MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent Institute of Manipal University)

MANIPAL - 576 104, KARNATAKA, INDIA

MARCH 2017

Rainfall Prediction Using Naive Bayes

By

Akash Mukul Ringe - 140911014
Daniel Pristine David - 140911044

Under the Guidance of

Ms. Anju R.
Mr. Nirmal Kumar Nigam
Mrs. Girija Attigeri

I. INTRODUCTION

Rainfall prediction is an important and integral part of meteorological weather forecasting. Quantitative precipitation forecast which is the plan plot of exact numeric amounts of rain, that is expected to fall over a specific area in a specific period of time, also involves prediction and analysis of previous data. However, keeping the numeric aspects aside, it is also possible to predict whether it will rain or not just with the use of a simple classification Naive Bayes algorithm.

The reason for the lack of perfect accuracy of the rainfall prediction for most of the prediction methods is that rainfall is a random event and the cause of its occurrence is very complex. Even under the same weather conditions, it may be possible that it will rain at this moment but not at another moment. The number of explanatory variables used as the input parameters may not be sufficient to capture all the necessary features for the 24-hr-period prediction.

II. PROBLEM DEFINITION

To find an as accurate as possible a measure, to determine if there is going to be rain or not, on the basis of other atmospheric factors which are likely to influence the rainfall of a particular area.

Despite their importance, representing monsoons remains a major challenge for global weather and climate models. One of the major sources of error is that the deep convective clouds (cumulonimbus) that generate rain are unresolved by the grid-spacings used in global models.

III. OBJECTIVE

To apply the Naive Bayes classifier, in order to classify the various weather attributes into a rainfall prone class or rainfall unlikely class. To train the program with a few split datasets, in order to classify the input vectors accurately into the class they belong to.

IV. SCOPE/IMPORTANCE OF PROJECT

There are several reasons why prediction of rainfall is necessary

1. In case of an almost guaranteed notice or prediction of heavy rainfall in the coming future, forecast becomes an important aspect in protecting life and property from damage.
2. In agriculture based economies, rainfall prediction is exorbitantly imperative in order to keep farmers well informed about rainfall patterns and knowledge about crops that can be grown in the predicted rate of rainfall.
3. The need for accurate rainfall predictions is readily apparent when considering the many benefits such information would provide for river control, reservoir operations, forestry interests, flash flood watches, etc.

The scope of the project is limited to a Naive Bayes approach. However, the technique does not require sophisticated weather and climate modeling infrastructures. Once datasets are made available, the technique can be used locally

V. METHODOLOGY

- ➔ Firstly, we plot the most important attributes pertaining to the atmosphere which are directly affiliated with rainfall and its occurrence. Some of these are, temperature, station level pressure, sea level pressure, wind speed and precipitation quotient.
- ➔ Once the acquired attributes are plotted, we procure a dataset containing these all these attributes. The last attribute however, should denote whether the attributes yielded to rainfall or not. This last attribute will denote our class.
- ➔ We divide the dataset into two classes. The first class will contain the non-rainfall dataset values for each attribute whilst the second class will contain the attribute values for which there has been rainfall.
- ➔ We summarize the classes, by finding mean and standard deviation for each attribute in both these classes.
- ➔ We divide the original dataset into two parts, a test set and a train set by random division of rows. After the training is done, we test each vector in the test set for probability of belonging to both classes, using normal distribution. Total probability of any test vector will be the product of probabilities of all attributes (Naive Bayes approach)
- ➔ We classify the vector into the class which it has a higher probability according to the normal distribution.
- ➔ We find the accuracy, precision and recall of algorithm by dividing the correctly predicted vectors with the total number of entries in the test data set.

VI. RESULTS

- An accuracy between 78-79% which is a relatively good number considering the fact that rainfall is not something objective, also occurs at most unpredictable times and at times does not occur even when it is most likely.
- A marvelous precision rate of 91-92% depicting when algorithm correctly predicted rain had occurred given the total number of times algorithm predicted that rain had occurred.
- A mediocre recall rate of around 45% implying that algorithm correctly predicted rain had occurred, given the total number of times rain had actually occurred.

$$PRECISION = \frac{No\ of\ times\ algorithm\ correctly\ predicted\ rainfall}{No\ of\ times\ algorithm\ predicted\ rainfall}$$

$$RECALL = \frac{No\ of\ times\ algorithm\ correctly\ predicted\ rainfall}{No\ of\ times\ rainfall\ actually\ occurred}$$

VII. CONCLUSION

Bayes is a good way to provide a solution to a difficult problem with a very simplistic approach. Rain forecasting is a difficult task but Bayes approach provides surprisingly good results and a really good precision despite considering all attributes independent of each other.

However the drawbacks of using Bayes is that there implies equal weightage to every attribute that needs to be taken into consideration even if that is not the case. Also there is a very strong unreliable assumption that each and every attribute is independent of each other.

Furthermore, the recall is not very good, indicating that the algorithm also gives a lot of false negatives and is not good in predicting when there is *no rain* albeit a good accuracy in correctly predicting the times when there is.

Overall, Naive Bayes is a good technique for simple prediction but should not be used in cases of intense weather forecasting or quantitative precipitation forecasts where there is a larger involvement of overall preciseness and numeric amounts.

VIII. REFERENCES

Ganesh P. Gaikwad, V. B. Nikam, "Different Rainfall Prediction Models And General Data Mining Rainfall Prediction Model", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 2 Issue 7

Modeling Rainfall Prediction Using Data Mining Method - A Bayesian Approach