Digital agriculture based on big data analytics: a focus on predictive irrigation for smart farming in Morocco

Loubna Rabhi[1], Noureddine Falih[2], Lekbir Afraites[3], Belaid Bouikhalene[4]
[1,3]Mathematics and Applications Laboratory, Faculty of Sciences and Technics, Sultan Moulay Slimane University, Beni Mellal, Morocco
[2,4]Laboratory of Innovation in Mathematics, Applications and Information Technologies, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco

Article Info

Article history:

Received Jun 24, 2021
Revised Aug 26, 2021
Accepted Aug 30, 2021

Keywords:

Agriculture 4.0
Big data analytics
Digital agriculture
Machine learning
Predictive irrigation
Smart farming

ABSTRACT

Due to the spead of objects connected to the internet and objects connected to each other, agriculture nowadays knows a huge volume of data exchanged called big data. Therefore, this paper discusses connected agriculture or agriculture 4.0 instead of a traditional one. As irrigation is one of the foremost challenges in agriculture, it is also moved from manual watering towards smart watering based on big data analytics where the farmer can water crops regularly and without wastage even remotely. The method used in this paper combines big data, remote sensing and data mining algorithms (neural network and support vector machine). In this paper, we are interfacing the databricks platform based on the apache Spark tool for using machine learning to predict the soil drought based on detecting the soil moisture and temperature.

Corresponding Author:

Loubna Rabhi
Mathematics and Applications Laboratory, Faculty of Sciences and Technologies
Sultan Moulay Slimane Universit
Beni Mellal, Morocco
Email: rabhi.lubna@gmail.com

1. INTRODUCTION

Agriculture is one of the most pillar sectors in Morocco. In some cities, it is the main occupation, which most people depend on for their livelihood. Nowadays, agriculture is becoming more productive because of advancements and the latest technologies as sensors, devices, and information technology that help generate a large amount of data [1]. However, irrigation scheduling can be more efficient by using new technologies based on accurate crop and irrigation models, weather forecasting, sensors, IoT, AI, and cloud computing [2]. Irrigation scheduling is considered one of the most prevalent problems of the agriculture system. The farmer must answer the questions “when do I water?” and “how long do I water?”. Starting an irrigation cycle too early and/or running an irrigation cycle too long is considered overwatering and a waste of money. Likewise, beginning an irrigation cycle too late and/or running an irrigation cycle too short is deemed underwatering. Both overwatering and underwatering can cause crop damage, reduced yields, and poor crop quality. The best way to avoid these damages and minimize their financial and practical impact on crops is to implement a Smart and optimal irrigation system [3] by making a decision based on big data analytics [4].

Moreover, big data in the irrigation sector designs a large amount of structured, semi-structured, and unstructured agricultural data (moisture, temperature, soil, and crop). Analytics of such data will be profitable for better decision-making. However, it is too difficult to process and analyze this massive data using traditional databases [5]. So, to get the best result out of big data analytics, other computer technics as machine learning are required. In this article, we are simulating an irrigation scheduling context on one hand. On the other hand, from a set of recorded data, we will predict whether the soil is dry or not based on its state of moisture and crop temperature. Thus, we will exploit algorithms of machine learning to predict the position of the pumping motor. If the soil is dry, the pump will be set to "ON" else to "OFF".

The organization of the paper proceeds as follows: section 2 provides the relevant literature on the smart agriculture concept based on big data analytics. Section 3 is an overview of the research method used in this paper. The results are discussed in section 4. And finally, the conclusion and future work.

2. RELATED WORKS

Before proposing a research method, we will study some recent researches on big data analytics in agriculture. The summary of success findings of smart agriculture based on big data analytics is presented in Table 1. It explains the background of recent studies that treat the same subject of big data analytics applied in digital agriculture in general and particularly in irrigation scheduling. Based on the results of this study of the existing, we will try to propose a methodology for our paper.

Table 1. Related works

Code	Authors	Problem-focused on	Work done	Outcome	Citation
1	Ramya *et al.*, 2020	IoT framework for smart irrigation using machine learning technique	Proposed a smart irrigation system to predict the irrigation requirement of the field using machine learning.	-Training methods in machine learning (ML) using the collected real-time data. -A low-cost prototype model with advanced technological features. -Observed results are optimal with 90% accuracy	[6]
2	Goldstein *et al.*, 2018	Applying ML on sensor data for irrigation recommendations.	Discussed different regression and classification algorithms applied on a dataset to develop models that were able to predict the weekly irrigation plan.	The best regression model was Gradient Boosted Regression Trees, with 93% accuracy, and the best classification model was the boosted tree classifier, with 95% accuracy.	[7]
3	Evert *et al.*, 2017	Big data for weed control and crop protection	Big data analytics models are outlined, together with algorithms for training them. Advanced tools to process big data for actionable information to farmers are reviewed.	- Adopting an appropriate machine learning model (a neural network). - Training the model using an appropriate algorithm (a gradient descent method).	[8]
4	Fernandez *et al.*, 2020	Smart soil monitoring and water conservation using irrigation on technology	Discussed the role of IoT in the irrigation sector. They used different sensors: soil moisture, humidity, temperature sensors to provide a user irrigation solution for soil monitoring and water conservation.	Web Application that helps farmer to check any time the water level (high or low). Therefore, he can set the pump on or off.	[9]
5	Shekhar *et al.*, 2017	Intelligent IoT based automated irrigation system.	K- nearest neighbor classification deployed for analyzing sensor data for prediction irrigating.	Predicting the soil condition based on trained data set using K-NN machine learning algorithm	[10]
6	Sanchez *et al.*, 2020	A decision system for irrigation: Analysis and modeling of various learning technics.	Discussed several learning techniques to determine the goodness and error relative to decisions in irrigation.	The results obtained lead to the linear regression (LR), the random forest regression (RFR), and the support vector regression (SVR) methods are valid engines to develop automatic irrigation scheduling.	[11]
7	Sagar and Cauvery, 2018	Agriculture data analytics in crop yield estimation.	Presented insights on various big data analytics methods applied to crop yield prediction.	Challenges and opportunities in the field of big data analytics. Agriculture using IT trends analytics is in its infancy.	[12]
8	Tantalaki *et al.*, 2019	Data-driven decision making in precision agriculture: the rise of big data in agricultural systems.	Developed a review of data science techniques, especially machine learning ML technics in agricultural systems.	Advanced ML methods like convolutional neural networks, Advanced neural networks, support vector machines, decision trees, random forests and deep learning techniques offer higher accuracy, robustness, flexibility, and generalization performance.	[13]
9	Fan *et al.*, 2015	Prediction of crop yield using Big Data	Provided a new method to predict the crop yield based on big-data analytics technique.	The nearest neighbor's method using MapReduce on weather data gives best accuracy and time processing.	[14]

Indeed, once the relevant data are prepared and stored, the big data analytics process extracts valuable insights. Conventionally, agricultural applications use traditional statistical methods like regression, analysis of variance, and principal component analysis methods. However, big data applications require new and advanced methods. Indeed, traditional statistics may be inadequate to deal with an outsized number of big data variables that might be relating in a complex, non-linear manner. The high complexity and non-linearity of problems faced in agriculture require methods ready to integrate data coming from different sources and exploit the insight contained in the reference samples. These methodologies are represented by ML techniques. Artificial neural networks, advanced neural networks, support vector machines, decision trees, and deep learning are common MLtechniques, frequently applied for agricultural management purposes:

- Artificial neural networks (ANN): it is among the most important algorithms that perform learning tasks. ANN are inspired by biological neural networks. Similarly, the ANN can learn from complicated data and provide insights through predictions or classifications [15]. With ANN, function f is modeled as a functional composition of basic computational elements, for example neurons, where each neuron consists of a linear activation function, parameterized by a weight vector w and a non-linear transfer function. ANN have been used in many agricultural use cases, including weed control and crop protection [8], and data-driven decision making in precision agriculture [12]. Large training datasets are needed for accurate results, which is a time-consuming procedure.
- Advanced neural networks: it arose to address several drawbacks of the conventional ANNs. They have good generalization performance and learn much faster than conventional ANNs. Advanced neural networks such as adaptive neuro-fuzzy inference systems (ANFIS) and extreme learning machines (ELM) avoid the fuzziness characteristic of agricultural data [16].
- Support vector machines (SVMs): these are a family of ML algorithms that have good intrinsic generalization ability and are relatively robust to noise in the training data. Given a training data set, SVMs make at each data point an error of at most 'e', where 'e' is a predefined small positive number. SVMs were initially intended for linear models but non-linear ones can be made by transforming input X through a non-linear mapping $\Phi(X)$ and then applying conventional linear SVMs over features $\Phi(X)$. According to the data, the performance of SVM is the same order, or even better, than a neural network (NN).
- Decision tree: the decision tree algorithm belongs to the family of supervised learning algorithms. It is considered as one of the predictive modeling approaches. The decision tree algorithm is useful for creating a training model that can be used for predicting the class or value of the target variable by learning simple decision rules inferred from prior data (training data) [17].
- Deep learning is one of the main machine learning technologies. It is a quite promising technique that extends classical ANN by adding more complexity ("depth") to the model. Deep learning requires a large volume of data and the calculating power of machines to work well [18].

3. SMART IRRIGATION BASED ON BIG DATA ANALYTICS AND MACHINE LEARNING

Nowadays the farming has become digitalized. Even small-scale farmers use the information collected from large datasets and precision analytics. The use of large information sets and the digital tools for collecting, preparing, and analyzing them together referred to as big data analytics. At this level, we begin by projecting the definition of big data in the field of irrigation before explaining the processing of this data using the techniques of big data analytics and machine learning.

3.1. Big data concept in irrigation

Today irrigation sector is embedded with advanced tools and sensors. These devices are connected to crops and soil to collect all the information in the real world by sensing techniques and then transform them into digital information. This is what is called the internet of things (IoT). These smart devices will generate impressive amounts of data which are considered big data. Authors in [5] define big data by 5Vs:

- Volume: using digital farming tools in irrigation, big data can be collected from different sources: manual acquisition, weather data (rainfall, temperature, and humidity), crop data (crop height and surface resistance), soil data (texture, structure and depth), aerial systems (drones and aircraft).
- Velocity: data gathered in near real-time. The periodicity may go down to one minute.
- Variety: gathered data can be in heterogeneous formats (photo, excel file, and graph) [19].
- Veracity: agricultural data must be correct and accurate for applying analytics process.
- Value: means the usefulness of data among this huge amount of data.

3.2. Big data analytics and machine learning for smart irrigation scheduling

Big data analytics is an approach that exploits devices connected to the internet of things (IoT) for collecting and exchanging data with each other. Through sensors placed directly on the plant and soil, internet-connected devices allow measures that concern soil and crop conditions [20]. Coupled with weather data (rain, temperature), these big data will be transmitted via wireless sensor network then stored and analyzed with advanced tools of big data analytics to extract useful insights [10]. Big data analytics aims to facilitate the task to the farmer by combining his experience with the results of the analysis of the collected massive data. He can have an estimation of daily crop irrigation needs by evaluating the historical climate data that are daily updated during the irrigation season, calculating the daily soil water balance, evaluating the soil water data, determining the impact of the weather on the plant and determining the driest areas that require watering them first [17].

Moreover, the main actor who can contribute to solving the irrigation scheduling problem is a data scientist who is an analyst with a deep mastery of analytical tools having statistics and mathematics skills. His goal is to make sense of the data being manipulated and extract value from it to help the company make strategic or operational decisions. The data scientist can be responsible for gathering agricultural data from different sources and sensors (ingestion), storing data (storage), preparing data by cleaning and filtering them (processing), analyzing data using smart tools and analytical techniques such as machine learning and deep learning (advanced processing), and finally restituting the result in an interactive manner (visualization) [21]. Among the technologies that contribute to the development of big data analytics is machine learning [22]. It is a computer program that can learn from data. It improves its performance at some tasks through experience.

4. RESEARCH METHOD

The data handled in the irrigation sector is sliced and stored in a distributed and redundant parallel file system which is databricks file system (DBFS). Moreover, the current dataset is analyzed by machine learning implemented by the support vector machine classifier. In this section, we present the technical environment of this work and we define the proposed research methodology based mainly on the following technologies.

4.1. Apache Spark and databricks platform

Apache Spark is a fast engine for large-scale data processing. It is an open-source tool and unified framework for big data analytics. It enables combining tools like structured query language (SQL), statistics, machine learning, and graph processing to gain meaningful insights. Spark is known for its parallel distributed processing on commodity hardware [23]. It enables fast speeds due to in-memory caching and directed acyclic graph (DAG) processing engine. It comes to correct the limitations of Hadoop. Indeed, Spark is one hundred times fast than Hadoop's MapReduce for in-memory computations. Moreover, It can sort 100 TB of data (1 trillion records) three times faster and using ten times fewer machines than Hadoop's MapReduce shows in Table 2 [24].

Databricks is a big data processing platform founded by apache Spark. It is created for data scientists, engineers, and analysts to provide an alternative to the MapReduce system and provides a just-in-time cloud-based platform for big data processing. Databricks as an open-source distributed computing framework provided automated cluster management. It supports different programming languages: Python, Scala, R and SQL as well as data science frameworks [25]. It created its file format, data bricks file system (DBFS) which is distributed and mounted directly on the unified platform to run all types of analytics workloads [26]. Moreover, it allows deploying advanced big data analytics through its virtual analytics platform.

Table 2. Spark vs Hadoop

Criterion	Hadoop	Spark
Data Size	Large-scale data	Large-scale data
Aspect	Object oriented	Functional
Main data structure	Class	RDD (resilient distributed datasets)
Usage	Disk (HDFS)	In-memory caching
Speed	Slower compared to Spark	100 times fast than Hadoop's MapReduce for in-memory computations and 10 times faster on disk
Written in	Java	Scala

4.2. Machine learning and data science models

In the demonstration, we will work with an existing dataset where information is collected from sensors (temperature and moisture). The used dataset will be stored in the databricks platform. This latter provides a simple and an user interface for uploading data from files and databases. Once, data are stored, we will create a cluster then create a notebook on databricks which is attached to this cluster shows in Figure 1.

As a research methodology, the machine learning technique is chosen to analyse these data then respond to irrigation use case. Moreover, the neural network and support vector machines are the proposed techniques to be modeled using Spark engine to predict the state of the pumping motor: whether it should be ON or OFF based on moisture and temperature conditions. In this section, we define the used dataset and discuss the application of a data science model, before presenting the generated results 597-60524].

Big data analytics can be done using machine learning technique by following as shown in Figure 2:

- Data ingestion: find appropriate data that may be stored in databases and comma separated values (CSV) files. We can use big data tools (apache sqoop and apache flume) to collect data. Then merge it into a single table.
- Data transformation: preprocessing and cleaning data. This step is for preparing data to be analyzed later. We can use Apache Spark, Hive, or Pig tools.
- Training data: prepare a dataset from which we learn. Training means the process of going over the training set and producing a model.
- Test data: prepare data set on which we check if our model works. For the experimentation, the dataset consisting of 200 records. 80% of the records are used for training the NN and SVM classifiers and 20% are used for testing. The data handled is sliced and stored into databricks file system DBFS.
- Data science model: a logical and sequential sequence of steps to solve a given problem. Once we receive an organized and cleaned table, we can model a set of algorithms based on analytical techniques, then choose according to a survey which one gives the best result for deployment.
- Prediction: based on SVM and neural network techniques applied to the current dataset, we can first predict if the soil is dry or not based on moisture and temperature conditions using a function that predicts target values of the test data given in parameter. Then, evaluate the prediction using a text report showing the classification metrics: accuracy, precision, recall (sensitivity), and f1-score.

In the demonstration, we will be based on the support vector machines and neural network techniques to predict the state of the pumping motor based on moisture and temperature. The implementation of the proposed models will be done using the notebook that we have already created.

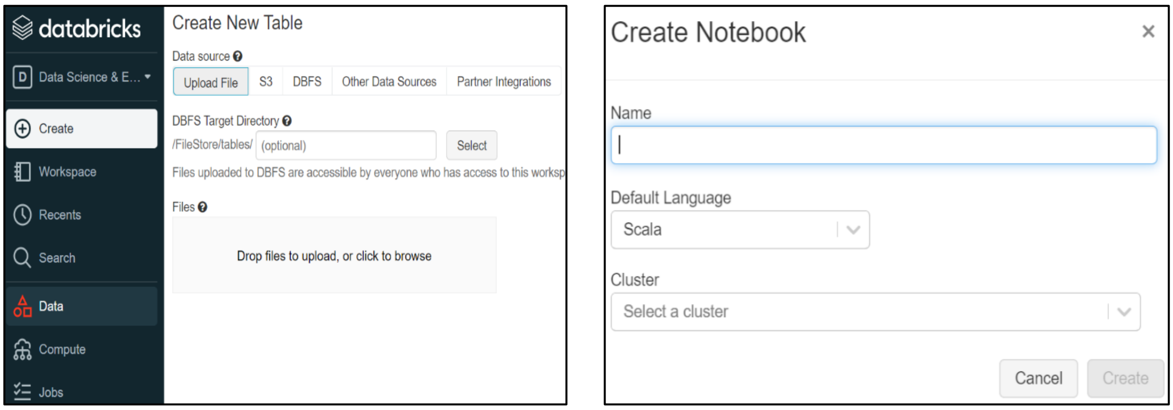

Figure 1. Interface of databricks

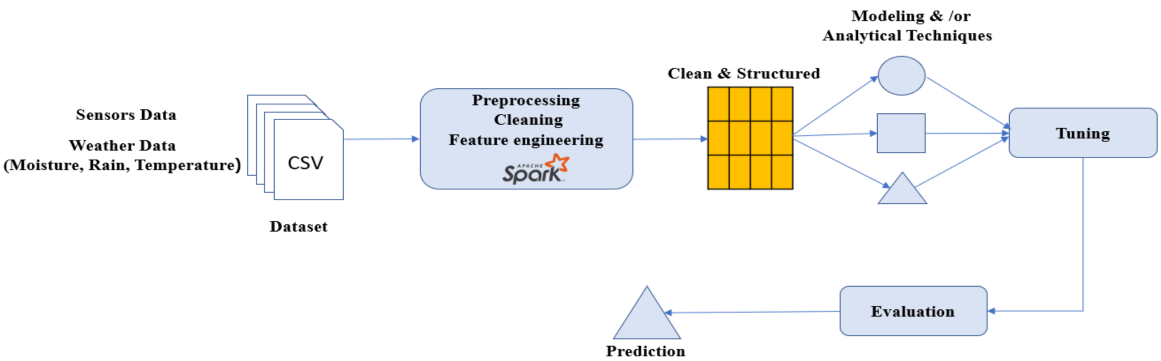

Figure 2. Data advanced processing using machine learning

SVMs is a set of machine learning algorithms useful to build and train a model using cell records, and classify them into specific classes. Support vector machines works by mapping data to a high-dimensional features space so that data points can be categorized, even when the data are not otherwise linearly separable (this gets done by Kernel function of SVM classifier). As a separator between the categories is found, the data is transformed in such a way that the separator could be drawn as a hyperplane. The classifcation of the SVM model is described by the equation [27]:

$$Y = sgn(\vec{x} \cdot \vec{w} - b)$$

where:
x: is the feature vector
w: is the model weights vector
b: is the bias value

Neural network (NN) tries to predict whether the pump will be ON or OFF. Indeed, the input layer contains data from moisture and temperature. Moreover, the output will be a number 0 or 1, representing how likely the pump is ON or OFF. One other hidden layer will be added between the input and output layers. The parameters (neurons) of those layers will decide the final output. All layers will be fully connected. Finally, the prediction will be given by the final (output) layer of the network [26].

5. RESULTS AND DISCUSSION

This section presents the results obtained from the experiment. We present the prediction rate of the class "pump" for the used algorithm. In this study, we used supervised learning because we have a labeled dataset, composed of a set of existing data including the target values. The targets can have two possible outcomes (pump ON or OFF). In this simulation, we used one of the most popular data mining algorithms: SVM and NN.

As mentioned earlier, we have chosen to use a big data analytics platform called "databricks". The used dataset is stored in distributed file system on the databricks platform (15,3 GB memory, 2 cores, 1 DBU, databricks file system). Moreover, the dataset contains a large number of data related to moisture, temperature, and position of the pumping motor. If the pump field contains 1 that is means the pumping is ON and the soil is dry. If the pump field contains 0 that is means the pumping is OFF and the soil is watered. To analyze the dataset in databricks, a cluster is first created to serve as a database. Next, a table is created and attached to the created cluster. Then, this table will be filled with our dataset by importing a CSV file containing data. The CSV file is composed of three columns, the two first columns contain the parameters (moisture and temperature), and the last one represents the corresponding class (pump) as shown in Figure 3.

moisture	temp	pump
638	16	1
522	18	1
741	22	1
798	32	1
690	28	1
558	23	1
578	12	1

Figure 3. Example of the dataset used in the simulation

The Figure 4 explain that the two classes (pump ON and pump OFF) are well clustered. They seem completely linearly separable (Figure 3). The next steps consist of modeling the support vector machines algorithm that should be able to find an optimal hard margin hyperplane that splits these two classes. The optimal hyperplane is defined by the plane that maximizes the perpendicular distance between the hyperplane and the closest samples. This perpendicular distance can be spanned with support vectors.

Spark MLlib offers an implementation of SVM with a linear kernel suitable for linearly separable data. A SVM constructs a hyperplane that can be used for classifying pump classes (ON or OFF) as shown in Figure 5. These are plots of eventual optimal separating hyperplane and the parallels to the separating hyperplane that pass through the support vectors. The discussed SVM is restricted to binary linearly separable data only. If there is a classification of non-linearly separable data, kernel-based SVMs techniques are used.

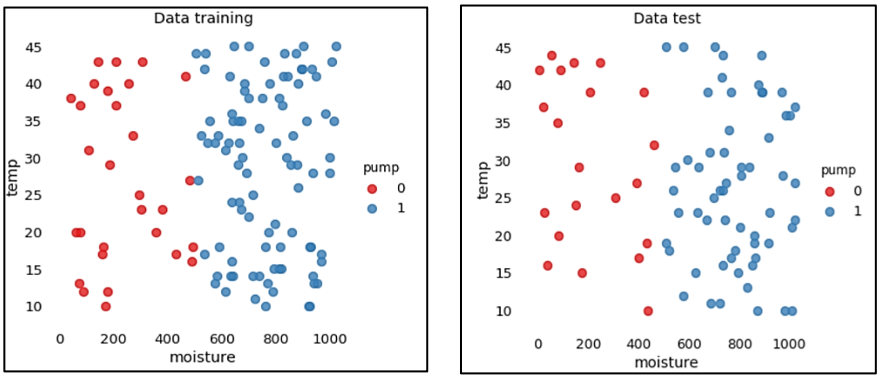

Figure 4. Distrubution of training set and test set

Figure 5. Linear SVM

Neural network (NN): We first start by creating the layers of the proposed model in the constructor. Then, we call the forward method that accepts the input (moisture and temperature) and allows it to flow through each layer. Finally, we apply the model to the training set. Moreover, we have used the TensorBoard interface to visualize the graph, debug, and optimize the model. It helps to track metrics like loss and accuracy as shown in Figure 6. The first graph in left describes the metric loss function that serves to minimize the loss function to make few errors by the model. We can observe that the plot of loss decreases from 0.030 to 0.016 that is means our model is improving. Similarly, the second graph in in right denotes accuracy. We can observe that the plot rises from 0.945 to 0.988 that is means our accuracy of the model is increasing gradually. However, the performance of the algorithms is expressed using classification metrics. We obtained a high prediction rate of pumping motor occurrence (98,50% for neural networks and 97,00% for SVM).

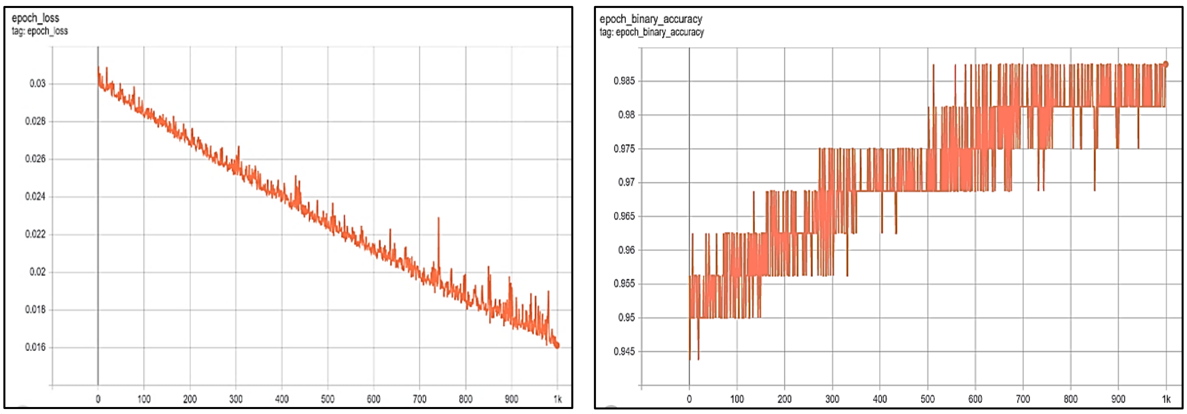

Figure 6. Loss function and NN accuracy for NN technique

Both used algorithms NN and SVM gave good results, but we can notice that neural networks outperformed SVM in terms of accuracy and efficiency shows in Table 3. To sum up, the results of the training suppor vector machine and neural network are used to detect if the soil is dry or not by knowing moisture and temperature conditions. Proposed detection methods have higher accuracy and are efficient to process large amounts of distributed agricultural data.

Table 3. NN Vs SVM accuracy

		Precision	Recall	F1 Score	Support
	Pump off (0)	1,00	0,95	0,97	9
NN	Pump on (1)	0,98	1,00	0,99	31
	Accuracy			0,985	40
	Pump off (0)	0,98	1,00	0,94	9
SVM	Pump on (1)	1,00	0,97	0,98	31
	Accuracy			0,970	40

6. CONCLUSION

In this paper, we dealt with a new relevant topic which is big data analytics for smart irrigation. In front of intelligent agriculture that embodies the enormous global use of internet of thing (IoT) analysis solutions, Irrigation scheduling is becoming more precise. It offers new technical skills to the farmer, less hardship on the farm, water resources for future generations, reduced losses, increased yields and improved product competitiveness. The present irrigation dataset is analyzed using NN and SVM algorithms based on moisture and temperature factors. The simulation was run using the big data platform "databricks" which implements the spark framework as an engine for big data processing. The models proved good results for SVM and neural network (SVM 97.00%, NN 98.50%). In the future work, we will propose an extension of the 19440 enterprise metamodel, adding specific constructs relating to advanced technologies (big data analytics tools, IoT, AI, cloud, and machine learning) and reflecting a new metamodel likely to be projected on a smart farm as part of an agriculture 4.0.

REFERENCES

[1] A. H. Ali, R. F. Chisab, and M. J. Mnati, "A smart monitoring and controlling for agricultural pumps using LoRa IoT technology," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 1, pp. 286–292, 2019, doi: 10.11591/ijeecs.v13.i1.pp286-292.

[2] M. M. Subashini, S. Das, S. Heble, U. Raj, and R. Karthik, "Internet of things based wireless plant sensor for smart farming," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 2, pp. 456–468, 2018, doi: 10.11591/ijeecs.v10.i2.pp456-468.

[3] M. K. I. A. Rahman, M. S. Z. Abidin, S. Buyamin, and M. S. A. Mahmud, "Enhanced Fertigation Control System Towards Higher Water Saving Irrigation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 3, pp. 859–866, 2018, doi: 10.11591/ijeecs.v10.i3.pp859-866.

[4] T. T. Gurmessa, "A Big Data Analytics Framework in Climate Smart Agriculture," *Computer Engineering and Intelligent Systems*, vol. 10, no. 6, pp. 1–6, 2019, doi: 10.7176/ceis/10-6-01.

[5] L. Rabhi, N. Falih, A. Afraites, and B. Bouikhalene, "Big Data Approach and its applications in Various Fields: Review," *Procedia Comput. Sci.*, vol. 155, pp. 599–605, 2019, doi: 10.1016/j.procs.2019.08.084.

[6] S. Ramya, A. M. Swetha, and M. Doraipandian, "IoT framework for smart irrigation using machine learning technique," *J. Comput. Sci.*, vol. 16, no. 3, pp. 355–363, 2020, doi: 10.3844/JCSSP.2020.355.363.

[7] A. Goldstein, L. Fink, A. Meitin, S. Bohadana, O. Lutenberg, and G. Ravid, "Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge," *Precis. Agric.*, vol. 19, no. 3, pp. 421–444, 2018, doi: 10.1007/s11119-017-9527-4.

[8] F. K. Van Evert, S. Fountas, D. Jakovetic, V. Crnojevic, I. Travlos, and C. Kempenaar, "Big Data for weed control and crop protection," *Weed Res.*, vol. 57, no. 4, pp. 218–233, 2017, doi: 10.1111/wre.12255.

[9] S. G. Fernandez *et al.*, "Smart soil monitoring and water conservation using irrigation on technology," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 19, no. 1, pp. 99–107, 2020, doi: 10.11591/ijeecs.v19.i1.pp99-107.

[10] Y. Shekhar, E. Dagur, S. Mishra, R. J. Tom, M. Veeramanikandan, and S. Sankaranarayanan, "Intelligent IoT based automated irrigation system," *Int. J. Appl. Eng. Res.*, vol. 12, no. 18, pp. 7306–7320, 2017.

[11] R. Torres-Sanchez, H. Navarro-Hellin, A. Guillamon-Frutos, R. San-Segundo, M. C. Ruiz-Abellón, and R. Domingo-Miguel, "A decision support system for irrigation management: Analysis and implementation of different learning techniques," *Water (Switzerland)*, vol. 12, no. 2, 2020, doi: 10.3390/w12020548.

[12] B. M. Sagar and N. K. Cauvery, "Agriculture data analytics in crop yield estimation: A critical review," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 12, no. 3, pp. 1087–1093, 2018, doi: 10.11591/ijeecs.v12.i3.pp1087-1093.

[13] N. Tantalaki, S. Souravlas, and M. Roumeliotis, “Data-Driven Decision Making in Precision Agriculture: The Rise of Big Data in Agricultural Systems,” *Journal of Agricultural & Food Information*, vol. 20, no. 4, pp. 344–380, 2019, doi: 10.1080/10496505.2019.1638264.
[14] W. Fan, C. Chong, G. Xiaoling, Y. Hua, and W. Juyun, "Prediction of Crop Yield Using Big Data," *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*, 2015, pp. 255-260, doi: 10.1109/ISCID.2015.191.
[15] H. Balti, I. Chebbi, N. Mellouli, I. R. Farah, and M. Lamolle, “A big remote sensing data analysis using deep learning framework,” *International Conferences Big Data Analytics, Data Mining and Computational Intelligence*, pp. 119–126, 2019, doi: 10.33965/bigdaci2019_201907l015.
[16] T. C. T. Chen, C. L. Liu, and H. D. Lin, “Advanced artificial neural networks,” *Algorithms*, vol. 11, no. 7, 2018, doi: 10.3390/a11070102.
[17] V. Jinubala, R. Lawrance, and P. Jeyakumar, “Classification of Rice Pest Data Using Decision Tree Algorithm,” *Int. J. Res. Advent Technol.*, vol. 7, no. 5S, pp. 148–154, 2019.
[18] N. Zhu *et al.*, “Deep learning for smart agriculture: Concepts, tools, applications, and opportunities,” *Int. J. Agric. Biol. Eng.*, vol. 11, no. 4, pp. 32-44, 2018, doi: 10.25165/j.ijabe.20181104.4475.
[19] H. Cadavid, W. Garzón, A. Pérez, G. López, C. Mendivelso, and C. Ramírez, “Towards a smart farming platform: From IoT-based crop sensing to data analytics,” *Communications in Computer and Information Science*, vol. 885, pp. 237–251, 2018, doi: 10.1007/978-3-319-98998-3_19.
[20] K. Singh, S. Jain, V. Andhra, and S. Sharma, “IoT based approach for smart irrigation system suited to multiple crop cultivation,” *Int. J. Eng. Res. Technol.*, vol. 12, no. 3, pp. 357–363, 2019.
[21] M. N. Islam Sarker, M. Wu, B. Chanthamith, S. Yusufzada, D. Li, and J. Zhang, "Big Data Driven Smart Agriculture: Pathway for Sustainable Development," *2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2019, pp. 60-65, doi: 10.1109/ICAIBD.2019.8836982.
[22] U. Moorthy and U. D. Gandhi, “A Survey of Big Data Analytics Using Machine Learning Algorithms,” pp. 95–123, 2017, doi: 10.4018/978-1-5225-2863-0.ch005.
[23] C. -Y. Lin, C. -H. Tsai, Ching-Pei Lee, and C. -J. Lin, "Large-scale logistic regression and linear support vector machines using spark," *2014 IEEE International Conference on Big Data (Big Data)*, 2014, pp. 519-528, doi: 10.1109/BigData.2014.7004269.
[24] K. Jin and S. Bagui, “Apache Spark SVM for Predicting Obstructive Sleep Apnea,” *Big Data Cogn. Comput.*, vol. 4, no. 4, 2020, doi: /10.3390/bdcc4040025.
[25] R. Singh and J. Woo, “Applications of machine learning models on Yelp data,” *Asia Pacific Journal of Information Systems*, vol. 29, no. 1, pp. 65–49, 2019, doi: 10.14329/apjis.2019.29.1.35.
[26] Y. O. Sayad, H. Mousannif, and H. Al Moatassime, “Predictive modeling of wildfires: A new dataset and machine learning approach,” *Fire Safety Journal*, vol. 104, pp. 130–146, 2019, doi: 10.1016/j.firesaf.2019.01.006.
[27] L. Wang, Y. Yu, L. Deng, and H. Pang, “A two-stage agriculture environmental anomaly detection method,” *Commun. Comput. Inf. Sci.*, vol. 763, pp. 779–789, 2017, doi: 10.1007/978-981-10-6364-0_77.