

Rainfall Prediction in Lahore City using Data Mining Techniques

Shabib Aftab, Munir Ahmad, Noureen Hameed, Muhammad Salman Bashir, Iftikhar Ali, Zahid Nawaz

Department of Computer Science
Virtual University of Pakistan
Lahore, Pakistan

Abstract—Rainfall prediction has extreme significance in countless aspects and scopes. It can be very helpful to reduce the effects of sudden and extreme rainfall by taking effective security measures in advance. Due to climate variations, an accurate rainfall prediction has become more complex than before. Data mining techniques can predict the rainfall through extracting the hidden patterns among weather attributes of past data. This research contributes by exploring the use of various data mining techniques for rainfall prediction in Lahore city. Techniques include: Support Vector Machine (SVM), Naïve Bayes (NB), k Nearest Neighbor (kNN), Decision Tree (J48) and Multilayer Perceptron (MLP). The dataset is obtained from a weather forecasting website and consists of several atmospheric attributes. For effective prediction, pre-processing technique is used which consists of cleaning and normalization processes. Performance of used data mining techniques is analyzed in terms of precision, recall and f-measure with various ratios of training and test data.

Keywords—Rainfall prediction; data mining; classification techniques

I. INTRODUCTION

Time series data mining is one of the hot research topics in the domain of knowledge discovery [19]. The data with time series approach is collected over a specific period of time such as daily, weekly, monthly, quarterly or yearly [13]. This data can be used for predictions in different domains such as finance, stock market and climate change etc. Data mining techniques are used to extract the hidden knowledge from time series data for future use [13], [17], [25], [28]. Weather prediction with time series data is beneficial but quite challenging task [16], [27], [29]. It comes with an array of complexities which needs to be tackled for optimal results [18]. The statistical weather data has a wide variety of fields which are called features such as humidity, pressure, wind speed, pollutants, concentrations etc. Data mining techniques can predict the weather on the basis of hidden patterns among these features [27], [29]. Rainfall prediction is an important aspect of climate forecasting. Accurate and timely rainfall prediction is crucial for the planning and management of water resources, flood warnings, construction activities and flight operations etc. [14], [15]. This study used 5 data mining techniques for rainfall prediction in Lahore, capital of Punjab province, Pakistan. In Lahore, development and construction activities are increasing exponentially, so timely rainfall prediction is crucial for better assessment of future requirements and planning. The used data mining techniques include: Support

Vector Machine, Naïve Bayes, k Nearest Neighbor, Decision Tree and Multilayer Perceptron. These algorithms belong to supervised data mining class where pre-classified data is required first for training purpose. During training, these algorithms make rules of classification for input dataset (test data) [20]-[25], [30]. In this research, dataset is obtained from weather forecasting website [10] from December 1, 2005 to November 31, 2017 (12 years), which contains several weather related attributes such as Temperature, Atmospheric pressure, Relative humidity etc. For rainfall prediction, a classification framework is used in which the dataset gone through cleaning and normalization process before classification. Cleaning is performed to deal with the missing values and the purpose of normalization is to keep the attribute values in a certain limits. These pre-processing activities are crucial for the smooth classification process as well as for good results [9], [12]. Prediction performance of used data mining techniques is evaluated in terms of precision, recall and f measure, which are the important metrics of information retrieval. Finally the results are shown in tables and graphs.

Further organization of this paper is as follows. Section II describes the related work. Section III discusses the materials and methods used in this research. Section IV presents results and discussion. Section V finally concludes the paper.

II. RELATED WORK

Many researchers have been working to achieve high accuracy in rainfall prediction using data mining techniques; some of the selected studies are discussed here. Researchers in [1] performed a comparative analysis of multiple classifiers for rainfall prediction in Malaysia. Classifiers include Naïve Bayes, Support Vector Machine, Decision Tree, Neural Network and Random Forest. Dataset was obtained from multiple stations of Selangor, Malaysia. Pre-processing tasks were applied before classification to deal with the noise and missing values. According to results, Random Forest performed better as with small training data it correctly classified large amount of instances. In [2], researchers presented Clusterwise Linear Regression (CLR) method, which is the combination of clustering and regression techniques. The proposed technique is used to predict monthly rainfall in Victoria, Australia, by using input data of 8 geographically diverse weather stations. To analyze the performance of proposed CLR, results were compared with other techniques such as: CLR using the maximum likelihood framework by the expectation-maximization algorithm, multiple linear

regression, artificial neural networks and the support vector machines. It was observed that proposed algorithm performed better than other methods in most of the locations. Researchers in [3] performed a comparative analysis of a modern technique named Markov Chain (extended with rainfall prediction) and six other well-known machine learning techniques: Genetic Programming, Support Vector Regression, Radial Basis Neural Networks, M5 Rules, M5 Model trees, and k-Nearest Neighbors. For prediction, rainfall time series data of 42 cities with different climatic features is used. The results reflected that machine learning techniques have the capacity to perform better than Markov Chain technique moreover this study has also pointed out the correlations between different climatic attributes and predictive accuracy. In [4], Artificial Neural Network (ANN) was used to develop one-month and two-month ahead forecasting models for rainfall prediction using monthly rainfall data of 141 years from various weather stations in the North India. In these models, Feed Forward Neural Network (FFNN) using Back Propagation algorithm and Levenberg-Marquardt training function was used. The performance of both the models was analyzed on the basis of Regression Analysis, Mean Square Error (MSE) and Magnitude of Relative Error (MRE). ANN showed optimistic results for both the models and found that one month ahead forecasting model performed better than two months model. Researchers in [5] proposed an algorithm which combined data mining and statistical techniques. The likely predictors with highest confidence level, based on association rules were selected. Those predictors were derived from local and global conditions. From local conditions: sea level pressure, wind speed, and maximum & minimum temperatures were recorded. On the other hand from global condition, southern oscillation and Indian Ocean dipole conditions were taken. The algorithm predicted the rainfall in five categories: Flood, Excess, Normal, Deficit and Drought. Researchers in [6] presented Wavelet Neural Network model (WNN) for rainfall prediction which is the combination of wavelet technique and Artificial Neural Network (ANN). Proposed WNN and ANN, both models were applied on monthly rainfall data of Darjeelin grain gauge station, west Bengal, India. Statistical methods were used to analyze the performance of both techniques and it was observed that WNN performed much better than ANN model. In [7], researchers implemented a rainfall forecasting model using Focused Time-Delay Neural Networks (FTDNN). The parameters for neural networks were taken from several experiments to perform prediction with one step ahead. For prediction, the daily rainfall data was obtained from Malaysia Meteorological Department (MMD) and then converted to monthly, biannually, quarterly and yearly basis. Models were trained and tested on each dataset and corresponding accuracies were evaluated using Mean Absolute Percent Error. According to results, most accurate forecasts were made with yearly rainfall dataset. The authors have pointed out that more parameters such as temperature, humidity and sunshine data should be incorporated into the neural networks to make the performance more accurate. Researchers in [8] presented a methodology to predict maximum temperature in the day, which followed the Support Vector Regression approach. Proposed technique performed prediction on the basis of several features, obtained from different measuring stations in

Europe. Weather related features included temperature, precipitation, relative humidity, air pressure, specifically synoptic situation of the day and monthly cycle. The proposed technique performed well when compared with other neural networks, multi-layer perceptron and an extreme learning machine.

III. MATERIALS AND METHODS

This research aims to analyze the performance of data mining techniques on rainfall prediction in Lahore city using a classification framework (Fig. 1). Dataset used in this research consists of several attributes along with the known output class. Output class is one which is going to be predicted on the basis of other available attributes. The reason of including the output class in dataset among other features is to analyze the performance and accuracy of data mining techniques [20], [24]. The output result after processing is compared with the known class and performance is measured in terms of precision, recall and f measure [1], [20], [21], [24], [26].

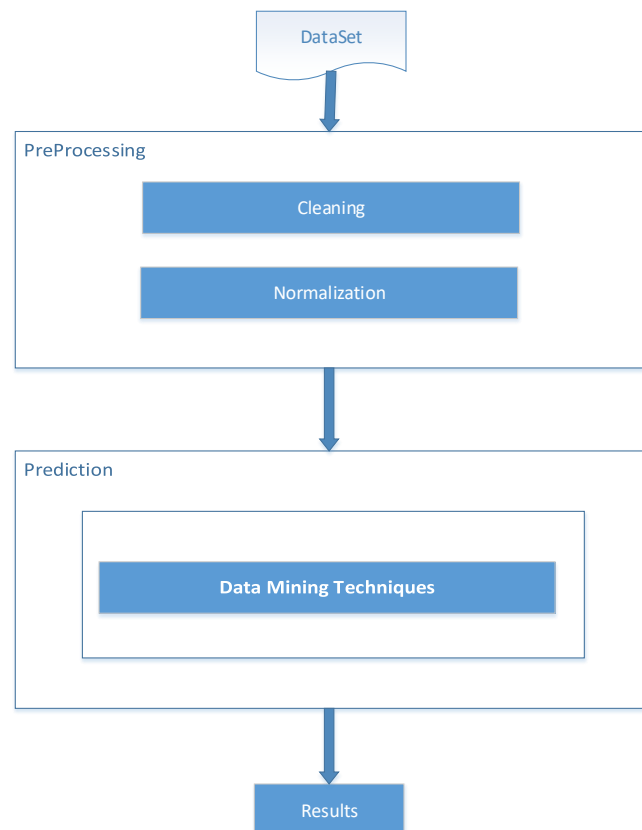


Fig. 1. Classification framework.

Weka [22], [23] is used in this study for classification and performance analysis. It is one of the extensively used data mining softwares. Weka is developed in Java language at the University of Waikato, New Zealand. It is famous and widely accepted tool among students and researchers due to its easy to use GUI interface, portability and General Public License.

The classification framework used in this research consists of four stages: Selection of appropriate dataset, Preprocessing, Prediction and Simulation of results. The input dataset for

rainfall prediction is obtained from weather forecasting website [10] and consists of several atmospheric attributes. Name, type and measurement unit of selected attributes are given in Table I.

TABLE I. DATASET ATTRIBUTES

Attribute Name	Attribute Type	Measurement
Temperature	Continuous	Degrees Celsius
Atmospheric Pressure (weather station)	Continuous	Millimeters of Mercury
Atmospheric Pressure (sea level)	Continuous	Millimeters of Mercury
Pressure Tendency	Continuous	Millimeters of Mercury
Relative Humidity	Continuous	%
Mean Wind Speed	Continuous	Meters per Second
Minimum Temperature	Continuous	Degrees Celsius
Maximum Temperature	Continuous	Degrees Celsius
Visibility	Continuous	Km
Dew Point Temperature	Continuous	Degrees Celsius

Dataset contained missing values as shown in Table II. The incomplete data can affect the accuracy of results as the attribute which has the missing value cannot fully participate in prediction process. Beside the missing values, dataset also contained noise where value resides below or exceeds from a certain limits. For effective data mining results it is recommended to keep the values in a certain limits [1], [11]. Pre-processing of input data is a crucial stage in classification framework which ensures the high accuracy of mining results. This stage consists of two activities: cleaning and normalization. Cleaning process deals with the missing values by using average mechanism. In this mechanism sum of all the instances of selected attribute is divided by the number of samples. On the other hand normalization process deals with the noise by limiting the values within a specific interval. Such interval can effectively facilitate the prediction process where the values will be mapped onto a particular range. In this research the normalization process is performed in Weka. Prediction is the final stage of classification framework where data mining algorithms perform classification by exploring the hidden patterns.

TABLE II. VALID RECORDS AND MISSING VALUES

Attribute Name	Valid Record	Missing Values
Temperature	25846.0	73
Atmospheric Pressure (weather station)	23689.0	2230
Atmospheric Pressure (sea level)	23714.0	2205
Pressure Tendency	11320.0	14599
Relative Humidity	25790.0	129
Mean Wind Speed	25890.0	29
Minimum Temperature	2415.0	23504
Maximum Temperature	4174.0	21745
Visibility	25829.0	90
Dew Point Temperature	25865.0	54

Performance of any supervised machine learning technique can be analyzed by comparing the output result with known class (pre-classified data). Performance evaluation of used data mining techniques is performed with 10 proportions (10:90-90:10) of training data and test data. For comparative analysis, three evaluation parameters of information retrieval are used: Precision, Recall and F Measure.

The aim of Precision is to evaluate the True Positive (TP) entities with respect to False Positive (FP) entities. It can be calculated as follows:

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (1)$$

TP is used for the entities, which are correctly classified, and FP is for those entities, which are wrongly classified.

The aim of recall is to evaluate the True Positive entities with respect to the (FN) False Negative entities, which are not classified at all. It can be calculated as follows:

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (2)$$

There may be a point where performance evaluation will not be possible with precision and recall, for example if one mining algorithm has higher precision but lower recall than another algorithm so the question arises that which algorithm is better. Solution to this issue is to use F-measure, which provides the average of precision and recall. F-measure can be computed as bellow:

$$\text{F-measure} = \frac{\text{Precision} * \text{Recall} * 2}{(\text{Precision} + \text{Recall})} \quad (3)$$

IV. RESULTS AND DISCUSSION

With SVM the results are almost same (Table III, Fig. 2) in all three accuracy parameters (Precision, recall and f-measure). Results for no-rain class with first seven proportions from 10:90 to 70:30 in precision, recall and f-measure are 0.941, 1, and 0.955 respectively however minor improvement were seen when proportions 80:20 and 90:10 were used. The notable point is that the result for rain class with all proportions in all accuracy parameters is 0, which means that this technique could not classify a single instance correctly for rain class even with 90:20 ratios.

TABLE III. SVM RESULTS

Proportion	Class	Precision	Recall	F-Measure
10:90	No Rain	0.914	1	0.955
	Rain	0	0	0
20:80	No Rain	0.914	1	0.955
	Rain	0	0	0
30:70	No Rain	0.914	1	0.955
	Rain	0	0	0
40:60	No Rain	0.914	1	0.955
	Rain	0	0	0
50:50	No Rain	0.914	1	0.955
	Rain	0	0	0
60:40	No Rain	0.914	1	0.955
	Rain	0	0	0
70:30	No Rain	0.914	1	0.955
	Rain	0	0	0
80:20	No Rain	0.919	1	0.958
	Rain	0	0	0
90:10	No Rain	0.919	1	0.958
	Rain	0	0	0

The results with Naive Bayes are shown in Table IV and Fig. 3. It can be seen that with no-rain class the 10:90

performed better in precision, 30:70 and 40:60 in recall and 90:10 in f-measure. With rain class 40:60 performed better in precision, 10:90 in recall and 50:50 in f-measure.

TABLE IV. NB RESULTS

Proportion	Class	Precision	Recall	F-Measure
10:90	No Rain	0.941	0.85	0.893
	Rain	0.215	0.435	0.287
20:80	No Rain	0.932	0.955	0.943
	Rain	0.353	0.258	0.298
30:70	No Rain	0.931	0.957	0.944
	Rain	0.356	0.254	0.296
40:60	No Rain	0.933	0.957	0.945
	Rain	0.366	0.264	0.307
50:50	No Rain	0.934	0.953	0.943
	Rain	0.362	0.283	0.318
60:40	No Rain	0.935	0.943	0.939
	Rain	0.331	0.299	0.314
70:30	No Rain	0.933	0.942	0.938
	Rain	0.316	0.282	0.298
80:20	No Rain	0.936	0.954	0.945
	Rain	0.34	0.268	0.3
90:10	No Rain	0.938	0.954	0.946
	Rain	0.355	0.286	0.317

The results with KNN are shown in Table V and Fig. 4. With no-rain class the 80:20 and 90:10 performed better in precision, 10:90 in recall and 90:10 in f-measure. With rain class 10:90 performed better in precision, 80:20 in recall and 90:10 in recall.

TABLE V. KNN RESULTS

Proportion	Class	Precision	Recall	F-Measure
10:90	No Rain	0.935	0.95	0.942
	Rain	0.358	0.295	0.324
20:80	No Rain	0.935	0.946	0.941
	Rain	0.35	0.306	0.326
30:70	No Rain	0.936	0.948	0.942
	Rain	0.365	0.317	0.339
40:60	No Rain	0.937	0.948	0.943
	Rain	0.368	0.321	0.343
50:50	No Rain	0.936	0.948	0.942
	Rain	0.363	0.314	0.337
60:40	No Rain	0.937	0.948	0.942
	Rain	0.364	0.317	0.339
70:30	No Rain	0.936	0.948	0.942
	Rain	0.361	0.313	0.335
80:20	No Rain	0.942	0.943	0.943
	Rain	0.35	0.346	0.348
90:10	No Rain	0.942	0.949	0.946
	Rain	0.373	0.343	0.357

The results with Decision Tree (J48) are shown in Table VI and Fig. 5. With no-rain class the 80:20 performed better in precision, 20:80 in recall and 90:10 in f-measure. With rain class 20:80 performed better in precision, 80:20 in recall and 80:20 in recall.

TABLE VI. DECISION TREE RESULTS

Proportion	Class	Precision	Recall	F-Measure
10:90	No Rain	0.936	0.976	0.955
	Rain	0.527	0.286	0.371
20:80	No Rain	0.926	0.993	0.958
	Rain	0.673	0.154	0.25
30:70	No Rain	0.931	0.988	0.959
	Rain	0.636	0.224	0.331
40:60	No Rain	0.933	0.98	0.956
	Rain	0.549	0.255	0.348
50:50	No Rain	0.932	0.985	0.958
	Rain	0.599	0.244	0.347
60:40	No Rain	0.934	0.985	0.959
	Rain	0.626	0.262	0.37
70:30	No Rain	0.934	0.984	0.959
	Rain	0.614	0.266	0.371
80:20	No Rain	0.939	0.981	0.96
	Rain	0.568	0.287	0.381
90:10	No Rain	0.938	0.985	0.961
	Rain	0.607	0.257	0.361

The results with MLP are shown in Table VII and Fig. 6. It can be seen that with no-rain class 80:20 performed better in precision, 90:10 in recall and 90:10 in f-measure. With rain class 90:10 performed better in precision, 60:40 in recall and 80:20 in f-measure.

TABLE VII. MLP RESULTS

Proportion	Class	Precision	Recall	F-Measure
10:90	No Rain	0.927	0.993	0.959
	Rain	0.709	0.173	0.278
20:80	No Rain	0.935	0.983	0.959
	Rain	0.61	0.28	0.384
30:70	No Rain	0.931	0.99	0.96
	Rain	0.681	0.221	0.334
40:60	No Rain	0.937	0.979	0.958
	Rain	0.579	0.303	0.397
50:50	No Rain	0.931	0.992	0.96
	Rain	0.724	0.22	0.337
60:40	No Rain	0.937	0.98	0.958
	Rain	0.587	0.304	0.4
70:30	No Rain	0.93	0.99	0.959
	Rain	0.67	0.212	0.322
80:20	No Rain	0.94	0.984	0.962
	Rain	0.619	0.296	0.401
90:10	No Rain	0.935	0.997	0.965
	Rain	0.846	0.21	0.336

A. Critical Analysis

Data mining techniques used in this study showed good results for no-rain class in all accuracy measures (Precision, recall and f-measure) however for rain class these techniques did not perform well and results are not accurate enough. F-measure is a high-quality accuracy measure as it provides the average of precision and recall. Table VIII is arranged according to highest f-measure score in each mining technique along with its class and proportion. There could be several reasons for the lower results with rain class such as, missing values as mean value cannot reflect the actual one, absence of one or more important climatic attributes and the most important is the lower rainfall rate in the city. Due to climate

variations, rainfall rate in most of the locations is not, what it used to be. Moreover the dataset does not include the rainfall quantity/measure, instead it only includes the rainfall polarity (yes/no). So the data is reflecting the number of times it rained but not how much. There might be only one rainy day in a week but that might have been catastrophic with extreme rainfall. With overall lower rainfall rate (number of times it rained), less patterns were provided to classification algorithms which resulted in poor performance with rain class whereas on the other hand in no-rain class, more patterns were available for training of classification techniques, resulted in high accuracy.

TABLE VIII. DM TECHNIQUES WITH HIGHEST F-MEASURE

DM Algorithm	Class	Proportion	F-Measure
SVM	No Rain	80:20	0.958
	Rain	90:10	0
NB	No Rain	90:10	0.946
	Rain	50:50	0.318
KNN	No Rain	90:10	0.946
	Rain	90:10	0.357
Decision Tree	No Rain	90:10	0.961
	Rain	80:20	0.381
MLP	No Rain	90:10	0.965
	Rain	80:20	0.401

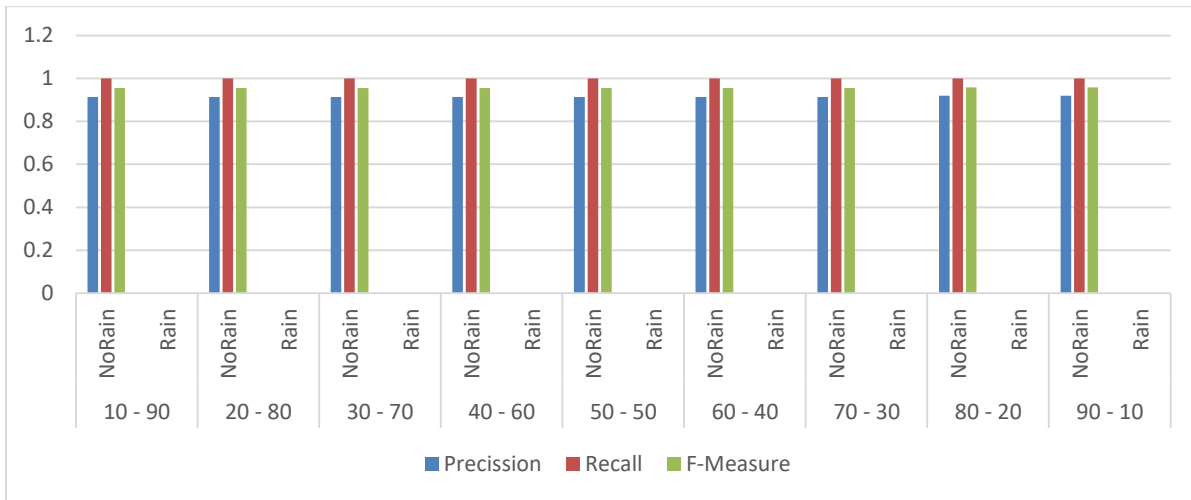


Fig. 2. SVM results.

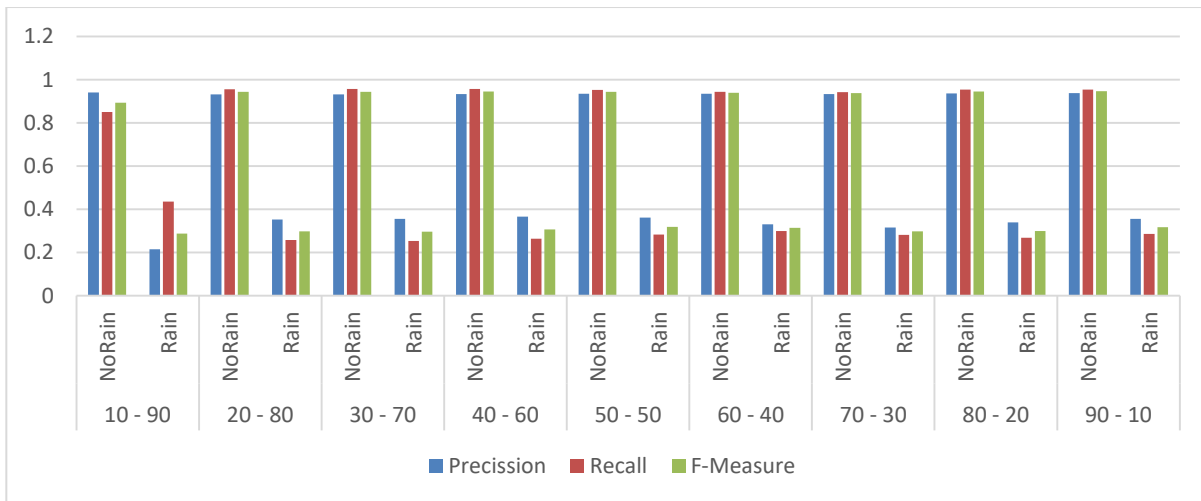


Fig. 3. NB results.

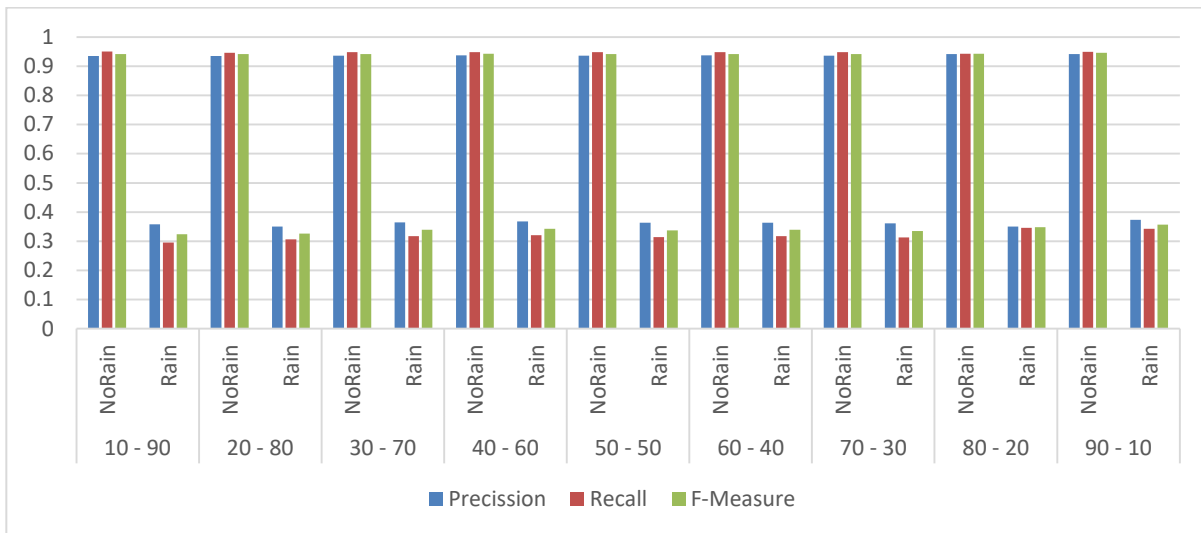


Fig. 4. KNN results.

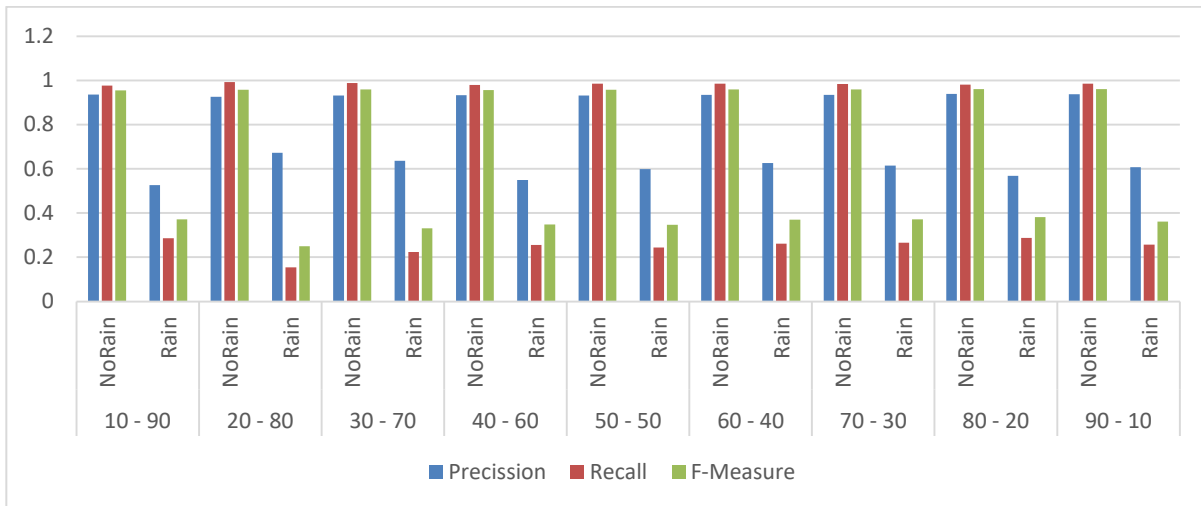


Fig. 5. Decision-tree results.

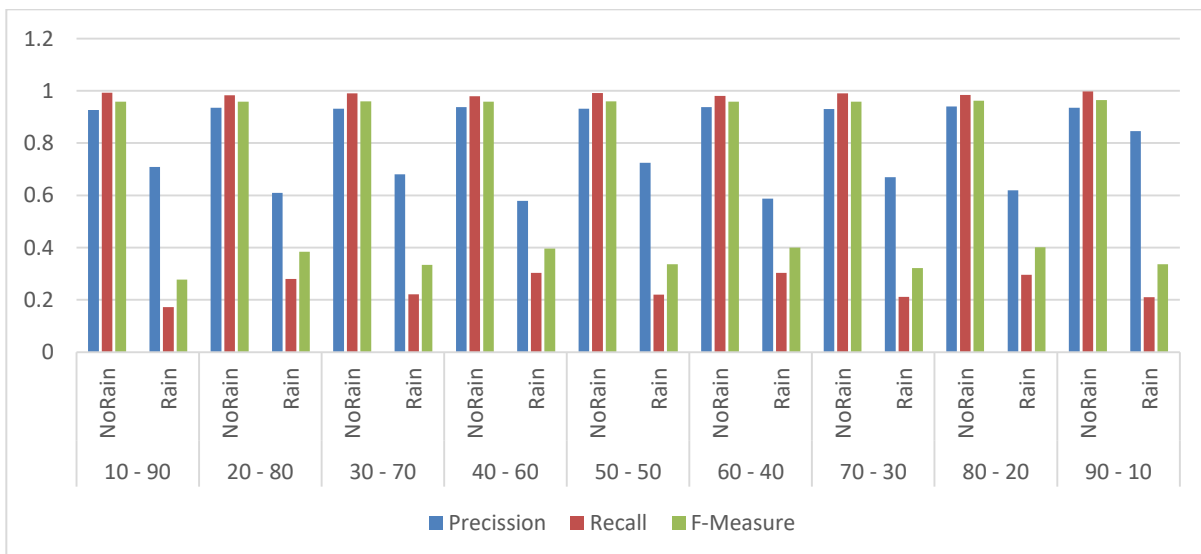


Fig. 6. MLP results.

V. CONCLUSION AND FUTURE WORK

This research performed rainfall prediction in Lahore city using five data mining techniques: Support Vector Machine, Naïve Bayes, k Nearest Neighbor, Decision Tree and Multilayer Perceptron. 12 years of past weather data from December 1, 2005 to November 31, 2017, is used for prediction in this research. Performance analysis of used data mining techniques is performed using three accuracy measures: precision, recall and f-measure and results are presented in tables and graphs. For effective prediction, a classification framework is used in which the input data went through a pre-processing stage and got cleaned and normalized before classification process. To analyze the performance dependency of classification techniques on training data, ten ratios of training and test data (training data: test data) are used from 10:90 to 90:10. According to results, used classification techniques performed well for no-rain class however for rain class, the techniques did not perform well. The reasons behind the lower accuracy in rain class may include: missing values, absence of important climatic attributes in dataset and overall lower rate of rainfall in the city. It is suggested for future work that further predictions should be performed by exploring more classification techniques and climatic attributes on different weather data.

REFERENCES

- [1] S. Zainudin, D. S. Jasim, and A. A. Bakar, "Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 1148–1153, 2016.
- [2] A. M. Bagirov, A. Mahmood, and A. Barton, "Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach," *Atmos. Res.*, vol. 188, pp. 20–29, 2017.
- [3] S. Cramer, M. Kampouridis, A. A. Freitas, and A. K. Alexandridis, "An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives," *Expert Syst. Appl.*, vol. 85, pp. 169–181, 2017.
- [4] N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data," *Int. J. Intell. Syst. Appl.*, vol. 10, no. 1, pp. 16–23, 2018.
- [5] H. Vathsala and S. G. Koolagudi, "Prediction model for peninsular Indian summer monsoon rainfall using data mining and statistical approaches," *Comput. Geosci.*, vol. 98, pp. 55–63, 2017.
- [6] R. Venkata Ramana, B. Krishna, S. R. Kumar, and N. G. Pandey, "Monthly Rainfall Prediction Using Wavelet Neural Network Analysis," *Water Resour. Manag.*, vol. 27, no. 10, pp. 3697–3711, 2013.
- [7] K. K. Htike and O. O. Khalifa, "Delay Neural Networks," *Int. Conf. Comput. Commun. Eng.*, no. May, pp. 11–13, 2010.
- [8] A. Paniagua-Tineo, S. Salcedo-Sanz, C. Casanova-Mateo, E. G. Ortiz-García, M. A. Cony, and E. Hernández-Martín, "Prediction of daily maximum temperature using a support vector regression algorithm," *Renew. Energy*, vol. 36, no. 11, pp. 3054–3060, 2011.
- [9] C. Sivapragasam, S. Liong, and M. Pasha, "Rainfall and runoff forecasting with SSA-SVM approach," *J. Hydroinformatics*, no. April 2016, pp. 141–152, 2001.
- [10] "http://ru8.rp5.ru/Weather_archive_in_Lahore." [Online]. Available: http://ru8.rp5.ru/Weather_archive_in_Lahore.
- [11] S. S. Monira, Z. M. Faisal, and H. Hirose, "Comparison of artificially intelligent methods in short term rainfall forecast," *Proc. 2010 13th Int. Conf. Comput. Inf. Technol. ICCIT 2010*, no. Iccit, pp. 39–44, 2010.
- [12] D. Isa, L. H. Lee, V. P. Kallimani, and R. RajKumar, "Text Document Preprocessing with the Bayes Formula for Classification Using the Support Vector Machine," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 9, pp. 1264–1272, 2008.
- [13] N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "A Comprehensive Survey of Data Mining Techniques on Time Series Data for Rainfall Prediction," vol. 11, no. 2, pp. 168–184, 2017.
- [14] K. W. Chau and C. L. Wu, "A hybrid model coupled with singular spectrum analysis for daily rainfall prediction," *J. Hydroinformatics*, vol. 12, no. 4, p. 458, 2010.
- [15] J. Wu, J. Long, and M. Liu, "Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm," *Neurocomputing*, vol. 148, pp. 136–142, 2015.
- [16] W. C.L. and K.-W. Chau, "Prediction of Rainfall Time Series Using Modular Soft Computing Methods," *Eng. Appl. Artif. Intell.*, vol. 26, no. 852, pp. 1–37, 2012.
- [17] D. Gupta and U. Ghose, "A Comparative Study of Classification Algorithms for Risk Prediction in Pregnancy," pp. 0–5, 2015.
- [18] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," *2012 IEEE Control Syst. Grad. Res. Colloq.*, no. Icsgrc, pp. 82–87, 2012.
- [19] M. A. Nayak and S. Ghosh, "Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier," *Theor. Appl. Climatol.*, vol. 114, no. 3–4, pp. 583–603, 2013.
- [20] M. Ahmad and S. Aftab, "Analyzing the Performance of SVM for Polarity Detection with Different Datasets," *Int. J. Mod. Educ. Comput. Sci.*, vol. 9, no. 10, pp. 29–36, 2017.
- [21] A. Sharma and S. Dey, "A boosted SVM based sentiment analysis approach for online opinionated text," *Proc. 2013 Res. Adapt. Converg. Syst. - RACS '13*, pp. 28–34, 2013.
- [22] G. Holmes, A. Donkin, and I. H. Witten, "WEKA: A Machine Learning Workbench." 1994.
- [23] "http://www.cs.waikato.ac.nz-mlweka." [Online]. Available: <http://www.cs.waikato.ac.nz/mlweka>.
- [24] M. Ahmad, S. Aftab, and I. Ali, "Sentiment Analysis of Tweets using SVM," *Int. J. Comput. Appl.*, vol. 177, no. 5, pp. 25–29, 2017.
- [25] M. Ahmad, S. Aftab, and S. S. Muhammad, "Machine Learning Techniques for Sentiment Analysis: A Review," *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 3, p. 27, 2017.
- [26] M. Ahmad, S. Aftab, M. S. Bashir, and N. Hameed, "Sentiment Analysis using SVM : A Systematic Literature Review," vol. 9, no. 2, pp. 182–188, 2018.
- [27] Y. Radhika and M. Shashi, "Atmospheric Temperature Prediction using Support Vector Machines," *Int. J. Comput. Theory Eng.*, vol. 1, no. 1, pp. 55–58, 2009.
- [28] K. Lu and L. Wang, "A Novel Nonlinear Combination Model Based on Support Vector Machine for Rainfall Prediction," *2011 Fourth Int. Jt. Conf. Comput. Sci. Optim.*, pp. 1343–1346, 2011.
- [29] V. B. Nikam and B. B. Meshram, "Modeling Rainfall Prediction Using Data Mining Method: A Bayesian Approach," *2013 Fifth Int. Conf. Comput. Intell. Model. Simul.*, pp. 132–136, 2013.
- [30] M. Ahmad, S. Aftab, I. Ali, and N. Hameed, "Hybrid Tools and Techniques for Sentiment Analysis: A Review," *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 3, 2017.