International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V70I2P204

Volume 70 Issue 2, 24-31, February, 2022

Original Article

Ada Naïve Bayesian Algorithm for Predicting the Intensity of Rain to Improve the Accuracy

S. Sakthivel[1], G. Thailambal[2]

[1]*Ph.D Research Scholar, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India*
[2]*Research Supervisor, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India*

[1]*sakthivelphd78@gmail.com*, [2]*thaila.scs@velsuniv.ac.in*

Abstract *— The Indian subcontinent is the globe's more vulnerable disaster area. A coast border for such area was approximately 7516 kilometres long, including 132 kilometres at Lakshadweep, 5400 kilometres in a major landmass, and 1900 kilometres on an Andaman. Almost 10percent of every major disaster that has developed across the globe has occurred in this territory. According to estimates from 2008 to 2021, a disaster impacted a total of over 370 thousand population across India. Throughout the pre-monsoon seasons, storm development is roughly 30percent on the Bay of Bengal & 25percent on an Arabian Ocean. Thousands of people died as a result of a hurricane, which also caused significant damage to governmental & corporate property. As a result, predicting the intensity of rain is increasingly necessary & critical. An XG Boost (severe Gradients Booster) technique is used in the next part of the study to forecast the development & intensity of rain around the Bay of Bengal, & their effectiveness was evaluated using SVM models. Several lives have been rescued as a result of great scientific growth & improvement at detecting rain far in ahead, such as Gaja & Fani, from IMD. During the last step of this study, the hybrids technique integrating a genetics algorithm and an XGBoost method (GA-XGBoost) is presented for forecasting the intensity for subtropical cyclones (TCs) at the Bay of Bengal (BoB) using data by the Indian Meteorology Department (IMD). Moreover, while comparing current methods, the outcomes of a hybrid approach for TC information are superior.*

Keywords *— AdaNaïve Bayesian, Classifiers, Disaster, Rainfall prediction.*

I. INTRODUCTION

Tornadoes were known described as "whirling rain " because they circle about a minimal core. From the dawn of the moment, natural calamities have been humanity's constant, though poorly constructed, companion. Hazardous disasters can strike unchecked & without warning, & their scale, uncertainty, frequency, and financial impact are now on the rise. Whenever such disasters occur within densely populated areas, they pose a threat to humans [1], institutions, economic assets, & to a lesser degree, to an environment.

Especially in comparison with the early 2008s, pecuniary expenditures were continuing this same rise toward alarming levels, having increased over as multiplier approaching 8. Throughout the entire planet [2], 90% of major catastrophic catastrophes, including 95% of most of those disasters, have tied with their formation for countries. This was due to this fact because almost the majority all this same country's greatest heinous disasters occur within between Tropics between summer and also Sagittarius. This area was occupied mainly this same country's poorest countries, wherein disaster-related difficulties were person due to perceived conflicts among basic human needs but also economic progress [3].

Volcanoes, hurricanes, droughts, and flooding are all particularly common across Asia's regions. Forests fires, earthquakes, including mudslides are common occurrences throughout northeastern today's Himalayas area [4]. Tragedies threaten the minimum total of 26 countries, 36 among them were territory. Such type such tragedy affects 50 billion persons around its nation each annually. Extinctions but instead earthquakes were environmental calamities that claimed numerous thousands throughout Ireland between 2008 through 2021. Through tonight's modern society, such scenarios have changed mainly part result by large combining among agricultural but also resources protection initiatives which it has decreased overall mortality impact inflicted by hunger but also floods. Strong gusts, rain, but also tremors accounted for about 98 percent of overall recorded casualties, with their percentages increasing over the previous ten decades. During 2008 through 2010, numerous disasters occurred linked throughout Asia that possesses rather significant mortality but also injury rate. Crop failures, volcanoes, rain, hurricanes, rockslides, and collapses were indeed prevalent environmental catastrophes occurring hit various areas throughout Asia on a regular basis [5].

Environmental catastrophes have a significant influence upon agriculture, irrigation, economy, cleanliness, healthcare, environmental ecosystem, especially agricultural stability every month. As such result, many emerging

countries are concerned about the current scenario. Primary but also tertiary effects include structural damages ageing affecting crops, but also infrastructures primarily result in consequence flooding various natural disasters across differing degrees [6]. Hunger has produced widespread animal although instead agricultural losses, although overall generally had little effect on economic capability or infrastructure. Rain and flooding may have a significant impact on agriculture's time cycles but also equipment. Volcanoes wreak significant harm to overall buildings but have minimal effect upon agriculture.

Asia was struck through notable catastrophic disaster after another almost every year, with a very high mortality burden and unfathomable tragedies. During this same century in 1990, when been recognized designated this same European Century towards Environmental Tragedy Prevention (2008), humanity witnessed another series of large devastating disasters that defied every attempt and contained disasters [7]. This same Latur (Maharashtra) earthquakes from 2010, which died approximately 3000 persons; these same Madhya States hurricanes both 2010 through 2011, which murdered approximately 1100 individuals annually; this same Gujarati storm in 2008, which dead over 3000 individuals; finally, this same Orissa monster storm on 2009, which dead approximately 1 million individuals. Apart from those genuine events were minor earthquakes near Bhopal, Taluks, among Chamoli, also well ongoing flooding inside the eastern northwest, Bengal, Uttarakhand, and Malabar [8]. Unfortunately, those environmental catastrophes were never regarded as teaching opportunities. Therefore, practices were never drawn from events to the same extent that they might be used to prevent subsequent calamities. Whatever happened throughout Ahmedabad before 2001, including how this was handled under international face increasing gloomy headlines, must be remembered concern improved [9].

Costs could be assessed, counted, defined, scaled by measures destroyed assets, people lost, even economic potential. Only someone, several of these same falls which occurred as a result of calamity events, could be accurately quantified. Absence from the sanctuary was caused by a variety of destitute factors [10], including disasters, severe shortage for sustenance, difficulties, and permanent absence from employment, accessibility, among macroeconomic activity. Many any losses incurred are a direct result as such effects were covered by catastrophe protection for assistance. This same creation of explicit prevention solutions was maintained separate from this same investigation but also reporting concerning sociological but also physiological components that were not related to these same socioeconomic components [11].

II. RELATED WORKS

Numerous agronomic advising methods have been developed throughout the research industry, helping improve the overall agricultural ecosystem with associated plants [12]. Most simulations that have been constructed thus much have been created having scholastic industry scientific considerations in mind. Furthermore, most created algorithms focused upon both any certain commodity but rather any specific watering challenge. However, within the course of each season, many crops could be grown upon any similar piece of ground. As a result, smaller producers have an extremely strong demand for easy ruling instruments. Manufacturers should examine this designer's accessibility while designing such economic strategy such, even producers using small pieces of agricultural ground would profit effectively [13].

Establishing one robust firewall among demands with statistics was another significant problem throughout trying that accomplish this same intended result. Under this current environment, scientists were mostly focused upon technical, analytical capabilities rather than their producers' perspective; nevertheless, using photographs shot using cell devices with whatever limited information supplied from growers was insufficient, enabling developing advanced metrics [14]. As an agricultural result, a more solid link connecting those groups was required to develop effective plantation information analysis tools. Because many simulations were harvested, therefore primarily concentrate upon economic benefits derived from overall growth on any single commodity, another concept was for combine 2, but rather many tiny simulations such enable them may communicate but instead being employed within non - linear even when non-ecosystems.

Severe precipitation but instead environmental changes have a very significant influence upon agriculture, which have a significant macroeconomic effect upon every country. On such a result, defining baseline criteria across various crops farmed within various places was necessary when constructing such simulation [15].

Agricultural production may likewise be boosted through employing technologically engineered elevated seedlings. However, whenever working there with these seedlings, someone must take into account both genetic markers for this plant also much its reaction towards external factors when making predictions. Additional elements that affect production include ground nitrogen but also moisture levels, as well as vegetation but also associated treatment. Statistics regarding precise farming may be gathered through even any variety of resources. Thus it's mostly unorganized. Devices may capture agriculture information such as warmth, ground humidity, and soils mineral concentration. Meteorological sensors, by this same contrary extreme, may gather precipitation, thermal irradiation, even air temperature. Records, such example harvest yields throughout previous decades, may become obtained through governmental institutions, economic groups, and other web pages through comprehensive statistically annual analysis. Several systematic studies on sediment category cluster analysis it has been introduced.

Ankita Patil's presentation focused on addressing crop management within arid locations. These devices were used to capture data such as heat, ground humidity, and soils PH, which could then be evaluated to enhance agricultural yield. Monthly precipitation patterns within its appropriate regional region were predicted using a conventional logistic model using an artificial networking approach. This information was taken through Hyderabad, Today's meteorological stations, where its monthly highest but instead lowest temperatures, moisture, and precipitation over its previous ten decades was investigated. Using the use of logistic extrapolation using artificial neural networks (ANN) can assist upcoming forecast precipitation better correctly [16].

Devices being utilized can detect ground humidity gathered using known average humidity values in each specific flower, allowing its quantity required irrigation each certain facility would being tracked using combined mix involving technology but also application. Their research was described through one location approach towards information treatment, during how information was separated between pairs, then another KNN technique was performed over these pairs that maintain comparable patterns enabling predictions. Another concept regarding any sensing architecture with agricultural equipment. Information was acquired employing cordless sensing but instead saved within digital internet service towards such purpose. Producers will be able to receive agriculture statistics through a smartphone application that will be launched at this time. Those devices were linked through IoT gates, which in turn became linked with another IoT business platform. Through these provider servers, various devices may be visually displayed but also watched, while their information can be retrieved for analysis.

Aside from that, a slew of tools and applications have been developed to aid farmers in their work. Certain storing architectures connected with agriculture characteristics, for its alternative extreme, have successfully been implemented under such special architecture. Under that, the scientists created that universal template enabling commercial computers can handle information with whatever format, i.e., diverse heterogeneity networked systems. Several efforts involving agricultural forecasting depending on various approaches have been investigated throughout any number of various methods. This research uses 4 vegetative spectrum indicators. Their study developed growing neuronal networks that anticipate wheat output utilizing four different vegetation markers: NDVI, GVI, SAVI, and PVI. This research implies whether integrating geographical with quantitative methodologies might help estimate agricultural yields. Utilizing past information, another statistics framework for agriculture was suggested to determine the potential impacts of climatic warming upon agricultural output. This statistics approach indicates overall agricultural reaction following changes between average moisture and average temperatures. The impact between worldwide heating but instead greenhouse gases upon 6 among the world's greatest popular commodities.

Another temperature prediction system regarding agricultural production across Europe was developed utilizing atmospheric information that indicates acceptable happenings. To determine overall effects from weather upon crops, this technique offers a very detailed examination using overall weather, freshwater, mixed agricultural production models. For Plains, another comparable quantitative model-based research concerning precise agriculture was provided. Another agricultural system was proposed that uses consecutive information integration to estimate yields from cornfield crops utilizing electronically obtained information regarding ground wetness but also green surface indices. About appropriate agricultural development environment but instead, production from Polish was explored using the AVHRR indicators concept. Another ensembles convolution filter-based information integration method was proposed that addresses inherent inaccuracy from overall agricultural production estimate caused by ambiguity with periodic precipitation dispersion. This usage of chloroplast concentration within its green size indices because primary photosynthetic activities marker was discussed throughout order that determines agricultural production [17].

This same influence between harsh winters and also summertime climate circumstances upon Mediterranean crops is discussed to recommend an appropriate annual harvest. Each of these books listed below demonstrates different methods for predicting that viable harvest depends upon its respective predictions. Throughout these papers, they present an alternative massively simultaneous software development language called Colonnaded Bayesian Hadoop capabilities offers one novel approach that codifies overall predicting procedure quicker. This operates throughout tandem using said inherent storing capabilities can save but instead recover each crop's relevant information [18].

III. ANALYSIS

A. Back Propagation Network (BPN)

Another among several categorization methods that may generate any judgment trees given any given information collection was called a choice forest. This same principle, called informational volatility, was applied throughout such a method. This same greatest feasible characteristic was chosen allows any group more observations can become effectively segregated. These characteristic separation parameters were chosen depending upon a balanced informational benefit or very significant probability differential for this specific component. Another judgment tree was created containing networks depending upon that criterion. These various characteristics called terminal characteristics provided every intermediate component were used to make these concluding conclusions regarding their

challenge. The overall assessment procedure using generating BPN method including various methods using that trained dataset including bagged. Table 1 shows empirical ambiguity matrices using BPN, which shows overall reliability for every category from given identical classification with particular.

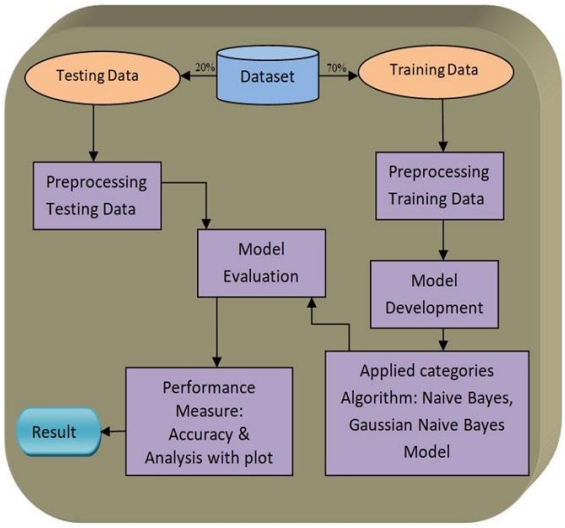


Fig. 1 Proposed Work

For every variable, categorization & weighing on category data were performed at the stage. Every variable was divided into 3 groups by an investigator (Table 1). Both lowest & highest readings of every variable were used to calculate weighted category ratings, which were dependent on the information which has been pre-processed.

Table 1. Categories & Weighting of class values of each parameter

Parameter	Class	Class Value
Average temperature (Celsius)	High Medium Low	> 27.1 > 25.2 - ≤ 27.1 ≤ 25.2
Average humidity (Percentage)	Fairy Damp Damp Very Damp	≤ 83.8 > 83.8 - ≤ 92 > 92
Sun exposure duration (hr)	Long Medium Short	≥ 7.2 ≥ 3.5 - < 7.2 < 3.5
Average wind speed (m/s)	Fast Medium Slow	≥ 2.4 ≥ 1.5 - < 2.4 < 1.5

B. Naïve Bayesian (NB)

A Naïve Bayesian technique was regarded as one of the best classifiers with greater information. An NB technique combines branch predictions like randomized vectors elements, which were dependent on every branch measured independently & have an identical dispersion of every branch within a forest. Because NB might be substantial with a high quantity of branches, forestry converging has a generalization mistake to the limitation. For the generalization errors or rather connection among a branch analyzer, the specific branch power was applied. These subgroups were chosen by this coaching information, which includes every tree being taught by the substitute. A remainder of coaching information is used to evaluate the relevance between parameters & mistakes. In residuals & another tree, several choices were applied to designate the category that was determined with averaged outcomes. The assessment procedure, reliability, & confused matrices for an NB method are detailed in Tables 2,3 and 4.

Table 2. The efficiency of the proposed method compared with an existing system

BPN:	
No. of instances correctly classified 460	74.21%
No. of instances incorrectly classified 170	25.9%
Kappa Statistics	0.42
Mean Absolute Error	0.33
Root Mean Squared Error	0.39
Total No. of instances	628
Naïve Bayesian algorithm:	
No. of instances correctly classified 547	86.8%
No. of instances incorrectly classified 83	13.01%
Kappa Coefficient	0.738
Mean Absolute Error	0.18
Root Mean Squared Error	0.298
Total No. of instances	628
AdaNaive Bayes algorithm:	
No. of instances correctly classified 346	53.81%
No. of instances incorrectly classified 285	46.17%
Kappa Statistics	0.1062
Mean Absolute Error	0.4697
Root Mean Squared Error	0.5134
Total No. of instances	628

Table 3. Confusing vector proposed method compared with an existing system

BPN Algorithm	x=24179	y=82208	Classified as x=yes, y=no
Naive Bayes Algorithm	a=25142	b=39293	Classified as a=yes, b=no
AdaNaive Bayes Algorithm	x=21289	y=9137	Classified as x=yes, y=no

Table 4 Reliability specifications of the proposed method compared with an existing system

BPN Algorithm

Performance metrics	TPR	FPR	F-measure	Recall	Precision	ROC area	Class
Average Weight	0.71 0.68 0.698	0.31 0.27 0.274	0.764 0.731 0.698	0.71 0.68 0.698	0.741 0.732 0.698	0.83 0.83 0.812	Yes No

Naive Bayes Algorithm

Performance metrics	TPR	FPR	F measure	Recall	Precision	ROC area	Class
Average Weight	0.86 0.91 0.90	0.12 0.16 0.15	0.88 0.89 0.875	0.87 0.91 0. 90	0.88 0.90 0.875	0.94 0.95 0.95	No Yes

AdaNaive Bayes Algorithm

Performance metrics	TPR	FPR	F measure	Recall	Precision	ROC area	Class
Average Weight	0.8 0.42 0.561	0.61 0.4 0.441	0.8 0.42 0.561	0.542 0.608 0.578	0.601 0.492 0.542	0.572 0.572 0.572	No Yes

C. Naïve Bayes Algorithm

A Naive Bayesian method was included within a possibility classification group, & Baye's theory reveals the notion. Within a possible categorization set, the Naive Bayes approach is used, & Baye's concept explains a concept. With the premise that factors were completely free on one another, a classifier was applied to calculate a criterion for category probabilities. For developing a classifiers system with the WEKA tools, judgment rules & possibility structure are integrated:

$$q(D_x|i_1,\ldots,i_n) = \frac{1}{P} q(D_x) \prod_{y=1}^{h} q(j_y|D_x) \quad (1)$$

$$= averagemaxq(D_x) \prod_{y=1}^{h} q(j_y|D_x) x \ni |1,\ldots\ldots h| \quad (2)$$

Table 5. Comparison of algorithms

Description	BPN Algorithm	Naïve Bayes	Random Forest
Classification of correct instances	462	347	552
Classification of correct instances	169	281	79
Time taken	0.14	0.23	0.27
Kappa Statistics	0.5015	0.1864	0.742
Mean Absolute Error	0.3764	0.4817	0.187
Root Mean Squared Error	0.4346	0.5164	0.322
F-measure	0.744	0.540	0.90
Precision	0.744	0.567	0.90

Eq. (1) describes a conditioned dispersion, with 'Q' being a scalability component & 'C' denoting a category element.

Eq. (2) represents a creation for the Bayes classifiers with the probabilistic concept. An effectiveness assessment, reliability statistics, & a confused vector of a Ada Naive Bayes method are presented.

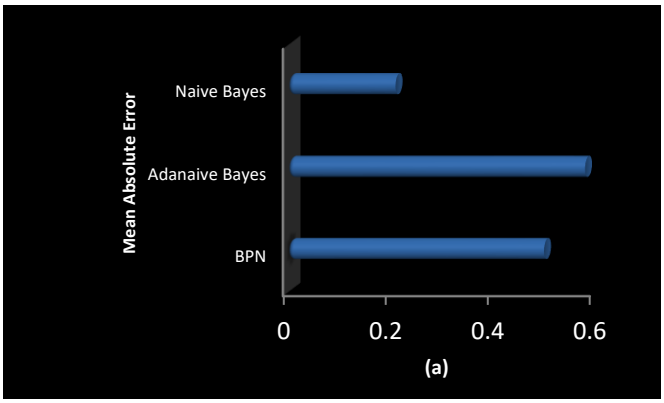
As a result of an efficiency evaluation, a deciding tree C4.5 (J48) method was demonstrated to have more suitable for usage in a subsequent development phase of studying climate information. During such information selecting stage, the information which was before applied a discoverable filter, which comprises 8 variables overall, with the highest moisture being eliminated since it includes sound that might damage the suggested program's quality.

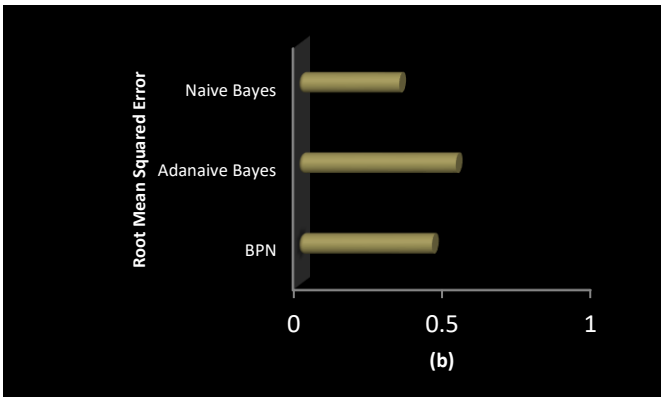
IV. RESULTS AND DISCUSSION

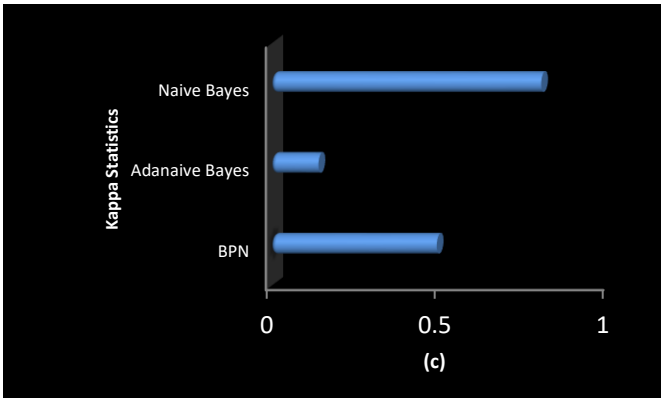
The efficiency assessment of choice tree, Naive Bayesian, &randomized forests methods was given in Table 5, & these were all shown in Fig. 2. As compared to decisions tree BPN & Naive Bayesian methods, an AdaNaive Bayes method produces higher results. This evaluation was dependent upon 7 different factors that the methods apply. A deployment of BPN to a climate information collection was the following phase. This procedure was presented as the intake resource of processed information, & an intended result was retrieved as the decisions tree. This BPN has a group of vertices, & every node may have a variety of characteristics, which was a criterion of network breaking during a categorization procedure.

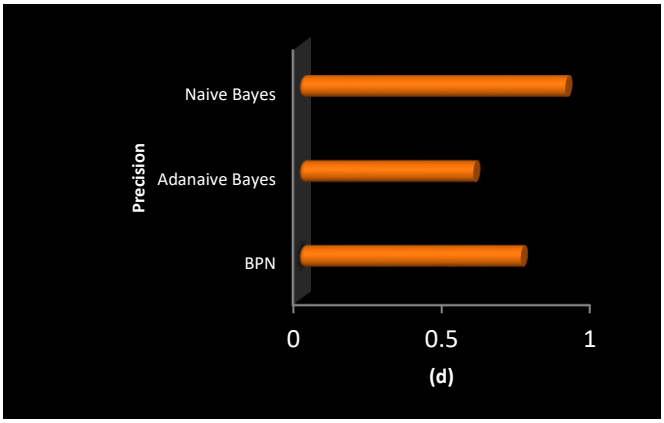
A bagged & ensembled approach, as well as a BPN, is applied to improve the efficiency of a suggested method, which is selected as such ultimate applications.BPN & Naive Bayesian. That period it takes to acquire the precision of Ada Naive Bayes, on the one hand, was a bit longer than the alternative methods.

It's also worth noting that increasing the characteristics again for provided information collection, which includes the guidelines lowest temperature, highest temp, average temp, highest moisture, least moisture, mean moisture, air velocity, rainfall, & fog, has an impact on the effectiveness of Naive Bayesian & BPN, but it is treated subtly by the Ada Naive Bayes method. As a result, while working on the climate information sets, an Ada Naive Bayes technique was preferable. Moreover, the approach could equalize an imbalanced information collection category populations mistake. This confused matrix was used to do an efficient evaluation of a meteorological information group. As just a consequence, the achievement of the Ada Naive Bayes classification would be stronger for such a huge quantity of information to recognize climate, & this offers a foundation for understanding the influential variable for climate transformations, namely, temperature (Min & Max), which, if this rises, causes deviations to establish fog shield and rain. By including the variable temperatures within the water, i.e., sea surface temperatures (SST), we may go much farther in forecasting hurricanes.









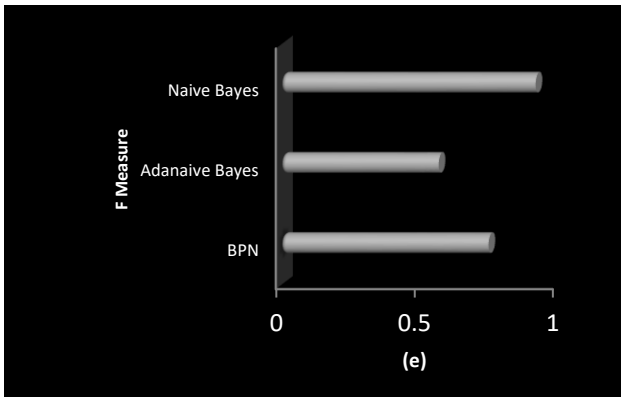


Fig 2. Proposed methods with existing systems based on performance metrics

Information & datasets were 2 sorts of information that were employed as trial substances. The information outcome on a well before stage using 329 information was called data-training. Information, on either hand, was the mixture for classes for every variable, which was never included within knowledge. In data-testing, the study employed variable category variations. Table 6 demonstrates the (1) the greatest parameter value of rain severity has been through the minimal category, indicating that rain severity seems to be lesser for situation frequency 1; (2) The greatest efficiency value of rain severity was within a lower group, indicating as a rain severity of criterion item 2 was weak; (3) The greatest efficiency value of rain severity was within a lower group, indicating as rain severity was lower in criterion item 3; (4) A maximum efficiency value of rain severity was within a lower group, indicating the rain severity was lesser under requirement item 4;& (5) the greatest efficiency value of rain severity was within a lower category, indicating that a rain severity of criterion item 5 was lesser.

RMSE, MSE and MAE of the proposed system with existing work are compared as shown in Fig. 3 to Fig. 5.

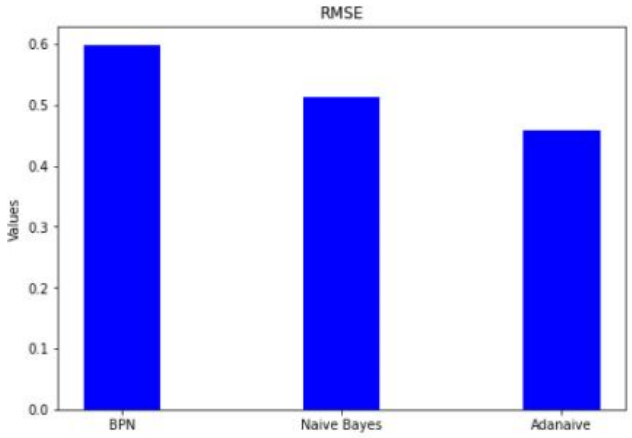


Fig 3. RMSE value of a proposed system with existing work

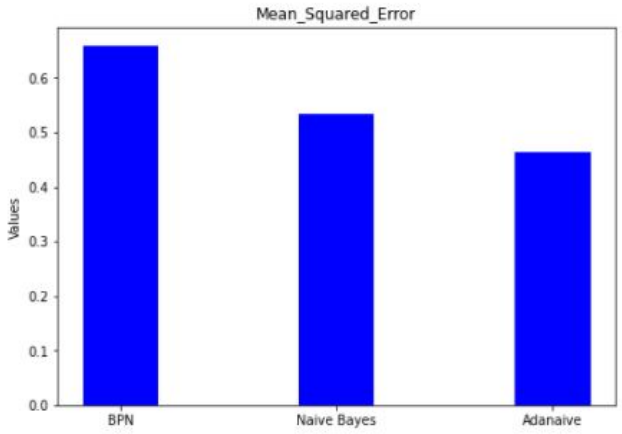


Fig 4. MSE value of a proposed system with existing work

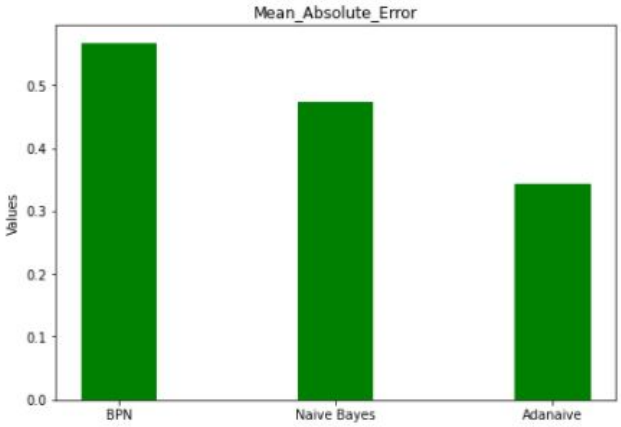


Fig 5. MAE value of a proposed system with existing work

TABLE 6. Rainfall intensity using the proposed system

Parameters					Entropy of Rainfall Intensity		
Number	Averg. Temperature	Averg. Humidity	Sun Exposure duration	Averg. Wind Speed	High	Medium	Low
1	High	Very	Long	Slow	0.000	0.000	0.001
2	Low	Damp	Long	Slow	10	20	48
3	Medium	Damp	Short	Slow	0.001	0.000	0.001
4	High	Fairy	Short	Slow	28	75	61
5	Low	Damp	Short	Medium	0.000	0.002	0.012
		Very			29	01	41
		Damp			0.000	0.000	0.005
		Damp			12	97	93
					0.002	0.005	0.008
					72	96	08

V. CONCLUSION

Inadequate estimate, the classifying structure was employed in that all incoming information was preprocessed, sanitized, & balanced before a classifying procedure. Some chosen information specimens were moved to the spreadsheets document for additional treatment to make them acceptable to information extraction methods. Every

information collection was balanced to reduce the impact of scale upon information & stored in CVS document formats. According to a finding, the applied classifications techniques worked effectively. However, the strongest information extraction methodology for predicting rain was ada Naive Bayes, which performs in 91.57 percent. When studying various categorization methodologies & climate aspects upon distinct meteorological information at various sections of districts, various projections of next research may be made.

REFERENCES

[1] Kalimuthu, M., Vaishnavi, P., & Kishore, M. Crop prediction using machine learning. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), (2020) 926-932.

[2] Karthick, S., Malathi, D., Sudarsan, J. S., & Nithiyanantham, S. Performance, evaluation and prediction of weather and cyclone categorization using various algorithms. Modeling Earth Systems and Environment, (2020) 1-9.

[3] Fayaz, S. A., Zaman, M., & Butt, M. A. Knowledge Discovery in Geographical Sciences—A Systematic Survey of Various Machine Learning Algorithms for Rainfall Prediction. In International Conference on Innovative Computing and Communications (2022) 593-608.

[4] Gupta, R., Sharma, A. K., Garg, O., Modi, K., Kasim, S., Baharum, Z., ... & Mostafa, S. A. WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics. IEEE Access, 9 (2021) 137869-137885.

[5] Shah, B., Thapa, S., Diyali, R. S., HK, S., & Maharjan, S. Rain Prediction Using Polynomial Regression for the Field of Agriculture Prediction for Karnataka. (2020), Available at SSRN 3635278.

[6] Raval, M., Sivashanmugam, P., Pham, V., Gohel, H., Kaushik, A., & Wan, Y. Automated predictive analytics tool for rainfall forecasting. Scientific Reports, 11(1) (2021) 1-13.

[7] Pham, B. T., Phong, T. V., Nguyen, H. D., Qi, C., Al-Ansari, N., Amini, A., ... & Tien Bui, D. A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve Bayes, and logistic model tree for flash flood susceptibility mapping. Water, 12(1) (2020) 239.

[8] Setiadi, T.,Noviyanto, F., Hardianto, H., Tarmuji, A., Fadlil, A., & Wibowo, M. Implementation of naïve bayes method in food crops planting recommendation. International Journal of Scientific and Technology Research, 9(2) (2020) 4750-4755.

[9] Singh, T. P., Nandimath, P., Kumbhar, V., Das, S., & Barne, P. Drought risk assessment and prediction using artificial intelligence over the southern Maharashtra state of India. Modeling Earth Systems and Environment, (2020) 1-9.

[10] Aiyelokun, O., Ogunsanwo, G., Ojelabi, A., & Agbede, O. Gaussian Naïve Bayes Classification Algorithm for Drought and Flood Risk Reduction. In Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation, (2021) 49-62.

[11] Murad, S. H., Mohammed, Y., & Salih, M. Comparable Investigation for Rainfall Forecasting using Different Data Mining Approaches in Sulaymaniyah City in Iraq. International Journal, 4(1) (2020) 11-18.

[12] Ganachari, P., & Vijetha, R. K. Machine Learning-Based Rainfall Analysis, (2021).

[13] Pham, B. T., Jaafari, A., Nguyen-Thoi, T., Van Phong, T., Nguyen, H. D., Satyam, N., ... & Prakash, I. Ensemble machine learning models based on Reduced Error Pruning Tree for prediction of rainfall-induced landslides. International Journal of Digital Earth, 14(5) (2021) 575-596.

[14] Anwar, M. T., Hadikurniawati, W., Winarno, E., & Widiyatmoko, W. Performance Comparison of Data Mining Techniques for Rain Prediction Models in Indonesia. In 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (2020) 83-88.

[15] Parashar, S., & Hurra, T. A Study on Prediction of Rainfall Using different Data mining Techniques, (2020).

[16] Rajkumar, K. V., & Subrahmanyam, K. ANALYSIS OF PRINCIPAL COMPONENTS AND CLASSIFICATION ENHANCEMENT FOR RAINFALL PREDICTION, (2020).

[17] Tamil Selvi, M., & Jaison, B. Lemuria: A Novel Future Crop Prediction Algorithm Using Data Mining. The Computer Journal. (2020).

[18] Jaison, B. Ada Lemuria: A progressive future crop prediction algorithm using data mining. Sustainable Computing: Informatics and Systems, (2021) 100577.