

An Efficient Approach for Weather Forecasting Using Ensemble Classification Technique

¹Nikhil Jadhav, ²Vishwajit Bokan, ³Vaibhav Joshi, ⁴Bhushan Bhavsar, ⁵Amrut A. Patil

^{1 2 3 4}BE 4th Year, ⁵Assistant Professor, ^{1 2 3 4 5}Department of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India

Abstract: Weather forecasting is a crucial application in meteorology and up till now it has been one of the most scientifically and technologically challenging problems around the world in the last century. To make the forecast of weather up to the mark; is one of the major tasks faced in today's world. Since ancient times, weather forecasting has been one of the most interesting and fascinating domain. Data mining techniques can be used for weather forecasting and climatic change scenarios. Recently in the area of machine learning the concept of combining classifiers is proposed as a new direction for the improvement of the performance of individual classifiers. We are investigating the use of data mining techniques in the prediction of maximum temperature, rainfall, evaporation and wind speed. The final prediction of the system will be selected upon majority voting of these predictions. This technique is known as ensemble classification.

IndexTerms - Classifiers, Datasets, Ensemble Learning, Machine Learning, Supervised Learning, Data Stream Classification

I. INTRODUCTION

Due to unusual weather conditions weather predations is a major challenge. Hence to predict weather accurately we need to build an efficient system that would do so. We can overcome this issue using ensemble learning. In our system, we are going to provide input as datasets to classifiers. Output generated by each of these classifiers will be an individual prediction. The final prediction of the system will be selected upon majority voting of these predictions. Typical weather forecasting system uses an algorithm for prediction. But the accuracy may differ from algorithm to algorithm. The term data mining refers to the techniques that are used to extract the required information from the given set of data that might be useful for statistical purpose or making predictions by learning patterns in data and correlation between different parameters. Data mining has now been adopted by many domains such as sports, banking, meteorological department, etc., and because of this, scientists, mathematicians and researchers have come up with a wide range of algorithms for finding solution. Ensemble learning [1] is the technique of combining set of learners together and predictive power of the model. Ensemble methods have been called the most dominant development in Data Mining [8] and Machine Learning [7] in the past decade.

II. RELATED WORKS

Weather forecasting is an important aspect in today's day-to-day life. Weather forecasting has been understood by machine learning techniques. Ensemble classification technique [1] is one of the most appropriate options to understand the predictions made by individual classifiers. A Survey on Ensemble Learning for Data Stream Classification in this paper authors have proposed that, taxonomy for data stream ensemble learning as derived from reviewing over 60 algorithms. Cluster Based Ensemble Classification for Intrusion Detection in this authors proposed that ensemble classifier [2] [6] outperforms other classifiers with 99.8 accuracy using AD Tree [5] and KNN.

An ensemble approach for multi-level classification of item click Sequences [3], authors tried to predict the probability of purchase for the given item. A study of Ensemble based evolutionary classifiers for detecting unsolicited Emails [4] in authors tried to give accurate prediction in weather forecasting domain. This paper is an application of data mining techniques in weather prediction and climatic changes. In this paper we have studied that by using data mining techniques weather forecasting can be possible. Ensemble algorithms are especially useful for data mining as they can be integrated with drift detection and incorporate dynamic updates, such as removal or addition of base classifiers. In this way our system proposes that, to make use of this system to predict accurate weather conditions in order to use in the fields of transport, events & disaster and to design a system for efficient weather forecasting.

III. PROPOSED SYSTEM

In our system, we are going to provide input as datasets to classifiers. Output generated by each of these classifiers will be an individual prediction. The final prediction of the system will be selected upon majority voting of these predictions. Figure below gives the overview of our system.

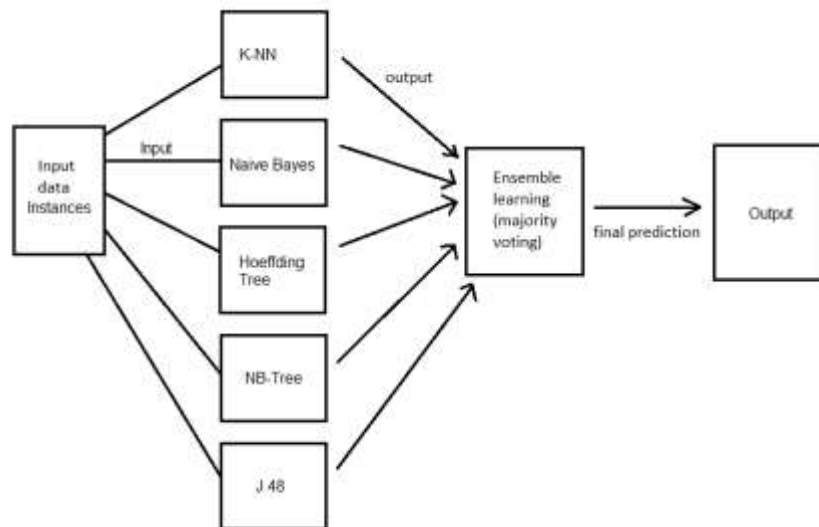


FIGURE 3.1 : SYSTEM ARCHITECTURE

K-nearest - neighbour is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). Naïve Bayes classifiers are a family of simple probability classifier based on applying Bayes theorem with strong (naïve) independence assumptions between the features. Hoeffding tree is an incremental, anytime decision tree induction algorithm that is capable of learning from massive data streams, assuming that the distribution generating examples does not change over time. NB-Tree Class is for generating a decision tree with Naive Bayes classifiers at the leaves. J48 is an ensemble learning method for classification, an open source Java implementation of the C4.5 decision tree algorithm.

IV. CONCLUSION

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to appropriately forecast the target class for each case in the data. A typical classification system uses a base classifier to classify data instances using a relevant data set as a reference. The system that we have proposed uses ensemble learning technique. Same input is provided to multiple classifiers and a majority voting is taken to finalize the output. The proposed system will increase the accuracy in prediction by 5-10 percent. If accuracy of base classifier is 85-90 percent, our system will try to increase it up to 90-95 percent.

REFERENCES

- [1] Heitor Murilo Gomes, Jean Paul Barddal, Fabricio Enembreck, "A Survey on Ensemble Learning for Data Stream Classification" Article 23, March 17.
- [2] M. A. Jabbar, Rajanikanth Aluvalu, S.Sai Satyanarayana Reddy, "Cluster Based Ensemble Classification for Intrusion Detection System" February 24-26, 2017.
- [3] Murat Yagci, Evfik Aytekin, Fikret S. Gurgun, "An ensemble approach for multi-label classification of item click sequences", September, 2015.
- [4] Shrawan Kumar Trivedi, Shubhamoy Dey, "A study of Ensemble based evolutionary classifiers for detecting unsolicited Emails" October, 2014.
- [5] M.A. Jabbar, BL Deekshatulu, P Chandra, " Alternating decision trees for early diagnosis of heart disease", 2014.
- [6] Salima Benqdara et al, "Ensemble of Clustering Algorithms for Anomaly Intrusion Detection System", 2014.
- [7] Introduction to Automata Theory, Languages and Computation.
- [8] Data Mining: Concepts and Techniques, Third Edition by Han, Kamber & Pei (2013).