Review on weather prediction using machine learning

[1]Rubhi gupta
[1]student
[1]jiwaji university

***Abstract* - In this paper, we have evaluated the machine learning techniques to predict weather with much accuracy. During this research process we have used following parameters to predict weather: temperature, rainfall, evaporation, sunshine, wind speed, wind direction, cloud, humidity and size of dataset. This research aims to compare the performance of some machine learning algorithms for predicting weather using weather data. From the collected weather data which contains some weather attributes, which are most relevant to weather prediction. In this paper, various Machine Learning Techniques have explored which includes Naive Bayes Bernoulli, Logistic Regression, Naive Bayes Gaussian and KNN. The experimental results show that Naive Bayes Bernoulli algorithm has good level of accuracy than other algorithms.**

***keywords* - Weather Forecast, Machine Learning Techniques: Naive Bayes Bernoulli, Logistic Regression, Naive Bayes Gaussian, KNN classification, Data pre-processing.**

1.Introduction

In today's information technology era, weather forecasting has become the most challenging and important technique which helps us to predict the weather of any location [1]. Weather prediction help in outdoor programming, crop cultivation, time management and other things that are concern for the mankind. From the last few decades the advancement and development in science and technology enable scientists to make better and precise weather prediction. More advance techniques and technologies are used by the scientists to analyse more accurate weather predictions. Number of methods and techniques are used by the scientists to forecast weather; some of these techniques are more accurate than others. There is huge amount of weather data available which is rich in information and can be used for weather prediction. Forecasting is a process of collecting data on weather conditions, which records the temperature, rainfall, evaporation, sunshine, wind direction, cloud, humidity wind speed and its direction. Various Machine Learning Techniques are applied on weather data to predict climate parameters like temperature, wind speed, rainfall, meteorological pollution [2]. Some of the most commonly used Machine Learning Techniques for weather prediction are Decision Trees, Artificial Neural Networks (ANN), Naive Bayes Networks, Support Vector Machines, Fuzzy Logic, Rule-based Techniques which includes Memory based reasoning Techniques and Genetic Algorithms.

2. Literature Review

In past years, many researchers work on weather prediction using different techniques. Some are explained in this section. In this research paper comparative study on weather prediction using ML Techniques data. Researcher analysis on different Machine Learning Algorithms. Firstly, describes weather prediction has many different problems. Even the simplest weather predictions are not perfect. Prediction of forecast varies from one to two degrees of the actual temperature. Although this accuracy of weather prediction is not bad, as predictions are made for further in time. Also, sometimes accuracy of weather prediction can be even worse. Furthermore, weather prediction in some areas where the climate is not consistence, is off by even more. Machine Learning Algorithms and many classifiers names Naive Bayes Bernoulli, Logistic Regression, Gaussian, support vector machine are uses for evaluate more accurate output.

3. Data collection and pre-processing

Data mining is a technique that changes raw data into a comprehensible format. Raw data (real world data) is always incomplete and cannot be sent through a model. Data mining process steps have been applied to pre-process the data and clean the collected raw weather data Understanding how the data is collected, stored, transformed, reported, and used is essential for the data mining process [3].

A. Data collection

For predict the weather we have collected weather data. For the prediction model, we have used weather data. In raw weather data maximum temperature, minimum temperature (in degree Celsius), humidity, rainfall, evaporation, sunshine, wind gust, wind direction (9am), wind direction (3pm), wind speed, Air pressure (9am), Air pressure (3pm), cloud (9am), cloud (3pm) and temperature (9am), temperature (3pm) above all are the parameters. For weather prediction, we have used the Average temperature, Average Humidity, Average air pressure, Average wind and Events features. We have ignored less relevant features in the dataset for better model computation and prediction [4].

B. Data Pre-processing and Data Cleaning

The main challenge in weather prediction is poor data quality and selection. For this reason, we have used preprocess data carefully to obtain accurate and correct prediction results. In this phase unwanted data or noise is removed from the collected data set which is done by removing the unwanted attributes and keeping the most relevant attributes that help in better prediction. Another major issue is to be rectified the missing values in the collected data set [5]. Missing values in the data set is filled by using various techniques.

Data mining is the process for extracting the useful data from dataset that will give us clean valuable dataset for model computation and better prediction. Most of the data mining algorithms would require data to be structured in a tabular format with records in rows and attributes in columns [5].

4. Research Methodology

There are two types of category in Machine Learning: supervised learning and unsupervised learning. In this work we have carried out research on supervised learning. Classification is a supervised learning approach which is based on training sample set. Machine Learning tool is used to build predictive models. We have implemented four classifications which are experimentally implemented and compared against each other.

These Classification algorithms are Naive Bayes Bernoulli, Logistic Regression, Naive Bayes Gaussian and KNN.

The methodology consists following stages for each study period data of weather parameters which are

(i) Computation of descriptive statistics.
(ii) Development of weather forecasting models and comparison of their predictive ability.
(iii) Identification of precise and reliable weather forecasting model.

4.1 Naive Bayes Bernoulli Algorithm

Naive Bayes classifier gives more accurate results when we use it for textual data analysis. Bayes approach is a method to classify events based on occurrence probability or not happening [6]. Naive Bayes shows proper results using native attribute when it receives primitive practice.

Bayes' theorem: -

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

4.2 Logistic Regression Algorithm

LR Algorithm calculate the relationship between the categorical dependent variable and one or more independent variables. LR gives the output in the form of binary classification.

We can evaluate the probabilities using a logistic function (sigmoid function)

1 / (1 + e^-value)

Where e is the base of the natural logarithms (Euler's number or the EXP() function) and value is the actual numerical value that you want to transform. Below is a scenario of the numbers between -5 to 5 and converted into the range of 0 to 1 using the logistic function.

4.3 Naive Bayes Gaussian Algorithms

Gaussian Naive Bayes algorithm is a particular type of NB algorithm. Naive Bayes Algorithm used when the features have continuous values. After completing the data preprocessing implement machine learning algorithm on it. We have built a Gaussian NB classifier. The classifier is trained using training data. After building a Gaussian NB classifier, our model is ready to make predictions using predict () method with test set features as parameters [7].

4.4 KNN

KNN makes predictions using the dataset. Probabilities are made for new instance (x) by searching through the data set for the K most similar instances and predict the output variable for those K instances.

5. Result and Discussion

Naive Bayes Bernoulli, Naive Bayes Gaussians, KNN, Logistics Regression are the classification model used for predicting the value. There are two group which are separated from the data set for training and testing the algorithms for classification. In this processor there is no more separation of data is from the loaded data [8].

By execute the classification algorithms, the Naive Bayes Bernoulli model gives highest accuracy as compare to other models. First set of result shows prediction accuracy while increasing training data by adding more data, and by adding more parameters [9]. Second set of results mostly emphasize the noticeable performance improvement of our models when different parameters are included in the training data.

Model	Train data Accuracy (%)	Test data Accuracy (%)
Naive Bayes Bernoulli	**1.00**	**1.00**
Logistic Regression	**0.9913**	**0.9697**
Naive Bayes Gaussian	**0.9520**	**0.9293**
KNN	**0.8865**	**0.9091**

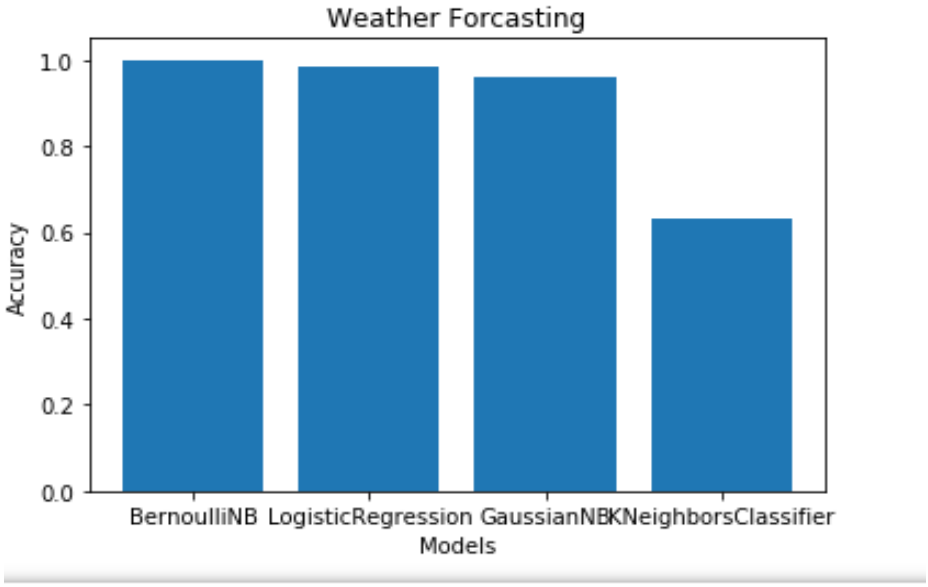


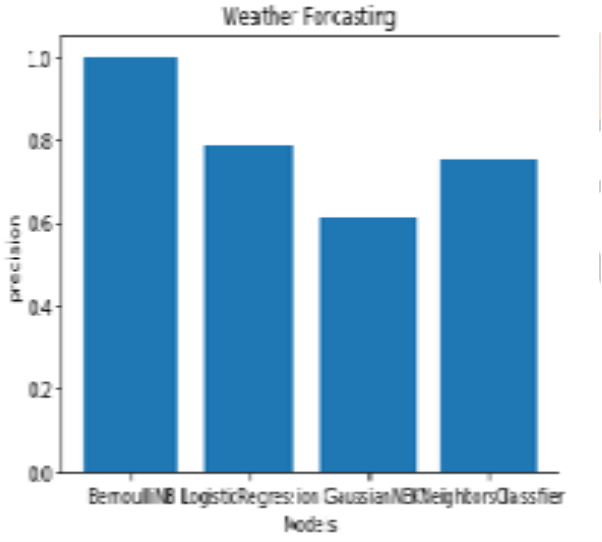
Figure 1. Model Comparison

Table 1. Accuracy Comparison

The results of various machine learning algorithms are compared based on Accuracy, Precision, Recall, and weighted average. Prediction accuracy and performance measures models are based on weather dataset which is shown in figure 2 and figure 3 and is also graphically observed in table 2.

Model	Precession	Recall
Naive Bayes Bernoulli	1.0000	1.0000
Logistic Regression	0.7857	1.000
Naive Bayes Gaussian	0.6111	1.0000
KNN	0.7500	0.2727

Table 2. Precision and Recall value comparison



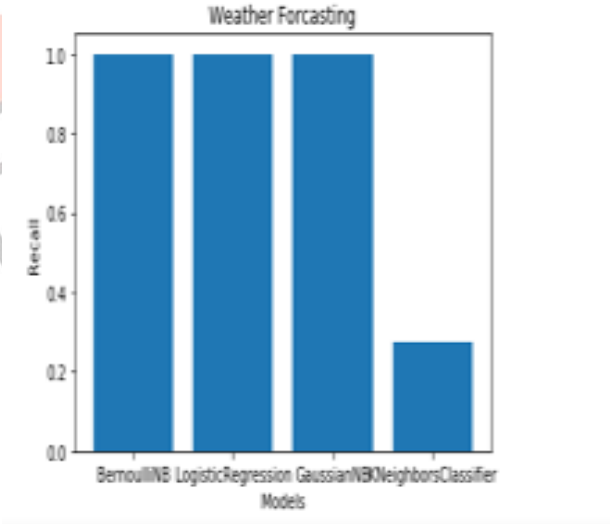


Figure 2. Precision Comparison **Figure 3. Recall Comparison**

It can be seen from the results that Naive Bayes Bernoulli Algorithms has the best prediction model as compared to other algorithms. The graph given in figure-1 shows the result of various classification algorithms and their performance measures. From figure 1, we can observe that accuracy value of Naive Bayes Bernoulli Algorithm is highest as compared to other Machine Learning Algorithm. The different values of Precision, Recall and Accuracy for given weather dataset is shown in table 2. It can be observed that out of four classification algorithms, Naive Bayes Bernoulli Algorithm exhibits highest values of Precision, Recall and Accuracy.

6.Conclusion

In this work we carried out an experimental work to compare popular Machine Learning Algorithms for weather prediction using various performance measures over weather data. The different measuring attributes play a pivotal role in giving precise weather prediction. We have observed that Naive Bayes Bernoulli gives the best weather prediction results with an accuracy of

100% and also exhibits highest values in Recall as compared to other classification algorithms. In our case, Naive Bayes Bernoulli approach proves to be an efficient and acceptable method for weather prediction. The level of accuracy and prediction are highly depending on the data being used as input for classification and prediction. Every algorithm has its advantages and limitations; it is difficult to choose the best algorithm. For weather dataset, it has concluded after analyzing various models of supervised learning that the Naive Bayes Bernoulli classification algorithm has appreciable level of accuracy and acceptance. In future, we have plans to utilize more low-cost devices, such as temperature and humidity sensors. Sensors could increase the number of parameters in the training dataset. This data will improve the performance of our prediction models.

References

[1] Olaiya, Folorunsho and Adesesan Barnabas Adeyemo, "Application of data mining techniques in weather prediction and climate change studies", International Journal of Information Engineering and Electronic Business, Vol. 4, No. 1, pp. 51, 2012.

[2] Sawaitul, D. Sanjay, K. P. Wagh and P. N. Chatur, "Classification and prediction of future weather by using back propagation algorithm-an approach", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, No. 1, pp. 110-113, 2012.

[3] P. Langley, W. Iba and K. Thompson, "An analysis of Bayesian Classifiers", in Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA, 1992.

[4] Kotu, Vijay and BalaDeshpande, Predictive analytics and data mining: concepts and practice with rapidminer, Morgan Kaufmann, 2014.

[5] D. Hand, H. Mannila and P. Smyth, "Principles of data mining", MIT, 2001.

[6] A. Mccallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification", Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98)-Workshop on Learning for Text Categorization, pp. 41-48, 1998.

[7] R. O. Duda and P. E. Hart, Pattern classification and scene analysis, John Wiley and Sons, 1973.

[8] Mohita Anand Sharma1 and Jai Bhagwan Singh2" COMPARATIVE STUDY OF RAINFALL FORECASTING MODELS", Department of Mathematics and Statistics, School of Basic and Applied Science, Shobhit University, Meerut, Uttar Pradesh, 250110, India.

[9] Razeef Mohd1 , Muheet Ahmed Butt2 and MajidZaman Baba3"Comparative Study of Rainfall Prediction Modeling Techniques", Student, 2 Scientist-D, Department of Computer Science, 3 Scientist-D, Directorate of IT&SS, University of Kashmir, Jammu and Kashmir, India.

[10] Mark Holmstrom, Dylan Liu, Christopher Vo" Machine Learning Applied to Weather Forecasting", Stanford University (Dated: December 15, 2016).

[11] E. B. Abrahamsen, O. M. Brastein, B. Lie*"Machine Learning in Python for Weather Forecast based on Freely Available Weather Data", Department of Electrical Engineering, Information Technology and Cybernetics University of South-Eastern Norway, N-3918 Porsgrunn.

[12] A H M Jakaria, Md Mosharaf Hossain, Mohammad Ashiqur Rahman" Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee", Tennessee Tech University Cookeville, Tennessee.

[13] Ahmed, Bilal, "Predictive capacity of meteorological data: Will it rain tomorrow?",Science and Information Conference (SAI), 2015, IEEE, 2015.