INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

WEATHER FORECASTING MODEL USING DATA MINING

SIVASANKARA BABU JONNA
Research Scholar
Department of Computer Science and Engineering
J.S. UNIVERSITY, Shikohabad, UP.

Dr. SARATH BABU DUGGIRALA
Associate Professor
Department of Computer Science and Engineering
J.S. UNIVERSITY, Shikohabad, UP.

Abstract:: Because of its importance in the field of meteorology and its status as one of the most pressing global scientific and technical challenges of our day, weather forecasting has aroused considerable interest among academics. In research, machine learning approaches were shown to be more effective than conventional statistical methods. Predicting the weather is challenging due to the complexity and chaos of the system, this implies that even little errors in the initial circumstances of a prediction may rapidly add up and affect its predictability. A deluge of real-time data is currently being generated by the Internet of Things (IoT), gadgets, radars, weather stations, and satellites. There is a strong need for big data analytics in Internet of Things applications. These two technologies are not new to the business and IT communities. Improved weather predictions may be achieved by evaluating and training machine learning algorithms and data mining approaches using large datasets. This research primarily intends to examine two prominent forecasting methods, namely the Naïve Bayes and the Support Vector Machine, in relation to weather forecasting via the use of data mining. This study's original contribution is to use and compare the two methods' accuracy and correlation using evaluation matrices based on weather confusion matrices; the method with the most accurate depiction of results will be the optimal algorithm for making predictions.

Keywords: Weather forecasting, Naïve Bayes, Support Vector Machine

ISSN: 2057-5688

I. INTRODUCTION

Nowadays, being able to predict the weather is crucial. Industries, communication, catastrophe management, agriculture, severe weather forecasting, drought detection, climate monitoring, and many more examples abound. A major part of metrology is weather forecasting, which is perhaps one of the most intriguing and captivating areas of study. In an effort to make sense of historical atmospheric behaviour, humans have looked for patterns and connections, hoping to draw conclusions about what the future holds. It is common practice to use Machine Learning (ML) in weather forecasting, which often involves processing massive amounts of data. It has become a difficult job to train the model with massive data sets, but there is a benefit for the model that is trained with these enormous datasets.

Weather forecasting, sometimes known as weather prediction, is the process of using scientific and technological methods to foretell the upcoming atmospheric conditions for a certain area and time frame. Agricultural and production, climate monitoring, severe weather prediction, communication, disaster management, and a host of other modern-day domains rely heavily on accurate weather forecasts using tools such as super computers, radar, satellites, weather stations, and aircraft.

Data mining entails collecting and managing massive amounts of data with the purpose of examining them. Sorting the controlled data makes it ready to be displayed in any order, since it is another step in the preparation process. In a consulting room, a team of experts reviews the data gathered by these technologies and inputs it into a model. The model then produces two types of forecasts: those for the near future and those for the far future. Both the short-term and long-term predictions provide outcomes for the now and the future, respectively.

II. LITERATUREREVIEW

In the past, a lot of people have tried to figure out how to forecast the weather using different methods. Here we will go over a few of them. This research study offers a comparative analysis of weather prediction methods that use machine learning techniques. Experts in the field of machine learning examine various algorithms. First of all, there are several problems with explaining weather forecast. The accuracy of even the most fundamental weather predictions may vary. The difference between the actual and predicted temperatures is one to two degrees. Nonetheless, the accuracy of weather forecasts is not terrible because they are made over a longer duration. In addition, there are instances where weather forecasts are far less accurate than expected. Additionally, different indicators and trends are used to analyse

and contrast different predictive analytic methodologies. Many different algorithms exist for classification and prediction. Machine learning algorithms and a plethora of classifiers, including support vector machines, naive bayes, logistic regression, and Galois are used to assess more precise output.

III. RESEARCHMETHODOLOGY

Supervised and unsupervised learning are the two main subfields of machine learning. In order to write this article, we engaged in supervised learning research. Using a machine learning technology, predictive models are created. Classification is a kind of supervised learning that uses a collection of training samples. We use two main methods, Naïve Bayes and Support Vector Machine (SVM), in this article.

- **Naive Bayes**

The Naive Bayes model is easy to build and is especially useful for large data sets. Because of its simplicity, Naive Bayes is known for surpassing even the most powerful classification algorithms. Using (c), $P(x)$, and $P(x|c)$, the Bayes theorem allows you to calculate the posterior probability (|) ( ), ( ), (|).Consider the following equation:

$$P(c|x))= \frac{P(c|x)P(c)}{P(x)}$$

The following formula is used to compute the conditional probability that an observation belongs in the class :

$$P\left(C_k|(x_1,x_2,\dots,x_p)\right)=\frac{\frac{P(C_k)\prod_{i=1}^{p}[P(C_k)]}{R}}{\sum_{k=1}^{K}\left(\frac{P(C_k)\left(\prod_{i=1}^{p}[P(C_k)]\right)}{R}\right)}$$

Where,

R is the regularization constant.

C_k stands for class k.

The input variable p is represented by Xp

- **Support Vector Machine (SVM)**

Machine learning's supervised learning models for analysis's regression and classification rely on support vector machines (SVMs) to execute the learning algorithms. Another way to manage the SVM classification issue is by using corresponding approaches, such random forests or decision trees. Support Vector Machines work by first trying to map the main data set X into a high-dimensional feature space F, and then using a non-linear mapping function, they build the best feasible hyper plane in this new space. Support vector machines are useful for both regression and classification. Classification is a perfect example of a hyper plane that neatly splits data into two categories. In contrast, regression calls for the construction of a hyper plane that passes near a large number of locations.

Hence classification models (In fig.1) such as Naïve Bayes and Support Vector Machine (SVM) are developed using classifier techniques and the purpose is to concentrate on difficult concepts. When combined with a classifier, SVM will produce superior classification by providing adequate training to poor learners.

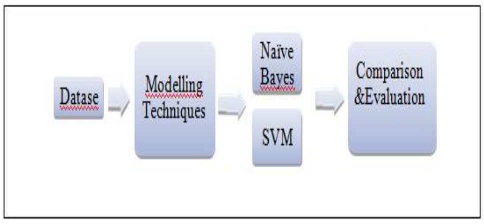

Fig.1. Model System Design

The Naïve Bayes classifier, which is based on the Naïve Bayes algorithm and is used to generate categorized data, takes a similar approach.

- **Dataset Description**

The data is in CSV format and includes several factors that are responsible for a location's weather. The prediction model was developed using weather data. Fog, hail, humidity, thunder, wind direction, temperature, humidity, cloud cover, precipitation, wet day frequency, and monthly totals for a given year are all utilised as record characteristics in raw meteorological data.

IV. RESULTS

In this section we have discussed results of prediction obtained, the inferences that can be drawn from those results and the related discussion over them. The results are shown using evaluation matrices.

- **Evaluation Matrices**

The quality of a statistical or machine learning model is measured using evaluation metrics. Any task requires evaluating machine learning models or methods. There are a variety of evaluation measures that may be used to test a model.

Different terms used in evaluation matrices are described below:

- **Precision**

Precision is defined as the ratio of successfully predicted positive observations to total expected positive observations. It's a good measure to use when the cost of a false positive is large.

Precision is linked to a low false positive rate. Better the precision, better the outcome.

Precision is formulated as

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

- **Recall**

The number of true positives that the model captures by labeling it as positive is calculated by recall (true positive). With this knowledge, it is possible to conclude that recall should be used as acriterionforpickingtheoptimalmodel whenthereisahighcostassociatedwithf alsenegatives. Recall is formulated as

$$= \frac{}{+}$$

- **Accuracy**

Accuracy refers to how close a measurement or a set of standards is to a measured value or as set of standards. In time series analysis, accuracy refers to the predicted value being very close to the actual value.

For accuracy, use the formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

- **Results & Comparison of two matrices**

The results obtained as follows:

Table 1. Summarization of results obtained

Algorithm	Accuracy	Precision	Recall
NaiveBayes	49%	0.64	0.51
SVM	91%	0.92	0.92

It is clear that the SVM is preferably

suited for our purpose of weather prediction as it scores higher on different evaluation metrics used.

➢ **Visualization of results obtained**

To visualize the results learned by the model for the dataset.

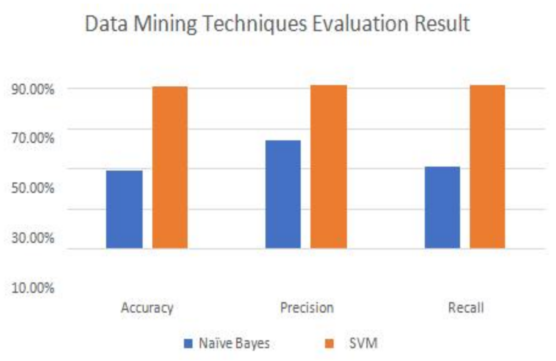

Fig.2.Clustered Column chart comparison of table1

V. CONCLUSION

This model makes use of two pre-existing methods for prediction: Support Vector Machine (SVM) and Naive Bayes. The goal of processing synoptic weather data using Support Vector Machine (SVM) or Naive Bayes algorithms is to provide the most accurate surface analysis (depth). To find out how reliable a model is, each given set of input data will be used to assess each method's performance model. Additionally, the optimal algorithm is determined by comparing the test results to find the method with the maximum accuracy. Our results show that SVM outperforms Naive Bayes, making it the optimal method for our dataset. Next time you need to forecast the weather, use the Support Vector Machine (SVM) technique instead of Naive Bayes. If we want to see if we can broaden the scope of the research to see if changing the method or adding more data to the same method produces different results that are something we can tackle in the future

REFERENCES

1) Fairoz Q. Kareem, Adnan Mohsin, Abdul Azeez and Dathar A. Hasan, "Predicting Weather Forecasting States Based on Data Mining Classification Algorithms". Asian Journal of Research in Computer Science, 9(3): 13-24, 2021; Article no.AJRCOS.68636. ISSN: 2581-8260
2) C. Karthikeyan, Gurram Sunitha, J. Avanija , K. Reddy Madhavi, and E.S. Madhan, "Prediction of Climate Change using SVM and Naïve Bayes Machine Learning Algorithms". Turkish Journal of Computer and Mathematics Education Vol.12 No.2 (2021), 2134 – 2139.
3) P. Kalaiyarasi and Mrs A. Kalaiselvi, "Data Mining Techniques Using Weather Prediction". International Journal of Computer Science Trends and Technology (IJCST) – Volume 6

Issue 3, May - June 2018.

4) Razeef Mohd, Muheet Ahmed and Majid Zaman, “Modeling Rainfall Prediction: A Naive Bayes Approach”. International Journal of Advances in Electronics and Computer Science, ISSN(p): 2394- 2835. Volume-5, Issue-12, Dec.-2018.
5) Urmay Shah, Sanjay Garg, Neha Sisodiya, Nitant Dube and Shashikant Sharma, “Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques”. Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), DOI: 10.1109/PDGC.2018.8745763. Conference: 2018
6) Ning Yang, Lewis Westfall and Ms. Preeti Dalvi, "A weather prediction model with Big Data”. Proceedings of Student-Faculty Research Day, CSIS, Pace University, May 4th, 2018.
7) Siddhant Revankar and Shaba Desai, “Weather Forecasting Using Data Mining”. International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), Vol 5, Issue 2, February 2018.
8) M. S. Bennet Praba, Antony John Martin, Siddharth Srivastava & Ajay Rana, “Weather Monitoring System and Rainfall Prediction Using SVM Algorithm”. International Journal of Research in Engineering, Science and Management Volume-1, Issue-10, October-2018, ISSN (Online): 2581-5792.
9) Sayali D. Jadhav and H. P. Channe, “Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques”. International Journal of Science and Research (IJSR) ISSN (Online): 2319- 7064.
10) Y. Radhika & M. Shashi, “Atmospheric Temperature prediction using Support Vector Machine”, International Journal of Computer theory & Engineering, Vol.1.No.1.April 2009.
11) Md Rashid Mahmood, Raj Kumar Patra, Rohit Raja and Prof. G. R. Sinha, “A Novel Approach for Weather Prediction Using Forecasting Analysis and Data Mining Techniques”. In book: Innovations in Electronics and Communication Engineering (pp.479-489), DOI: 10.1007/978-981-13-3765-9_50.