Rainfall Prediction: Analysis of Machine Learning Algorithms and Ensemble Techniques

Abhinav Sharma
Dept. of Computer Engineering
VESIT
Chembur, India
abhinavrks012@gmail.com

Akshay Khanna
Dept. of Electronics and Telecommunication Engineering
VESIT
Chembur, India
2018.akshay.khanna@ves.ac.in

Muskaan Bhargava
Dept. of Electronics and Telecommunication Engineering
VESIT
Chembur, India
2018.muskaan.bhargava@ves.ac.in

Rutwik Pendse
Dept. of Instrumentation Engineering
VESIT
Chembur, India
2018.rutwik.pendse@ves.ac.in

2021 7th International Conference on Signal Processing and Communication (ICSC) | 978-1-6654-2739-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICSC53193.2021.9673275

***Abstract*—In 2018, approximately 60% of land in India was reported as Agricultural Farmland. Most of the farmers in India depend on rainfall as the primary source of irrigation which implies that rainfall is directly linked to the yield. The quantity of rain a land receives helps in planning which crop to sow and also in which month the farmer should begin farming. Precipitation is not only limited to agriculture, even the infrastructure development sector has to keep track of the monsoon season and amount of rainfall that a construction site will receive because of the influence it has on construction projects. It also serves as one of the most important sources of freshwater for all living things on the planet. As rainfall prediction model gives data on the impact of various climatological variables on rainfall amounts it has many applications across many sectors. We propose a study with analysis of various machine learning algorithms for rainfall prediction. An accuracy of 95% was obtained with XGBoost, which is a gradient boosting framework. This also serves to inform the robusticity of the proposed model in comparison.**

Keywords*—Rainfall prediction, Decision tree, Logistic Regression, Random Forest, XGBoost, Ensemble model, Catboost, Light GBM, Weather parameters, MICE Imputation, Wrapper method, Chi square.

I. Introduction

India Rainfall implies yield and yield implies life. It is the most important factor in human establishment not just for survival but also for human development and advancement. Precise precipitation data is critical for the planning and management of water assets, as well as for repository activity and floods forecasting. Moreover, the hydroelectric method of electricity generation greatly depends on the rainfall rate in the catchment area. Furthermore, precipitation has an influence on traffic, sewer systems, and other human activities in urban areas. On the contrary, rainfall acts as a cause of various natural disasters like landslides, one of the major catastrophes found to be commonly occurring all around the globe.

As essential as it is, precipitation is also the most unpredictable component of the hydrologic cycle due to the factors it depends on which include intricacies in the climatic conditions. This has been actively attracting the attention of various government institutions and industries along with scientific communities. Attempts to accurately forecast precipitation has been carried out by horticulture areas. Even though there are various hardware instruments that may be used to forecast rainfall based on physical parameters such as temperature, pressure, and humidity, these approaches fail to produce reliable results. Statistical approaches also fail to give accuracy for rainfall forecasting due to the dynamic nature of Earth's atmosphere and climatic conditions.

Using machine learning algorithms is an effective way to get the desired accuracy and eliminate errors. We can simply accomplish that by analysing past rainfall data in order to forecast rainfall for future seasons. Different approaches, such as classification and regression, can be used depending on the needs, and error between the real and predicted values can also be determined, enhancing the accuracy even further.

In this research, multiple critical factors of a region which greatly affect the precipitation trends have been pinpointed, few of them being temperature, atmospheric pressure and sunshine. A total of seven machine learning algorithms have been implemented and a comparison has been made to find out which is the most suitable model for deployment, thereby providing us with precise and accurate data. The models executed are Logistic Regression, Decision Tree, Neural Network, Random Forest, Light GBM, CatBoost and XGBoost.

Contributions

Various studies have been conducted in an attempt to predict rainfall accurately, and this is what distinguishes our study from the others. We have made use of MICE Imputation and Wrapper method for filling out the missing data. Outliers have been removed and checking of multi-collinearity has been done between features. Chi-Square Value has been used for feature selection and training has been done on ensemble models like LightGBM, Catboost and XGBoost.

II. Literature Survey

Ozlem Terzi in 2012[1] collected monthly rainfall data from Turki's Isparta Uluborlu, Egirdir, Senirkent, and Yalvac district weather bureaus, and it was used to create a precipitation estimation model. Decision Table, Rainfall estimation models were created using the KStar, Multi-linear Stagnation, M5 Laws, Multi - layer perceptron, back propagation neural network, Randomized Subspace, and Straightforward Linear Regression., and the estimated significance level (R2) and underlying error (RMSE), are perhaps the most well-known and extensively used order to determine the strength, have been used to judge the efficacy of these models. He created the MLR model that delivers the best results by combining multiple data combos supplied onto the previously existing concepts. A K-medoid algorithm for grouping shapes/peaks has been studied. The different categorization and association rule extraction approaches were addressed in this article. Instead, they've chosen all of the catchments in their study area where, for short time frames, a significant chunk of rainfall events is obtainable.The enhanced data was therefore exposed to several standard criteria, like the amount of precipitation which had to occur in a brief period on the specific day to constitute as an unprecedented event. A huge proportion of rainstorm events may be acquired in short time frames. The Root mean square error for LSTM was set to 2.55 and for ConvNet, 2.44 by Aswin S[2]. The dataset consisted of monthly precipitation data at global level from 1979 to 2018. ConvNet was concluded to be promising at 100 epochs. MAPE for ConnNet was 1.7281, which was better than LSTM.

Pratap Singh Solanki et al. [3] examined the use of data extraction techniques in the area of water management. Water transmission has long been viewed as one of the most rigorous, engaging, and fascinating issues on the globe. Researchers can use a variety of techniques to assess precipitation, flooding warnings, fluid input, water resources, and needs, among other things, based on enormous amounts of information. They investigated just how machine learning algorithms could be used to evaluate influx, famine risk, weather forecasts, temperature, humidity, heat, wind velocity, as well as other aspects in this study. The research involves a review of literature and also work done by researchers that used a different algorithm and numerical modelling, including frequent patterns, categorization, grouping, selection trees, and neural networks, among others. Table I below is a pithy of the past work since.

III. Methodology

We used weather parameters like sunshine, temperature, atmospheric pressure, etc. as features to predict rainfall in an area. 7 machine learning algorithms- both traditional and ensemble models are implemented and compared to find the most suitable model for deployment. The predicted values of rainfall can further be used to predict other disasters like rainfall and flash floods.

Figure 1 illustrates flow diagram for the proposed system. First, in data exploration we check whether all the data types are same or not and are there any missing values in data, then we handle the class imbalance in data between the number of days rainfall occurred and no rainfall days. There are some missing values in data which are handled in Imputation and transformation step using two different methods, then by analyzing the pair plot graph we select the best features from out data for training and these features are used as input for training different models.

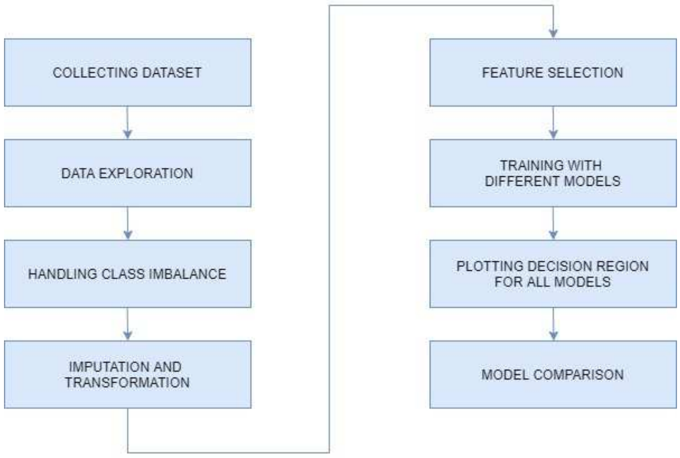

Fig. 1. Flow diagram of proposed work.

The graphs for all the trained models are compared along with other evaluation metrics to find the most suitable model

A. System Setup

The machine learning algorithms are executed on Google Colaboratory using the Tesla K80, having 2496 CUDA cores GPUs for runtime, which outperformed CPUs runtime by 17 times and in-built modules and function from Scikit-learn package in Python 3.6 are used for training and testing purposes. The saved model file is deployed on a website using Google Firebase and API services for sending features and receiving predictions.

B. Dataset Collection

This dataset contains about 10 years of daily weather observations from many locations across Australia. Daily Weather Observations list observations of a number of weather elements each day for a month. Included are daily minimum and maximum temperature, rainfall, strongest wind gust, evaporation, and sunshine, together with values at 9am and 3pm of temperature, humidity, wind, cloud and pressure.

C. Data Pre-processing

In data pre-processing, we start with checking the size of our dataset i.e., the number of rows and column. The size of dataset is 87927 rows and 24 columns. In the dataset both "RainToday" and "RainTomorrow" are object values i.e., Yes/No. We converted them into Boolean i.e., 1/0 for our convenience. Next, we checked whether the dataset we had was balanced or imbalanced. The reason behind this was that if the dataset is not balanced, we need to oversample minority or under-sample majority to balance it. Imbalanced data can give lower accuracy in results. Figure 2 is the distribution of the dataset. We can observe that numbers of "0" and "1" are in the ratio 78:22. This implies that there is a class imbalance between "0" and "1". We used oversampling of the minority class to address class imbalance. Because the data set is relatively small, under-sampling of the majority class won't have any positive impact. Figure 3 is the distribution of the dataset after handling class imbalance.

TABLE I. RECENT WORKS ON RAINFALL PREDICTION.

Author	Database	Methodology	Performance
Cramer, et al. (2017) [4]	This dataset comprised 20 conurbations from Europe and 22 conurbations from the United States. This information was acquired with the collaboration of NOAA and NDC.	Genetic programming approaches include Support Vector Regression, K-Nearest Neighbours, M5 Model Trees, Radial Basis Neural Networks.	Error reduction of around 70% was achieved.
Pham, B. T., et al. (2018) [5]	The dataset includes data for temperature and rainfall for the Hoa Binh province, Vietnam.	Particle Swarm Optimization (PSO), Artificial Neural Networks, Support Vector Machines, Adaptive network based fuzzy inference system (ANFIS)	Accuracy of 80.44% was achieved.
Manandhar, et al. (2019) [6]	-	Supervised Machine Learning Technique, Evaluation Metrics, Down sampling Technique	The model had predicted around 79.6% accuracy.
Deepak Ranjan Nayak, et al. (2013) [7]	The dataset has parameters for rainfall in San Francisco Bay Area	SVM, Feedforward Network, SOM, Radial basis function Networks,	The poll conducted reveals that SVM, Feedforward Network, SOM, Radial basis function Networks, and forecasts techniques are better at determining rain than in other numeric approaches.
Venkata Ramana, et al. [8]	The dataset consists of data of rainfall for the city of Darjeeling	Artificial Neural Networks, Levenberg-Marquardt method, Discrete Wavelet Transform	More than 94% efficacy was achieved.
Sahai, A. K., et al. (2000). [9]	Data concerning midsummer & monsoon precipitation in India during 1871 through 1994 for the period of June to September.	Artificial Neural Networks	Results, the monsoon evolves as a sophisticated structure. The monsoon stands out as a dynamic system.
Hernández, E., et al. (2016).[10]	The dataset is a compilation of Statistics from the department of information in Manizales' downtown region.	Auto Encoder, MLP	According to the findings, our suggested architecture outperforms existing techniques in terms of MSE and RMSE.
Chau, K. W., et al. (2010) [11]	Between January 1st, 1988, and December 31st, 2007, rainfall depth data was studied at six rain gauge facilities located well above Yangtze research watershed.	Fuzzy C-means, Singular spectroscopic analysis, ANN, Support Vector Regression (SVR)	There appears to be a significant gain in precipitation prediction precision with their approach
Nikhil Sethi, et al. (2020) [12]	Monsoon, low clouds, mean temperature, and moisture content are utilized as indicators in research examining regional rainfall data in Udaipur, Rajasthan, India.	Multiple Linear Regressions	Rainfall can be predicted by knowing climatic factors and an accuracy of 97% is achieved.
Nikam, V. B., et al. (2013) [13]	Pune city Data in June to Nov from Indian Meteorological Department (IMD) Pune	Bayesian Prediction Algorithm	An average accuracy of above 90% was obtained.
Neelam Mishra, et al. (2018) [14]	Pune's statistics from 1871 to 2012, courtesy of the Indian Meteorological Department.	Artificial Neural Network	A Multilayer Feedforward neural network (FFNN) was made by combining the Back - propagation learning technique as well as the Levenberg-Marquardt training algorithm.
Agboola A.H., et al. (2013) [15]	The data originate from the municipality of Akure, which serves as the state capital of Ondo.	Fuzzy Logic	An exactness of 90% was achieved
Ghamariadyan, M., et al. (2020).[16]	The dataset comprises data of rainfall from Australia.	Discrete wavelet transforms, ANNs model, Coupling ANN with wavelet	DWT method achieved an effectiveness of 90%
Chatterjee, S., et al. (2018) [17]	The dataset consists of rainfall data from the South Western part of West Bengal.	Hybrid Neural Network	An average accuracy of 84.26% was achieved by this method.
Varghese, L. R., et al. (2020) [18]	Dataset contains data of rainfall of the district of Ernakulam from the year 2008 to 2018.	Regression Coefficient	The model precisely predicted rainfall.

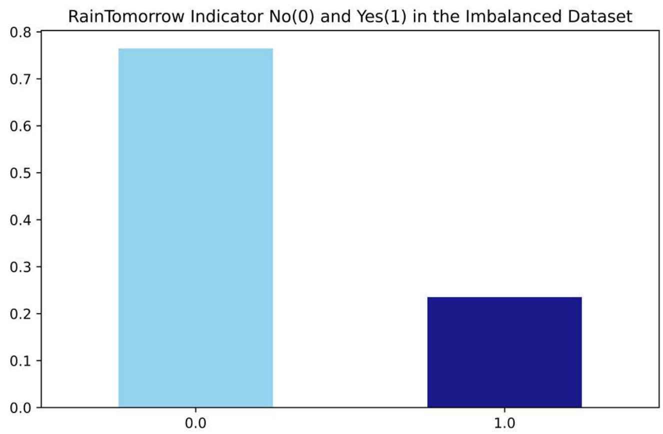

Fig. 2. Distribution of imbalanced dataset.

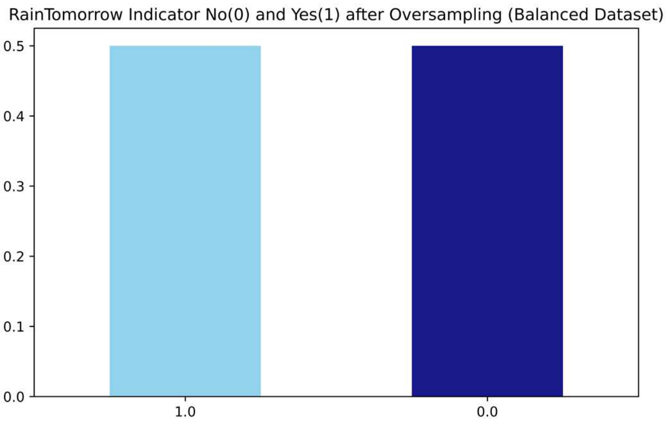

Fig. 3. Distribution of balanced dataset.

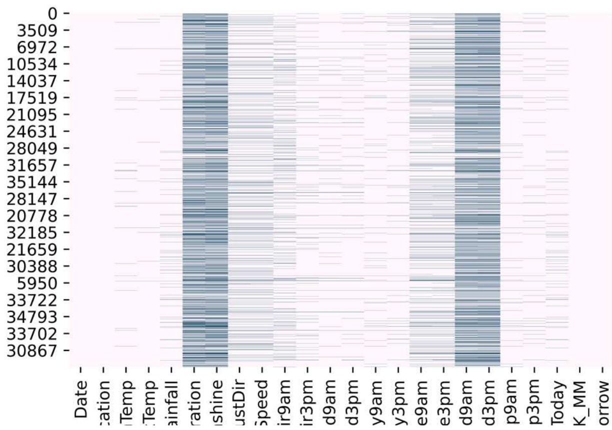

Fig. 4. Plot of missing data pattern.

After handling class imbalance, we looked for the missing data pattern in the dataset. 'Evaporation,' 'Sunshine,' 'Cloud9am,' and 'Cloud3pm' are the characteristics with the highest missing percentage. So, we generated the missing data specifics for these four characteristics. Figure 4 depicts the missing data pattern in the dataset. We see that all four characteristics have 50% missing data. Instead of rejecting them entirely, we included them in our model with correct imputation.

We have utilized mode to impute categorical columns, and then used label encoder to transform them to numeric ones. Once all of the columns in the entire dataframe have been transformed to numbers, we have used the MICE programme to fill in the missing values (Multiple Imputation by Chained Equations). Following that, we have used Inter-Quartile Range to find outliers and remove them from the final workable data set. Finally, we have examined the correlation between various weather parameters (features), and when we discover any pair of strongly linked features, we have eliminated one while retaining the other.

Further, we checked if all missing “NaN” values have been imputed correctly. As a result of MICE imputation, the data frame no longer contains any “NaN” values. Based on the Inter-Quartile Range, we then discovered and deleted outliers from the data set.

The original dataset had 87927 rows and 24 columns. The dataset now has 86065 rows and 24 columns after running the outlier-removal code snippet. As a result, the dataset is free of 1862 outliers. We now look for multi-collinearity, or whether one trait is strongly linked to another. The following pairs of features are having high correlation between them:

1. Temp9am and Temp3pm
2. Pressure9am and Pressure3pm
3. MaxTemp and MinTemp
4. Evaporation and MaxTemp
5. MaxTemp and Temp3pm

In no situation, however, is the correlation value a complete "1". As a result, no feature is being eliminated.

However, by looking at the pairplot we may learn more about pairwise correlation among these highly correlated traits. Each pairplot displays highly distinct groups of RainTomorrow "yes" and "no" responses. There is only a small amount of overlap between them.

D. Feature Extraction

We have used both wrapper method and filter method for feature selection. Description of both methods are as follows:

1. Feature Selection by Filter Method (Chi-Square Value): We must first normalize our data before proceeding. To avoid negative numbers, we're using MinMaxScaler instead of StandardScaler. In comparison to other features, we can see that "Sunshine," "Humidity9am," "Humidity3pm," "Pressure9am," and "Pressure3pm" are more important.
2. Feature Selection by Wrapper Method (Random Forest): Except for RISK MM, all feature importance’s are approaching 0, which is extremely fascinating. This can happen in one of two ways. Either when all of the features are highly correlated, or when the characteristics have extremely low relative feature importance’s in regard to the target variable. We know the first possibility isn't true because we've already plotted correlation. We'll use Permutation Importance to see if the second choice is correct.

With the exception of "RISK MM," all features have very low relative value (all close to zero) in relation to the goal variable "RainTomorrow." "RISKMM” is the quantity of rainfall in millimeters for the next day," says Joe Young, the dataset's originator. Rain, drizzle, hail, and snow are all examples of precipitation that reaches the ground. It was this column that was used to determine whether or not it rained when the binary target was created. Because it explicitly contains information about the target variable, incorporating it would leak future data to our model" (Quoted from his comment). As a result, "RISK MM" is not included in the model. "Date" is likewise left out of the model for the apparent reason that it adds no value.

E. Model Architecture

We have divided the full data set into two parts: training (75%) and testing (25%) correspondingly. We'll standardize our X train and X test data to improve our findings (i.e., features without target for training and testing data sets). We have implemented the following 7 models [19][20][21][22][23][24]:

1. Logistic Regression penalized by Lasso
2. Decision Tree
3. Neural Network (Multilayer Perceptron)
4. Random Forest
5. Light GBM
6. CatBoost
7. XGBoost

First 4 models are standard Machine Learning models and last three models are ensemble models, used for achieving higher accuracy as they will fit better to data.

IV. Result And Analysis

The comparison of evaluation metrics of all the models used are shown in Table No. 2. Figure 5, 6, 7, 8, 9, 10, 11 are area under curve and confusion matrix of Logistic Regression, Decision Tree, Neural Network, Random Forest, Light GBM, CatBoost and XGBoost respectively. We can see how different models, including the ensemble model, have distinct class borders (plotting is done considering the training data only). In comparison to the other models, CatBoost has a specific geographical border. XGBoost and Random Forest models, on the other hand, have a far lower amount of misclassified data points than other models. Figure 12 depicts class borders of all models.We must now determine which model performed the best based on accuracy, ROC AUC, Cohen's Kappa, and total execution time. A point worth mentioning here is that we could have used F1-Score instead of accuracy to measure model performance, but we had previously balanced the data set, thus using accuracy as a criterion to choose the best model was is justified. For taking a more accurate decision, we used "Cohen's Kappa," which is an excellent metric for determining the optimal model in the situation of unbalanced datasets. Let's take a look at which model has done well on which front. Figure 13 is a plot of Accuracy and Time taken for execution and Figure 14 is a plot of AUC_ROC and Cohen's Kappa.

B. Limitations

Along with recognizing the benefits of this study, it is only fair to note its shortcomings. There are multiple abrupt climate events like cloud bursts which can cause unpredicted. Sudden increase/ decrease of pressure and temperature can result in rainfall. If there is even a single error in the input values entered for prediction, result will contain incorrect output.

TABLE II. Comparison Of Evaluation Metrics.

Model	Accuracy	Roc area under curve	Cohen's kappa	Training time taken (seconds)
Logistic Regression	0.79	0.77	0.75	0.61
Decision Tree	0.87	0.87	0.73	0.12
Neural Network	0.86	0.87	0.72	88.05
Random Forest	0.93	0.94	0.87	6.61
Light GBM	0.89	0.89	0.77	2.18
CatBoost	0.94	0.94	0.88	223.78
XGBoost	0.95	0.95	0.89	70.02

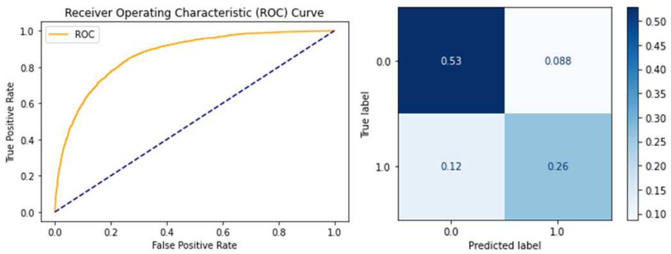

Fig. 5. AUC_ROC and Confusion matrix for LR.

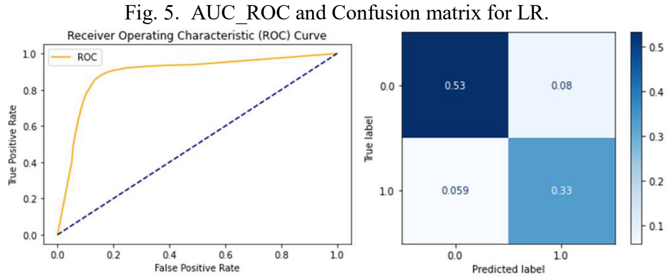

Fig. 6. AUC_ROC and Confusion matrix for DT.

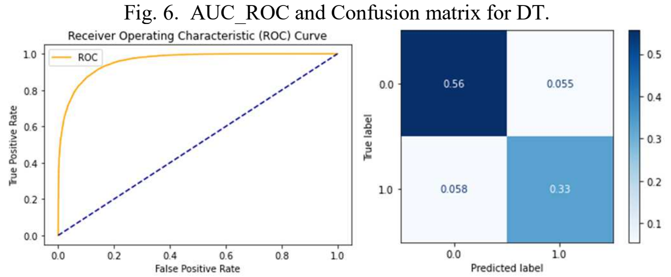

Fig. 7. AUC_ROC and Confusion matrix for NN.

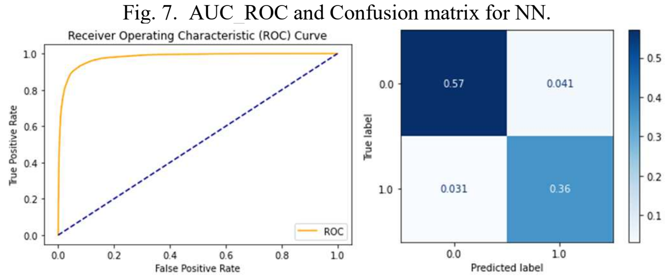

Fig. 8. AUC_ROC and Confusion matrix for RF.

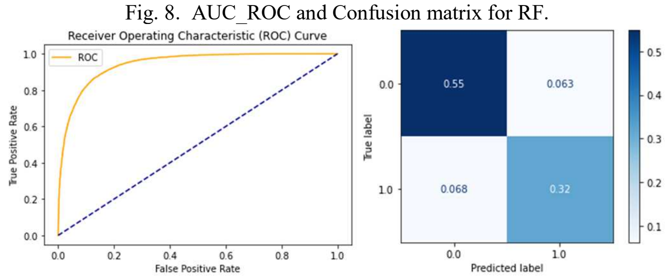

Fig. 9. AUC_ROC and Confusion matrix for Light GBM.

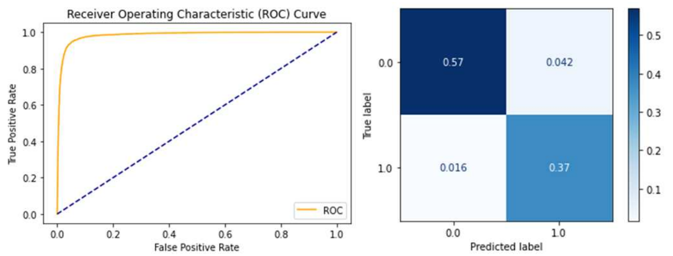

Fig. 10. AUC_ROC and Confusion matrix for Catboost.

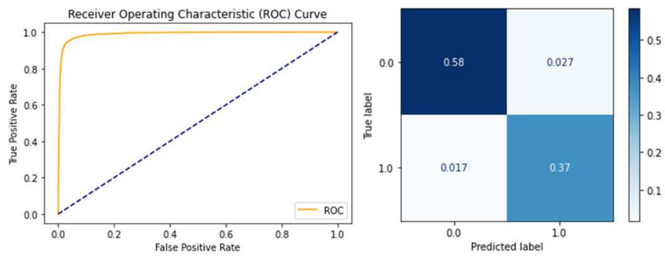

Fig. 11. AUC_ROC and Confusion matrix for XGBoost.

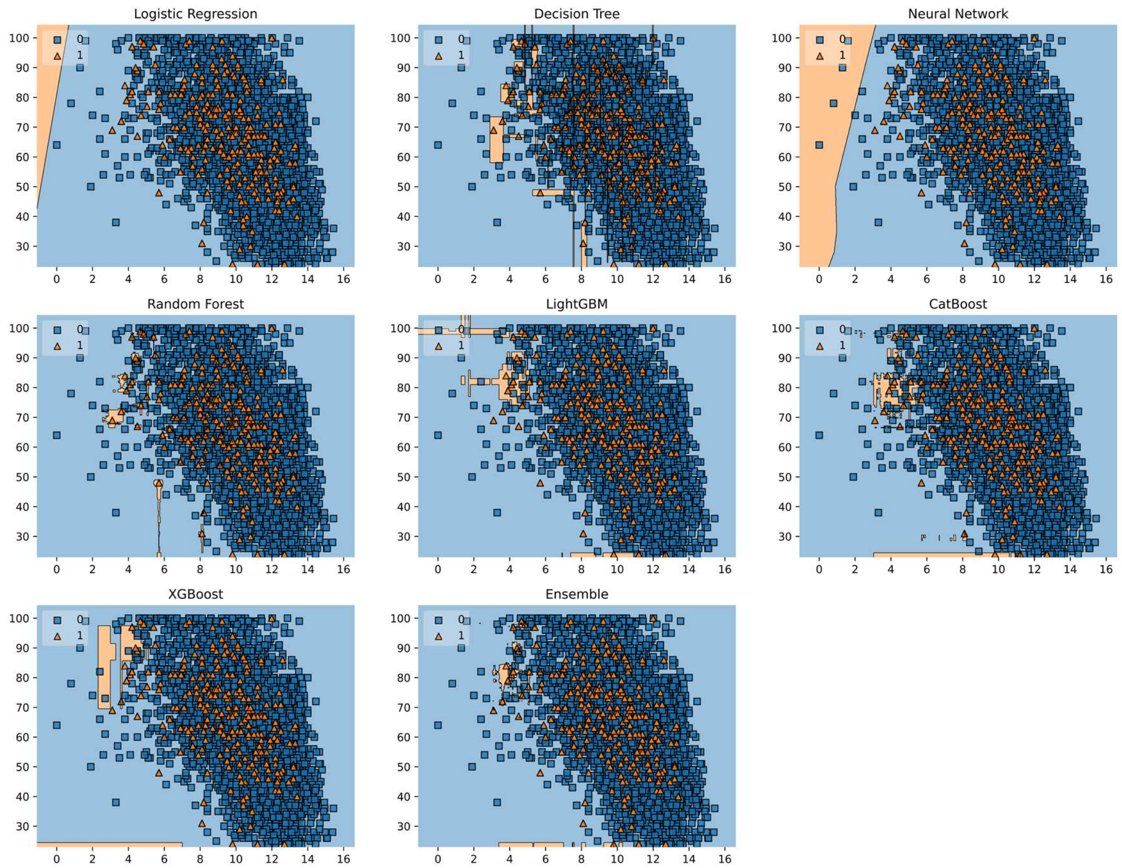

Fig. 12. Plot of class borders for all models.

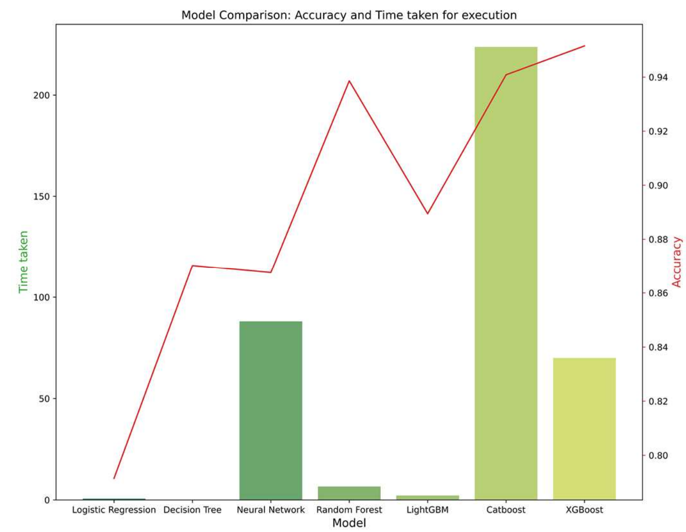

Fig. 13. Plot of Accuracy and Time taken for execution.

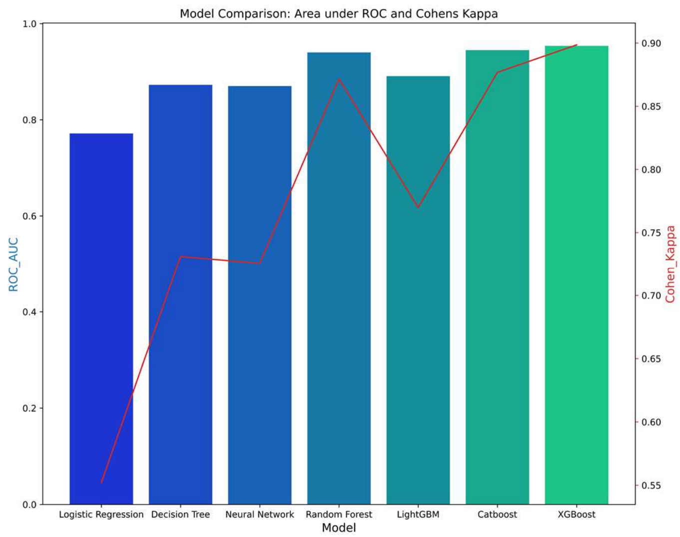

Fig. 14. Plot of Area under curve and Cohen's Kappa.

C. Future Scope

1. Including inputs from various weather satellites to find whether the inputs are similar or not
2. Using the output in prediction of other disasters like landslide, glacial lake outburst and flash floods
3. Giving automatic alerts to stakeholders if there is prediction of high rainfall.

V. Conclusion

We have proposed a solution which considers weather parameters to predict rainfall. We are taking data to test this system from weather satellites using API. It can be observed that XGBoost, CatBoost and Random Forest have performed better compared to other models. After analyzing the results, we conclude that CatBoost performs well only in the case where we use categorical variables in the data and the parameters are tuned efficiently. The only drawback of XGBoost is that it is quite slow. However, if speed is an important thing to consider, we can stick to Random Forest instead of XGBoost or CatBoost.

Acknowledgment

The work was performed on Google Colaboratory and this research would not have been possible without the publication of weather dataset by Indian Metrological Department.

Refernces

[1] Ozlem Terzi, "Monthly Rainfall Estimation Using Data-Mining Process", Hindawi Publishing Corporation Applied Computational Intelligence and Soft Computing, Volume 2012, Article ID 698071, 6 pages DOI:10.1155/2012/698071.

[2] S. Aswin, P. Geetha, and R. Vinayakumar, ''Deep learning models for the prediction of rainfall,'' in Proc. Int. Conf. Commun. Signal Process. (ICCSP), Apr. 2018, pp. 657–661.

[3] Pratap Singh Solanki, R. S. Thakur '" A Review of Literature on Water Resource Management Using Data Mining Techniques", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391 Volume 5 Issue 7, July 2016 www.ijsr.net Licensed Under Creative Commons Attribution CC BY

[4] Cramer, S., Kampouridis, M., Freitas, A. A., & Alexandridis, A. K. (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Systems with Applications, 85, 169–181.

[5] Pham, B. T., Le, L. M., Le, T.-T., Bui, K.-T. T., Le, V. M., Ly, H.-B., & Prakash, I. (2020). Development of advanced artificial intelligence models for daily rainfall prediction. Atmospheric Research, 104845.

[6] Manandhar, S., Dev, S., Lee, Y. H., Meng, Y. S., & Winkler, S. (2019). A Data-Driven Approach for Accurate Rainfall Prediction. IEEE Transactions on Geoscience and Remote Sensing, 1–9.

[7] Deepak Ranjan Nayak,Amitav Mahapatra, Pranati Mishra. A Survey on Rainfall Prediction using Artificial Neural Network. International Journal of Computer Applications (0975 – 8887) Volume 72– No.16, June 2013

[8] Venkata Ramana, R., Krishna, B., Kumar, S. R., & Pandey, N. G. Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. Water Resources Management, 27(10), 3697–3711.

[9] Sahai, A. K., Soman, M. K., & Satyan, V. (2000). All India summer monsoon rainfall prediction using an artificial neural network. Climate Dynamics, 16(4), 291–302.

[10] Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J., & Duque, N. (2016). Rainfall Prediction: A Deep Learning Approach. Hybrid Artificial Intelligent Systems, 151–162.

[11] Chau, K. W., & Wu, C. L. (2010). A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. Journal of Hydroinformatics, 12(4), 458–473.

[12] Nikhil Sethi, Dr.Kanwal Garg, Exploiting Data Mining Technique for Rainfall Prediction,) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3982-3984

[13] Nikam, V. B., & Meshram, B. B. (2013). Modeling Rainfall Prediction Using Data Mining Method: A Bayesian Approach. 2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation.

[14] Neelam Mishra, Hemant Kumar Soni, Sanjiv Sharma, A K Upadhyay,Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data,Published Online January 2018 in MECS.

[15] Agboola A.H., Gabriel A. J., Aliyu E.O., Alese B.K Development of a Fuzzy Logic Based Rainfall Prediction Model,International Journal of Engineering and Technology Volume 3 No. 4, April, 2013.

[16] Ghamariadyan, M., & Imteaz, M. A. (2020). A Wavelet Artificial Neural Network method for medium-term rainfall prediction in Queensland (Australia) and the comparisons with conventional methods. International Journal of Climatology.

[17] Chatterjee, S., Datta, B., Sen, S., Dey, N., & Debnath, N. C. (2018). Rainfall prediction using hybrid neural network approach. 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom)

[18] Varghese, L. R., & Vanitha, K. (2020). A Time-series based Prediction Analysis of Rainfall Detection. 2020 International Conference on Inventive Computation Technologies (ICICT).

[19] Abramovich, F., Grinshtein, V., & Levy, T. (2021). Multiclass Classification by Sparse Multinomial Logistic Regression. IEEE Transactions on Information Theory, 67(7), 4637–4646. doi:10.1109/tit.2021.3075137

[20] Geetha, A., & Nasira, G. M. (2014). Data mining for meteorological applications: Decision trees for modeling rainfall prediction. 2014 IEEE International Conference on Computational Intelligence and Computing Research. doi:10.1109/iccic.2014.7238481

[21] Murao, H., Nishikawa, I., Kitamura, S., Yamada, M., & Pingping Xie. (n.d.). A hybrid neural network system for the rainfall estimation using satellite imagery. Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan). doi:10.1109/ijcnn.1993.716761

[22] Suresh, N., Ramesh, N. V. K., Inthiyaz, S., Priya, P. P., Nagasowmika, K., Kumar, K. V. N. H., … Reddy, B. N. K. (2021). Crop Yield Prediction Using Random Forest Algorithm. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). doi:10.1109/icaccs51430.2021.9441

[23] Diao, L., Niu, D., Zang, Z., & Chen, C. (2019). Short-term Weather Forecast Based on Wavelet Denoising and Catboost. 2019 Chinese Control Conference (CCC). doi:10.23919/chicc.2019.8865324

[24] Liao, X., Cao, N., Li, M., & Kang, X. (2019). Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days. 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). doi:10.1109/icitbs.2019.00167