METEOROLOGICAL FORBODE USING ADABOOST ALGORITHM

Pavithra R[1], Ramya B[2] „Mrs Malathi A[3]

[1]*Student, BE Computer Science Engineering, Anand Institute of Higher Technology, Tamil Nadu, India*

[2]*Student, BE Computer Science Engineering, Anand Institute of Higher Technology, Tamil Nadu, India*

[3]*Assistant Professor, ME Computer Science Engineering, Anand Institute of Higher Technology, Tamil Nadu, India*

ABSTRACT

Owing to the increasing evidence of the climatic changes worldwide is becoming the reason to understand a lot more about the weather like what's going to happen tomorrow or the next day. To forecast the weather, we need to analyse a large set of data's, therefore use of big data in climatic predictions will provide accurate predicting of seasonal weathering. Big Data is a field to analyse systematically to extract information's from data sets. In enduring system, a concept of data mining called Linear Regression is used to predict the climate, it doesn't have a solid software that recovers from a component failure and also leads to immense output fields and consuming vast compute and storage resource. The planned system explored various application domains that could benefit from weather forecasting using big data and ADABOOST Algorithm. The algorithm is a supervised learning that uses both classification and regression challenges with high accuracy and less computation power. WEKA tool is implemented which uses a GUI software and sanctioned with GNU General Public License. It is kind of Create and Load Database with IA-32, X86-64, Java SE platforms. Data volumes and variety are growing at very fast rate and this is becoming a great challenge in weather forecasting, as now difficulty is to mix these data to provide correct forecast. Some examples of these domains include Forecasting solar power for Utility operations, large scale crop production forecasts for global food security, in precision agriculture for future farming and space weather forecasting. In order to know how these applications could impact normal operations this project defines various climate predictions and challenges.

In future work, day to day predictions should be implemented with the help of API's and accuracy should be added for more efficient purposes.

Key Words – *Big data, ADABOOST Algorithm, WEKA TOOL, Linear Regression.*

1. INTRODUCTION

Big data is a term that describes the large volume of data both structured and unstructured, it refers to the huge data sets obtained from various sources such as social media, sensor data, public data. Big data analytics is the process of examining large data set to uncover hidden patterns, unknown correlations.

The objective of this model development is to minimize the cost and to provide more accuracy in weather prediction system. The goal is to provide "seamless prediction" on a varying temporal scale.

Weather forecasting system is a complicated task because large amount of elements are taken into consideration. Until now the weather forecasting result is not accurate it is based on calculation and prediction. The traditional weather prediction system is based on numerical weather prediction model (NWP).

Data mining knowledge are also being used in the weather forecasting problem probabilistic graphical models (Bayesian networks) in meteorology as data mining technique. Data mining involves the use of complicated data analysis tools to discover previously unknown, interesting patterns and relationships in large data set.

This Southeast Asia region is characterized by complex terrain and land-water contrasts, tropical forests, and many islands. The climate of SE Asia and especially the maritime SE Asia is mainly tropical–hot and humid all

year round with plenty of rainfall. Majority of the Southeast Asia region is influenced by monsoon and much of the region is affected by extreme weather events, particularly tropical cyclones, droughts and floods.

2. LITERATURE SURVEY

To blend growing amounts of renewable energy into utility grids requires accurate estimates of the power from those resources for both days ahead planning and real-time operations [1]. This requires predicting the wind and solar resource on those timescales. Accurate prediction of these meteorological variables is a big data problem that requires a multitude of disparate data, multiple models that are each applicable to a specific time frame, and application of computational intelligence techniques to successfully blend all of the model and observational information in real-time and deliver it to the decision-makers at utilities and grid operators. Considering that the capacity of renewable energy continues to grow an additional challenge includes selecting and archiving data for continuous retraining of machine learning algorithms.

Wind power forecasting can enhance the value of wind energy by improving the reliability of integrating. The National Center for Atmospheric Research (NCAR) has collaborated with Xcel Energy [2].to develop a multifaceted wind power prediction system. Both the day-ahead forecast that is used in trading and the short-term forecast are critical to economic decision making. The system utilizes publicly available model data and observations as well as wind forecasts produced from an NCAR-developed deterministic mesoscale wind forecast model with real-time four-dimensional data assimilation and a 30-member model ensemble system, which is calibrated using an Analogue Ensemble Kaltman Filter & Quantile Regression. The model forecast data are combined using NCAR's Dynamic Integrated Forecast System (DI Cast).

Cloud computing has been growing rapidly since Amazon brought this idea to the public [3]. Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. From the survey conducted by, Right Scale 96% of the respondents from different organizations have adopted cloud in various ways. Indeed, cloud computing brings huge advantage to the firm, but moving to the cloud computing is not an easy task. The contribution of this work is to present an overview of cloud computing. Research findings of this paper can be utilized to acquire an understanding of cloud computing and its applications in different sectors

Geo Sensor is a novel, open-source system that enriches change detection over satellite images with event detection over news items and social media content [4]. Geo Sensor combines these two orthogonal operations through state-of-the-art Semantic Web technologies. At its core lies the open-source, semantics-enabled Big Data infrastructure developed by the EU H2020 Big-Data Europe project. This allows Geo Sensor to offer an on-line functionality, despite facing three major challenges of Big Data: Volume (a single satellite image typically occupies a few GBs), Variety (its data sources include two different types of satellite images and various types of user-generated content) and Veracity, as the accuracy of the end result is crucial for the usefulness of our system.

Cloud services is used by many organizations and it has captured a major segment of the competitive market today [5]. The green, or eco-friendly, aspect of the cloud is one of the most multifaceted advantages of cloud computing. The environmental advantages of cloud services include: reducing a firm's carbon footprint, data center efficiency, dematerialization, saving green, educed electricity use and so on. Even with its unprecedented growth, the question of security is also of paramount concern among the users of cloud services. There is a huge demand for new protocols and tools in order to enhance and assess the security strength of its service. Once the authentication of the user is verified successfully dual encryption is performed on the cloud stored files using El-Gamal cryptosystem and Hyper Elliptical Curve Cryptography (HECC).

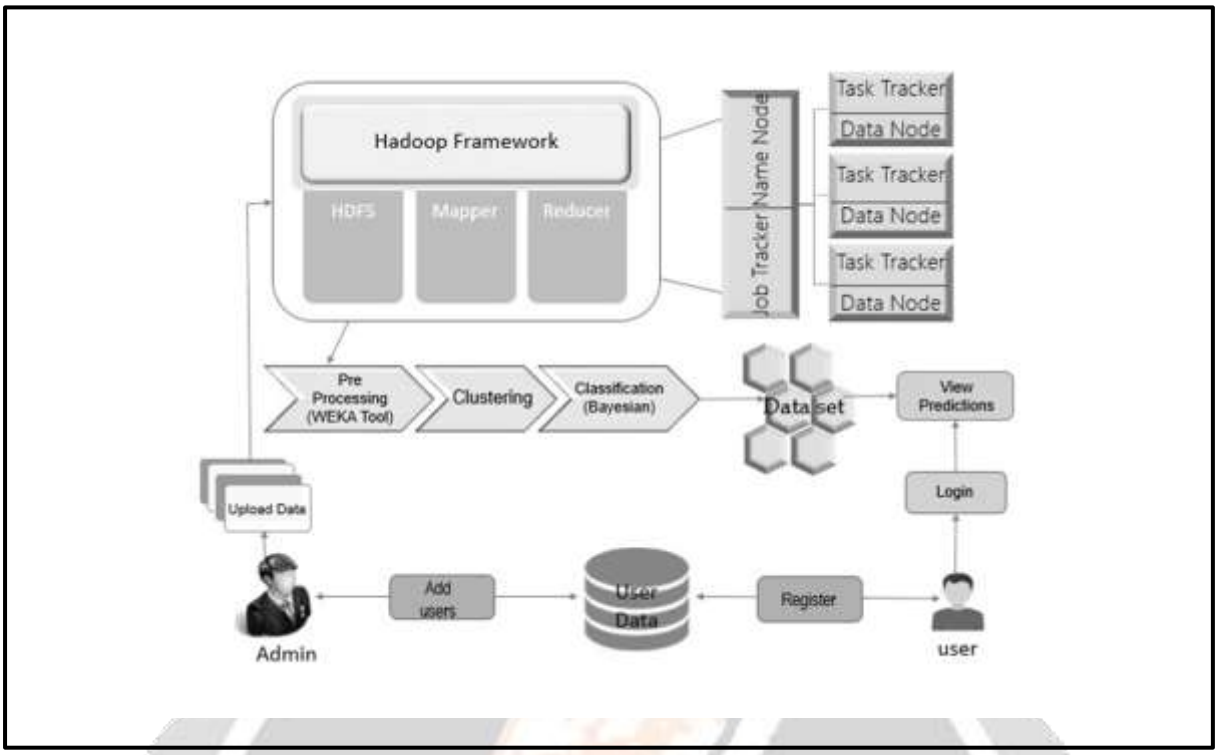

Figure 1: Core Architecture

3. SYSTEM DESIGN

The system architecture is described in Figure 1. System Architecture is a generic discipline to handle objects (existing or to be created) called "systems", in a way that supports reasoning about the structural properties of these objects. System architecture is a response to the conceptual and practical difficulties of the description and the design of complex systems. This is the most important step of the feasibility study this study helps us predict the operational ability of the system that is being developed. This study also helps us analyze the approach towards which the system must be developed by which development effort is reduced. Proposed system is beneficial only if they can be turned into information systems that will meet the organization requirements. This system supports in producing good results and reduces manual work. Only by spending time to evaluate the feasibility, do we reduce the chances from extreme embarrassments at larger stager of the project. Effort spend on a feasibility analysis that results in the cancellation of a proposed project is not a wasted effort.

4. MODULES

The modules which we are developing using different strategies to give the accurate prediction for the user.

4.1 User Testament
4.2 Data Hook-Up
4.3 Data Bewitching
4.4 Data Clumping
4.5 Report Proclaim

4.1 USER TESTAMENT

Every last client login the page at that point makes the exchange and utilize this application. Validness is confirmation that a message, exchange, or other trade of data is from the source it cases to be from. Validness includes verification of character. We can check validness through confirmation. Enroll and login choice in landing page. Every single client needs to enlist as the new client for login. Client need to Fill the all prerequisite for security reason just, so fill the all subtle elements unique points of interest. Every one of the

subtle elements spared in various ways. Make new table for every client and spare points of interest in like manner table.

4.2 DATA HOOK-UP

The readied informational index will store the Hadoop document framework. HDFS occurrences are partitioned into two segments: the name node, which keeps up metadata to track the arrangement of physical information over the Hadoop case and data nodes, which really store the information. The information stacking to the HDFS utilizing HDFS URL way. The transferred information will be kept up by the name hub and information hubs. The informational index traits are kept up in the name hub like record name, measure, get to consent, and so forth. The crude information kept up by the information hubs. The information hub controlled by the name node. The transferred information can't change any qualities in light of the fact that HDFS have compose once perused many time properties.

4.3 DATA BEWITCHING

Information preprocessing is an information mining method that includes changing crude information into a reasonable arrangement. True information is frequently inadequate, conflicting, as well as ailing in specific practices or drifts, and is probably going to contain numerous blunders. Information preprocessing is a demonstrated strategy for settling such issues. The transferred information recover from the HDFS. The recovered information going to the Map Reduce calculation and information will be composed into organized configuration. In this procedure have expelling the unusable qualities from the informational indexes. The information diminishment process is lessened portrayal of the information in an information distribution center.

4.4 DATA CLUMPING

To gather that information into those bunches whose climate information class has been as of now characterized. In this way it develops a procedure to anticipate the promoting of the up and coming days. This one procedure gathering the information and ought to be made out of focuses isolated by little separations, in respect to the separations between groups. The information will be gathering in light of the value, open, high, low, shut and time. In this bunching will apply the map reduce with ADABOOST approach. It gives more productive in high volume information bunching process.

4.5 REPORT PROCLAIM

The cluster value will be different ranges. Those values are gathered and compared to each other's. Finally, we will get the low and high result based on the calculation. The predicted values will be given the graphical representation graph

5. MODULE EVALUATION

5.1 USER TESTAMENT

The user registers their details here like name, password, email and DOB. Once they have registered the same username and password must be remembered and can sign in their account to view the prediction The figure 2 shows the registration details.

Figure 2: User Testament

5.2 DATA HOOK-UP

The admin can login with their necessary ID and password and uploads the data's based on the weather. The datasets are collected and segregated so that it will be convenient to the users to view the prediction. The figure 3 shows the data uploading.

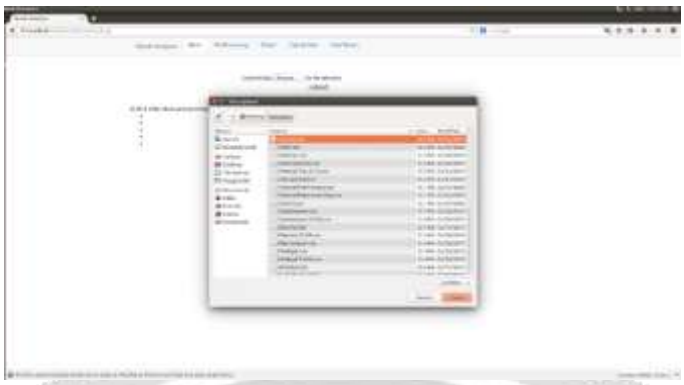

Figure 3: Add location details

5.3 DATA BEWITCHING

The data is being bewitched and removes the unwanted data by machine learning and also WEKA tool is implemented in order to create the database tables and upload the data's inside the table. It is an preprocessing format .The figure 4 shows the bewitched details.

Figure 4: Data bewitching.

5.4 DATA CLUMPING

All the data's have been gathered and provides the required details to the users using the Clustering only after this process the actual data's have been segregated and gives the clear view to the user. The figure 5 shows the grouped data.

Figure 5: Data clumping

5.5 REPORT PROCLAIM

The weather has been observed and predicted accurately in two forms:

(i) Month prediction
(ii) Week prediction.

The figure 6 shows the predicted weather to the users.

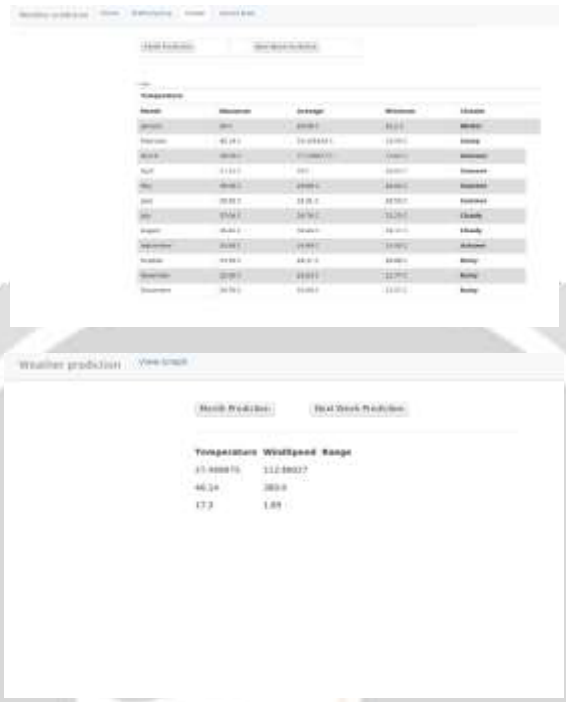

Figure 6: Report Proclaim

5.6 PERFORMANCE

By the use of this graph it is easy for the users to predict the weather on the yearly basis too.The figure 7 shows the performance of the weather that is predicted.

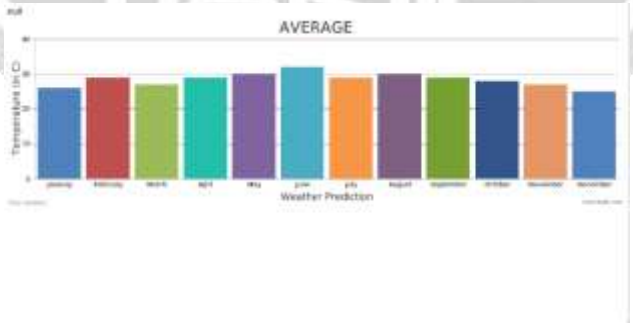

Figure 7: Performance

6. ALGORITHM USED

We are proposing an algorithm for the best prediction for the users. This algorithm can give accurate and best solutions for the users.

- ADABOOST ALGORITHM

6.1 ADABOOST ALGORITHM

ADABOOST refers to a particular method of training a boosted classifier. A boost classifier is a classifier in the form. where each is a weak learner that takes an object as input and returns a value indicating the class of the object. For example, in the two-class problem, the sign of the weak learner output identifies the predicted object class and the absolute value gives the confidence in that classification. Similarly, the classifier is positive if the sample is in a positive class and negative otherwise.
Each weak learner produces an output hypothesis, for each sample in the training set. At each iteration, a weak learner is selected and assigned a coefficient such that the sum training error of the resulting -stage boost classifier is minimized. Here is the boosted classifier that has been built up to the previous stage of training, is some error function and is the weak learner that is being considered for addition to the final classifier.

Given: $(x_1, y_1), \ldots, (x_m, y_m); x_i \in \mathcal{X}, y_i \in \{-1, +1\}$
Initialise weights $D_1(i) = 1/m$
For $t = 1, ..., T$:

- Find $h_t = \arg\min_{h_j \in \mathcal{H}} \epsilon_j = \sum_{i=1}^{m} D_t(i)[\![y_i \neq h_j(x_i)]\!]$
- If $\epsilon_t \geq 1/2$ then stop
- Set $\alpha_t = \frac{1}{2}\log(\frac{1-\epsilon_t}{\epsilon_t})$
- Update

$$D_{t+1}(i) = \frac{D_t(i)exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

Output the final classifier:

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

7. CONCLUSION

Data volume and variety are growing at very fast rate and this is becoming a great challenge in weather forecasting; as now difficulty is to mix these data to provide correct forecasts. Since we have tremendous amount of complex data, so transport, storage and management of data is becoming a problem and is also increasing the overheads. Complex too large required to manage this much amount of data, which can cause millions of dollars of overhead. If in case weather forecasting goes wrong it could lead to shut down of many businesses causing severe loss to the economy and society.

So if we could understand the nature of these applications and challenges, we can identify the optimal solutions to these challenges and can have more efficient and reliable applications which can save lives, improve quality of life and business, reduce risks and enhance profitability.

REFERENCES

[1] S. E. Haupt and B. Kosovic, "Big Data and Machine Learning for Applied Weather Forecasts: Forecasting Solar Power for Utility Operations," *2015 IEEE Symposium Series on Computational Intelligence*, Cape Town, 2015, pp. 496-501.
[2] W. P. Mahoney *et al.*, "A Wind Power Forecasting System to Optimize Grid Integration," in *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 670-682, Oct. 2012.
[3] Ngo Yang Chong, Y. Fan, L. Tong and W. Hu, " Cloud Computing Challenges In A General Perspective," *2019 IEEE International Conference on Web Services (ICWS)*, Milan, Italy, 2019, pp. 242-246.

[4] Nikiforos Pittaras1, 2, George Papadakis2, George Stamoulis2, Giorgos Argyriou2, Efi Karra Taniskidou2, Emmanouil Thanos2, George Giannakopoulos1 and Manolis Koubarakis2. *"*Geosensor: Semantifying Change And Event Detection Over Big Data*"*. 960-962. 10.1145/3019612.3019876.

[5] T. Devi, R. Ganesan, Ekoloji, 2019 *"Environmental Benefits Of Enhanced Hecc- Elgamal Cryptosystem For Security In Cloud Data Storage Using Soft Computing Techniques"*. Proceedings of the 19th International Conference on World Wide Web, WWW '10. 1029-1038. 10.1145/1772690.1772795.

[6] M. Ashraf, J. C. Loftis, K.G. Hubbard, "Application of geostatistics to evaluate partial weather station networks", Agricultural and Forest Meteorology, vol. 84, pp. 255-271, 1997.

[7] G. Urban, "An attemp to assess the results of air temperature measurements from automatic weather stations in comparison to glass thermometer measurements in the context of weather types", Bulletin of Geography Series, no.9,pp.67-79,2015.

[8] S. AL-Yahyai, Y. Charabi, A. Gastli, S. Al-Alawi, "Assessment of wind energy potential locations in Oman using data from existing weather stations", Renewable and Sustainable Energy Reviews, vol. 14, pp. 1428-1436,2010.

[9]S. Abbatea, M. Avvenutia, L. Carturan, D. Cesarinia, "Deploying a communicating Automatic Weather Station on an Alpine glacier", The 3rd international Workshop on Sensor Networks for Intelligence Gathering and Monitoring (SNIGM) Procedia Computer Science,vol.19,pp.1190-1195,2013.

[10] R. Kusmierek-Tomaszewska, J. Zarski, S. Dudek, "Meteorological automated weather station data application for plant water requirements estimation", Computers and Electronics in Agriculture, vol. 88, pp. 44-51,2012.

[11] Lyons, W. A.; Tremback, C. J.; Pielke, R. A.; "Applications of the Regional Atmospheric Modeling System (RAMS) to provide input to photochemical grid models for the Lake Michigan Ozone Study (LMOS)". J. Appl.Meteor.,34,1995,pp1762-1786.

[12] Hock Beng Lim, Keck Voon Ling, Wenqiang Wang, Yuxia Yao, Mudasser Iqbal, Boyang Li, Xiaonan Yin, Tarun Sharma. "The National Weather Sensor Grid".ACMSenSys'07.Nov.2007.

[13]. H. Saini, A. Thakur, S. Ahuja, N. Sabharwal, N. Kumar, "Arduino based automatic wireless weather station with remote graphical application and alerts", 2016 3rd InternationalConference on Signal Processing and Integrated,Networks.(SPIN),pp.605-609,Feb2016.

[14] A. Lage, J. C. Correa, "Weather station with cellular communication network", 2015 XVI Workshop on Information Processing and Control (RPIC),pp.1-5,Oct2015.

[15] A. V. Bosisio, M. P. Cadeddu, "Rain detection from ground-based radiometric measurements: Validation against rain sensor observations", 2015 IEEE InternationalGeoscience and Remote Sensing Symposium (IGARSS), pp. 2323-2326, July 2015.