Indonesian Journal of Electrical Engineering and Computer Science
Vol. 31, No. 1, July 2023, pp. 451~458
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v31.i1.pp451-458

Precipitation forecasting using machine learning in the region of Beni Mellal-Khenifra

Hamza Jdi, Noureddine Falih
LIMATI Laboratory, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco

Article Info

Article history:

Received Jan 12, 2023
Revised Mar 24, 2023
Accepted Mar 26, 2023

Keywords:

Agriculture
Apache spark
Big data analytics
Machine learning
Precepitation forecast

ABSTRACT

Agriculture in the region of Beni Mellal-Khenifra, Morocco relies on irrigation from rain and dams, but recently there has been a lack of precipitation which may negatively affect crop growth. This has made accurate precipitation forecasts even more important for farmers, as they need this information to make informed decisions about their crops. However, a lack of data-driven research utilizing past data presents a challenge for the development of such research and leaves farmers relying solely on weather forecasts from TV, which cannot relied upon in systems such as irrigation. The objective of this paper is to propose various approaches for forecasting precipitation in the region of Beni Mellal-Khenifra using big data analytics and machine learning techniques. The study made use of Apache Spark, a big data analytics tool, and five machine-learning algorithms: Lasso regression, ridge regression, elastic net, auto regressive integrated moving average, and random forest. These algorithms were applied on dataset of daily rainfall from 2000 to 2015 to forecast the amount of precipitation in the region. The results of the study showed that the random forest algorithm had the lowest mean absolute error, making it the most effective at forecasting precipitation in the region.

Corresponding Author:

Hamza Jdi
LIMATI Laboratory, Polydisciplinary Faculty, Sultan Moulay Slimane University
Beni Mellal, Morocco
Email: hamzajdi@mail.com

1. INTRODUCTION

In the agricultural sector, machine learning (ML) and internet of things (IoT) technologies are implemented progressively by farmers to plan their farming activities. These technologies help get insights about the weather conditions, water use optimization [1], and early plant disease recognition [2], [3]. Farmers are keener to use ML to produce more yield, crops and products as they see the benefit of scientific and data driven ways to make use of the massive amounts of data that are put in their hands via sensors, which became in recent years accurate, dependable and most importantly affordable.

The combination of ML and IoT is complementary to each other in the sense that IoT technologies provide the infrastructure for collecting data and ML algorithms provide the means for analyzing and extracting insights from that data. In the context of agriculture [4], IoT sensors [5] can be used to monitor various aspects of agricultural processes, provide decision support, and automate irrigation. The data collected through these IoT systems can then be fed into ML algorithms and big data analytics tools [6], [7] to generate useful insights [8] for farmers. The complementary relationship between ML and IoT allows for the development of more efficient and effective agricultural practices.

Agricultural irrigation optimization [9]-[11] is a very complex task because it depends on various variables. One variable depend on the changing weather conditions, which needs to be anticipated by predicting weather conditions through answering for instance the simple question: “Will the rain fall?” another variable is soil that needs to be understood with the data provided by sensors. Therefore, according to the levels of pan evaporation, soil moisture reserves and other soil parameters [12], we can make automated, data driven decisions to use water in an optimized way [13]. Thus the need for precipitation forecast as it can help greatly in the optimization of water use [14], so we can prevent over watering and in some cases reduce irrigation amounts in anticipation of rain. The agricultural sector in the region of Béni Mellal-Khenifra region depends heavily on rain and dams as sources of water. However, the dam levels in the region have been insufficient for meeting the water needs of agriculture. To address this issue and reduce costs, it is critical to optimize the use of water through effective planning. Precipitation prediction helps in scheduling the release of water from dams, utilizing alternative sources, and providing valuable insights into current and future state of the dams. This can be particularly useful in determining how much water should be released and when alternative sources ought to be used. It is possible through precipitation prediction to ensure sufficient water supply for agriculture while also preserving resources. It should be noted that data-driven studies that capture the characteristics of the region are currently lacking.

Let us look at some recent studies that shade the light on precipitation forecasting and ML. Shaari *et al.* [15] examined the effectiveness of using auto regressive integrated moving average (ARIMA) and empirical wavelet transform in forecasting drought, based on clustering analysis, using the standard precipitation index. The research utilized daily rainfall data from Arau, Perlis from 1956 to 2008. Yin *et al.* [16] a real-time hourly precipitation forecast in japan is presented. According to the authors, this real-time forecast is crucial for early flood detection. The chosen methods are support vector machine (SVM), quantile-mapping (QM) and CDF-transform (CDFt). The authors combined different methods to improve accuracy. SVM improved the spatial representation of precipitation while QM and CDFt failed in this task. This is the reason why the authors were encouraged to combine these methods. As a result, a higher accuracy is obtained. Ramsundram *et al.* [17], we take a look at a comparison between decision tree (DT) and artificial neural network (ANN) in the matter of predicting rainfall, taking into account climatic variables as features. The findings show a huge difference in terms of performance in favor of DT as it outperformed ANN in predicting future rainfall. Mohammed *et al.* [18] proposes a comparison of SVM, linear regression and multiple linear regression (MLR) in the matter of rain prediction, based on a dataset from 1901 to 2015. The dataset is split 70% for training the model and 30% for testing it. The comparison, based on means absolute error (MAE), shows clear advantage in using SVM.

Jdi and Falih [19] with the help of Sliding Window Algorithm, Hadoop, and MapReduce, the authors predicted weather conditions for the full year of 2019 by using collected data of the year 2018. Smith *et al.* [20] make a comparative study between MLR and RF in an attempt to find out which is best suited for neuroscience prediction; the data is split 90% for training and 10% for testing. In general, MLR performed better than RF. Samadianfard *et al.* [21] attempt to predict precipitation. The data used is of a period spanning from 2004 to 2015. Dataset is split 70% for training and 30% for testing. The authors forecast precipitation using RF, logistic model tree, J48, and predictive association rule trees (PART). The results show that PART performed well. Rachmawati *et al.* [22] create a method for implementing Lasso regression (Lasso). Accordingly, sixteen predictor variables, such as temperature, humidity, and sun, are used in rainfall intensity modeling. Lasso model effectively narrows to nine the set of variables to be used. Meenal *et al.* [23] employed the conventional temperature-based empirical models and machine learning algorithms, including linear regression to forecast the weather parameters of precipitation, relative humidity, wind speed, and solar radiation. The results indicated that the machine learning based methods performed better in terms of prediction accuracy compared to the physics-based conventional models, with a mean square error of 0.1397 and a correlation coefficient of 0.9259. Mom *et al.* [24] propose a new rain attenuation prediction model for tropical locations based on the rain cell concept is proposed in this study. The International Telecommunications Union's model has a research gap, that this new model fills. The study's findings revealed that the proposed rain attenuation model (RAM) predicted signal availability correctly at seven of the thirteen monitored stations.

The present literature review shows that in the matter of precipitation prediction, numerous algorithms based on machine learning and artificial neural network are used frequently. However, algorithms based on DT, such as random forest (RF), are rarely used. In this paper, we will investigate further into Lasso regression, ridge regression (RR), elastic net (EN), auto regressive integrated moving average and random forest and compare them to reach a conclusion. These machine-learning algorithms, supplied with data provided by the National Climatic Data Center (NCDC), are used to compute daily precipitation of the whole rainy season in Beni Mellal-Khenifra.

2. METHOD

Beni Mellal is at the foot of the middle and high atlas. It is the main city and the capital of the region of Béni Mellal-Khenifra. Beni Mellal takes advantage of its status as an administrative capital, the richness of its agricultural land, and its new status as a university town. The amount of precipitation varies roughly between 13 inches and 25 inches depending on the year. Showed in Figure 1 is the average yearly precipitation in the region of Beni Mellal for the last 22 years, presented in a bar chart from the dataset to give an overview of precipitation. The year of 2010 had the highest PRCP with 20.89 inches. Concerning the last seven years (2016-2022), Beni Mellal had clear prolonged shortages in precipitation.

We endeavor to predict the entire rainy season (November-March) in the city of Beni Mellal. Weather parameters provided online by the NCDC are used. The ML models are trained using algorithms namely Lasso, RR, EN, ARIMA and RF These algorithms are run in databricks platform using Apache Spark [25] version 3.2.1. We choose to work with the programming languge Phyton as it is considered one of the top three Spark options when it comes to ML. Preparation of data in ML is an important step as it shapes the whole process ahead in terms of accuracy. Let us look at the dataset that contains various weather parameters, dates and information about the station location. The following data are of interest to us the most in terms of the scope of the paper as far as Lasso, RR, EN and RF: day wise total precipitation for a day in inches (PRCP), mean temperature for the day in Fahrenheit (TEMP), maximum temperature reported during the day (MAX), minimum temperature reported during the day (MIN), mean dew point for the day (DEWP), mean visibility for the day in miles (VISIB), maximum sustained wind speed reported for the day in knots (MXSPD), and mean wind speed for the day (WDSP). Features and labels the are two terms used in ML to defrentiate between descriptive attributes and predictive variable(s). In our case, as far as Lasso, RR, EN, and FR concern, the descriptive attributes are TEMP, DEWP, VISIB, WDSP, MXSPD, MAX and MIN; the predictive variable is PRCP. As for ARIMA we will be using a time series of past daily precipitation values (i.e., lagged values) as predictors.

We invest 5,356 records in the training phase and reserve the last rainy season, consisting of 211 records, for testing. We can already see huge difference because Lasso, RR, EN, and FR need no further processing of the data. On the contrary, ARIMA needs the data to be stationary wich was verifed using the augmented Dickey-Fuller. To identify the most suitable ARIMA model, we utilize the akaike information criterion (AIC) [26] and compared the models based on their MAE in order to determine which model would provide the most accurate predictions. After completing the prediction phase, which yield five different readings, we analyze the results and compared them using MAE.

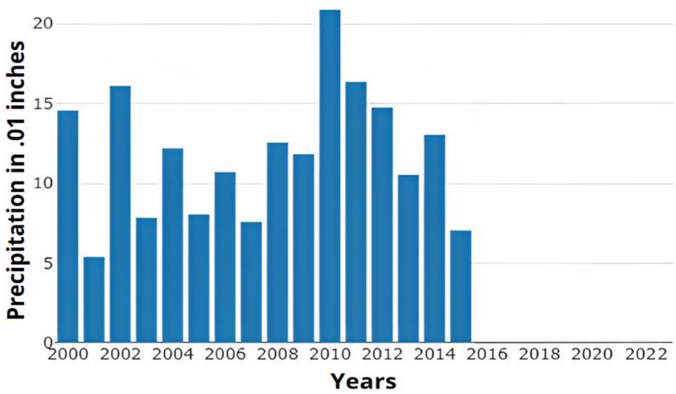

Figure 1. Average yearly precipitation Beni Mellal 2000-2022

3. RESULTS AND DISCUSSION

Using 5,356 records to train the four models, we compare the performance of the five algorithms to identify the most adequate to predict precipitation in the region of Beni Mellal. The algorithms are evaluated using MAE. It is worth mentioning that the first three algorithms Lasso, RR and EN, inspired by ordinary least squares, make it possible to set the convergence tolerance of iterations, while RF enables us to choose the number of trees and depth to be used. Table 1 shows the optimal parameters achieved for each algorithm.

Lasso [27] is an adjustment of linear regression. In this algorithm, the loss function is reformed with the aim to reduce model complexity by restricting the summation of the absolute values of the model coefficients. The penalty parameter alpha is what reduces some weight values to zero to clear the way for non-zero coefficients. Figure 2 shows the comparison between predicted and actual average daily precipitation values using Lasso. We can extract the following patterns concerning Lasso. During the two first months and in a consecutive sequence, the prediction is good. This leads us to conclude that Lasso algorithms predict well within the limits of two months, afterwards the prediction accuracy declines considerably. Therefore, the months that present good prediction are namely September and October. Please note that November 05, 2014 has an abnormal high amount of precipitation that Lasso did not predict well because the value predicted for that day was 0.187 whereas the actual value was 1.22.

Table 1. The parameters used per algorithm

Algorithm	Parameters used
Elastic Net	setMaxIter(1000) setElasticNetParam(0,23) setTol(0.0000001)
Ridge Regression	setMaxIter(1000) setElasticNetParam(0) setTol(0.0000001)
Lasso Regression	setMaxIter(1000) setElasticNetParam(1) setTol(0.0000001)
ARIMA	p = 2 d = 3 q = 3
Random Forests	setMaxDepth(30) setNumTrees(10000)

RR [28] is an addition to linear regression. In this algorithm, the loss function is reformed to reduce model complexity by restricting the summation of the absolute values of the model coefficients. It is worth noting that overfitting is a result of a low alpha value, whereas under-fitting is caused by a high alpha value. Figure 3 shows the comparison between predicted and actual values using RR. Based on our analysis, it appears that Lasso is more effective at predicting outcomes within a two-month period as evidenced by the consistent success of the predictions during this timeframe. This suggests that the use of this algorithm may be particularly useful for forecasting during a limited timeframe. That is why September and October are the two months with a good prediction. As for November 05, 2014, the RR didn't predict well because the value predicted for that day was 0.507 and the actual value was 1.22. In that day, RR performed better compared to Lasso prediction.

EN [29] takes advantage of both previously explained algorithms by bringing L1-norm and L2-norm into play to penalize the model. Figure 4 demonstrates that, within a two-month period, EN is the most effective algorithm for predicting percipitation. This is supported by the consistently successful predictions made during this period. The months of September and October in particular saw relatively accurate predictions by EN. On November 5, 2014, EN outperformed both Lasso and RR, with a predicted value of 0.707 that was relatively close to the actual value of 1.22. These findings suggest that EN is a valuable tool for short-term forecasting. It can be inferred from the Figures 1-3 that Lasso, RR, or EN were not successful in predicting periods with no precipitation. Among the three algorithms, Lasso and RR had the worst performance in predicting these periods, while EN had somewhat better results.

Built on decision trees, the RF modeling technique is used for behavioral analysis and modeling predictions. It has numerous decision trees, each of which represents a different instance of how the classification of data is done when entered into the RF. The forecast chosen by the RF technique is the one that receives the most votes after taking into account each case separately. Figure 5 shows that RF algorithm was effective in predicting daily precipitation levels from September 1, 2014 roughly up to January 11, 2015. This period saw a high level of accuracy in the algorithm's forecasts. However, the period from December 14, 2014 to January 11, 2015 was particularly noteworthy, as the RF algorithm was the only one to accurately predict a non-rain period. This demonstrates the superior performance of the RF algorithm in comparison to other forecasting methods used. The RF algorithm is able to maintain good prediction accuracy over the course of four months. It also made a close alignment between the predicted value of 0.952 and the actual value of 1.22 seen on November 5, 2014. This further reinforces its reliability as a tool for forecasting precipitation levels.

ARIMA [30] is a statistical model used for time series forecasting. It is a combination of an autoregressive model and a moving average model, with an additional term to account for differencing of the data. The model is fit by specifying the orders of the autoregressive and moving average terms and the order of differencing to apply to the data. The ARIMA model demonstrated good performance in predicting future values of the percipitation throughout the six month period, with a particularly close match between the predicted value of 0.916 and the actual value of 1.22 observed on November 5, 2014. It was able to capture the trends and patterns in the data to a certain degree, resulting in mostly accurate predictions. One notable point of the ARIMA model is that it is the only algorithm that produces negative values as predictions. This can be seen in Figure 6, where there was a dry period with sinusoidal negative and positive values close to 0.

EN and RF offer several advantages over ARIMA model in terms of feature selection, handling correlated predictors, handling non-linear relationships and handling non-stationary data. Indeed, EN and RF have feature selection mechanisms that can aid in identifying the most relevant predictors for the model, which can improve the model's performance and reduce overfitting. Additionally, both EN and RF can handle correlated predictors by assigning small or zero coefficients to irrelevant predictors, which can improve the model's interpretability and stability. Furthermore, both models can handle non-linear relationships between the predictors and the response, and can be used to model non-stationary data by using differencing and/or polynomial transformations of the predictors. In contrast, ARIMA is designed to work with stationary data and may require differencing to make the data stationary.

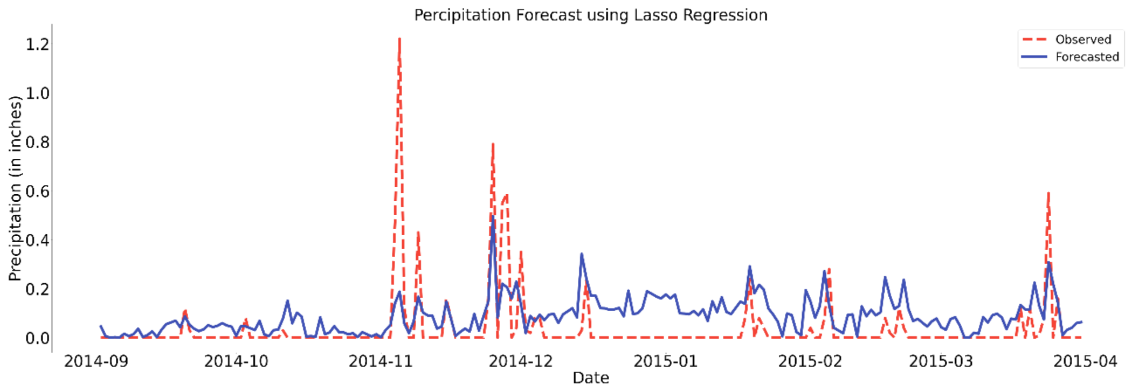

Figure 2. Analyzing the difference between forecasted and observed daily precipitation levels from September 2014 to March 2015 using Lasso

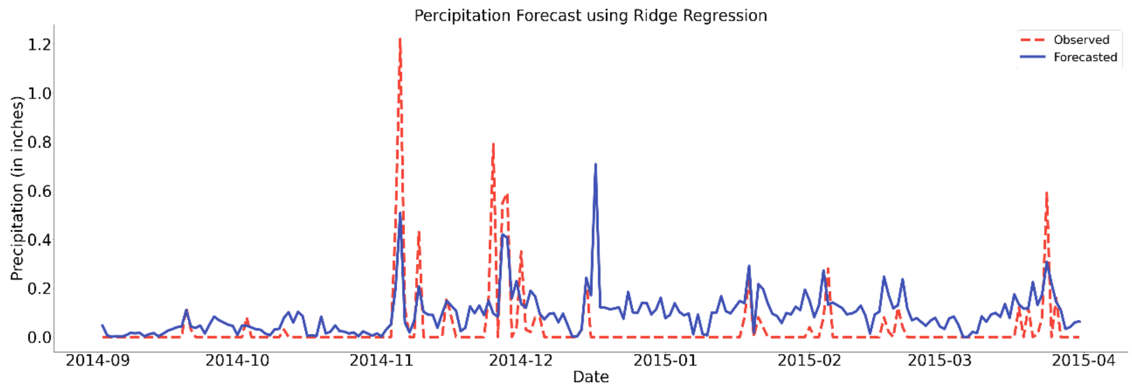

Figure 3. Comparing predicted and actual daily precipitation levels from September 2014 to March 2015 using RR

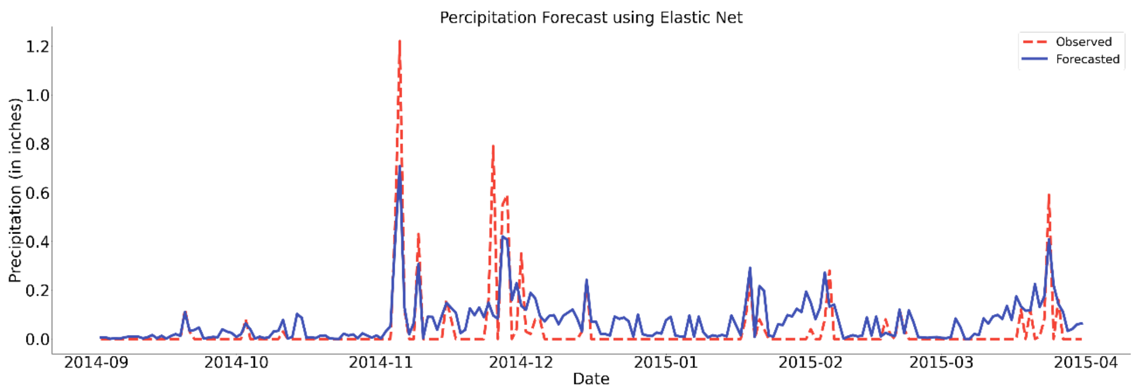

Figure 4. Examining the discrepancy between predicted and actual daily precipitation levels from September 2014 to March 2015 using EN

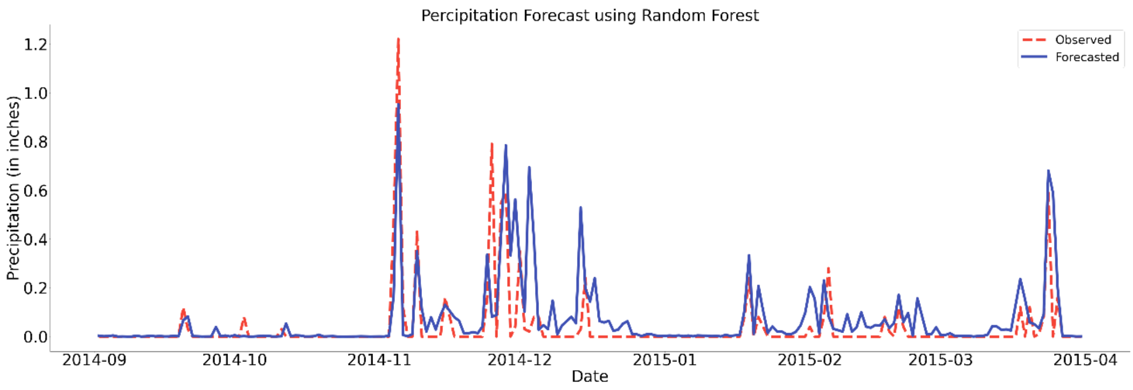

Figure 5. Evaluating the accuracy of daily precipitation forecasts from September 2014 to March 2015 through comparison with observed values using RF

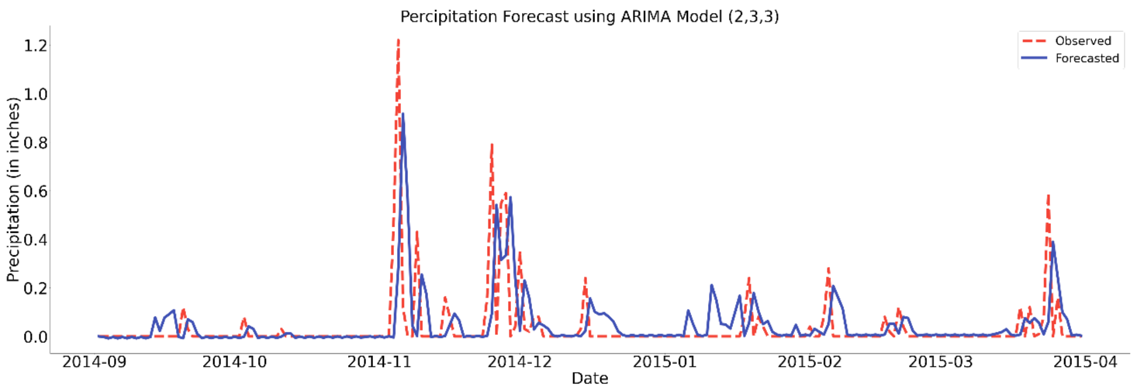

Figure 6. Evaluating the accuracy of daily precipitation forecasts from September 2014 to March 2015 through comparison with observed values using ARIMA

Based on the information provided in Figures 2-4, it appears that the Lasso, RR, and EN algorithms demonstrate good prediction capabilities for the first two months. This is evident from the consecutive sequence of good predictions using these algorithms during this time period. However, after the first two months, the prediction accuracy of these algorithms declines considerably. In contrast, the RF algorithm demonstrated good prediction capabilities for the first four months. This is a significant advantage. It also suggests that the RF algorithm may be more robust and capable to maintain its prediction accuracy for longer periods of time. Overall, these findings suggest that the RF algorithm may be a more reliable choice for predicting outcomes over a longer timeframe. On November 4, 2014, the region of Beni Mellal experienced sudden high precipitation, reaching a value of 1.22 inches. When analyzing the performance of different algorithms under these conditions, it was found that the RF algorithm had the best performance, followed by ARIMA, EN, RR and Lasso. This suggests that the RF algorithm may be particularly effective at predicting outcomes related to heavy rain events.

The weather of the period from December 16, 2014 to January 20, 2015 caracterises as dry. The RF and ARIMA demonstrated the best performance when there was lack of percipitation. In contrast, the performance of the other algorithms fluctuated during this time period. This suggests that RF and ARIMA are particularly effective at predicting outcomes in periods with no precipitation. The period from December 14 to December 28 makes it easy for us to make the following point: machine-learning algorithms are inclined to perform poorly when switching prediction from a period with rain to a rainless point in time. Case in point, when it is raining on December 14, 2014, it took RF fourteen days to adjust gradually its precipitation forecast to give correctly a no-rain reading, which can be explained by the fact that precipitation, in its nature, is nonlinear; that is why Lasso, RR and EN have difficulty reflecting the lack of rain during the period of dry weather. Based on Table 2, it appears that RF is a more effective in terms of MAE compared to ARIMA, Lasso, RR, and EN. Specifically, RF outperforms Lasso by 65.115%, RR by 69.565%, ARIMA by 25.4% and EN by 6.687%. These results suggest that RF is a highly effective algorithm for precipitation prediction in the region of Beni Mellal. The random forest is the algorithm of choice amongst the algorithms chosen, because it displays the lowest MAE. Furthermore, the high accuracy precipitation measurement acquired by RF over six months demonstrates that percipitaiton is correlated with the features used.

Table 2. Shows the MAE obtained using RR, Lasso, RF and EN

Algorithm	MAE
Lasso Regression	0,08489047943
Ridge Regression	0,08717834648
ARIMA	0.06447176951
Elastic Net	0,05485114914
Random Forests	0,05141272751

4. CONCLUSION

Getting low MAE when it comes to precipitation prediction is challenging and complicated. On one hand, the region of Béni Mellal has a particularly difficult pattern to predict as it rains irregularly and intermittently. On the other hand, precipitation prediction in general is highly nonlinear. Therefore, to get better predictions, the model needs to dig deeper to find correlations between features, which explains why RF has been successful. In future work, we will gather extended data about periods containing prolonged precipitation shortages in the past (before 2000) in order to train our models with the help of ML and ANN to predict the end of the prolonged precipitation shortage of the period spanning from 2016 to 2022.

REFERENCES

[1] P. Tangwannawit and K. Saengkrajang, "An internet of things secosystem for planting of coriander (Coriandrum sativum L.)," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 5, p. 4568-4576, Oct. 2021, doi: 10.11591/ijece.v11i5.pp4568-4576.

[2] P. Chandana *et al.*, "An effective identification of crop diseases using faster region based convolutional neural network and expert systems," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, p. 6531-6540, Dec. 2020, doi: 10.11591/ijece.v10i6.pp6531-6540.

[3] H. Ahmed, M. A. Hossain, I. Hossain, S. S. Akhi, and I. J. Lima, "Detection and classification of plant diseases in leaves through machine learning," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 28, no. 3, p. 1676-1683, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1676-1683.

[4] R. Gómez-Chabla, K. Real-Avilés, C. Morán, P. Grijalva, and T. Recalde, "IoT applications in agriculture: a systematic literature review," *In ICT for Agriculture and Environment: Second International Conference, CITAMA 2019, Guayaquil, Ecuador, January 22-25, 2019, Proceedings. Cham: Springer International Publishing* 2019, pp. 68-76, doi: 10.1007/978-3-030-10728-4_8.

[5] J. Xu, B. Gu, and G. Tian, "Review of agricultural IoT technology," *Artificial Intelligence in Agriculture*, vol. 6, pp. 10-22, Jan. 2022, doi: 10.1016/j.aiia.2022.01.001.

[6] I. K. Nti, J. A. Quarcoo, J. Aning, and G. K. Fosu, "A mini-review of machine learning in big data analytics: Applications, challenges, and prospects," *Big Data Mining and Analytics*, vol. 5, no. 2, pp. 81-97, Jun. 2022, doi: 10.26599/BDMA.2021.9020028.

[7] A. H. Gandomi, F. Chen, and L. Abualigah, "Machine learning technologies for big data analytics," *Electronics*, vol. 11, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/electronics11030421.

[8] R. Akhter and S. A. Sofi, "Precision agriculture using IoT data analytics and machine learning," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 5602-5618, Jun. 2022, doi: 10.1016/j.jksuci.2021.05.013.

[9] V. Ramachandran *et al.*, "Exploiting IoT and its enabled technologies for irrigation needs in agriculture," *Water*, vol. 14, no. 5, no. 5, Jan. 2022, doi: 10.3390/w14050719.

[10] L. Rabhi, N. Falih, L. Afraites, and B. Bouikhalene, "Digital agriculture based on big data analytics: a focus on predictive irrigation for smart farming in Morocco," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 24, no. 1, pp. 581-589, Oct. 2021, doi: 10.11591/ijeecs.v24.i1.pp581-589.

[11] A. Singh, "Irrigation planning and management through optimization modelling," *Water Resources Management*, vol. 28, no. 1, pp. 1-14, Jan. 2014, doi: 10.1007/s11269-013-0469-y.

[12] J. Zhang *et al.*, "Sustainable irrigation based on co-regulation of soil water supply and atmospheric evaporative demand," *Nature Communications*, vol. 12, no. 1, p. 5549, Dec. 2021, doi: 10.1038/s41467-021-25254-7.

[13] S. Velmurugan, "An IOT based smart irrigation system using soil moisture and weather prediction," *International Journal of Engineering Research & Technology (IJERT)*, vol. 8, no. 7, 2020.

[14] O. Abrishambaf, P. Faria, L. Gomes, and Z. Vale, "Agricultural irrigation scheduling for a crop management system considering water and energy use optimization," *Energy Reports*, vol. 6, pp. 133-139, Feb. 2020, doi: 10.1016/j.egyr.2019.08.031.

[15] M. A. Shaari, R. Samsudin, and A. S. Ilman, "Forecasting drought using modified empirical wavelet transform-ARIMA with fuzzy c-means clustering," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 11, no. 3, p. 1152-1161, Sep. 2018, doi: 10.11591/ijeecs.v11.i3.pp1152-1161.

[16] G. Yin, T. Yoshikane, K. Yamamoto, T. Kubota, and K. Yoshimura, "A support vector machine-based method for improving real-time hourly precipitation forecast in Japan," *Journal of Hydrology*, vol. 612, p. 128125, Sep. 2022, doi: 10.1016/j.jhydrol.2022.128125.

[17] N. Ramsundram, S. Sathya, and S. Karthikeyan, "Comparison of decision tree based rainfall prediction model with data driven model considering climatic variables," *Irrigation and Drainage Systems Engineering*, vol. 05, no. 03, 2016, doi: 10.4172/2168-9768.1000175.

[18] M. Mohammed, R. Kolapalli, N. Golla, and S. S. Maturi, "Prediction of rainfall using machine learning techniques," *International Journal of Scientific and Technology Research*, vol. 9, no. 01, pp. 3236-3240, 2020.

[19] H. Jdi and N. Falih, “Weather forecast using sliding window algorithm based on hadoop and mapreduce,” *In Advances in Information, Communication and Cybersecurity: Proceedings of ICI2C’21. Springer International Publishing,* 2022, pp. 122-132, doi: 10.1007/978-3-030-91738-8_12.
[20] P. F. Smith, S. Ganesh, and P. Liu, “A comparison of random forest regression and multiple linear regression for prediction in neuroscience,” *Journal of Neuroscience Methods*, vol. 220, no. 1, pp. 85-91, Oct. 2013, doi: 10.1016/j.jneumeth.2013.08.024.
[21] S. Samadianfard, F. Mikaeili, and R. Prasad, “Evaluation of classification and decision trees in predicting daily precipitation occurrences,” *Water Supply*, vol. 22, no. 4, pp. 3879-3895, Apr. 2022, doi: 10.2166/ws.2022.017.
[22] R. N. Rachmawati, A. C. Sari, and Yohanes, “Lasso regression for daily rainfall modeling at Citeko Station, Bogor, Indonesia,” *Procedia Computer Science*, vol. 179, pp. 383-390, 2021, doi: 10.1016/j.procs.2021.01.020.
[23] R. Meenal, K. Kailash, P. A. Michael, J. J. Joseph, F. T. Josh, and E. Rajasekaran, “Machine learning based smart weather prediction,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 28, no. 1, p. 508-515, Oct. 2022, doi: 10.11591/ijeecs.v28.i1.pp508-515.
[24] J. Mom, S. S. Tyokighir, and G. Igwue, “Development of a new rain attenuation model for tropical location,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 24, no. 2, p. 937-948, Nov. 2021, doi: 10.11591/ijeecs.v24.i2.pp937-948.
[25] E. Azhir, M. Hosseinzadeh, F. Khan, and A. Mosavi, “Performance evaluation of query plan recommendation with apache hadoop and Apache Spark,” *Mathematics*, vol. 10, no. 19, Art. no. 19, Jan. 2022, doi: 10.3390/math10193517.
[26] P. Hingley and G. Dikta, “Use of information criteria for finding box-jenkins time series models for patent filings counts forecasts,” *IAENG International Journal of Applied Mathematics*, vol. 52, no. 1, 2022.
[27] J. Ranstam and J. A. Cook, “LASSO regression,” *Journal of British Surgery*, vol. 105, no. 10, pp. 1348-1348, Aug. 2018, doi: 10.1002/bjs.10895.
[28] G. C. McDonald, “Ridge regression,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 1, pp. 93-100, Jul. 2009, doi: 10.1002/wics.14.
[29] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, pp. 301-320, Apr. 2005, doi: 10.1111/j.1467-9868.2005.00503.x.
[30] S. L. Ho, M. Xie, and T. N. Goh, “A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction,” *Computers and Industrial Engineering*, vol. 42, no. 2-4, pp. 371-375, Apr. 2002, doi: 10.1016/S0360-8352(02)00036-0.

BIOGRAPHIES OF AUTHORS

Hamza Jdi has a M.Sc. in business intelligence, received in 2019 from the faculty of science and technologies, University Sultan Moulay Slimane, Beni Mellal, Morocco. Currently, he is a Ph.D. student in the polydisciplinary faculty of Beni Mellal. His research interest is about big data analytics, business intelligence, computer science, digital agriculture, and deep learning. He can be contacted by email at: hamzajdi@gmail.com.

Noureddine Falih has a Ph.D. on computer science, received from the faculty of sciences and technologies of mohammedia, Morocco in 2013. He is an associate professor since 2014 in the polydisciplinary faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco. He has 18 years of professional experience in several renowned companies. His research topics are about information system governance, business intelligence, big data analytics and digital agriculture. He can be contacted by email at: nourfald@yahoo.fr.