Weather Prediction Using Machine Learning

Sana Khan
Student,
Department of Computer Science and Engineering,
Institute of Technology and Management

Mani Priya Mishra
Student,
Department of Computer Science and Engineering,
Institute of Technology and Management

Rukaiya Khatoon
Student,
Department of Computer Science and Engineering,
Institute of Technology and Management

Ritika Singh
Student,
Department of Computer Science and Engineering,
Institute of Technology and Management

Abstract

In this research paper, we have research on the machine learning technique to prediction of weather with much accurate. Ins this research process we have used following parameters to predict weather: temperature, rainfall, evaporation, sunshine, speed, wind direction, cloud, humidity and size of datasets. This research aims to compare the performance of many machine learning algorithms for predicting weather using weather data. From the collected weather dataset which contains some weather attributes, which are most relevant for weather prediction. In this review, various Machine Learning Techniques have used which includes Naive Bayes Algorithm, Logistic Regression. The results show that the Naive Bayes algorithm has given good level of accuracy than other algorithms.

Keywords: Weather Forecast, Machine Learning Techniques-Naive Bayes Algorithm, Logistic Regression, Data pre-processing

---

1.INTRODUCTION: Weather forecasting prediction has been one amongst the foremost regressive with variables. It improves the static difficult issues round the multivariate analysis and thus initially its uses have increased from last century. The nature of present time weather forecasting is not only very complex but also highly quantitative. However, the system of differential equations that given this physical model is unstable under perturbations, and uncertainties in the initial measurements of the atmospheric conditions and an incomplete understanding of complex atmospheric processes restrict the extent of accurate weather forecasting to a 10 day period, beyond which weather forecasts are significantly unreliable .Weather forecasting is use for predicting weather by determining some factors that include temperature, humidity, pressure, wind, etc. There are various technique and ways to find weather and to analyses its results, which include statically format known as Statistical Weather Prediction, Numerical Weather Prediction and Graphical Weather Prediction. Among them we have come across and worked on Numerical Weather Prediction (NWP).

2.RELATED WORKS: Related works included many different and interesting techniques try to perform weather predcition. While many of present time forecasting technology involves many simulations based on different differential equations, many other approaches from artificial intelligence used machine learning techniques, mostly neural networks while some draw on probabilistic models such as Bayesian networks. In many paper on machine learning for weather prediction we found, many of them used neural networks some others use support vector machines. Navie Bayes is a popular machine learning model and it is good choice for weather prediction because of the ability to capture the non-linear attributes of past weather conditions and future weather conditions. This provides the advantage of not assuming linear dependencies of all features over on models. other two neural network models, one [3] is used for a hybrid model that is used neural networks to model the structure behind weather prediction while the other [4] one applied learning more directly to predicting weather conditions.
Many other proposals for weather forecasting using Bayesian networks. other model [2] used Bayesian networks to model and make weather predictions but used a machine learning algorithm to find the rightest Bayesian networks and parameters which was quite computationally expensive because of the large amount of different dependencies but performed very well. Another approach [1] focused on a more specific case of predicting severe weather for a

specific geographical location which limited the need for fine tuning Bayesian network dependencies but was limited in scope.

3.DATASET: The pressure, humidity, max temperature, min temperature, and weather condition for each day in the years 1996-2016 for Delhi, India were obtained from Weather Underground. there were other weather classifications: clear, cloud, fog, overcast, rain, tornado, thunderstorm, and snow. Since these classifications are same and some are sparsely populated, these were reduced to four weather classifications by combining clouds and partly cloudy into moderately cloud; mostly cloudy, fog, and overcast into very cloudy; and rain, thunderstorm, and snow into precipitation. The data from the first four years were used to train the algorithms, and the dataset from the last year was used is a test set.

4.METHODLOGY: These are following step in our project methodology: -

- Setup
- Data Collection
- Data Preprocessing
- Training Module

4.1 SETUP: In this step we complete dataset analysis and then we preprocessing dataset in Anaconda's Python Environment. We using libraries like Pandas, Sklearn, NumPy, Matplotlib. for training purposes and we used Python 3.6 environment provided by Jupyter Notebook.

4.2 DATA COLLECTION: In order to analyze different machine learning algorithm. We have collected different weather dataset of Delhi. The datasets include several attributes Air temperature at 2-meter height above the earth's surface, Atmospheric pressure at weather station level, Atmospheric pressure reduced to mean sea level, Relative humidity at a height of 2 meters above the earth's surface, Mean wind direction at a height of 10-12 meters above the earth's surface, Total cloud cover, Horizontal visibility, Dew point temperature at a height of 2 meters above the earth's surface. This dataset has data of previous 12 years 1996-2016. Data for some days are missing in this dataset but this can be improved by the large size of the dataset.

4.3 DATA PREPROCESSING: In this step we preparing raw data into making suitable for a machine learning model. It is the first of creating a machine learning model. When creating a machine learning based project, it is not always a case that we find the clean and formatted dataset. And while working any operation with dataset, it is mandatory to clean it and put in a formatted way.

4.3.1 Data Cleaning: No real-world dataset is complete so we have to fill empty values and columns, texts need to mapped with numbers before feeding dataset to different models. First, we have a tendency to assign completely diffrent numbers to different wind direction and map it to wind directions columns. All columns have numerical values for columns saturation point and rain we fill not accessible rows with zero representing no condensate and rain respectively. We then used linear interpolation to fill other columns.

4.3.2 Data Normalization: The data normalization is to change the values of columns in the dataset on as common scale, without affecting differences in the other ranges of values or losing any information in dataset. Normalization help in faster training of models. We have a tendency to normalized the dataset aside form the temperature in 0-1 range.

4.4 TRAINING MODULE: We used different machine learning techniques for weather forecasting in our project. For analysis purpose we have a tendency to predicted temperature but identical algorithms can be used for predicting alternative attributes of weather like precipitation, pressure, snowfall etc.

4.4.1 Artificial Neural Network: A typical feedforward with back propagation network should have at least three layers- an input layer, a hidden layer, and an output layer. We have selected the appropriate number of hidden layers and neurons in each of them.

4.4.2 Naive Bayes Algorithm: Naïve Bayes Algorithm is a classification technique which based on Bayes Theorem. Naïve Bayes is simple to create and extremely helpful for large datasets. By exploitation the Naïve Bayes equation, we will notice the future probability [12]. The Equation is as follows:

P(c|x) =P(x|c) *P(c)/P(x)

Where, (c|x) is future probability of class (c, target), P(c) is the prior probability of the class, P(x|c) is the likelihood which is probability of predictor of given class, P(x) is the prior probability of predictor. The condition of predicting weather forcasting of our project is as follows:

Class:

C1: Weather Forecasting = 'Good',

C2: Weather Forecasting = 'Bad'.

To find the class, Ci that maximizes P (X| Ci) *P (Ci)compute:

P (Weather = Good | x) ∝ P (Weather=Good) · [P (O = s | Weather = Good) · P (T = c | Weather = Good) · P (H = h | Weather = Good) · P (W = t | Weather = Good)]

P (Weather =Bad | x) ∝ P(Weather=Bad) · [P (O = s | Weather = Bad) · P (T = c | Weather = Bad) · P (H = h | Weather=Bad) · P (W = t | Weather =Bad]

IF P (Weather= Good| X) < P (Weather = Bad | X), so classify X as Weather= Bad

Otherwise, Classify X as Weather=Good.

Thus exploitation higher than probability prediction of the long run possibilities of weather good or bad will be easy.

5. DESIGN AND ANALYSIS: This topic incorporates the methodology within the design of the system. This system analyzes and measures weather data. The architecture is given in fig.3, clarifies the working model of the project. The design characterizes the behavior, structure and views of our system.

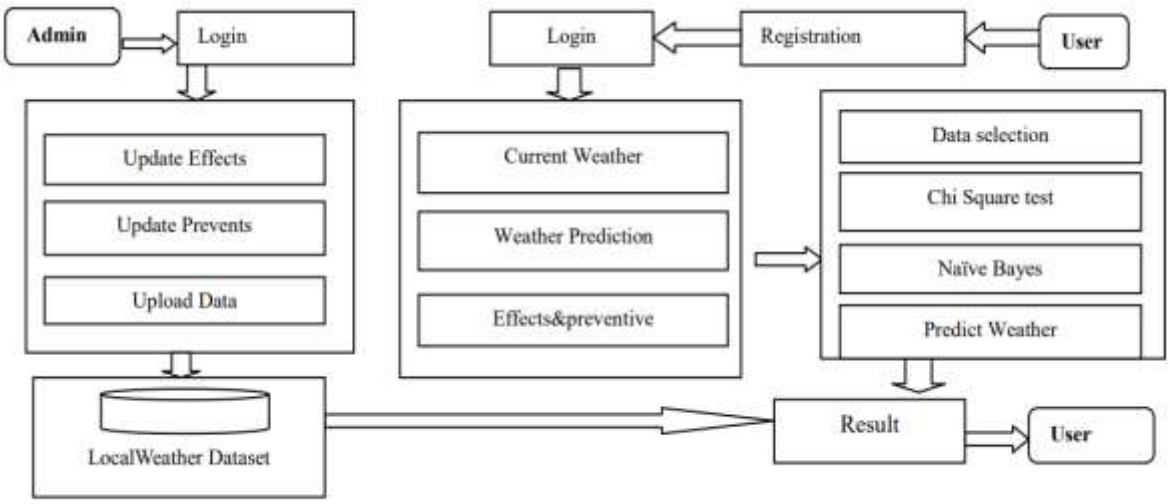
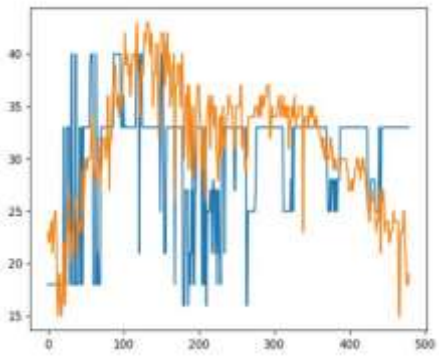


Fig. 3: System Architecture

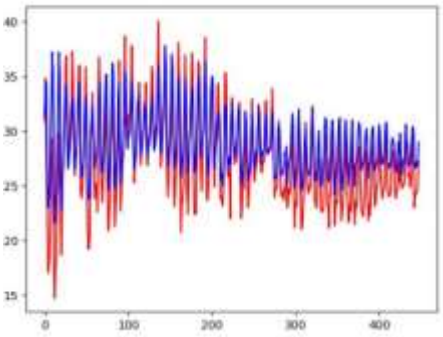
In this project, the job admin is to upload the data, for example, impact of the Outlook, Temperature, and Humidity, Windy and preventive measure and upload the dataset to system. nonetheless at a customer side client need to register to the application. when the login client gets this present weather condition. To predict the weather conditions in the proposed system, the data mining algorithms has been utilized. So as to predict following weather condition or future weather condition the system needed to take input of the weather conditions, based on the client input generate the next possible outcome of weather condition. To partition the information and find the weather condition Naive bays are used here. At that time final prediction of weather condition (Good or bad) are performed.

5. RESULT: The rms error for linear regression and the variation on functional regression are shown in table III. The rms error for a weather forecasting service is additionally enclosed within the same table. Since data regarding the accuracy of weather forecasting services in Stanford, CA weren't accessible, the data were instead taken from weather forecasts for Melbourne, VIC by the Australian Bureau of Metrology's Victorian Regional Forecasting Centre. [5]

In figure shows temperature on the y-axis and sequence no of test data on the x-axis, the temperature predicted by Naive Byese is represented by blue color and the actual temperature is represented by orange



In figure shows temperature on y-axis and sequence no of test data on the x-axis, the temperature predicted by ANN is represented by red color and the actual temperature is represented by blue color.



6.CONCLUSION: In this paper, we work with mixture of Naïve Bayes algorithm to predict weather condition. The constant information i.e. time-series information is assembled associated analysis is performed on this dataset utilizing an interface named Weather Prediction System, developed utilizing Java using Eclipse tools. This framework arranges the given knowledge into various classifications associated what is more predicts the danger of the weather prediction of obscure example is given as an input. The system will be crammed in as training tool for Meteorology Students. This methodology can decide the nonlinear relationship that exists between the historical data (temperature, wind speed, humidity, and so forth.,) provided to the system during the training phase and on that premise, build a prediction of what the weather would be in future. The Future work of this project is to include a lot of attribute of weather condition to predict and to work with other classification algorithm to become a lot of correct in prediction.

References:

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

[2]. Cof? no, Antonio S., et al.” Bayesian networks for probabilistic weather prediction.”15th European 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, Loi Lei, et al.” Intelligent weather forecast. “Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on. Vol. 7. IEEE, 2004.

[5]. Stern, H. (2008), The accuracy of weather forecasts for Melbourne, Australia. Met. Apps, 15: 65?71. doi:10.1002/met.67

[6]. Maqsood, I., M. R. Khan, and A. Abraham. "An ensemble of neural networks for weather forecasting." Neural Computing & Applications 13.2 (2004): pp 112- 122.

[7]. Ranzato, M., Y., Boureau, 1., Chopra, S., &LeCun, Y. "A unified energy-based framework for unsupervised learning," In Proc.Conference on AI and Statistics (AIStats), vol. 20, 2007.

[8]. Stern, H. (2008), The accuracy of weather forecasts for Melbourne, Australia. Met. Apps, 15: 65?71.