ISSN: 2366-1313

Machine Learning Based Weather Prediction

[1]K. SAIKIRAN, [2]D.RAM MOHAN REDDY

[1]MTech Student, Dept of CSE, Newtons Institute of Engineering, Macherla, Guntur District

[2]Associate Professor, Dept of CSE, Newtons Institute of Engineering, Macherla, Guntur District

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: Machine learning techniques, weather forecast, classification, ensemble learning.

I. INTRODUCTION

The use of science & technology to forecast atmospheric conditions at certain locations and periods is known as weather prediction. Numerous people attempt to predict the weather several years Farmers can use rainfall forecasts. A model is built using machine learning, a data science approach, from a training dataset. Machine learning makes weather forecasting more convenient, accurate, real time, and less time-consuming. Some of the most current research has. Making use of historical weather data is required for machine learning techniques to predict weather conditions. Only the data used to train the models will determine how accurate the forecasts are. As a result, it is crucial that any machine learning model be trained on extremely accurate data. The information gathered from numerous sources is not necessarily reliable. So, preprocessing of the data is required. The preprocessing of the data includes eliminating useless columns that are unrelated to the model's predictions, eliminating zero values, grouping related columns, and performing many other pre-processing operations. Some of the earlier research on machine learning-based weather forecasting has produced remarkably precise findings. They demonstrated that for predictions of up to seven days, professional weather forecasting services outperformed both

ISSN: 2366-1313

models. A predictive algorithm for identifying the changing trends of weather conditions based on data mining was put out. In the suggested data model, weather observations are extracted using k-means clustering and predictions are made using a hidden markov model. Grover et al. explored weather forecasting using a hybrid technique that blends deep neural networks that simulate the combined statistic of a collection of weather-related data with discriminatively trained predictive models.

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).

II. 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.

To classify weather conditions for autonomous cars as either unfavorable or normal, Qasem Abu Al-Haija et al[8]. offer a deep learning (DL)-based detection system. Three deep convolutional neural networks (CNNs) are characterized in the proposed framework: SqueezeNet, ResNet-50, and EfficientNet. The system takes use of transfer learning methods and the capability of the Nvidia GPU. The created models have been tested on the most recent DAWN2020 and MCWRD2018 weather imaging datasets. All models showed excellent classification ability in experiments, however, the ResNet-50 CNN model trained on weather data had the highest detection accuracy (98.48%), precision (98.51%), and sensitivity (98.41%). Sebastian Scher et al. [9] proposed a deep learning-based solution employing artificial convolutional neural networks and weather forecast data. A new weather event adds a scalar confidence value to medium-range predictions initialised from the supplied atmospheric state to indicate if predictability is greater or lower than normal. Although ensemble weather prediction models predict forecast uncertainty better, the suggested technique is computationally economical and outperforms previous non-numerical methods. For two years of data with a 15-minute precision from 16 PV plants in Hungary, Dávid Markovics et al[10]. examined 24 machine-learning models for deterministic day-ahead power forecasting using NWP. The advantages of hyperparameter adjustment and the results of the predictor selection are also assessed. With a maximum prediction skill score of 44.6%, kernel ridge regression, and multilayer perceptron emerge as the two most effective models. With the use of historical data from various weather stations, A H M Jakaria et al.[11] demonstrated a method of weather prediction that may produce meaningful

ISSN: 2366-1313

predictions regarding particular weather conditions in the near future in a very short amount of time using basic machine learning models. These models don't need very powerful computers to operate. The evaluations demonstrate that the models are accurate enough to be utilized in conjunction with the most cutting-edge methods currently available. The authors also demonstrated the value of using data from weather stations in nearby regions rather than just those in the area being forecasted.

III. MATERIALS AND METHODS

A. Dataset

This study uses the weather dataset which contains the features like precipitation, maximum and minimum temperature, wind speed, etc. Target weather conditions are drizzle, rain, sun, snow, and fog.

B. Method

1) Data Pre-processing:

In order to verify that all of the characteristics are on the same scale, we do some preliminary processing on the data once it has been gathered. In order to make the data more acceptable for training machine learning algorithms, this step is crucial. Imbalance check and correction using SMOTE: Dealing with unbalanced datasets is a frequent difficulty when utilizing machine learning for weather prediction. In such datasets, the performance of the machine learning algorithms may suffer because one class may have a disproportionately large number of occurrences compared to the other classes.

The SMOTE (Synthetic Minority Over-sampling Technique) method may be used to solve this problem. By extrapolating between existing samples, SMOTE creates synthetic samples for the under sampled class. Following are the steps for using SMOTE to address the imbalance in weather forecast datasets:

a) Determine the class distribution: We must first analyze the dataset to find out how the classes are distributed. One way to do this is totally up the number of instances belonging to each class.

b) Check for imbalance: Oversampling may be used to rectify an unbalanced dataset in which classes are unevenly distributed. In this instance, we shall use SMOTE.

c) Apply SMOTE: To implement SMOTE, we create simulated samples for the underrepresented group by extrapolating from the majority group's data. The oversampling percentage determines how many artificial samples are created. d) Combine the original and synthetic

ISSN: 2366-1313

samples: In order to establish a new, more representative dataset, the original samples are joined with the synthetic ones once they have been made.

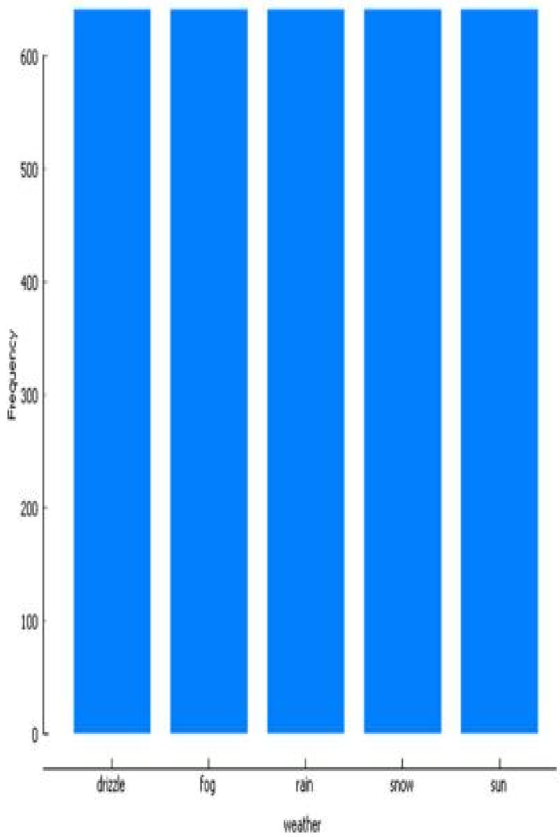

Fig.1 Class distribution after SMOTE

SYSTEM ARCHITECTURE

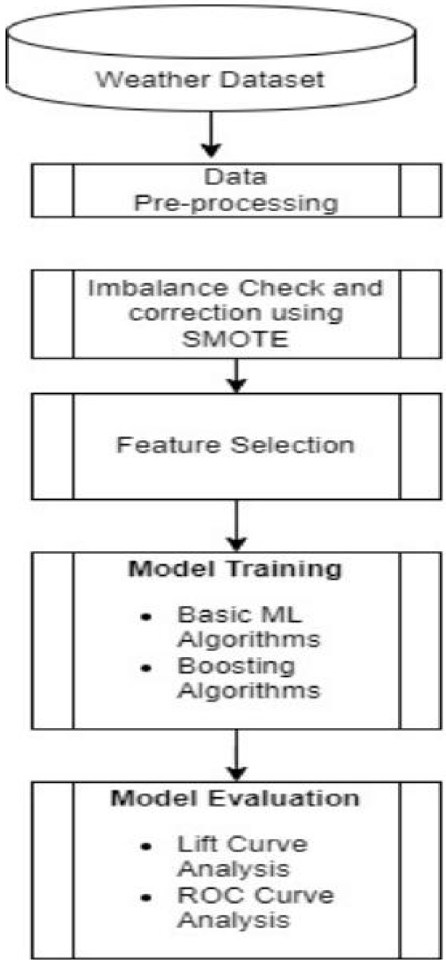

Fig.1 The flow of proposed system

In the fig. 1 demonstrates how well-balanced the dataset is. 2) Feature Selection: After the data has been cleaned and organized, the attributes most useful for predicting the weather are chosen. Temperature, humidity, wind speed, rainfall, and air pressure are all possible examples of such factors. In order to increase the efficiency of the machine learning algorithms, feature selection is a crucial process. 3) Algorithm Selection: In this research, we compare the efficacy of several machine learning algorithms used for forecasting weather. Among them are

ISSN: 2366-1313

boosting algorithms like AdaBoost and XGBoost[3], as well as more traditional methods like decision trees and random forests[14]. These algorithms were chosen because of their capacity to process big datasets and their applicability to the field of weather forecasting. 4) Dataset Splitting: The effectiveness of the machine learning algorithms is measured using a 10-fold crossvalidation procedure in this experiment. It reduces the likelihood of the model overfitting and improves predictions for future data[15]. 5) Model Training: We next use a stratified k-fold crossvalidation method to train the chosen algorithms on the cleaned and sorted data. This method guarantees that the performance of the algorithms is measured on a fair representation of the data and helps to avoid overfitting. Model Evaluation: We use measures like accuracy, precision, recall, and F1 score to assess the models' performance after training. To further assess the efficacy of the algorithms, we use ROC curve analysis and lift curve analysis. This process aids in finding the most reliable and productive weather forecast algorithms.

IV. RESULTS

The calculations for determining the classification accuracy (CA), precision, recall, F1 score, and area under the curve (AUC) of machine learning classification methods [16] are as outlined in Table I

Table.1 Comparison between various algorithms

Model	AUC	ACC	F1	Precision	Recall
XGBoost	0.98	0.88	0.88	0.88	0.88
AdaBoost	0.97	0.87	0.87	0.87	0.87
Random Forest	0.97	0.86	0.86	0.86	0.86
CatBoost	0.97	0.84	0.84	0.85	0.84
Gradient Boosting	0.96	0.82	0.83	0.83	0.82
Tree	0.91	0.81	0.81	0.81	0.81
kNN	0.94	0.80	0.79	0.80	0.80
Neural Network	0.92	0.70	0.70	0.70	0.70
Naive Bayes	0.88	0.64	0.64	0.65	0.64

V. CONCLUSION

In this paper, we use a combination of simple tools to acquire knowledge and improve algorithms for weather forecasting. According to our results, XGboost and Adaboost are the most useful algorithms with accuracy of 87.86% and 87.33%, respectively. Lift curve analysis and ROC curve evaluation established the results, providing additional evidence of the superior overall performance of these techniques. Since accurate weather forecasting is crucial for agriculture, transportation and emergency presentations, as well as many sectors and sports, our results have great implications for the field of weather forecasting. XGboost and Adaboost are examples of system awareness techniques that can be

ISSN: 2366-1313

used to improve the accuracy of weather forecasts and provide users with more honest information. Overall, our work demonstrates the promise of the technology for improving climate forecasting and paves the way for further research in this area. Further studies are needed to better understand how these algorithms perform on different datasets and discover new aspects that can improve the accuracy of climate prediction models.

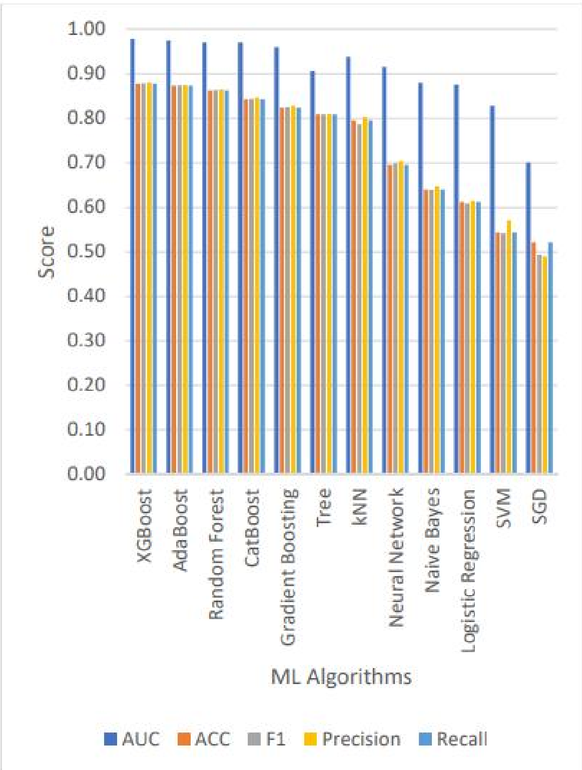

Fig.2 Performance comparison of M models

A comparison of the degrees to which various machine learning approaches achieve accurate categorization is shown in fig. 4

REFERENCES

1. Venkata Ramana, Chandra Mouli, Aileni Eenaja "Network Intrusion Detection By SVM & ANN With Feature Selection" (2020) IJCRT | Volume 8, Issue 6 June 2020 | ISSN: 2320-2882
2. H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, (2016).
3. P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.
4. M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, (2015), pp. 513–517.
5. M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, (2010).

ISSN: 2366-1313

6. A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, (2011).
7. M. Zamani and M. Movahedi,arXivpreprintarXiv:1312.2177, (2013). DOI: https://doi.org/10.48550/arXiv.1312.2177
8. N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," IJCBR ISSN (Online), Volume 4 Issue 2 May (2013) pp. 2229– 6166, DOI: https://link.springer.com/chapter/10.1007/978-981-16-8012-0_4
9. Vijayakumar, Alazab, Soman, Poornachandran, Al-Nemrat, S." A Deep Learning Approach for Intelligent Intrusion Detection System". IEEE Access (2019). DOI: https://doi.org/10.1109/ACCESS.2019.2895334.
10. Raghavendran C.V., Pavan Venkata Vamsi C., Veerraju T., Veluri R.K. (2021) Advances in Intelligent Systems and Computing, vol 1280. Springer, Singapore. DOI: https://doi.org/10.1007/978-981-15-9516-5_13.