WEATHER PREDICTION USING MACHINE LEARNING

Article · August 2022

CITATION
1

READS
3,766

1 author:

Uday CHANDRAKANT Patkar
Bharati Vidyapeeth's College of Engineering Lavale Pune
95 PUBLICATIONS **75** CITATIONS

SEE PROFILE

WEATHER PREDICTION USING MACHINE LEARNING

DR. UDAY PATKAR, MR. SANSKAR MASKE, MR. SAFFA AHMAD,
MR. RUSHIKESH MENGADE, MR. GAURAV SADAWARTI

(COMPUTER DEPARTMENT, BVCOEL)

ABSTRACT: In this paper, we assessed machine learning algorithms for predicting weather with high accuracy. We employed the following parameters to predict weather during this research process: temperature, rainfall, evaporation, sunshine, wind speed, wind direction, cloud, humidity, and dataset size. The purpose of this study is to assess the performance of various machine learning algorithms for predicting weather based on weather data. From the weather data that has been collected, some weather attributes that are most relevant to weather prediction have been identified. Various Machines are discussed in this paper. Naive Bayes Bernoulli, Logistic Regression, Naive Bayes Gaussian, and KNN are some of the learning techniques that have been investigated. The experimental results suggest that the Naive Bayes Bernoulli algorithm outperforms other algorithms in terms of accuracy.

KEYWORDS: Machine Learning Techniques: Naive Bayes Bernoulli, Logistic Regression, Naive Bayes Gaussian, KNN classification, Data Pre-processing.

1. INTRODUCTION

Weather forecasting has become the most difficult and crucial skill in today's information technology era, allowing us to anticipate the weather of any area [1]. Weather forecasting aids in outdoor programming, crop cultivation, time management, and other areas of significance to humanity. Scientists can now create more accurate and precise weather predictions thanks to advances in science and technology over the last few decades. Scientists employ more advanced tools and technologies to analyse more accurate weather predictions. Scientists utilise a variety of methodologies and approaches to forecast weather, some of which are more accurate than others. There is a vast amount of weather data available that is rich in information and can be utilised to forecast the weather. Forecasting is the process of gathering data on meteorological conditions such as temperature, rainfall, evaporation, sunshine, wind direction, cloud, humidity, wind speed, and wind direction. On weather data, several Machine Learning Techniques are used to forecast climatic parameters such as temperature, wind speed, rainfall, and meteorological pollution [2]. Decision Trees, Artificial Neural Networks (ANN), Naive Bayes Networks, Support Vector Machines, Fuzzy Logic, Rule-based Techniques (including Memory-based Reasoning Techniques), and Genetic Algorithms are some of the most often used Machine Learning Techniques for weather prediction.

2. WEATHER PREDICTION USING MACHINE LEARNING

2.1. Prof, Uday Patkar[1], Mr. Sanskar Maske [2] and Mr. Saffa Ahmad [3], Mr. Rushikesh Mengade[4] Mr. Gaurav Sadawarti[5]

[1] Department of Computer, BVCOEL, Lavale, Pune. (Email ID: Uday.Patkar@bharatividyapeeth.edu)
[2] Department of Computer, BVCOEL, Lavale, Pune. (Email ID: sanskarmaske982@gmail.com)
[3] Department of Computer, BVCOEL, Lavale, Pune. (Email ID: saffaahmad7@gmail.com)
[4] Department of Computer, BVCOEL, Lavale, Pune. (Email ID: mengaderushikesh@gmail.com)
[5] Department of Computer, BVCOEL, Lavale, Pune. (Email ID: gauravsadawarti123@gmail.com)

3. LITERATURE REVIEW

Many scholars have worked on weather prediction using various methodologies in recent years. This section explains some of them. A comparative analysis on weather prediction using ML Techniques data is presented in this research article. A study of different Machine Learning Algorithms by a researcher. To begin, weather prediction has a wide range of issues. Even the most basic weather forecasts are not flawless. Forecast accuracy is from one to two degrees above or below the actual temperature. Although this weather prediction accuracy is not poor, as predictions are produced for a longer period of time. Furthermore, the accuracy of weather prediction can be significantly poorer at times. Furthermore, weather forecast is off by considerably more in some locations where the climate is inconsistent. Machine Learning Algorithms with a variety of classifiers such as Naive Bayes Bernoulli, Logistic Regression, Gaussian, and support vector machine are used to assess more accurate output.

4. DATA COLLECTION AND PRE-PROCESSING

Data mining is a process for converting raw data into an understandable format. Raw data (data from the real world) is always incomplete and cannot be processed by a model. Steps in the data mining process were used to pre-process the data and clean the gathered raw weather data. Understanding how data is gathered, stored, converted, reported, and used is critical for data mining.

4.1. Data collection

We gathered weather data in order to forecast the weather. We used weather data to train the prediction model. Maximum temperature, minimum temperature (in degrees 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) are the parameters in raw weather data. We used the Average temperature, Average Humidity, Average air pressure, Average wind, and Events characteristics to forecast the weather. For better model calculation and prediction, we excluded less relevant features in the dataset [4]

4.2. Data Pre-processing and Data Cleaning

Poor data quality and selection are the fundamental challenges in weather prediction. As a result, we employed pre-process data with care to produce accurate and reliable prediction findings. Unwanted data or noise is eliminated from the collected data during this step. Data set, which is accomplished by reducing unneeded attributes while retaining the most relevant attributes that aid in prediction. Another significant difficulty is the correction of missing values in the gathered data set [5]. The data set's missing values are filled by utilising various ways.

Data mining is the process of collecting relevant data from a dataset in order to produce a clean, valuable dataset for model computation and prediction. The majority of data mining techniques would require data to be organised in a tabular fashion, with records in rows and attributes in columns [5].

5. RESEARCH METHODOLOGY

In Machine Learning, there are two types of categories: supervised learning and unsupervised learning. We conducted research on supervised learning in this study. Classification is a supervised learning technique that uses a training sample.

set. To create predictive models, a machine learning tool is employed. We have implemented four classifications, which are as follows: executed experimentally and compared to one another Naive Bayes Bernoulli, Logistic Regression, Naive Bayes Gaussian, and KNN are the Classification algorithms.

For each study period, the approach consists of the following stages:

The approach comprises of the following stages for each study period data of weather parameters:

(I) computation of descriptive statistics.

(ii) Development of weather forecasting models and comparison of their predictive ability.

(iii) Development of a precise and dependable weather forecasting model.

5.1. Naive Bayes Bernoulli Algorithm

When used for textual data analysis, the Naive Bayes classifier produces more accurate results. The Bayes approach is a strategy for classifying events based on their likelihood of occurring or not occurring [6]. When given primitive practise, Naive Bayes produces correct results utilising the native attribute.

Bayes' theorem: -

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

5.2. Logistic Regression Algorithm

The LR Algorithm computes the link between one or more independent factors and the category dependent variable. The output of LR is in the form of binary classification.

A logistic function (sigmoid function) can be used to calculate the probability.

1 / (1 + e^-value)

Where e is the natural logarithm base (Euler's number or the EXP () function) and value is the actual numerical value to be transformed. The logistic function was used to turn a situation with numbers ranging from -5 to 5 into a range of 0 to 1.

5.3. Naive Bayes Gaussian Algorithms

Gaussian The Naive Bayes algorithm is a subset of the NB algorithm class. When the features have continuous values, the Naive Bayes Algorithm is applied. After finishing the data pre-processing, apply the machine learning algorithm to it. We created a Gaussian NB classifier. Training data is used to train the classifier. Our model is complete after constructing a Gaussian NB classifier to make predictions with the predict method and the test set features as arguments.

5.4. KNN

The dataset is used by KNN to make predictions. Probabilities for new instances (x) are calculated by scanning the data set for the K most comparable examples and predicting the output variable for those K occurrences.

6. RESULT AND DISCUSSION

Naive Bayes Bernoulli, Naive Bayes Gaussians, KNN, and Logistics are examples of naive Bayes Bernoulli and naive Bayes Gaussians. Regression is a classification model that is used to predict

the value. There are two groups isolated from the data set for training and testing classification systems. There is no longer any separation of data from loaded data in this processor.

When the classification algorithms are executed, the Naive Bayes Bernoulli model has the highest accuracy when compared to other models. The first set of results indicates prediction accuracy as training data is increased by adding more data and parameters. The second set of findings emphasises our models' substantial performance improvement when different parameters are added 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

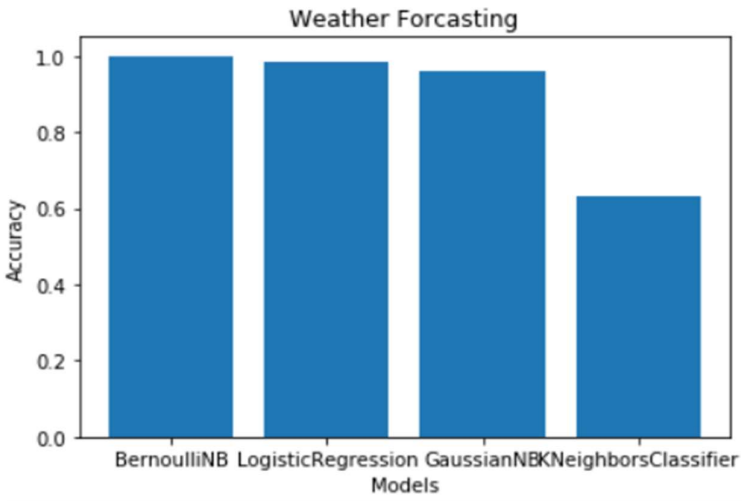
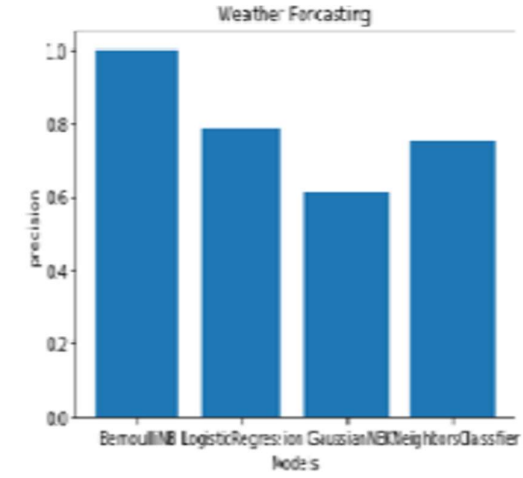


Figure 1. Model Comparison

Table 1. Values for Precision and Recall are compared.



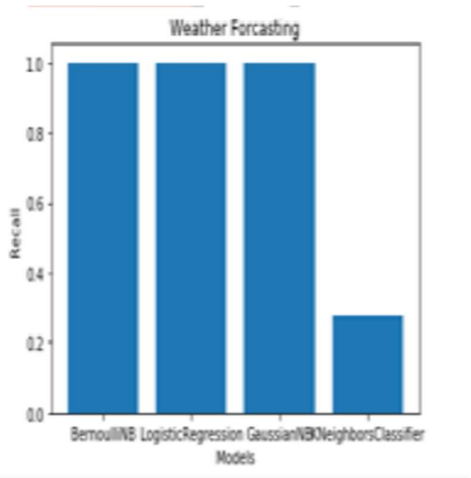


Figure 2. Precision Comparison **Figure 3. Recall Comparison**

The results show that Naive Bayes Bernoulli Algorithms have the best prediction model when compared to other algorithms. Figure 1 depicts the results of various categorization algorithms and their performance indicators.

Figure 1 shows that the accuracy value of the Naive Bayes Bernoulli Algorithm is the highest when compared to other Machine Learning Algorithms. Table 2 displays the various Precision, Recall, and Accuracy values for a specific meteorological dataset. It can be shown that, of the four classification methods tested, the Naive Bayes Bernoulli Algorithm has the highest Precision, Recall, and Accuracy values.

7. CONCLUSION:

We conducted an experiment to compare famous Machine Learning Algorithms for weather prediction using multiple performance indicators over weather data in this paper. The many measuring properties are critical in providing accurate weather predictions. We discovered that Naive Bayes Bernoulli produces the best weather prediction results with a 100% accuracy and the greatest values in Recall when compared to other classification algorithms. In our scenario, the Naive Bayes Bernoulli strategy for weather prediction seems to be an efficient and appropriate solution. The amount of accuracy and prediction depends heavily on the data provided as input for categorization and prediction. Every algorithm has advantages and disadvantages; selecting the ideal algorithm is tough. After studying numerous supervised learning models for the weather dataset, it was determined that the Naive Bayes Bernoulli classification algorithm had a commendable level of accuracy and acceptance. We intend to use more low-cost devices in the future, such as temperature and humidity sensors. The number of parameters in the training dataset could be increased by sensors. This information will help our prediction models perform better.

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 SouthEastern 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