Proceedings

Towards smart big weather data management

Chouaib EL HACHIMI[1,*], Salwa BELAQZIZ[1,2], Saïd KHABBA[1,3] Abdelghani CHEHBOUNI[1,4]

1 Mohammed VI Polytechnic University (UM6P), Center for Remote Sensing Applications (CRSA), Benguerir, Morocco.
2 UIZ University, Faculty of Science, LabSIV Laboratory, Department of Computer Science, Agadir, Morocco
3 Cadi Ayyad University, Faculty of Sciences Semlalia, LMFE, Department of Physics, Marrakesh, Morocco
4 Université de Toulouse, Centre d'Etudes Spatiales de la Biosphère (CESBIO), Toulouse, France.
* Correspondence: chouaib.elhachimi@um6p.ma

Abstract: Smart management of weather data is pivotal to achieving sustainable agriculture since weather monitoring is linked to crop water requirement estimation and consequently to efficient irrigation systems. Advances in technologies such as remote sensing and the Internet of Things (IoT) have led to the generation of this data with a high temporal resolution which requires adequate infrastructure and processing tools to gain insights from it. To this end, this paper presents a smart weather data management system composed of three layers: the data acquisition layer, the data storage layer, and the application layer. The data can be sourced from station sensors, real-time IoT sensors, third-party services (APIs), or manually imported from files. It is then checked for errors and missing values before being stored using the distributed database MongoDB. The platform provides various services related to weather data: i) forecast univariate weather time series, ii) perform advanced analysis and visualization, iii) use machine learning to estimate and model important climatic parameters such as the reference evapotranspiration (ET_0) estimation using the XGBoost model (R^2=0.96 and RMSE=0.39). As part of a test phase, the system uses data from a meteorological station installed in the study area in Morocco.

Keywords: Artificial intelligence; big data management; smart agriculture; weather forecasting; MongoDB; NoSQL databases.

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Journal Not Specified* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Since the industrial revolution, our planet is facing real challenges related to climate change [1]. Effects of this latter are going to get worse as we go into the future, especially in the context of a high rate of population growth [2] which reveals another type of issue related to food security. Agriculture is the concerned sector, and agricultural management practices must be optimized to meet the increasing food demand. One step toward this is optimizing water resources usage, which is used mainly in irrigation. There are several types of irrigation, such as surface irrigation, which is the most prevalent mode of irrigation worldwide [3] and pressurized irrigation that requires energy in order to supply water. Whatever the type of irrigation, efficiency can't be achieved without knowing the right amount of water to apply and when to apply it [4]. To achieve this, estimation of the evapotranspiration is pivotal. It is the sum of the evaporation from the soil surface and the transpiration from plants. The evapotranspiration of a crop (ET_c) is estimated by multiplying the crop coefficient (K_c) and the reference evapotranspiration (ET_0) which represents the rate of evapotranspiration for a reference crop (grass) with known properties obtained by monitoring some meteorological parameters such as air temperature, net solar radiation, humidity, and wind speed. Therefore, weather monitoring is essential for accurately managing irrigation water and, in turn, achieve sustainable agriculture.

Thanks to advances in science and technology, we are currently able to collect weather data at high spatial and temporal resolution and with low-cost [5–11]. This huge amount of generated data is unprecedented, and it brings the notion of big data to this field, which requires adequate infrastructure and processing methods.

This paper presents a web platform to store and process this abundant resource (data) using big data analytics and artificial intelligence algorithms.

2. Materials and Methods

2.1. Study area and data

The experimental site (Figure 1) is located 40 km east of Marrakesh city in Morocco (31°39′33.68"N, 7°36′23.586"W, 582 m above mean sea level using the World Geodetic Coordinate System (WGS84[12])). It is an irrigated area of about 2800 ha, which is almost flat. It has a Mediterranean semiarid climate, with around 250 mm of average annual rainfall [13,14] generally recorded between November and the end of April and average annual evapotranspiration (ET_0) of 1600 mm [15].

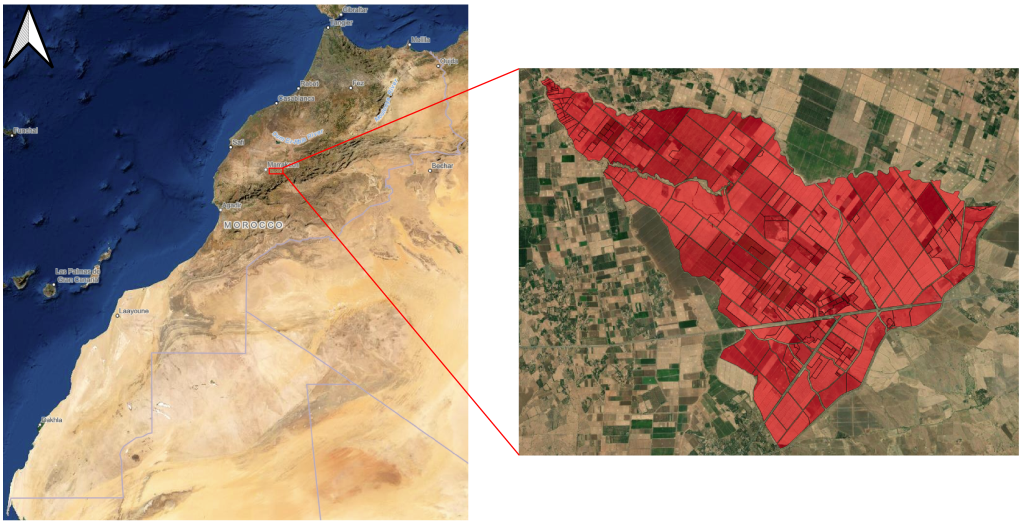

Figure 1. The location of the R3 district study area in Morocco [16].

The collected data from the weather station installed in the study area covers the period between January 3, 2013, to December 31, 2020 (Table 1) at half-hour scale. A detailed description of the different sensors used in the station can be found in previous work conducted in the same study area [17,18].

Table 1. Meteorological station data description.

Variables	Description	Unit
R3_Dv	Wind direction	Degree
R3_Hr	Relative humidity	No unit
R3_Rg	Global solar radiation	$W\ m^{-2}$
R3_Tair	Air temperature	°C
R3_Vv	Wind speed	$m\ s^{-1}$
R3_P30m	Rainfall	mm

2.2. Proposed smart weather data management platform

2.2.1. Overview

The proposed platform adopts a service-oriented architecture providing services that cover all four types of data analytics from descriptive data analysis to prescriptive data analysis (Figure 2).

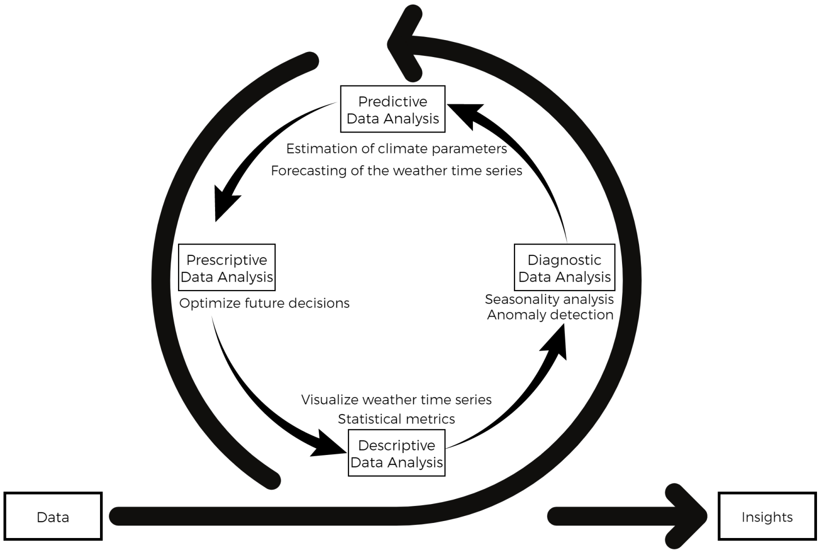

Figure 2. Data analytics cycle.

59 It consists of three layers: the data acquisition layer, the data storage layer, and the
60 application layer (Figure 3). The data can be collected from heterogeneous sources such
61 as meteorological station sensors, real-time IoT-based weather stations, reanalysis data,
62 third-party meteorological services or Application Programming Interfaces (APIs), or
63 manually imported from physical files (CSV, Excel, etc.). This data is then checked and
64 preprocessed to handle missing values, before being stored using the MongoDB NoSQL
65 database. This database represents a suitable solution for our use case, given its ability
66 to handle real-time data, querying, and retrieving large volumes of data, in addition to
67 its scalability and schema-less characteristics. In contrast, fault tolerance issues are a
68 weakness of MongoDB and are the case for almost all distributed databases as well.

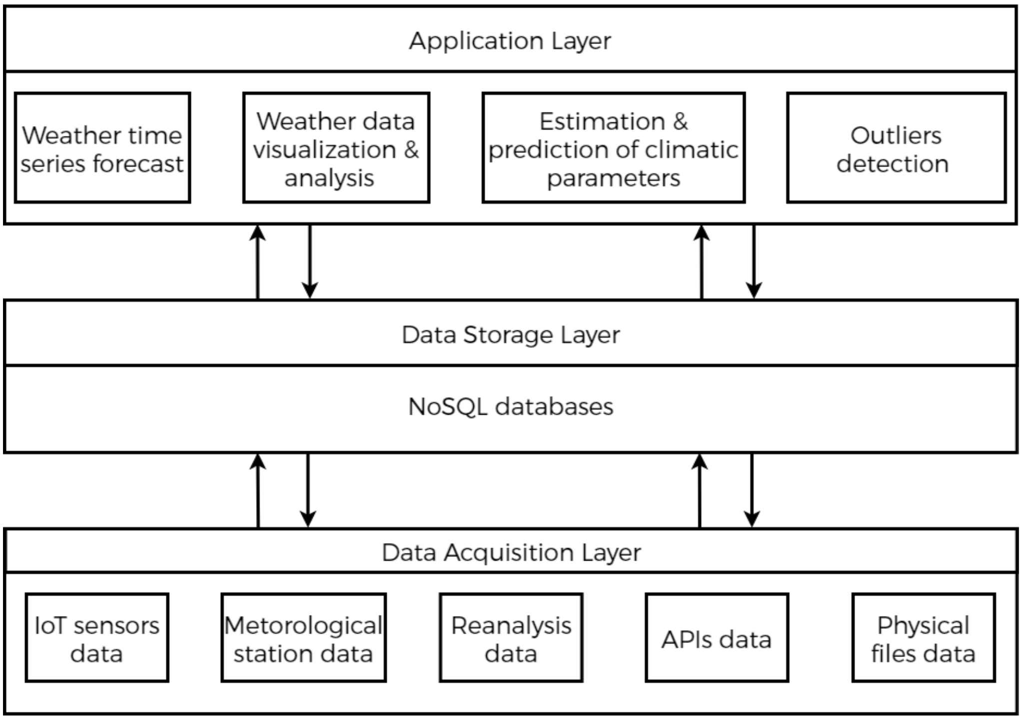

Figure 3. The layered architecture of the platform.

2.2.2. Forecasting service

The forecasting service helps predict the future evolution of weather (air temperature, solar radiation, relative humidity, etc.). It initially uses the Facebook Prophet model [19] given the performance it shows when compared with a Long Short-Term Memory (LSTM [20]) neural network architecture when applied to the same meteorological station data we used [21].

The platform enables the user to specify the number of years needed to train and test the model performance and the time period for the forecast.

Results for univariate time series forecasting were evaluated on the data available in the platform database described in table 1. The evaluation metrics used were: the Coefficient of determination (Equation 1) and Root Mean Squared Error (Equation 2). Results are presented in section 3.1.

$$R^2 = 1 - \frac{\sum_1^n (y_i - \hat{y}_i)^2}{\sum_1^n (y_i - \bar{y}_i)^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_1^n (y_i - \hat{y}_i)^2} \tag{2}$$

2.2.3. Weather data analysis and visualization service

This service provides multiple visualization options such as line charts for the evolution of weather time series and also provides data analysis of stored data. One example of such analysis is the generation of the Pearson correlation coefficient matrix calculated using equation 3 which is then used to investigate the relationship between weather variables.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{3}$$

In the proposed platform, the user choose the target variables and the correlation matrix will then be generated automatically (Figure 4).

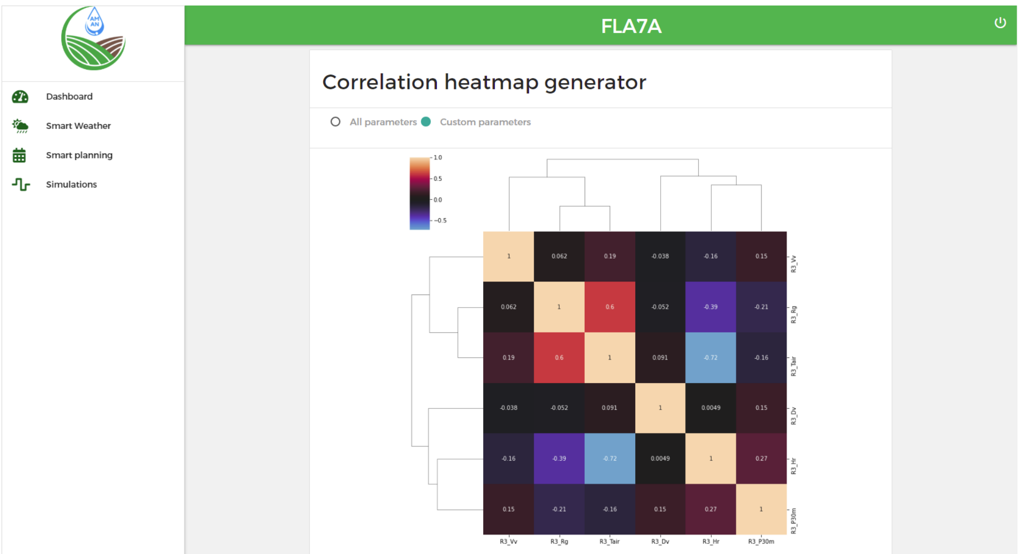

Figure 4. A screenshot of the platform's data analysis service.

Additionally, the platform uses machine learning for modeling and estimation of important parameters used in agriculture such as the evapotranspiration. The flowchart for figure 5 presents an approach that uses the FAO Penman-Monteith ET_0 [22] as a reference method to learn to estimate the reference evapotranspiration from raw station metrological data. Results of this approach are presented in section 3.2.

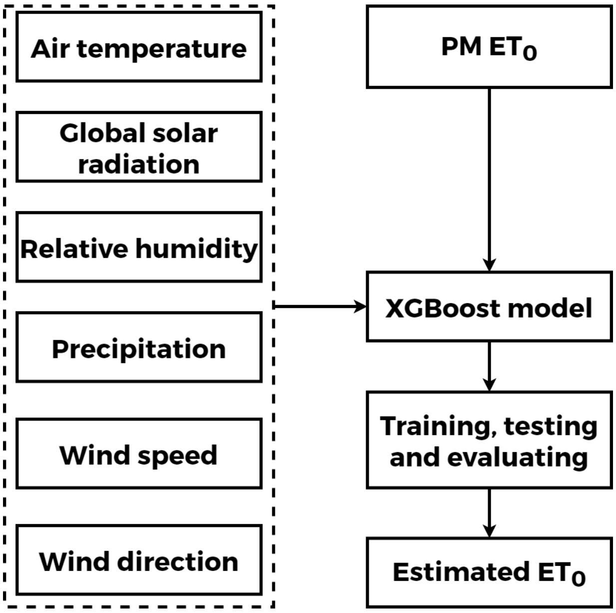

Figure 5. The flowchart of ET_0 estimating using machine learning.

3. Results and discussion

3.1. Univariate time series forecasting service

Table 2 presents results for a one-year forecast of the meteorological time series (2020) given the historical data (from 2013 to 2019) at an hourly scale. The concerned parameters were: Air temperature (T_a), Global solar radiation R_g, Relative humidity H_r, and Wind speed.

Table 2. Performance of univariate forecasting service.

Metric / Parameter	T_a	R_g	H_r	Wind speed
R^2	0.82	0.83	0.48	0.18
RMSE	3.61	114.23	16.73	0.94

3.2. Estimation of climatic parameters using machine learning

To estimate the ET_0, the performance of the XGBoost [23] machine learning model was evaluated on the data available on the platform, which is used to generate the training (80%) and testing (20%) split sets. This model gives a high coefficient of determination R^2=0.96. This means that 96% ET_0 is explained by the trained model. In turn, the RMSE was 0.39. Also, figure 6 shows a positive linear scatter which indicates a good fit for the model.

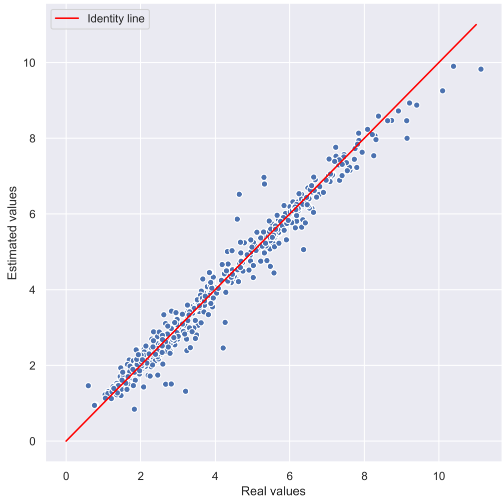

Figure 6. Scatter plot showing the relationship between real and estimated ET_0.

Despite the good results obtained, more studies can be conducted to investigate different machine learning models for both forecasting and regression tasks and integrate them into the platform. In the same sense, current proposed services can be optimized and enriched depending on the challenges and problems faced in the deployment phase. In data storage, for example, other technologies can be investigated when it comes to scaling up the system or in dealing with batch processing tasks. A solution to investigate could be the Hadoop ecosystem (Spark, MapReduce, Hive, etc.).

4. Conclusions

In this paper, we proposed a platform based on artificial intelligence and big data analytics as a way to manage weather data efficiently to assist decision-making in agriculture. The platform offers various services for both farmers and policymakers such as weather data visualization, analysis and estimation. It was tested using data from the meteorological station installed in our study areas, covering the period between 2013 and 2020. The platform will be enriched in future work with other services related to implementing smart agricultural practices.

Author Contributions: Platform and machine learning models development, C.E. S.B.; methodology, C.E., S.B., S.K. and A.C.; writing—original draft, C.E.; writing—review and editing, S.B., S.K. and A.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This study was supported by and conducted within the Center for Remote Sensing Applications (CRSA) (https://crsa.um6p.ma/), at the Mohammed VI Polytechnic University (UM6P) in Morocco.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wade, M., Hoelle, J., Patnaik, R., 2018. Impact of Industrialization on Environment and Sustainable Solutions – Reflections from a South Indian Region. IOP Conf. Ser. Earth Environ. Sci. 120, 012016. https://doi.org/10.1088/1755-1315/120/1/012016
2. Bongaarts, J., 2009. Human population growth and the demographic transition. Philos. Trans. R. Soc. B Biol. Sci. 364, 2985. https://doi.org/10.1098/RSTB.2009.0137
3. Frisvold, G., Sanchez, C., Gollehon, N., Megdal, S.B., Brown, P., 2018. Evaluating Gravity-Flow Irrigation with Lessons from Yuma, Arizona, USA. Sustain. 2018, Vol. 10, Page 1548 10, 1548. https://doi.org/10.3390/SU10051548
4. Nafchi, R.A., 2021. Evaluation of the Efficiency of the Micro-irrigation Systems in Gardens of Chaharmahal and Bakhtiari Province of Iran. http://www.sciencepublishinggroup.com 6, 106. https://doi.org/10.11648/J.IJAE.20210603.12

5. Math, R.K.M., Dharwadkar, N. V., 2019. IoT Based low-cost weather station and monitoring system for precision agriculture in India. Proc. Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2018 81–86. https://doi.org/10.1109/I-SMAC.2018.8653749
6. Majumdar, P., Mitra, S., 2021. IoT and Machine Learning-Based Approaches for Real Time Environment Parameters Monitoring in Agriculture: An Empirical Review. Agric. Informatics 89–115. https://doi.org/10.1002/9781119769231.CH5
7. Kumar, S., Ansari, M.A., Pandey, S., Tripathi, P., Singh, M., 2020. Weather Monitoring System Using Smart Sensors Based on IoT. Lect. Notes Networks Syst. 106, 351–363. https://doi.org/10.1007/978-981-15-2329-8_36
8. Kodali, R.K., Mandal, S., 2017. IoT based weather station. 2016 Int. Conf. Control Instrum. Commun. Comput. Technol. ICCICCT 2016 680–683. https://doi.org/10.1109/ICCICCT.2016.7988038
9. Mittal, Y., Mittal, A., Bhateja, D., Parmaar, K., Mittal, V.K., 2016. Correlation among environmental parameters using an online Smart Weather Station System. 12th IEEE Int. Conf. Electron. Energy, Environ. Commun. Comput. Control (E3-C3), INDICON 2015. https://doi.org/10.1109/INDICON.2015.7443621
10. Djordjević, M., Jovičić, B., Marković, S., Paunović, V., Danković, D., 2020. A smart data logger system based on sensor and Internet of Things technology as part of the smart faculty. J. Ambient Intell. Smart Environ. 12, 359–373. https://doi.org/10.3233/AIS-200569
11. Djordjevic, M., Dankovic, D., 2019. A smart weather station based on sensor technology. Facta Univ. - Ser. Electron. Energ. 32, 195–210. https://doi.org/10.2298/FUEE1902195D
12. Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems. Second Edition, 1991.
13. Er-Raki, S., Chehbouni, A., Duchemin, B., 2010. Combining Satellite Remote Sensing Data with the FAO-56 Dual Approach for Water Use Mapping In Irrigated Wheat Fields of a Semi-Arid Region. Remote Sens. 2010, Vol. 2, Pages 375-387 2, 375–387. https://doi.org/10.3390/RS2010375
14. Belaqziz, S., Khabba, S., Kharrou, M.H., Bouras, E.H., Er-Raki, S., Chehbouni, A., 2021. Optimizing the Sowing Date to Improve Water Management and Wheat Yield in a Large Irrigation Scheme, through a Remote Sensing and an Evolution Strategy-Based Approach. Remote Sens. 2021, Vol. 13, Page 3789 13, 3789. https://doi.org/10.3390/RS13183789
15. Er-Raki, S., Chehbouni, A., Guemouria, N., Duchemin, B., Ezzahar, J., Hadria, R., 2007. Combining FAO-56 model and ground-based remote sensing to estimate water consumptions of wheat crops in a semi-arid region. Agric. Water Manag. 87, 41–54. https://doi.org/10.1016/J.AGWAT.2006.02.004
16. Tiles (C) Esri – Source: Esri, USGS | GEOMATIC, Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, © OpenStreetMap contributors, HERE, Garmin, FAO, NOAA, USGS | Earthstar Geographics, and the GIS User Community
17. Le Page, M., Toumi, J., Khabba, S., Hagolle, O., Tavernier, A., Hakim Kharrou, M., Er-Raki, S., Huc, M., Kasbani, M., Moutamanni, A. El, Yousfi, M., Jarlan, L., 2014. A Life-Size and Near Real-Time Test of Irrigation Scheduling with a Sentinel-2 Like Time Series (SPOT4-Take5) in Morocco. Remote Sens. 2014, Vol. 6, Pages 11182-11203 6, 11182–11203. https://doi.org/10.3390/RS61111182
18. Er-Raki, S., Chehbouni, A., Khabba, S., Simonneaux, V., Jarlan, L., Ouldbba, A., Rodriguez, J.C., Allen, R., 2010. Assessment of reference evapotranspiration methods in semi-arid regions: Can weather forecast data be used as alternate of ground meteorological parameters? J. Arid Environ. 74, 1587–1596. https://doi.org/10.1016/J.JARIDENV.2010.07.002
19. Taylor, S.J., Letham, B., 2018. Forecasting at Scale. https://doi.org/10.1080/00031305.2017.1380080 72, 37–45.
20. Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
21. El Hachimi, C., Belaqziz, S., Khabba, S., Chehbouni, A., 2021. Towards precision agriculture in Morocco: A machine learning approach for recommending crops and forecasting weather. Proc. - 2021 Int. Conf. Digit. Age Technol. Adv. Sustain. Dev. ICDATA 2021 88–95. https://doi.org/10.1109/ICDATA52997.2021.00026
22. PENMAN, H.L., 1948. Natural evaporation from open water, hare soil and grass. Proc. R. Soc. Lond. A. Math. Phys. Sci. 193, 120–145. https://doi.org/10.1098/RSPA.1948.0037
23. Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 13-17-August-2016, 785–794. https://doi.org/10.1145/2939672.2939785