International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 08 | August -2020 ISSN: 2582-3930

Weather Data Analysis Using Data Science

Ms. Bhavana P[1], Ms. Jawaria Afshan[2], Ms. Nisha T P[3], Ms. Jyothi S[4], Ms. Pavithra D[5]

[1]*Assistant Professor,Dept.Of Computer science & Engineering, Cambridge Institute of Technology*

[2,3,4,5]*B.E,Dept.Of Computer science & Engineering, Cambridge Institute of Technology*

Abstract -The entire world is afflicted with the dynamical clement and their facet, to cut back this facet effects we use several techniques and algorithms via which we will predict the weather on the ready reference together with the respective context of given information from previous years sample temperature, dew, humidity air pressure and wind direction. When doing the analysis of existing data we inculcated the proposed scheme or techniques which have a tendency to state that, machine learning paradigm permits us to research the given set of knowledge and draw out the helpful information from the given dataset, consequently so as to grasp the unsteady patterns of climatic conditions, an advance model is also persuaded. In this paper, we have a tendency to explore progressive statistical Naive Bayes and Random Forest techniques of machine learning that teams' constant kind information sets along and to prefigure the weatherprediction.

Keywords:Machine Learning, dataset, climatic conditions, Naive Bayes, Random Forest.

1. INTRODUCTION

In today's world we have several barriers in handling more servers which results in failures. Many companies are forced to remove their most essential data because the cost of storing is too high. We all wish that all of our trade transactions occurs smoothly and continuously without defects and the data gets stored in a structured and well organized way. Hadoop with HDFS and MapReduce succeeds in dealing these problems and it is designed to persist to work in the face of system failures andcompute huge datasets in parallel. Hadoopincludes MapReduce, a programming model for processing and generating big data sets on a cluster. A HadoopMap Reduce job mostly has two user-defined functions: map function and reduce function. The input to HadoopMap Reduce job should have key-value pairs(k, v) and map function is known for each of these pairs. There are two types of nodes that controls the execution process: a jobtracker and task trackers. The jobtracker synchronizes and schedules all task to run on tasktrackers. It in turn sends progress reports to the jobtracker. If a task fails, the jobtracker can reschedule that on a different tasktracker. The importance of Hadoop map Reduce is that the users just define map and reduce functions. Hadoopmap Reduce makes use of the HadoopDistributed File System to perform I/O performance. An enhanced automated prediction technique, that is based on Hadoopframework for productive and scalable weather data analysis and forecasting system helps in finding solutions to problem such as climate change induced food insecurity, predicting and mitigating the impact of extreme weather events on global finance. We take two algorithms, Naïve Bayes clustering and Random Forest algorithm for predicting and comparing the accuracy among them.

1.1 Objectives

The main objective of this system is to we have a tendency to explore progressive statistical Naive Bayes and Random Forest techniques of machine learning that teams' constant kind information sets along and to prefigure the weather prediction.

2. PROPOSED SYSTEM

We are implementing the machine learning techniques like naive bayes and random forest which is done inside the hadoop framework. The algorithms gives the good result with moreaccuracy.

3. SYSTEM DESIGN

Various design features are followed to develop this system. The design specification explains the features of the system, the components or elements of the system and their appearance to end-users.

Input Design:

The input Design is that the process of converting the user-oriented inputs into the computer-based format. The goal of designing input file is to form the automation as easy and free from errors as possible. Providing an honest input design for the appliance easy data input and selection features are adopted. The input design requirements like user friendliness, consistent format and interactive dialogue for giving the proper message and help for the user at right time also are considered for the event of the project. Input design may be a a part of overall system design which requires very careful attention. Often the gathering of input data is that the costliest a part of the system, which must be route through number of modules .It is the purpose where the user able to send the info to the destination machine together with known IP address; if the IP address is unknown then it's going to susceptible to error.

Output Design:

A standard output is one, that meets the requirements of the top user and presents the knowledge plainly. In any system results of process are communicated to the users and other systems through outputs, it's most vital and direct source information to

the user. Efficient and intelligent output improves the systems relationship with source and destination machine. Outputs from computers are required primarily to urge same packet that the user has send rather than corrupted packet and spoofed packet. They also want to provide to permanent copy of those results for later consultation.

The MVC Design Method:

Swing actually makes use of a simplified form of the MVC design called the model-delegate. This design combines the view and therefore the controller object into one element that pulls the component to the screen and handles GUI events referred to as the UI delegate. Communication between the model and therefore the UI delegate becomes a two-way street. Each Swing component has a model and a UI delegate. The model is liable for maintaining information about the component's state. The UI delegate is responsible for maintaining information about the way to draw the component on the screen. The UI delegate (in conjunction with AWT) reacts to varied events that propagate through the component.

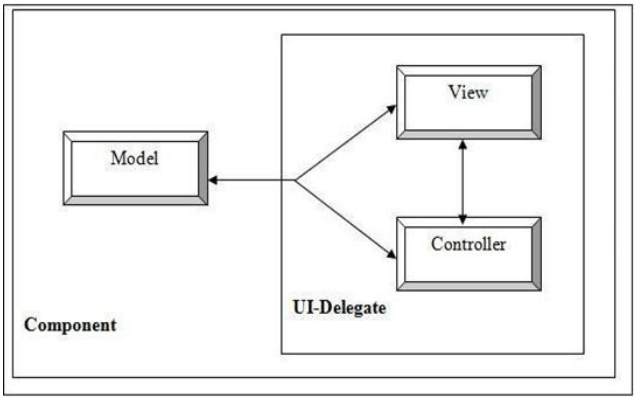

Fig -1: Combination of View & Controller into a UI delegate object

The design method that has been followed to style the architecture of the system is MVC design pattern. Swing uses the model-view-controller (MVC) architecture because the fundamental design behind each of its components. Basically, MVC breaks GUI component into three elements. Each of those elements plays an important role in how the component behaves. The MVC design pattern separates a software component into three well defined pieces: a model, a view, and acontroller.

Model

The model is that the piece that represents the state and low-level behavior of the component. It manages the state and conducts all transformations there on state. The model has no specific understanding of any of its controllers or its views. It encompasses the state data for each component. There are different models for various kinds of components. For example, the model of a scrollbar component might contain data about its current position of its adjustable "thumb", its minimum and maximum values, and thus the thumb's width. A menu on the other hand, may simply contain a listing of the menu items the user can select from. The system itself maintains relation between model and views and notifies the views when the model changesstate.

View

The view refers to how you see the component in the screen. It is the part that manages the visual display of the state represented by the model. Almost all window frames will have a title bar occupying the top of the window. However the title bar may have a close box on the left side or on the right side. These are the examples of different types of views for the same window object. A model can consist of more than one type of view.

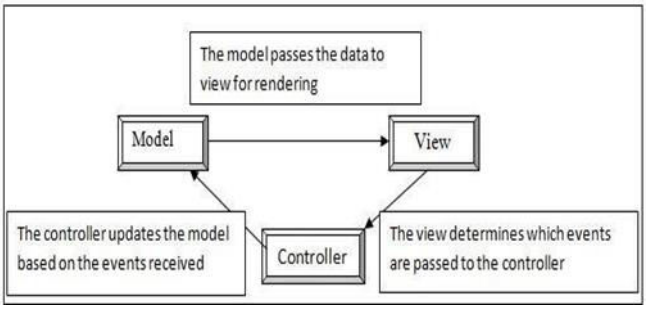

Fig -2:Communication through the MVC architecture

Controller

The controller is the part that manages user interactions along with the model. It produces the mechanism by which causes achange that are reflected to the state of the model. It is the detail description of the user interface that explains how the component interacts withevents.

The view cannot render the scrollbar rightly without obtaining information from the model first. In this case the scrollbar will not know where to draw its "thumb" unless it can obtain its current position and width relative to the minimum and maximum. Likewise the view determines if the component is the recipient of user events, such as mouse clicks. The view is used to pass these events on to the controller, which later decides how to handle the events at their best. Based on the controller's decision the values in the model may need to be altered. If the user drags the scrollbar thumb, the controller will react by incrementingthe thumb's position in the model. At that point the whole cycle canrepeat.

The JFC user interface component can be broken down into a model, view, and controller. The view and controller are combined to one piece, a common adaptation of the basic MVC pattern. They form the user interface for the component.

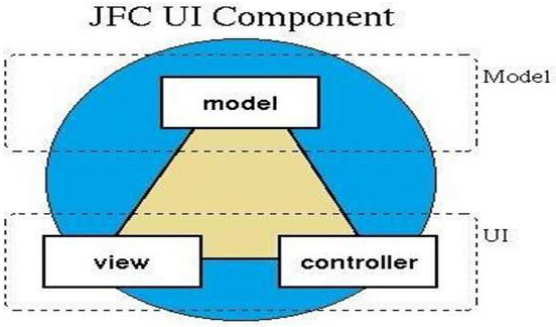

Fig -3: JFC user interface component

4. IMPLEMENTATION

The implementation of this system provides the system components or building blocks and provides a plan from which products can be procured, and systems developed, that will work together to implement the overall system.

The system consists of three importantcomponents

1.1.1 InformationExtraction
1.1.2 FeatureSelection
1.1.3 PredictiveModelling

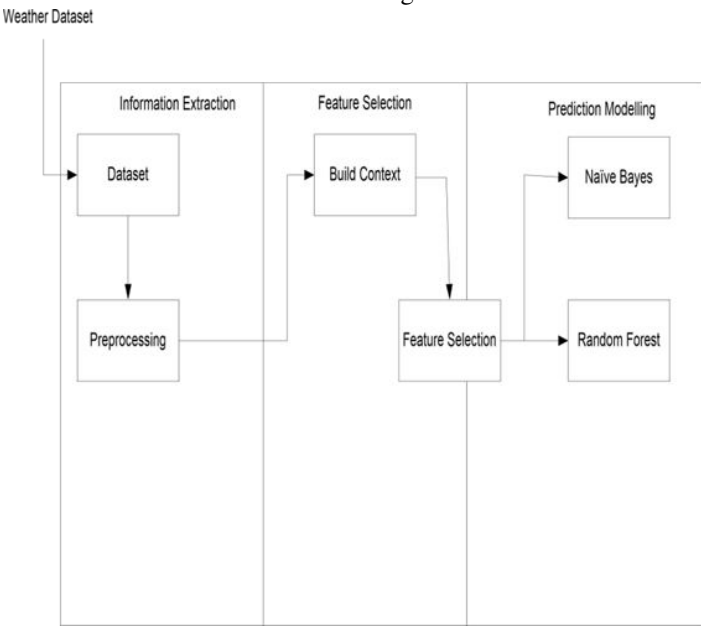

Fig -4: system architecture

Information Extraction:

In this Weather Dataset records are collected from varioussources. Apply preprocessing where data cleaning and data transformation has done.

Feature Selection:

The most important features for Weather and modelling is extracted among all the attributes collected for dataset . Predictive Modelling: The predictive model which is been constructed are Random Forest and Naive bayesmodel . The Big Data Hadoop Platform is been runned on predictive modelling part.

Naive Bayesian Classification: It is based on the Bayesian theorem .Naive Bayes is mainly used when the highest dimensionality of the inputs are present. Naive Bayes models uses the method of maximum likelihood for parameter estimation .Weka Classifier: It is a collection of machine learning algorithms which performs data mining tasks. The produced algorithms can either be called through our own Java code or can be applied directly to a dataset .

Weka which is a collection of machine learning has tools for classification, data pre-processing, clustering, regression, visualization and association rules. Weka is best-suited for developing new machine learning schemes.

5. RESULT

The world is plagued by the active element and their facet, to lessen this facet effects up to some extent there are several techniques and algorithms through which we will forecast the weather based on the ready reference along with respective context of given details from previous years for instance winddirection,temperature, air pressure and humidity. When doing the analysis of existing data from past years we inculcated the proposed scheme or techniques which have a tendency to conclude that, machine learning paradigm permits us to research the given set of knowledge and extract the helpful details from the given dataset , thus to grasp the unstable patterns of climatic conditions, a predictive model is also persuaded. During this paper , we have a tendency to explore progressive statistical Random Forest techniques and Naive Bayes of machine learning that groups constant kind detailed sets along and to foreshadow the weather prediction.

The following snapshots define the results or outputs that we will get after step by step execution of all the modules of the system.

Interpretation:

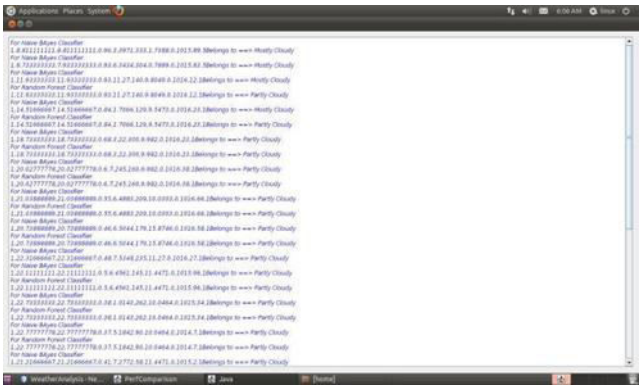

Fig -5: output of the weather dataset

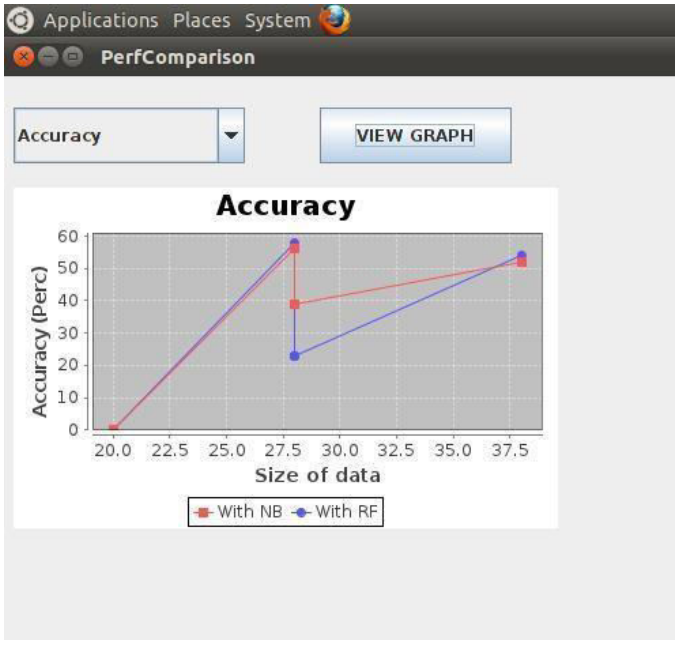

Fig -6: Accuracy comparison of the weather data

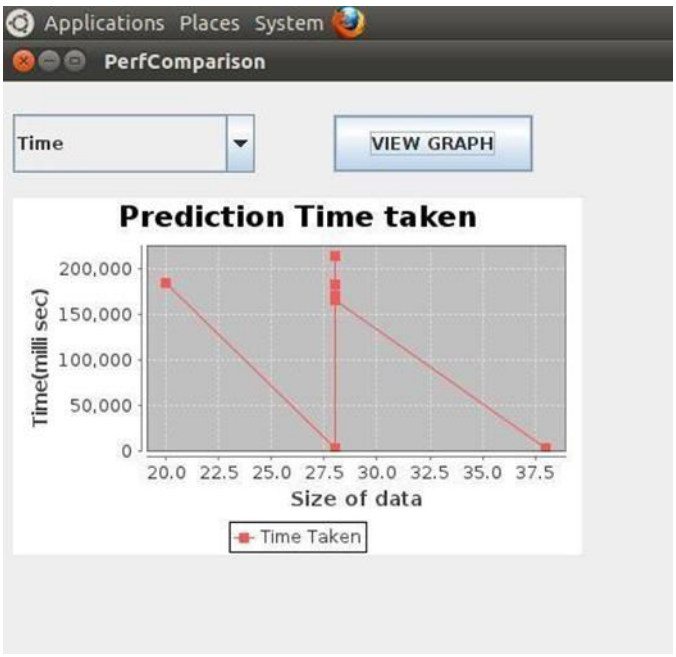

Fig -7: Prediction comparison of the weather data

This gives a brief interpretation of the expected and obtained result when each and every module is executed in their propersequence.

6. CONCLUSION

Both machine learning algorithms using Hadooplead realistic perfection and outwit by proficient climate or weather determining directions. Results are intrinsically high and accurate as illustrated as it is steady for exceptions and forecasting, so one approach to enhance the straight setback show is by accumulation of more information using Naive Bayes and Random Forest. Showing that the model was efficient and effective that its expectations can be enhanced by promote accumulation of information under the proposedscheme.

7. FUTURE SCOPE

Prediction using traditional Weather Reports models usually includes a machine learning algorithm and especially a supervised learning algorithm with the use of training data with labels to train the model. In the test set, Weather data can be classified into groups. These models are worth in clinical situations and are widelystudied.For future scope the same can be incorporated over Apache spark for simultaneous prediction of weather whereas the same can be used to differentiate with the results obtained from sensors.

Acknowledgement

We would like to put on record our profound feeling of appreciation to Shri. D. K. Mohan , Chairman, Cambridge Group of Institution, Bengaluru, India for giving great infrastructure academic environment at Cambridge Institute of Technology without which this work would not have been possible. We are incredibly appreciative to Dr. Suresh L., Principal, CITech, Bengaluru, for giving us the scholastic feeling and never-ending inspiration to complete this work and shaping our careers. We offer our earnest thanks to Dr. Shashikumar D. R., HOD, Department of Computer Science and Engineering, CITech, Bengaluru, for his stimulating guidance, constant consolation and inspiration over the span of present work. We additionally wish to stretch out our gratitude to Prof. Bhavana P., Assistant Professor, Department of Computer Science and Engineering, CITech, Bengaluru, for his direction and impressive technical suggestions to complete our paper publication.

REFERENCES

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

[2] Cofno, Antonio S., et al. ”Bayesian networks for probabilistic weather prediction.”15th Eureopean Conference on Artificial Intelligence (ECAI).2002.

[3] Krasnopolsky, Vladimir M., and Michael S. FoxRabinovitz. ”Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction.”Neural Networks19.2 (2006):122-134.

[4] Lai, LoiLei, et al. ”Intelligent weather forecast.”Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on. Vol. 7. IEEE,2004.

[5] Ng, Andrew. ”CS229 Lecture Notes Supervised Learning” 2016.

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