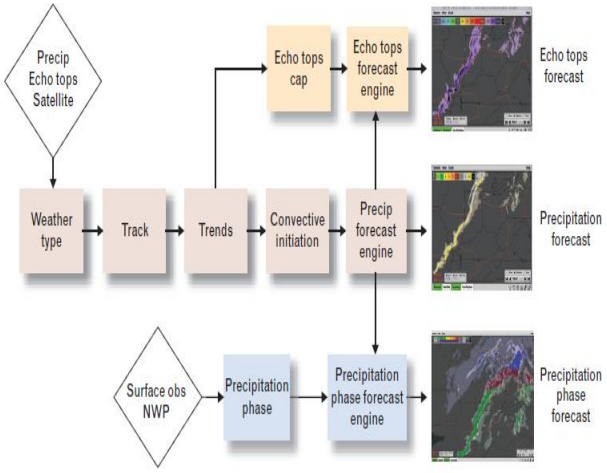

Fig 1. Convective weather forecast functional flow.

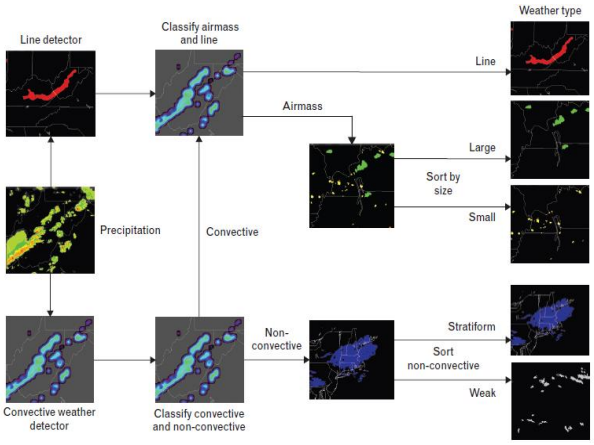

Fig 2. A simplified flow diagram for the weather type algorithm.

IV. TIME SERIES ANALYSIS

A Time series (TS) is a sequence of observations in ordered time. Mostly these observations arecollected at equally spaced discrete time intervals. The basic assumption in any time seriesanalysis modeling is that some aspects of the past pattern will continue to remain in the future.Also under this setup often the time series process is assumed to be based on past values of themain variable but not on explanatory variables which may affect the system. So the system act asa black box and we may

only be able to know 'what' will happen rather than 'why' it happens.

Here it is deliberately assumed that information about the past is available in the form ofnumerical data. Ideally at least 50 observations are necessary for performing time seriesanalysis/modeling as propounded by Box and Jenkins who were pioneers in TS modeling [9].Time series analysis can be used more easily for weather forecasting purposes because historical sequences of observations are readily available from published sources. All these successive observations are statistically dependent and Time series modeling is concerned with the techniques for analysis of such dependencies [9].

V. RESULTS AND DISCUSSIONS

The temporal weather forecasting technique based on BP/GA technique has been implemented by taking different population sizes.

Table 1 Selection of appropriate NN architecture

Population Size	Hidden Neurons	Iterations	MAPE
30	1	110	1.0
60	2	140	0.86
90	3	220	0.42
120	4	282	1.47
150	5	425	1.85

For each value of population, the program has been executed and the error has been calculated. Table 1 shows the variations in population size, number of neurons in hidden layer and the corresponding mean absolute percentage error values.

From the table 4, it is clear that the MAPE value is the lowest corresponding to population size 90 and number of hidden neurons as 3. So the present setup will use this population size for further research. The error vs. iteration graph corresponding to population size 90 is shown in figure. The error values corresponding to mean air temperature are shown in fig. 3 along with the Series 1, Series 2 and Series 3 of the inputs are shown. Error values are shown after 200 epochs, 400 epochs and 600 epochs. Clearly Series 3 shows the minimum error values in all the cases and it shows the lowest value after 400 epochs.Below is shown the actual prediction of mean air temperature using the proposed method along with the desired output as recorded with the help of instruments in fig 3.It is clearly shown in the fig 4 that the time series based temperature prediction model using integrated BP/GA technique is suitable to predict the temperature.

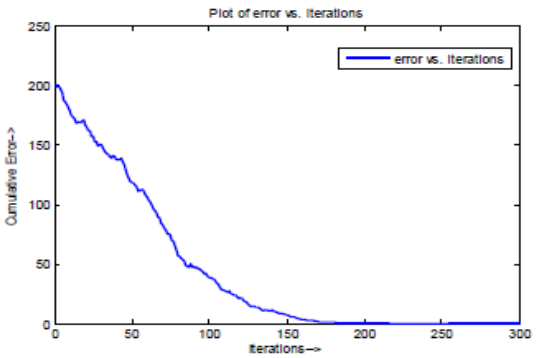

Fig 3. The cumulative error values corresponding to iterations for population size 90.

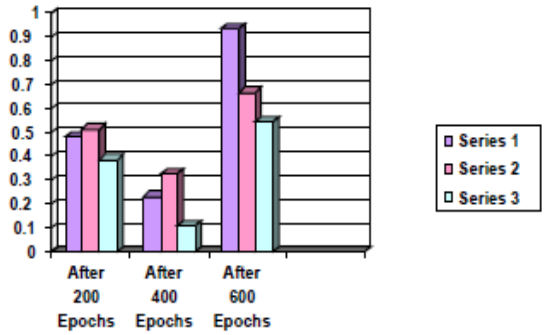

Fig 4. Temporal prediction for mean air temperature.

Secondly, dependence of weather parameters on the time series data. For the same weather parameter, the error values come out to be different when the network is trained with different data series. Thirdly the effect of under training the network through 200 epochs and over training the network through 600 epochs is easily visible as the error values are lowest after 400 epochs.

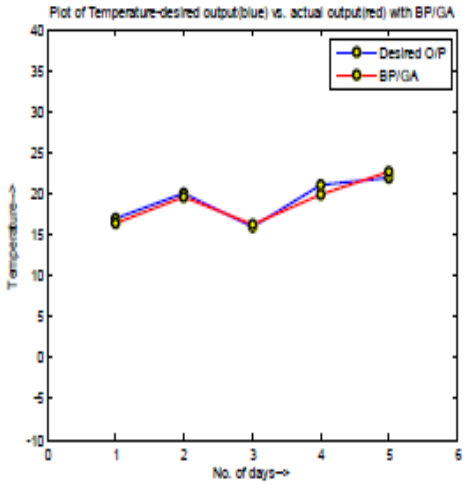

Fig 5. The five-day mean air temperature prediction using the proposed model.

Fig 6. Aviation Weather Assistance System

Fig 7. Aviation Weather Assistance System Data.

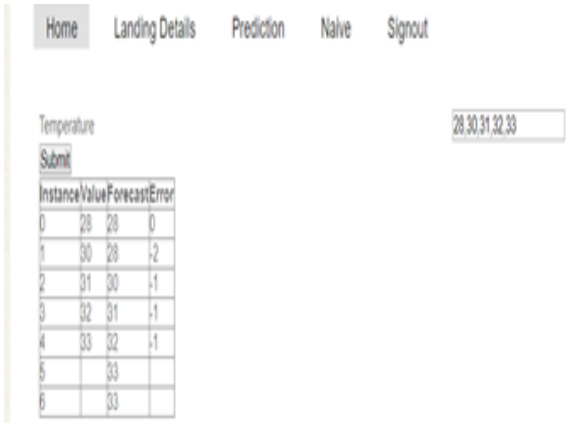

Fig 8. Dataset Forecasting.

VI. CONCLUSIONS

Flight delays are an important subject in the literature due to their economical and environmental impacts. They may increase ticket costs to customers and operationalcosts to airlines. Apart from outcomes directly related to passengers, delay predictionis crucial during the decision-making process for every player in the air transportation system.

In this context, researchers created flight delay models for delay prediction over thelast years and this work contributes with an analyzes of these models from a Data Science perspective. We developed a taxonomy scheme and classified models in respectof detailed components.

Applications built using the concept of fuzzy logic can easily determine the competence of a lecturer. So the research produced keeps perfect. The process of determining the thesis adviser is similar to the conventional process. It occurs since the algorithm comes from the natural way of human reasoning. Fuzzification can adjust the lecturer's knowledge level, so the given topic is not to deviate from the background of the problem. However, using this application, the academic authorities will get more consistent without having the burden to analyze which specification is the most appropriate.

Mainly, the taxonomy includes domain and Data Science branches. The formerbranch categorizes the problem (flight delay prediction) and the scope. The last branchgroups methods and data handling. It was observed that the flight delay predictionis classified in three main categories, such as delay propagation, delay innovation, andcancellation analysis. Besides, the scope determines one of the three specific extents:airline, airport, en-route airspace or an ensemble of them.

Additionally, considering Data Science branch, we aimed at the datum, by categorizingdata sources, dimensions that can be used in the models, and data management techniques to preprocess data and improve prediction models efficiency. We also studiedand divided the main methods into five categories: statistical analysis, probabilistic models, network representation, operations research, and machine learning. Those categorieshave been grouped as their use on specific forecast models for flight delays.

Besides the taxonomic scheme, we also presented a timeline with all articles to spottrends and relationships involving the main elements in the taxonomy. In the light of thedomain-problem classification, this timeline showed a dominance of delay propagationand delay innovation over cancelation analysis. Researchers used to focus at statisticalanalysis and operational research approaches in the past. However, as the data volumegrows, we noticed

the usage of machine learning and data management is increasingsignificantly. This clearly characterizes a Data Science trend.

VII. FUTURE SCOPE

To further explore the dataset, we decided to conduct an extraexperiment on a subclass of the dataset. We focused on the SanFrancisco Airport (SFO) and built models to predict flight delay ofthat airport. Besides the features mentioned in the previous section, we triedto look at the average arrival delay of the previous day for each flight. Intuitively, flight delay is greatly in effluence by the 'ripple'effect and we expect the delay situation of yesterday would havegreat impact on the flight delay.As shown in table 3, all models show significantly improvedaccuracy and F1 scores for delayed flights compared with the totaldataset. The weighted LR model reaches the best F1 score of 0.44and with an accuracy of 0.672. Though the extra experiment is conductedon the subset of flights at SFO airport, the improvement inperformance can also show the importance of the feature "AverageArrival Delay of yesterday".

REFERENCES

[1] Beniwal, S. &Arora(2012) ―Classification and feature selection techniques in data mining‖, International Journal of Engineering Research & Technology (IJERT), 1(6).
[2] Rebollo and Balakrishnan (2013) ―Survey of clustering algorithms‖, Neural Networks, IEEE Transactions on, 16(3), 645-678.
[3] Kit Yan Chan,'' Neural-Network-Based Models for Short-Term Traffic Flow Forecasting Using a Hybrid Exponential Smoothing and Leven berg–Marquardt Algorithm‖ , IEEE trans on intelligent transportation system, VOL. 13, NO. 2, pp.644-646,JUNE 2012.
[4] Narasimha Prasad, Prudhvi Kumar and Naidu MM, ―An Approach to Prediction of Precipitation Using Gini Index in SLIQ Decision Tree‖, 4th International Conference on Intelligent Systems, Modeling and Simulation, 2013.
[5] Mark Ian Animas, Yung-CheolByun, Ma.Beth Concepcion and Bobby D. Gerardo, Decision Support System for Agricultural Management Using Prediction Algorithm‖, 2013.
[6] Jehangir Ashraf Awan and Onaiza Maqbool, Application of Artificial Neural Networks for Monsoon Rainfall Prediction‖, Sixth International Conference on Emerging Technologies, 2010.
[7] Dong Li-li, XuShu-qin, Meng Fan-Xiang and Yang Xu, ―Application of Time –series Model in the Chahayang Farm of Rainfall Prediction‖, 2010.
[8] S. Kannan and S. Ghosh, A Novel Nonlinear Combination Model Based on Support Vector Machine for Rainfall Prediction‖, Fourth International Joint Conference on Computational Sciences and Optimization, 2011.
[9] S. Kannan, SubimalGhosh, Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output‖, Springer-Verlag, July- 2010.
[10] Sarah N. Kohail, Alaa M. El-Halees, Implementation of Data Mining Techniques for Meteorological Data Analysis‖, IJICT Journal Volume 1 No. 3, 2011.
[11] A Mucherino et al. Neural Networks: A Comprehensive Foundation‖, Second Edition, Prentice Hall International, 2012.
[12] Due R. A., ―A tatistical Approach to Neural Networks for Pattern Recognition‖, 8th edition. New York: John Wiley and Sons publication, 2013.
[13] Mucherino, A. Urtubia, Feature Selection for Datasets of Wine Fermentations, I3M Conference Proceedings, 10th International Conference on Modeling and Applied Simulation (MAS11), Rome, Italy, September 2014. Riul A Jr, de Sousa HC, Malmegrim RR, dos Santos DS Jr, Carvalho ACPLF, Fonseca FJ, Oliveira Jr ON, Mattoso LHC Wine classification by taste sensors made from ultra-thin films and using Neural Networks. Sens Actuators B98:77– 82, 2014.
[14] Brudzewski K, Osowski S, Markiewicz:T Classification of milk by means of an electronic nose and SVM neural network. Sens Actuators B98: 291–298, 2014.
[15] Druszler, A.: Meteorological effects of the land cover types changes during the twentieth century in Hungary. PhD Thesis, University of West Hungary, Sopron, p. 137 (2015)
[16] Pasztor, L., Szabo, J., Bakacsi, Zs: Digital processing and upgrading of legacy data collectedduring the 1:25.000 scale Kreybig soil survey. Acta Geodaet. Geophys. Hung. 45, 127–136(2015)
[17] Bretherton, C.S., Park, S.: A new moist turbulence parameterization in the communityatmosphere model. J. Clim. 22, 3422–3448 (2009)
[18] Hong, S.-Y., Dudhia, J., Chen, S.-H.: A revised approach to ice microphysical processes forthe bulk parameterization of clouds and precipitation. Mon. Weather Rev. 132, 103–120(2014)
[19] Chen, F.J., Dudhia, J.: Coupling and advanced land surface-hydrology model with the PennState-NCAR MM5 modeling system. Part I.model implementation and sensitivity. Mon.Weather Rev. 129, 569–585 (2011)
[20] Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A.:Radiative transfer for in homogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res. 102, 16663–16682 (2015)