IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 9 (September. 2018), ||V (I) || PP 01-10

Integrated Design For Energy Generation And Weather Forecasting Using Different Base Learners on Wind & Rain Dataset

Mr. Sameer Kaul , Roheela Amin, Er. Sushma Gupta,
Dr. Majid Zaman Dr. Muheet Ahmed
Corresponding Author: Mr. Sameer Kaul

Abstract: Weather datasets areenriched with indispensable information which isconcealedwithin the data. This information can be discovered using various machine learning algorithms and thereby, can be prolific in decision making and weather prediction. Forecasting weather conditions is considered to be an imperative application of meteorology which has been an extremely challenging problem round the globe. There are numerous data mining techniques that have been employed in the realm of forecasting atmospheric conditions. The meteorological associations deploy high intensive computing machines to execute weather forecasting model. Therefore to address this subject matter, we proposed and implemented diverse models based on application of different mining methods. However, in this paper we have applied four data mining methods viz. j48, naive bayes, IB1 and bayes netacross our dataset to formulate a successful model for predicting annual rainfall pertaining to Jammu and Kashmir. Moreover, when these different models supplied with the unknown class labels were tested, j48 among other models achieved an outstanding accuracy of 97.14% in predicting the annual rainfall.

Keywords: Data Mining, Classification, Naive Bayes, J48, Bayes Net, IB1

Date of Submission: 26-08-2018 Date of acceptance: 07-09-2018

I. INTRODUCTION

Forecasting weather conditions in advance for an explicit region is a significant application of science and technology. From the past decade, massive data pertaining to meteorological databases have been swelling in magnitude.These databases comprise of valuable and prolific information that can be extremely useful in forecasting the atmospheric conditions. Moreover, to examine such enormous data is impracticable for researchers and the stakeholders. Consequently, it becomes imperative to employ techniques such as data mining and knowledge discovery to unearth hidden patterns from meteorological data sources, so as to formulate a better forecasting strategy [1, 2, 3].

Data mining is an interdisciplinary field that can be applied across multiple areas including academics, business, medicine, marketing, production and, so on and so forth. The atmospheric conditions can be forecasted using various data mining techniques and algorithms. In addition,the meteorological prediction can serve multiple purposes in different walks of life [4]. Furthermore, accurate weather forecasting can be indispensable in saving time, capital and more positively lives of human beings [5].

Weather forecasting has been largely classified into various categories of prediction viz. current cast, small range, intermediate range and extendedrange. The current cast typically makes predictions up to few hours. However, methods such as smalland intermediate ranges can make weather predictions up to 3 days and 7 days respectively. In addition, there has not been drawn any stringent line on the number of days to be predicted in extended category. Therefore, as a subject matter of consequence, it permits this specific range to forecast from one month to a seasonal prediction [6].

The contemporary prediction system entails a supercomputing mechanism to predict climatic services related to several demographic provinces. The Indian meteorological department has persistently been enlarging and boosting its infrastructure for meteorological inspection. From its inception, Indian meteorological department has mutually fostered the development of climatic and atmospheric science at various places in India. Furthermore, there is an assortment of climatic research and forecasting models viz. Global data forecasting model, Generic forecasting model and Seasonal weather forecasting models utilized for predicting climatic conditions. Conversely, data mining models are subjected to historical data for making predictions, and various models are based on concepts of probability, induction, similarity patterns and so on. The model works in identical approach for entire prediction categories to deliver reasonable accuracy [7].

The Indian atmospheric department has initiated various measures to gather climatic observations,and in this realm several telemetry instruments have been installed to examine data linked to several areas such as agricultural harvest, phases of their development, and impact of pathogenic constituent's on them. Moreover, analysing the customary stacking data can be a challenging task. To process this mounting data call on researchers to employ data mining techniques, so as to explore the recent trends or patterns hidden in climatic data sources. Therefore, in this study we have focused on applying various mining techniques to uncover the hidden patterns from climatic background pertaining to Kashmir, and subsequently attempts will be made to forecast the annual rainfall of various regions including north, south and central Kashmir. Furthermore, our real dataset was subjected to procedures such as testing and training on 10 folds cross validation to corroborate its accuracy in forecasting the annual rainfall.

II. LITERATURE REVIEW

Data mining is the method of extracting implicitly hidden patterns from massive data repositories [8]. Data mining techniques have been applied in diverse areas viz. academics, prognosis in medicine, genetic engineering, meteorological casting, agriculture, and so on[9,10]. From the last decade, a number of thriving research attempts has been conducted in this direction, to improvise weather forecasting after application of various data mining techniques.

Petre (2009), conducted a research wherein decision tree was employed to predict atmospheric conditions. The researcher applied decision tree across its dataset for its straightforward understanding and simple demonstration. In addition, the model used in this study achieved a significant degree of accuracy to forecast the average temperature of the month, and successful experimental results were implemented in WEKA machine learning tool [11].

Kumar (2013), made distinction among several classifiers including neural networks, decision trees, naive bayes, rule based, bayesian belief networks, memory based analysis and support vector machines. In this study, the researcher also exposes prominent categories of decision trees such as iterative dichotomiser 3 (ID3), descendant of ID3 (C4.5) and chi-squared automatic interaction detector (CHIAD). Eventually, he puts forth notable observations of his revision on dependent variables rain and fog, and arrives to a conclusion that decision tree can be very prolific in forecasting and exploring the potential of large repositories of data [12].

Holmes (2008), applied decision tree and neural networks with the purpose to identify considerable elements responsible for stern storm,and subsequently corroborate variables and results which have previously been recognised by meteorological departments from radar datasets. The decision tree was evaluated on information gain to determine the split criteria for the nodes at each level. Furthermore, to prevent over-fitting of data, procedures such as pruning was applied to eliminate feeble nodes in tree[13].

Khan and Hayat (2014), employed several data mining methods viz. decision tree, k nearest neighbour and naive bayes, to forecast climatic states based on existing dataset. The researchers observed that decision tree attained promising results in weather forecasting in contrast to other data mining models. The decision tree achieved notable classification accuracy of 82.62% in predicting the correct instances from the resultant class[14].

A research team comprising of eminent researchers including Dalibor, Milan, Slavisa, Shahaboddin, and Shervin(2015) conducted a significant study, wherein they applied adaptive neuro fuzzy inference system (ANFIS) to examine predominant climatic factors on vaporisation from herbs and soil surfaces[15]. The researchers collected data from 12 meteorological stations in Serbia over a period of 1980 to 2010 and this data was applied as input to obtain noteworthy results. Moreover, the parameters that were taken under examination comprised of air temperature, wind speed, relative humidity, factual vapour pressure and daylight hours. The ANFIS model formulated, assisted in identifying patterns and acquiring significant change in environment outcomes. The model was further evaluated on various performance estimates such as pearson correlation coefficient, root mean squared error, and coefficient of determination to attain better prediction outputs.

Jillella's (2015)research was based on two conceptual techniques such as artificial neural networks and exploration techniques[16]. The study was conducted over a period from 2001 to 2005, in which artificial neural networks produced exceptional prediction accuracy of 95% in forecasting the environmental conditions. Another researcher who also contributed in the realm of weather forecasting using techniques such as neural networks included Viswambari(2014)[17]. In this study classification was chiefly achieved by means of backward propagation method.

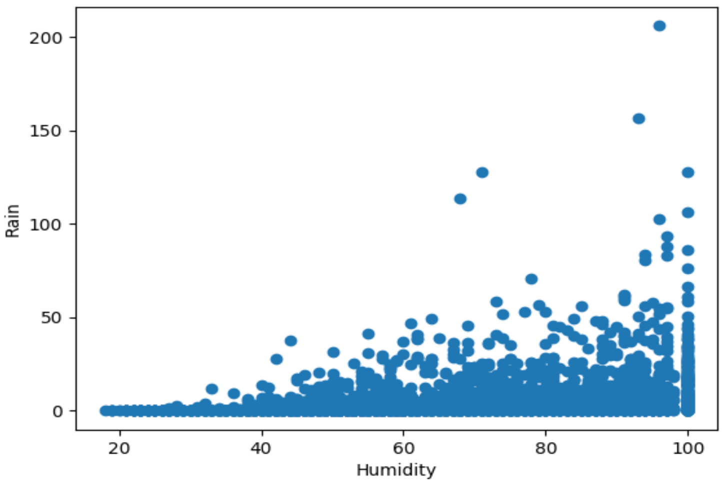
Kotsiantis et al. (2007) performed significant study, wherein the prominent researchers deployed six data mining techniques viz. k-nearest neighbour, m5 rules, decision tree, linear least squares regression, feed forward back propagation and instance based learning, on a dataset comprising of attributes humidity, temperature and rainfall [18]. The dataset consisted of relevant information pertaining to Patras city of Greece over a period of 4 years from 2002 to 2005. Moreover, they successfully predicted minimum and maximum temperature using diverse mining techniques.

Kalyankar and Alaspurkar (2013),applied clustering technique on weather dataset pertaining to Gaza. The generated results were then analysed and subsequently visualised to comprehend the factors responsible for accurate weather forecasting [19]. Furthermore, the study reviews several techniques of clustering to discover other means of improving results in the realm of weather forecasting.

From the past literature, it is noticeable that indispensible work has been accomplished,and most extensive techniques that have been deployed in forecastingweather conditions comprise of neural networks, decision tree and clustering methods. However, it becomes imperative on researchers to explore other possible learning algorithms in predicting the atmospheric conditions.

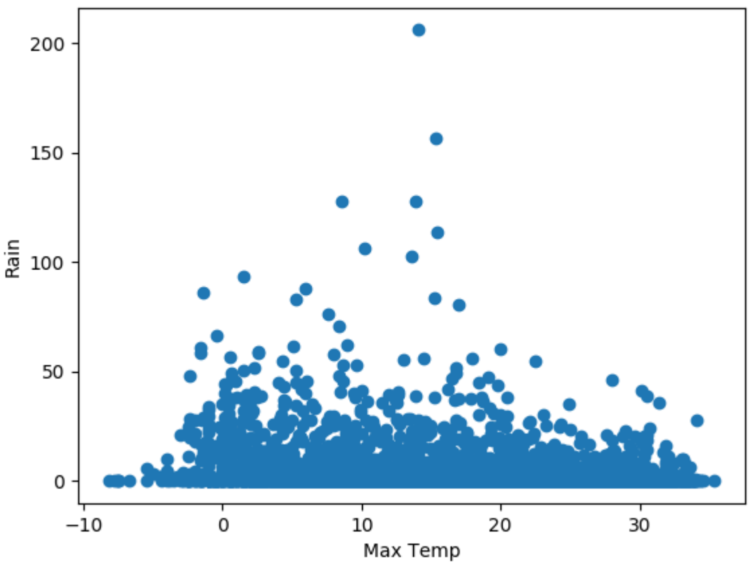
III. RAIN DATA ANALYSIS

Given below are images of data analysis done on various parameter of rain data



Above image depicts relationship between rainfall and humidity.

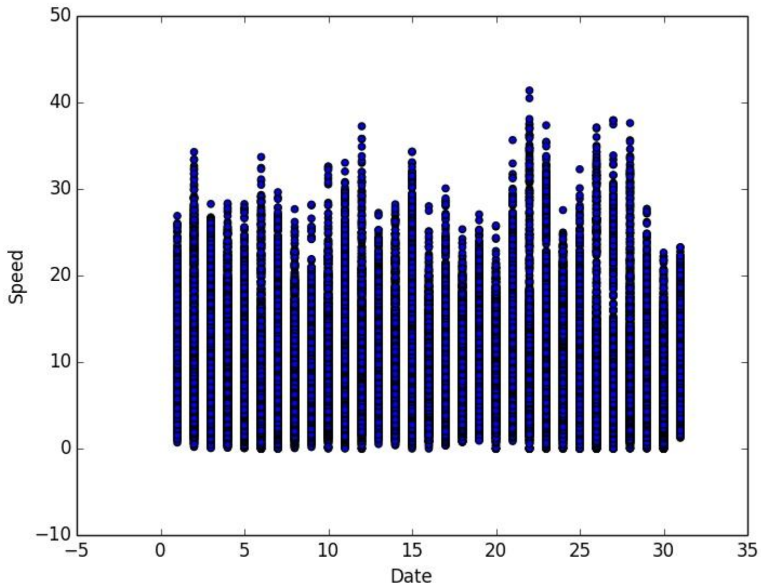
There is a defined pattern between two parameters of weather, which is rain and humidity. Graphically it is visible that rise in humidity causes increase in rainfall. This proves interdependence between these two parameters.



Above image depicts relationship between rainfall and Max Temp.
From the above image, it is clear that rise in temperature can result in higher rainfall; this further depicts interdependence between these two parameters. Resultantly Rain, Humidity and Temperature are interdependent and can be used for energy generation and weather forecasting.

IV. WIND DATA ANALYSIS

Given below is image of data analysis done on wind data.

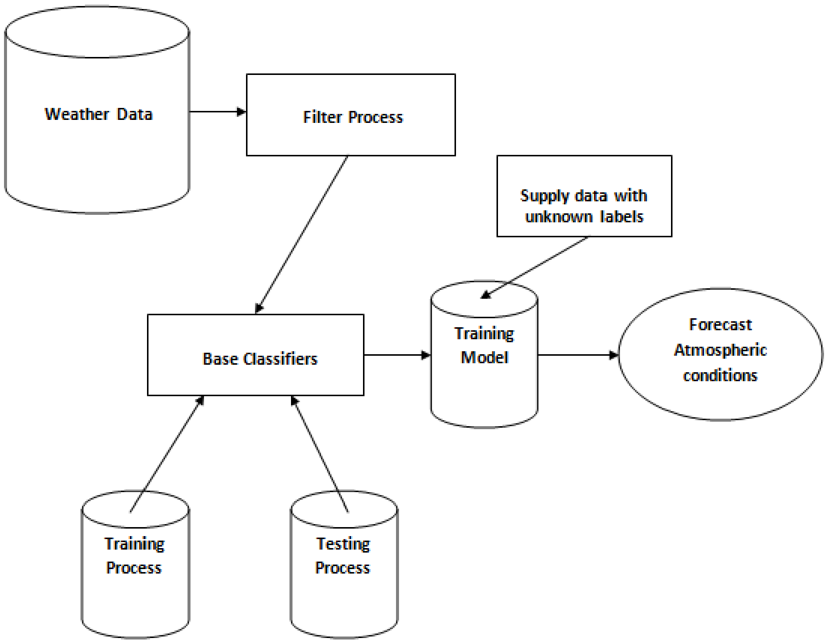


Above image depicts wind Speed across month.
From the above image, it is clear that wind speed is predictable across the month. This data can be used for harness of wind energy. Further, a relationship can be established between rain, temp, humidity and wind speed

V. AN INTEGRATED DESIGN OF FRAMEWORK FOR PREDICTING ATMOSPHERIC CONDITIONS

In this study, meticulous efforts were put forward to predict the meteorological data based on dataset comprising of diverse zones of Jammu and Kashmir. The figure 1 proposes an integrated procedure for predicting the weather conditions. To produce accurate predictions is a considerable challenge faced by various meteorologists around the globe. As, precise predictions have a significant role in different areas of life including agriculture, food safety, science, irregular events (Storms, floods and hurricanes) and so on, which would in turn assist in adapting precautionary measures at times of hazardous climatic phenomena.
From the figure, it is relatively eloquent that primarily atmospheric dataset is subjected to filtering process for eliminating any inconsistencies in data. Subsequently, the pre-processed data is then deployed to the custom classifiers, wherein the specific classifier goes through the process of training and testing the instances. Moreover, the model produced during the method of training and testing can be successfully employed for forecasting the unknown instances.

Figure 1 shows proposed design for forecasting meteorological instances



VI. PRE-PROCESSING OF DATA

The dataset that was subjected to examination using various data mining techniques comprised of 17 attributes. The weather dataset employed in our study contained information pertaining to different areas of Jammu and Kashmir (J&K) including north, south and central J&K. Moreover, the annual rainfall demonstrated in the underlying figure 1 is measured in millimetres (mm). Furthermore, prior to application of learning algorithms, the dataset was subjected to process of filtering to eliminate the inconsistencies that may occur while train and testing the algorithms. Therefore, in this study we have applied a filter namely "replace missing values" to fill values for misplaced attributes in the dataset.

Figure 2 exhibits the different rainfall estimates

State	District	Year	January	February	March	April	May	June	July	August	September	October	November	December	Annual_Rainfall	Rainfall_Status
Jammu and Kashmir	Anantnag	2004	125.1	78.3	13.3	145.8	65.9	75.6	52.7	65.6	28.5	81.1	27.8	17	776.7	AVG
Jammu and Kashmir	Anantnag	2005	178.1	379.6	148.3	93.1	100.2	39.1	127.2	30.8	31.6	16.8	29.4	1.1	1175.3	AVG
Jammu and Kashmir	Anantnag	2006	204	93.5	108	58.7	31.1	68.6	120.2	141.6	151.8	34.5	61.8	133.8	1207.6	AVG
Jammu and Kashmir	Anantnag	2007	24.6	65.5	281.1	7.8	60.5	82.6	59.1	87	44.2	0.3	0	55.5	768.2	AVG
Jammu and Kashmir	Anantnag	2008	139.7	155	15.9	84.4	101.5	72.2	62.2	73.1	66	37.7	31.2	112.5	951.4	AVG
Jammu and Kashmir	Anantnag	2009	166.1	128.2	76.5	107.7	82.9	92.7	78.4	47.2	43.9	5.6	67.4	27.9	924.5	AVG
Jammu and Kashmir	Anantnag	2010	53.9	110	65.2	142.9	266.9	100.1	91.2	107.2	87.7	62.6	25.6	57.2	1170.5	AVG
Jammu and Kashmir	Doda	2004	272.8	117.8	4.7	116.5	92.6	124.7	93.4	73.5	14.6	119.7	43.4	40.8	1114.5	AVG
Jammu and Kashmir	Doda	2005	206.9	576.4	231.1	70.5	64.1	39.3	266.3	60.6	72.9	6.2	11.4	0	1605.7	HIGH
Jammu and Kashmir	Doda	2006	326.7	117.6	187.9	63.9	53.7	70	240.6	214.4	283	67	12.2	37.9	1674.9	HIGH
Jammu and Kashmir	Doda	2007	4.9	151.9	353.2	14	62.3	66.9	88.7	100.4	28.8	0.9	0	83.9	955.9	AVG
Jammu and Kashmir	Doda	2008	25.1	227	7.7	122.7	65.6	102.2	102.6	72.8	48	44.8	6.6	131.7	956.8	AVG
Jammu and Kashmir	Doda	2009	160	151.9	82.7	141	59.6	50.6	112.2	51.7	59.1	7.1	81.2	19.1	976.2	AVG
Jammu and Kashmir	Doda	2010	48	240.8	41.6	79.2	177.6	109.2	167.2	189	110	36.2	29.8	73.8	1302.4	AVG
Jammu and Kashmir	Jammu	2004	110.2	25.9	0	42.5	25.2	103.2	350.9	164.6	78.5	48.5	4.4	25.1	979.0	AVG
Jammu and Kashmir	Jammu	2005	123	139.5	111.6	12.7	12.2	24.6	340.1	162.8	41.6	1.8	0	0	969.9	AVG
Jammu and Kashmir	Jammu	2006	70.8	4.2	44.6	10.7	16.1	150.2	379.9	269.4	123.8	56	7	32.4	1165.1	AVG
Jammu and Kashmir	Jammu	2007	0	95.9	243.6	3	28	185.2	274	333	42.4	0	3.7	8.9	1217.7	AVG
Jammu and Kashmir	Jammu	2008	98	23.4	2.2	56.4	42.8	228.2	384.9	225.8	20	21.1	0	20.3	1123.1	AVG
Jammu and Kashmir	Jammu	2009	69.5	34.1	15.6	41.1	11.6	18.6	244.4	273.4	26.1	7.2	4.2	0.6	746.4	AVG
Jammu and Kashmir	Jammu	2010	7.5	26.9	4.9	6.2	39.1	67	272.9	586.1	84.4	31.5	2.6	52.8	1181.9	AVG
Jammu and Kashmir	Kathua	2004	119	6.7	0	42	21.3	83.8	363.6	214.8	66.2	46.8	11.6	15.6	991.4	AVG
Jammu and Kashmir	Kathua	2005	45.8	174.4	65.6	12.8	7	16.4	404.4	277.8	210.2	1.4	0	0	1215.8	AVG

VII. CLASSIFIERS EMPLOYED

In our empirical results various classifications algorithms have been applied to construct models that would effectively forecast the annual rainfall of J&K. Therefore, various classifiers that have been used on our dataset included j48, naive bayes, instance base learning (IB1) and bayes network.

- J48

J48 is a classification technique that corresponds to one of the variants of decisions trees and is rooted on C4.5 classification algorithm. The C4.5 classifier recursively segregate the data, until a complete decision tree with entire of its instances classifiedis generated for the given dataset. The choices are advancing in a tree using depth- first approach. Moreover, j48 adopts two significant strategies to prune a tree viz. sub-tree replacement and sub-tree rising, so as to mitigate the size of the tree and more imperatively to eliminate the weak branches in a tree, and as a consequence an optimal tree is generated.

- Naive bayes

This classifier exercises probabilistic approach where in bayes theorem is employed to achieve the desired results. Furthermore, the classifier involves astraightforward mechanismwhich is based on conditional probability, and has been applied by various researchers effectively in producing substantial results. Moreover, for a given class attribute, the classifier presumes that the value of a specific attribute is independent of the value of each attribute, regardless of the strong correlation among the different variables of the dataset.

- IB1

The classifier symbolizes to class of lazy classifiers. This specific classifier is an instance based learner that computes its distance using Euclidean distance. During training, it uses this distance measurement (Euclidean distance) for determining which instance is more close to the given original instance, and endeavours to predict the correct class. Furthermore, if the distance of more than one training instance were found to be same while the classifier undergoes the process of training and testing, then in such conditions, the instance which was searched earlier is given preference.

- Bayes Network

This classifier is also known as Belief network or Bayesian network, which belongs to a group of probabilistic graphical models. These graphical formations symbolize information over an indecisive sphere. Each node in the graph demonstrates an arbitrary attribute and the edges among the nodes correspond to probabilistic dependencies between the arbitrary attributes. These models can be deployed in wide range of disciplines viz. predictive analytics, explanatory analytics, prognostic analytics and so on.

VIII. RESULTS AND DISCUSSIONS

In this subsection, the data was exposed to empirical investigations on our weather dataset using miscellaneous algorithms viz. j48, naive bayes, IB1 and bayes net, to forecast the yearly rainfall of different regions associated with Jammu and Kashmir. The rationalization of each classifier has been statistically demonstrated while forecasting different classes including low, average and high rainfall, in the underlying figures 2, 3, 4, and 5 respectively. As per the figure 2, j48 has shown exceptional classification accuracy of 97.14% in predicting the annual rainfall of diverse regions. The classifier has also explained low statistics of incorrectly classified instances (2.8%) in this case. In addition, other accuracy estimates linked with the classifier have exposed significant results viz. TP rate, FP rate, Precision, Recall and so on, as is evident from the underneath figure 2. Furthermore, the confusion matrix reveals which instances have been erroneously interpreted by the classifier as part of different given classes (low, high, avg). From the confusion matrix, it is noticeable that 42 instances have been correctly classified as average and one class has been misclassified as low by the classifier. Moreover, class labelled with high, has publicized one occurrence incorrectly classified as average and in case of class low, all the instances have successfully been classified in their respective precise classes while forecasting the annual rainfall.

Figure 3 **demonstrates j48 and various measurements associated with it.**

```
Correctly Classified Instances          68               97.1429 %
Incorrectly Classified Instances         2                2.8571 %
Kappa statistic                          0.9469
Mean absolute error                      0.019
Root mean squared error                  0.138
Relative absolute error                  5.2632 %
Root relative squared error             32.5818 %
Total Number of Instances               70

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.977    0.037    0.977      0.977    0.977      0.940    0.970     0.968     AVG
                 1.000    0.016    0.889      1.000    0.941      0.935    0.992     0.889     HIGH
                 0.947    0.000    1.000      0.947    0.973      0.964    0.974     0.962     LOW
Weighted Avg.    0.971    0.025    0.973      0.971    0.972      0.946    0.973     0.957

=== Confusion Matrix ===

  a  b  c   <-- classified as
 42  1  0 |  a = AVG
  0  8  0 |  b = HIGH
  1  0 18 |  c = LOW
```

Furthermore, the dataset was subjected to classifier viz. naive bayes and the classifier revealed substantial results in forecasting the annual rainfall. The classifier articulated considerable accuracy of 94.28% in correctly classifying the instances as can be apparently seen in figure 3. Nevertheless incorrectly classified instances (5.71%) have exposed small augmentation in contrast to j48. Additionally, in figure 3, various accuracy measurements associated with the classifier such as TP rate, FP rate, Precision, recall and so on, have shown significant results while forecasting the yearly precipitation. Moreover, as per the confusion matrix, 41 classes have been correctly predicted as average whereas two classes have incorrectly been classified as high and low. However, in classes high and low, one class from each category has been incorrectly predicted as average class.

Figure 4 illustrates naive bayes results

```
Correctly Classified Instances          66               94.2857 %
Incorrectly Classified Instances         4                5.7143 %
Kappa statistic                          0.8934
Mean absolute error                      0.0478
Root mean squared error                  0.1901
Relative absolute error                 13.1969 %
Root relative squared error             44.8732 %
Total Number of Instances               70

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.953    0.074    0.953      0.953    0.953      0.879    0.941     0.916     AVG
                 0.875    0.016    0.875      0.875    0.875      0.859    0.974     0.923     HIGH
                 0.947    0.020    0.947      0.947    0.947      0.928    0.975     0.968     LOW
Weighted Avg.    0.943    0.053    0.943      0.943    0.943      0.890    0.954     0.931

=== Confusion Matrix ===

  a  b  c   <-- classified as
 41  1  1 |  a = AVG
  1  7  0 |  b = HIGH
  1  0 18 |  c = LOW
```

The observations in figure 4 that were accomplished after IB1 was successfully deployed on our weather dataset exposed significant results, even though less effective than other classifiers employed with the same dataset. IB1 attained classification accuracy of 71.42% in correctly predicting the unknown instances while performing the process of training and testing. The instances that were incorrectly classified turned out to 28.57% which is again higher than remaining classifiers. In addition, average weighted accuracy as well as individual precision of all parameters (TP rate, FP rate, precision, recall and so on)for all the three classes including average, high and low is also notable. However, the number of misclassified instances as per the confusion matrix in class average are 7 and 5 respectively, where 7 have been incorrectly predicted as instances belonging to class high and 5 to class low. Moreover, from class high, entire instances have successfully been predicted whereas from class low, 8 instances have erroneously predicted as average. Therefore, the amount of instances that were misclassified in IB1 has amplified.

Figure 5 shows results of IB1

```
Correctly Classified Instances          50               71.4286 %
Incorrectly Classified Instances        20               28.5714 %
Kappa statistic                          0.4293
Mean absolute error                      0.1905
Root mean squared error                  0.4364
Relative absolute error                 52.6316 %
Root relative squared error            103.0326 %
Total Number of Instances               70

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.814    0.444    0.745      0.814    0.778      0.383    0.685     0.720     AVG
                 0.125    0.000    1.000      0.125    0.222      0.335    0.563     0.225     HIGH
                 0.737    0.157    0.636      0.737    0.683      0.556    0.790     0.540     LOW
Weighted Avg.    0.714    0.316    0.744      0.714    0.689      0.424    0.699     0.615

=== Confusion Matrix ===

  a  b  c   <-- classified as
 35  0  8 |  a = AVG
  7  1  0 |  b = HIGH
  5  0 14 |  c = LOW
```

Eventually, we applied bayes net across our weather dataset to further ascertain indispensible factors responsible for supporting atmospheric prediction procedures. We therefore examined that after application of bayes net, substantial accuracy of 92. 85 % was achieved in correctly classifying the instances. In addition, the erroneous classified instances proved to be 92.85% which is relatively noteworthy. Moreover, weighted average accuracy of each parameters associated with the bayes net classifier have exhibited considerable results. Furthermore, from the confusion matrix it is conspicuous that classes such as average, high and low have also demonstrated exceptional performance in predicting the annual precipitation. In class average, 1 instance has been incorrectly classified as high and 3 instances have been misclassified as low, whereas in class low, 1 instance has been wrongly classified as average. Nevertheless, entire instances pertaining to class high rainfall have been accurately classified.

Figure 6 displays results of bayes net

```
Correctly Classified Instances          65               92.8571 %
Incorrectly Classified Instances         5                7.1429 %
Kappa statistic                          0.8624
Mean absolute error                      0.0592
Root mean squared error                  0.1974
Relative absolute error                 16.3573 %
Root relative squared error             46.6017 %
Total Number of Instances               70

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.977    0.148    0.913      0.977    0.944      0.850    0.972     0.982     AVG
                 0.875    0.000    1.000      0.875    0.933      0.928    0.998     0.986     HIGH
                 0.842    0.020    0.941      0.842    0.889      0.853    0.978     0.961     LOW
Weighted Avg.    0.929    0.096    0.931      0.929    0.928      0.860    0.977     0.977

=== Confusion Matrix ===

  a  b  c   <-- classified as
 42  0  1 |  a = AVG
  1  7  0 |  b = HIGH
  3  0 16 |  c = LOW
```

IX. CONCLUSION

The underlying principle behind this study was to forecast the annual rainfall using an assortment of classifiers including j48, naive bayes, IB1 and bayes on meteorological dataset. It was examined that the discrepancy in predictingtarget values while training various classifiers on different models, j48 attainted significant accuracy of 97.14% in predicting the annual rainfall. J48classifier fetchedconsiderable accuracy which is somewhat better than naviebayes (94.28%) and bayes net (92.85%). The outcome of these models corroborated to be exceedingly auspicious and momentous. However, IB1 divulged poor performance in disparity to other predictive models on our weather data throughout the process of testing and training the model in classifying the correct atmospheric instances. The classifier accomplished an accuracy of 77% in predicting the exact instances. Furthermore, these types of studies can be extremely notable if the atmospheric forewarnings are communicated to populace well in advance of natural catastrophe.

From the past studies, it is explicable that there is a substantial gap, where researchers have not exploited deep learning techniques for making further strides and discoveries in the realm of meteorological prediction. Moreover, hybrid models and dedicated meteorological techniques are excellentresearch studies which must be given due course of significance, as it can further assist in meteorological forecasting.

REFERENCES

[1]. A. K. Chaudhary, J .A. Harding and M. K Tiwari, Data mining in manufacturing: Review based on the kind of knowledge, Journal of Intelligent Manufacturing, Volume 20, Issue 5, pp. 501-521 October 2009.

[2]. L. A. Kurgan, and P. Musilek, A survey of Knowledge Discovery and Data Mining process models, The Knowledge Engineering Review, Cambridge University Press, USA, Vol. 21(1), pp. 1-24, March 2006.

[3]. R. Kumar, Dr. S.R.Verma, KDD Techniques: A survey,International Journal of Electronics and Computer Science Engineering ,IJECSE,Vol. 1,.pp 2042- 2046,2008.

[4]. M. A. Kalyankar, Prof. S. J. Alaspurkar, Data Mining Technique to Analyse the Meteorological Data, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, pp. 114-118, February 2013.

[5]. A. Shahi, R. Bintiatana ,An Effective Fuzzy C-mean and Type- 2 Fuzzy Logic for Weather Forecasting, Journal of Theoretical and Applied Information Technology ,pp.510-511,2009.

[6]. Ganesh P. Gaikwad, V. B. Nikam, "Different Rainfall Prediction Models And General Data Mining Rainfall Prediction Model",International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 2 Issue 7, July – 2013.

[7]. FolorunshoOlaiya, Owo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies", IJIEEB Vol.4, No.1, February 2012.

[8]. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data mining," Pearson Education, 2007.

[9]. T. Asha, S. Nataranjan, K.N.B. Murthy, "Association rule based Tuberculosis disease diagnosis," Proc.SPIE7546, second International conference on Digital Image Processing., (February 26, 2010);doi:10.1117/12.853291.

[10]. T . Asha, S. Nataranjan, K.N.B. Murthy, "Diagnosisof Tuberculosis using ensemble methods," 2010 3rd IEEE International conference onComputer Science and Information Technology(ICCSIT),year:2010,volume:8,pages:409-412DOI:10.1109/ICCSIT.2010.5564025.

[11]. Elia Georgiana Petre,"A decision tree for weather prediction", SeriaMatematic□ - Informatic□ – Fizic□, No.1, pp. 77 – 82, 2009.

[12]. Rajesh Kumar, "Decision tree for the weather forecasting", International Journal of Computer Applications (0975 – 8887) vol.2, August 2013.

[13]. Ron Holmes, "Using a decision tree and neural net to identify severe weather radar characteristics".

[14]. ZaheerUllah Khan and Maqsood Hayat, "Hourly based climate prediction using data mining techniques by comprising entity demean algorithm", Middle-East Journal of Scientific Research 21 (8): pp. 1295-1300, 2014.

[15]. Dalibor Petkovic, Milan Gocic, SlavisaTrajkovic, ShahaboddinShamshirband, ShervinMotamedi, "Determination of the most influential weather parameters on reference evapotranspiration by adaptive neuro-fuzzy methodology", Computers and Electronics in Agriculture□114 (2015) 277–284.

[16]. Prasanth Rao Jillella.S.S .,"Weather Forecasting Using Artificial Neural Networks And Data Mining Techniques" , International Journal Of Innovative Technology And Research Vol No:3, pp.2534 – 2539, 2015.

[17]. M. Viswambari., Dr. R. AnbuSelvi . , "Data Mining Techniques To Predict Weather: A Survey", International Journal Of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, pp. 2348 – 7968, 2014.

[18]. S. Kotsiantis and et. al., "Using Data Mining Techniques for Estimating Minimum, Maximum and Average Daily Temperature Values", World Academy of Science, Engineering and Technology 2007 pp. 450-454

[19]. Meghali A. Kalyankar, S. J. Alaspurkar, " Data Mining Technique to Analyse the Metrological Data", International Journal of Advanced Research in Computer Science and Software Engineering 3(2), 114-118, February – 2013.

IOSR Journal of Engineering (IOSRJEN) is UGC approved Journal with Sl. No. 3240, Journal no. 48995.

Mr. Sameer Kaul " Integrated Design For Energy Generation And Weather Forecasting Using Different Base Learners on Wind & Rain Dataset." IOSR Journal of Engineering (IOSRJEN), vol. 08, no. 9, 2018, pp. 01-10.