Indonesian Journal of Electrical Engineering and Computer Science
Vol. 35, No. 2, August 2024, pp. 1139~1150
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v35.i2.pp1139-1150

Forecasting livestock feed sales using machine learning techniques: an analysis of the Moroccan market

Mohamed Amine Nebri[1], Abdellatif Moussaid[2], Belaid Bouikhalene[1]
[1]LIMATI Laboratory, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco
[2]ENSIAS, Mohammed V University in Rabat, Rabat, Morocco

Article Info

Article history:

Received Feb 25, 2024
Revised Mar 22, 2024
Accepted Apr 6, 2024

Keywords:

Climate
Fodder
Livestock
Machine learning
Sales prediction

ABSTRACT

Agriculture, especially cereals, is important in sustaining economies and food security globally. This study delves into the Moroccan agricultural landscape, specifically focusing on predicting livestock feed sales to assist cereal company producers in optimizing production, streamlining supply chain operations, and enhancing customer satisfaction. Data collected from various markets across Morocco, including sales dates and locations, was combined with climate data and analyzed using advanced machine learning techniques, particularly the gradient boosting regression (GBR) algorithm, which achieved high accuracy with a mean absolute error (MAE) of 0.0203 and a root mean square error (RMSE) of 0.0281. The evaluation of multiple regression models revealed promising results, demonstrating the effectiveness of predictive models in accurately forecasting sales. These findings contribute valuable insights to sales forecasting in the cereal industry by considering weather conditions, production methods, and livestock-related variables, highlighting the importance of leveraging advanced machine learning techniques for optimizing production processes and meeting market demands efficiently in the agribusiness sector.

Corresponding Author:

Mohamed Amine Nebri
LIMATI Laboratory, Polydisciplinary Faculty, Sultan Moulay Slimane University
Beni Mellal, Morocc
Email: mohamedamine.nebrifpb@usms.ac.ma

1. INTRODUCTION

Agriculture, encompasses the cultivation of crops, the husbandry of livestock, and the production of sustenance for human consumption [1]. It not only guarantees food security but also fosters the development of rural areas and mitigates socioeconomic disparity. Fodder is one of the key products within agricultural systems, playing an essential role in global animal food [2]. However, the dynamics of livestock feed sales are closely related to a variety of factors with significant influence.

Cereal sales are subject to many interconnected variables [3], among which livestock consumption patterns play an important role [4]. Livestock serves not only as a source of meat but also as a consumer of cereals, with consumption patterns varying based on factors such as the size and quantity of livestock [5]. For instance, larger animals like cows typically consume more cereals compared to smaller ones like goats or sheep [6]. Additionally, the demand for specific types of fodder fluctuates based on factors like livestock size and age, further influencing the supply-demand dynamics within the agricultural market [7].

Moreover, fodder sales are influenced by climate conditions [8]. In regions that endure extended periods of drought or inadequate rainfall, there is typically a reduction in pasture growth and availability of forage, as highlighted by Ouédraogo *et al.* [9] in West Africa, typically observed. As elucidated by Abdi *et al.* [10], the scarcity of natural feed prompts farmers to explore supplementary feed options for their livestock, necessitating the adoption of enhanced forage technologies and feed conservation techniques. Consequently, the demand for commercially produced feed surges during these periods, leading to high sales volumes and potential price inflation due to limited supply [11]. Conversely, during favorable weather conditions and abundant rainfall, pasture grows actively, reducing reliance on purchased feed, and resulting in lower sales in the feed market [12]. In general, fluctuations in weather patterns directly impact feed sales dynamics, highlighting the link between environmental conditions and agricultural markets [13].

Furthermore, the importation of cereals plays a crucial role, particularly in countries where domestic production may not meet the demand [14]. This is exemplified in China, as explained by Gale [15], where for most of China's increase in corn supply, the yield of corn is insufficient, necessitating imports from the United States and other countries to fill the deficit. Imports supplement local supplies, shaping market dynamics and influencing pricing mechanisms [16]. Understanding the complex relationship between imports and sales is crucial for developing effective strategies to secure the agri-food economy, especially concerning animal feed.

The advancements in agricultural technologies, like artificial intelligence and robotics, are transforming traditional farming practices [17]. Extensive research into predicting agricultural products has spurred the creation of diverse models, particularly highlighting the role of artificial intelligence algorithms [18]. We began with Kohli *et al.* [19], who employed regression techniques to forecast agricultural product sales performance, emphasizing the importance of predictive accuracy measured by metrics like root mean square error (RMSE) and mean absolute percentage error (MAPE). Results indicated that linear regression outperformed k neighbors regressor (KNN-R), achieving an RMSE of 1898.91 and MAPE of 22.065. However, recognizing model limitations, Mohamed-Amine *et al.* [20] introduced additional regression models such as decision trees (DT), KNN, random forest (RF), and gradient boosting regression (GBR) LASSO, and RF, expanding the study to forecast phytosanitary sales in Morocco with climate data. The GBR notably yielded promising results, with an MAE of 0.0035 and an RMSE of 0.0066. In a similar context of predicting sales products based on weather conditions, Rincon-Patino *et al.* [21] applied various machine learning techniques to forecast avocado sales in the United States. Their study yielded a multivariate regression prediction model that performed well, achieving a relative absolute error (RAE) metric score of 7.812%. In another way, Bayona-Oré *et al.* [22] emphasized the preference for neural network-based algorithms due to their accuracy and precision in price prediction of agricultural products at harvest time, compared to other types of regression models. Similarly, Paul *et al.* [23] highlighted the exceptional performance of the generalized regression neural network (GRNN) in forecasting agricultural market prices, particularly when dealing with large volumes of data.

On the other hand, time series models are employed for sales forecasting, employing a distinct approach from traditional machine learning methods [24]. For instance, autoregressive integrated moving average (ARIMA) models, as demonstrated by Mgaya [25] in forecasting the consumption of livestock products, yield a MAPE value of 4.833 for ARIMA (0,2,1). While the ARIMA model is suitable for short-term forecasting, it struggles with daily fluctuations [26]. Conversely, the recurrent neural network (DT, KNN, RF, and GBR), as demonstrated by Weng *et al.* [27], exhibits improved accuracy, particularly for daily and weekly forecasts, achieving significantly lower errors with an average absolute error of only 0.15 and an average relative error of 8.82%. In the same context, Yoo and Oh [28] employed seasonal long short-term memory (SLSTM), a type of RNN designed for time series forecasting agricultural product sales volumes. These studies demonstrate various applications of time series forecasting, underscoring the significance of choosing suitable models and taking into account external factors like climate data and daily consumption to enhance prediction accuracy.

Given the limited literature on the correlation between weather patterns and livestock feed sales, and existing research on agricultural product forecasting, this study employs various regression models such as linear regression, DT, KNN, RF, and GBR. Due to our small dataset, machine learning regression proves more computationally efficient than deep learning models [29], which demand higher computational resources. Time series models, although effective in certain contexts, rely heavily on historical data and may overlook external factors, potentially leading to less precise demand forecasts [30]. Therefore, to ensure accuracy in forecasting results from product sales data, we will evaluate them using the MAE and RMSE methods, which are crucial for optimizing supply chain management and effective decision-making in the agri-food industry. The paper is structured as follows: it starts by detailing the data sources, variables, and methodology section. Following that, the results and discussion section presents and analyzes the experimental outcomes.

2. METHOD

2.1. Geographic scope

Morocco, situated in North Africa, boasts a diverse landscape that significantly shapes its climate, agriculture, and livestock practices [31]. The country experiences a range of climates, from the Mediterranean along the coast to arid and desert conditions inland. The Atlas Mountains are key in creating distinct ecosystems [32]. Despite the challenging conditions, Morocco has successfully developed a robust agricultural sector. The fertile coastal plains are particularly suitable for cultivating crops such as wheat and barley [33]. Livestock farming holds paramount importance for Morocco. As of 2022, the ministry of agriculture and fisheries reported substantial numbers presented in Figure 1, with over 22 million sheep, 6 million goats, 3.18 million cows, and 62.4 thousand camels. While sheep and goats are prevalent throughout the country, cows are found everywhere except in specific areas like deserts, where their presence is scarce or nonexistent. Conversely, camels are common in these desert regions [34]. Morocco's agricultural and livestock sectors showcase a harmonious blend of traditional practices and modern techniques, substantially contributing to the nation's economy and livelihoods.

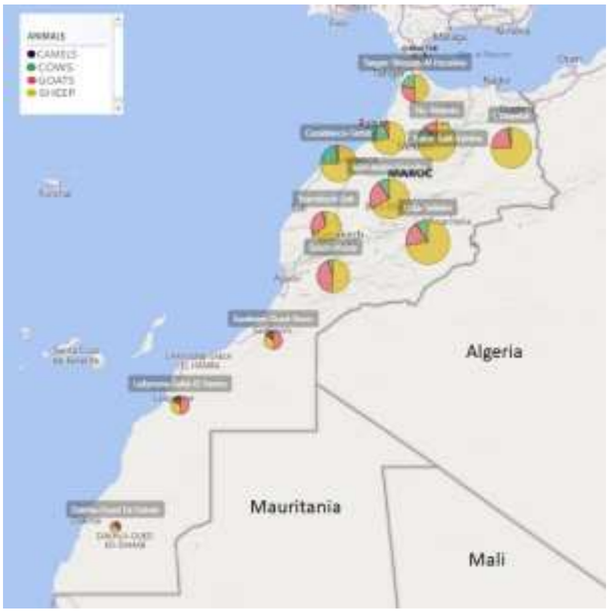


Figure 1. Mapping livestock in Morocco

2.2. Dataset
2.2.1. Data collection

Data collection is gathering for business decisions, strategic planning, and research. This is an important phase in any data science project because it serves as the foundation for successful data analysis and machine learning projects [35]. Data is typically collected from a variety of sources, cleaned, and prepared to provide the information required to analyze business performance and forecast future trends.

Our dataset consists of two parts: sales data, collected from various Moroccan markets specializing in the trade of forage and livestock, and climate data obtained from an open-source platform. The first set of data includes daily sales transactions from 2013 to 2020. This dataset provides details such as the date of the sales transaction, the specific region where the sales activity occurred, the city or locality within the specified region where the sales took place, and the product name, which encompasses three main products (Corn, Soybean, and Wheat). Additionally, it includes the target sales quantity in kilograms that we aim to forecast. The dataset comprises 5 columns and 1,578 rows, resulting in a total length of 7,890, as presented in Table 1.

Table 1. Representation of livestock feed sales transactions

Sales date	Region	Cities	Name	Sales qty
15012013	MARRAKECH-SAFI	CHICHAOUA	SOYBEAN	100
16012013	SOUSS-MASSA	INZEGANE	CORN	420
…	…	…	…	…
31122020	SOUSS-MASSA	OULED TEIMA	WHEAT	310

The second phase of the analysis incorporates climate data sourced from the open-source data website (https://weatherandclimate.com/). This dataset comprises daily information collected for two time periods: T-1 (previous day), and T-7 (one week ago). The data encompasses five crucial climate factors: temperature, dew point, humidity, wind speed, and pressure. The dataset includes 10 columns, capturing data collected for the times T-1, and T-7. The period of data collection spans from January 8, 2013, to December 31, 2020. The goal of gathering this data is to understand better how the weather influences the sales of our products. The details about the factors are detailed in Table 2. This dual data set approach improves our ability to forecast feed sales by taking into account market dynamics and climate impacts.

Table 2. Description of climate factors

Climate	Description
Dew point	This temperature is the point at which the air becomes saturated and dew forms. High dew points may have an impact on animal feed consumption. Measured in (°C).
Humidity	High humidity can cause heat stress in livestock and affect feed intake. Expressed as a percentage
Wind speed	Wind can significantly impact the perceived temperature, particularly in cold weather. Recorded in (km/h)
Precipitation	Continuous precipitation, especially over a long period, can impact the quality of forage and feed crops. Mesuread in (mm).
Temperature	Temperature, be it hot or cold, can influence the feed consumption of livestock. measured in (°C).

2.2.2. Data processing

To make a strong prediction model, we wanted to use all the information from the data sources. We combined two sets of data: the first dataset focused on transactions related to the sales of livestock feed, where the focus variable was the quantity sold. The second dataset focused on climate data. By combining these datasets, we created a new dataset with 69,810 rows and columns covering eight years of sales data. Figure 2 outlines these factors and explains how we organized them to create the field data.

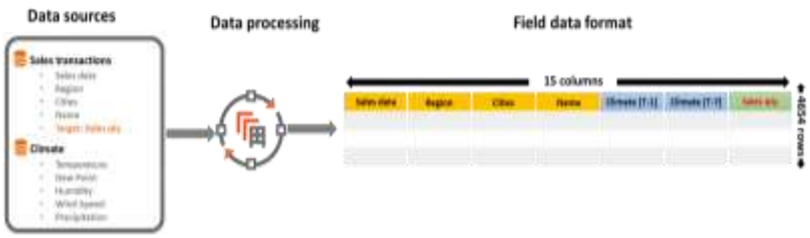


Figure 2. Final data format

2.2.3. Data exploration

Before using machine learning to make predictions, it's important to explore the data. This helps to understand the variables better and make any needed adjustments to improve the features. In our study, the distribution of the target variable (sales quantites) spanning from 2013 to 2020, exhibits a normal distribution pattern, as shown in Figure 3. This characteristic is advantageous for machine learning models.

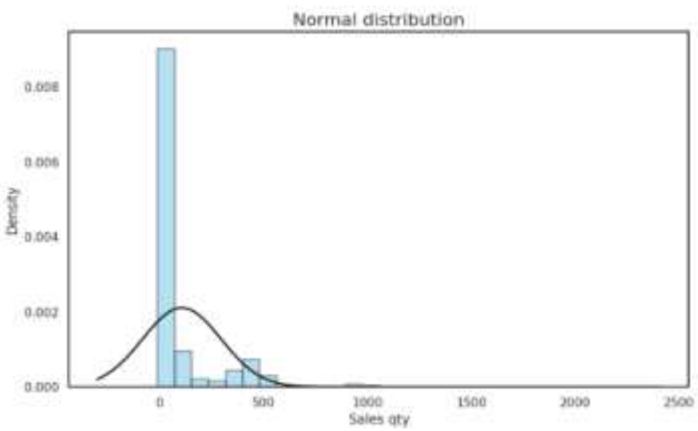


Figure 3. Sales quantity distribution

The first graph in Figure 4 illustrates the quantity sold for three products: corn, wheat, and soybean. For corn and wheat, most sales values fall within the range of up to 300 kilograms. In contrast, soybean typically has sales below 300 kilograms. This suggests that sales transactions of corn and wheat exceed those of soybeans.

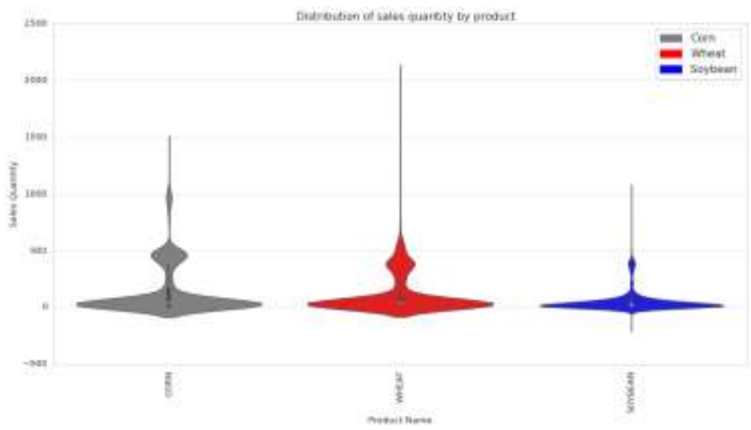


Figure 4. Comparison of product sales quantities

The second plot in Figure 5 shows how the sales of corn, wheat, and soybeans changed over the years. These changes are influenced by different things like low prices, market instability, and how much of each product is available. Corn sales are always high because the government supports farmers [36]. When corn prices are low, farmers want to buy more, as seen in 2013 and 2019. Furthermore, wheat sales experience fluctuations driven by market scarcity. In years with sufficient wheat production, such as 2016, 2019, and 2020, sales quantities are higher. Conversely, during periods of market shortage, as observed in 2015 and 2018, sales decreased accordingly [37]. The increasing presence of soybeans in the market also contributes to the rising sales observed over the years.

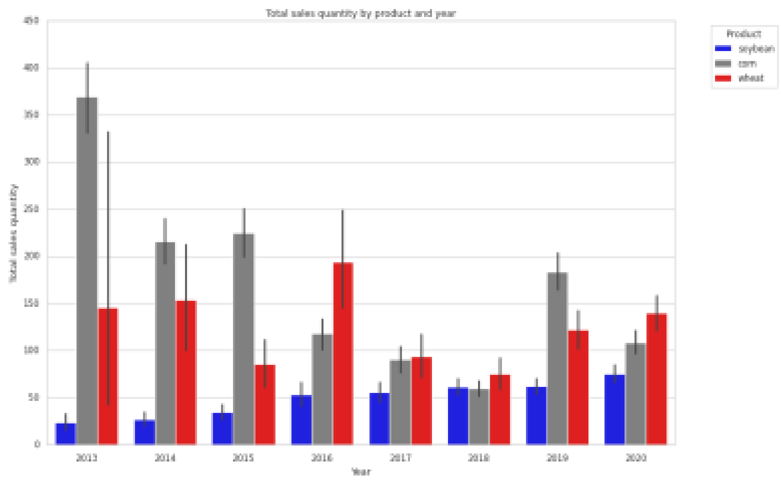
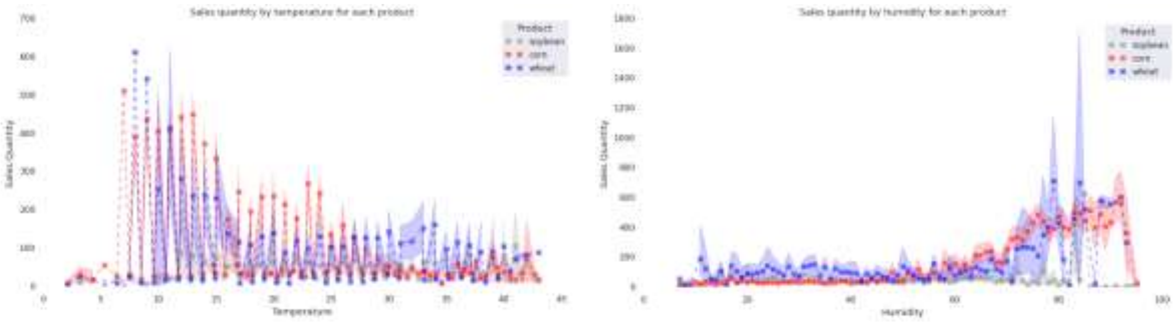


Figure 5. Fodder sales trends over the years

Figure 6 illustrates the impact of temperature and humidity on feed sales [38]. Figure 6(a) depicts the scenario when the temperature is low, at 7 °C, while Figure 6(b) shows humidity as high as 80%. This combination indicates wet and cold weather conditions, leading to an increase in feed demand of over 300 kg. Conversely, when the air temperature rises to around 40 °C and humidity drops to 20%, it signifies warm weather, resulting in reduced feed sales. Therefore, incorporating climate variables as inputs in our dataset highlights their substantial influence on animal nutrition transactions.



(a) (b)

Figure 6. Temperature (a) and humidity and (b) impact on feed sales

To summarize the data, we created a correlation matrix. This matrix shows how different factors are connected and how they affect the target [39]. Figure 7 shows this relationship using colors on the graph and helps us understand the relationship between the input variables. As observed in the heatmap, the majority of variables exhibit positive correlation (red color), while others show negative correlation (blue color). This suggests a significant relationship between them. Notably, climatic data demonstrates a correlation with the sales quantity. The interpretation of these findings highlights a strong correlation between animal fodder transactions and climate. This is underscored by the fact that the quantity sold depends on temperature, which is correlated with climatic conditions [40]. Recognizing this dependence is crucial for effective training of prediction algorithms and provides opportunities for valuable feature engineering.

Figure 7. Correlation matrix between variables

2.3. Our approach

In our project, as presented in Figure 8, the objective is to predict the monthly sales volume of livestock feed for the last year of 2020. To achieve this, we used various methods in data engineering and machine learning to create models that work well. We started by gathering data from various Moroccan markets specializing in the business of animal nutrition products, and climate data from open source platforms. Subsequently, we took steps to ready the data, dealing with missing values, encoding categorical features, and scaling each variable to a defined range. This ensures consistent representation and enhances the performance of our models. To ensure a dependable evaluation, we divided our dataset into two segments: the training set and the test set. We utilized data from the years 2013 to 2019 to train our models, and subsequently, we tested them on the data from the year 2020.

After splitting the data, we proceeded to the model training phase. We initialized several algorithms, with a primary emphasis on machine learning algorithms recognized for their strong performance in recent years [41]. In our scenario, we specifically selected machine learning algorithms to predict the volume of animal fodder sold in Morocco, using the collected and prepared dataset.

Intending to predict the number of products sold, we faced a regression challenge from a technical perspective [42]. Specifically, this challenge took the form of a supervised learning problem focused on regression prediction. Our initial approach involved employing linear regression, a simple yet effective method commonly used in addressing prediction challenges. The method involves selecting optimal alpha α and β values to best fit a line to the data. This process includes assessing squared differences between observed and predicted values, summing them, computing the gradient, and solving for the optimal values [43]. In the linear regression equation, we use y to represent what we're trying to predict, x for the factor we think influences it, α for the slope of the line, and β for where the line starts on the y-axis. In (1) illustrates the connection with a straight line.

$$y = \alpha x + \beta \tag{1}$$

Moreover, we utilized decision trees (DTs) as regression models for predicting continuous target variables. They use a top-down approach to organize and label observations within the dataset. The model divides the training data set into three distinct regions, where the values of the predictor variables are more closely related [44]. The tree-building process partitions the feature space into J distinct regions, denoted as $R_1, R_2, \ldots, R_j$. The objective is to identify rectangles $R_1, R_2, \ldots, R_j$ that minimize the residual sum of squares (RSS), calculated by finding residuals, squaring them, and summing across each region as outlined in (2). The tree-building process continues, creating splits down the tree as the summation decreases.

$$\sum_{j=1}^{J} \sum_{i \epsilon R_j} (y_i - \hat{y}_{R_j})^2 \tag{2}$$

Furthermore, we use KNN-R as the initialization model and exploit its robustness to handle small data sets [45]. Despite the similarity between regression and classification algorithms, the key distinction lies in the labels, as regression deals with real numbers rather than categories. In the KNN-R method, a training set consisting of examples $\{x_i, y_i\}$ is used, which x_i represents the attribute values of the examples and y_i represents the real-valued target. Subsequently, when a test point is specified for the prediction, the algorithm starts calculating the distance to all training points y_i. It identifies the k nearest instances x_i, retrieves their corresponding labels y_i, and does not perform voting or determine the most common class, but instead computes the average of these labels. Specifically, KNN-R calculates the arithmetic mean of the labels of the k nearest neighbors and uses this mean as the prediction for a given point x, as shown in (3).

$$\hat{y} = f(x) = \frac{1}{k} \sum_{j=1}^{k} y_{i_j} \tag{3}$$

Moreover, a recent approach in learning involves ensembling techniques. This method creates various models with different settings to improve the accuracy of predictions. Two primary ensembling methods are boosting and bagging [46]. In the boosting method, models are trained sequentially, continually learning until they encounter errors. As for the bagging method, models are trained simultaneously, each working on a smaller part of the data. The final model is then determined by a vote among the predictions made by all models [47]. We utilized the random forest as an ensemble learning technique, which is a model constructed using multiple trees, and the final decision is derived by averaging B predictions. In RFR, a set of decision trees is assembled by bootstrapping the data (D) and selecting a random subset of features m in B iterations. For a new data point x, the prediction (y), as defined in (4), signifies the average of the predictions made by each tree. This approach effectively minimizes overfitting and enhances accuracy.

$$\hat{y}(x) = \frac{1}{\mathrm{B}} \sum_{\mathrm{b}=1}^{\mathrm{B}} \mathrm{T_b}(\mathrm{x}) \tag{4}$$

The GBR, a powerful ensemble learning algorithm, utilizes multiple weak regression models like decision trees to predict continuous variables such as sales quantity. These weak models are trained sequentially, with each subsequent one aimed at correcting errors from its predecessor. Through an iterative process, the algorithm optimizes the loss function by following the gradient to reduce prediction errors, ultimately enhancing the model's predictive accuracy [48].

$$\mathrm{F_m}(x) = F_{m-1}(x) + h_m(x) \tag{5}$$

Where $F_m(x)$ represent the ensemble model's prediction for input x at iteration m, $F_{m-1}(x)$ denote the ensemble model's prediction for input x at iteration $m-1$, and $h_m(x)$ indicate the weak learner model for input x at iteration m.

In the context of our approach, the last phase in constructing a machine learning model involves implementing cross-validation with a tuning parameter algorithm. This step aids in determining the optimal hyperparameters for the model and safeguards against overfitting of the data [49]. To avoid overfitting and identify the most suitable algorithm, we utilized a K-fold cross-validation technique with 5 folds throughout the training and validation stages [50]. In each validation cycle, we applied the GridSearchCV technique to determine the optimal parameters for each algorithm. Following five iterations, we derived an average score, serving as the reliable validation score. Ultimately, we tested the models on the test set, comprising entirely new data, to obtain the prediction score. Our methodology is based on powerful machine learning technologies that are capable of adapting and learning from data, resulting in more accurate and robust sales forecasts compared to conventional methods. This contributes to enhancing supply chain operations and helps agricultural companies make better decisions, improving efficiency in the agricultural sector.

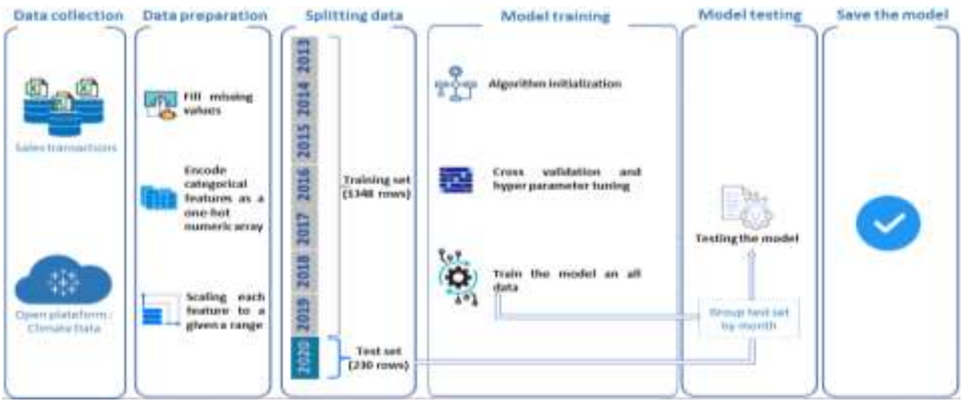


Figure 8. Data engineering and machine learning workflow

3. RESULTS AND DISCUSSION

This section is divided into two subsections that delve into the topic of study and present the corresponding outcomes. The first subsection delves into the testing and evaluation methods employed. These efforts lead to the generation of valuable results. The subsequent subsection discusses the simulation results, followed by an evaluation of the findings.

3.1. Evaluation metrics

In these equations, $\hat{y}_i$ represents the predicted value, y_i represents the actual value, and n represents the total number of data points. The evaluation metrics utilized in our study included the MAE and RMSE. These metrics were employed to evaluate the accuracy of prediction models. MAE calculates the average absolute difference between predicted and target values, while RMSE involves the square root of the mean squared difference between the predicted and real values. This metrics widely employed across diverse fields like business, engineering, and industry [51], these metrics are represented by (6) and (7).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2} \tag{7}$$

3.2. Simulation results and evaluation

Using 1,348 records to train the five models, we compared their performance to identify the most suitable algorithm for forecasting feed livestock sales in Morocco. We conducted cross-validation using a k-fold technique with k = 5 and parameter tuning using the GridSearchCV algorithm. The new models were trained on all training datasets and their performance was evaluated on new data specifically from the year 2020 using MAE and RMSE.

To assess the significance of sales data and the added value of climate data, we tested the models on two subsets of data: the first subset contained indicators from transaction data alone, while the second subset included both transaction data and weather information. Table 3 illustrates the scores obtained by the selected algorithms, showing close alignment. A comparison of regression algorithms including DT, GBR, KNN, LR, and RF models revealed that the GBR model had the least error MSE, performing better for most data series patterns than other regression models. The results demonstrated that the GBR model outperformed other algorithms for most datasets in this study, highlighting its forecasting capability. Specifically, the GBR algorithm exhibited favorable prediction scores with MAE of 0.0226 and RMSE of 0.0290. After integrating climate data, these scores improved to MAE of 0.0203 and RMSE of 0.0281, showcasing the powerful combination of these datasets for predicting quantity sold.

Table 3. Sales prediction results including MAE and MSE metrics

Category	Model	Sales data		Sales + Climate data	
		MAE	RMSE	MAE	RMSE
All products		0.0307	0.0440	0.0255	0.0354
		0.0226	0.0290	0.0203	0.0281
		0.0232	0.0311	0.0206	0.0279
		0.0255	0.0339	0.0209	0.0291
		0.0249	0.0342	0.0237	0.0318

The hyperparameters listed in Table 4 guide various aspects of the gradient boosting regression algorithm, including learning rate (learning_rate), maximum tree depth (max_Depth), segmentation quality evaluation criteria (criterion), and additional parameters. Adjusting these hyperparameters can notably influence the model's performance and ability to generalize effectively.

Table 4. Optimal hyperparameter selection for GBR

Parameters	Value
criterion	friedman_mse
learning_rate	0.1
max_depth	3
min_samples_leaf	1
n_estimators	100
random_state	2247
subsample	1.0
subsample	1.0

To analyze the results obtained, we showcase a representative graph utilizing the percentage error metric (8). To reinforce the validation of our approach and derive scores for each product, we conducted random sampling by selecting twelve months of data from each product in the year 2020.

$$\text{Percentage error} = \frac{|\text{Predicted value} - \text{True value}|}{\text{True value}} \tag{8}$$

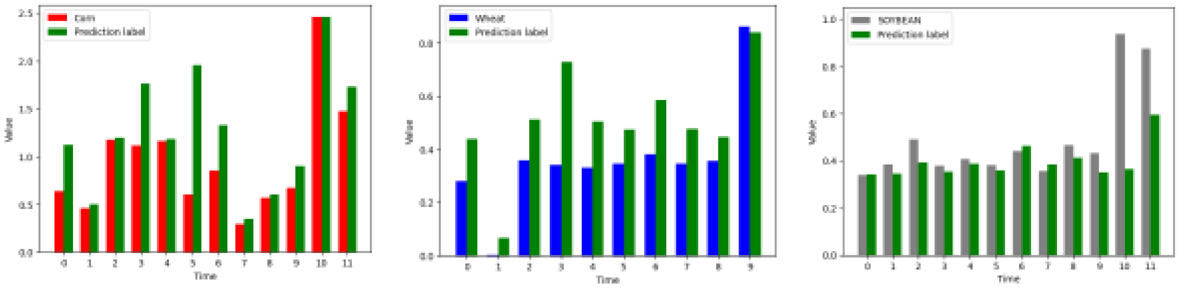
Figure 9 presents a bar chart comparing the actual quantity sold (depicted in green) with predicted values for corn, wheat, and soybean, offering insights into the predictive accuracy across different feed products. Specifically, Figure 9(a) focuses on corn, Figure 9(b) on wheat, and Figure 9(c) on soybean. This graphical representation provides a visual understanding of how well the predictive models perform for each feed product, highlighting any discrepancies between actual and predicted values.

In November, corn demonstrated the highest predictive performance, with the predicted value (2.46) closely aligning with the actual sales quantity (2.40), resulting in a low error rate of 2.44%. Similarly, for wheat, the highest predictive accuracy is observed in November, where the predicted value (2.47) closely aligns with the actual sales quantity (2.39), showcasing a low error rate of approximately 3.25%. Subsequently, for soybeans, the most accurate prediction occurs in January, with the predicted value of 0.34 closely matching the actual sales quantity (0.29), yielding a low error rate of 2.03%. These findings underscore a commendable level of accuracy in predicting sales for each respective product, suggesting an increased demand for livestock feed during these specific months.

In July, a moderate performance is observed for corn, wheat, and soybean, with percentage errors of approximately 19%. However, in April, both corn and wheat exhibit a percentage error of 15%, and in March, Soybean shows a percentage error of 10.5%. This implies that sales of these goods are slightly lower in these months.

Contrastingly, the lowest performance is noticeable in August for corn and wheat, leading to a significant discrepancy and a percentage error of approximately 29%. Similarly, for Soybean in December, there is the least accurate prediction with a percentage error of approximately 31.9%. This suggests that the demand for fodder is very minimal during these periods.

These results indicate that the GBR is a highly effective algorithm for forecasting feed livestock sales in Morocco. Among the selected algorithms, GBR stands out as the algorithm of choice due to its lowest MAE and RMSE, which are further supported by the percentage error. Despite these favorable outcomes, we find overall satisfaction in these results and recommend agro-companies and livestock feed producers to proactively collect market data to enhance their understanding of feed livestock demand.



(a) (b) (c)

Figure 9. Predicted vs. actual fodder sales for; (a) corn, (b) wheat, and (c) soybean

4. CONCLUSION

The forecasting of livestock feed in Morocco employs robust machine learning techniques with the primary objective of maximizing production to meet market demands, streamlining supply chain operations, and enhancing overall customer satisfaction. Various data sources, comprising fodder sales transactions from the Moroccan market and weather dataset, are utilized to enable sales forecasting, thereby establishing a rich and comprehensive dataset. This methodology lays the groundwork for precise and insightful analysis. This approach establishes a foundation for accurate and insightful analysis. Utilizing the capabilities of machine learning, specifically regression models, this study presents a dependable forecasting approach. The GBR algorithm proves effective, showcasing its ability to comprehend intricate sales trends. It exhibits superior accuracy, validated by measures such as MAE and RMSE, ensuring the reliability of the chosen models. By integrating data engineering, model training, cross-validation, and hyperparameter tuning, our research has delivered promising outcomes. Well-organized visualizations facilitate a clear understanding of sales dynamics. Hence, it signifies a new era of precision and efficiency in managing animal feed sales within the agricultural business sector.

Our study exhibits strong performance, particularly with smaller datasets. However, processing very large datasets using our current method can indeed take a significant amount of time. In future endeavors, we aim to enhance our sales forecasting capabilities by incorporating additional external factors to enrich our dataset. Additionally, we plan to explore more advanced techniques, including deep learning methods, to further improve the accuracy and efficiency of our sales forecasting efforts.

REFERENCES

[1] R. S. Upendra, I. M. Umesh, R. B. Ravi Varma, and B. Basavaprasad, “Technology in Indian agriculture – a review,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 2, pp. 1070–1077, Nov. 2020, doi: 10.11591/ijeecs.v20.i2.pp1070-1077.

[2] L. Mbambalala, Z. T. Rani, T. D. E. Mpanza, M. S. Mthana, L. Ncisana, and N. R. Mkhize, “Fodder radish as a potential alternative feed source for livestock in south Africa,” *Agriculture (Switzerland)*, vol. 13, no. 8, p. 1625, Aug. 2023, doi: 10.3390/agriculture13081625.

[3] C. Sage, “The interconnected challenges for food security from a food regimes perspective: energy, climate and malconsumption,” *Journal of Rural Studies*, vol. 29, pp. 71–80, Jan. 2013, doi: 10.1016/j.jrurstud.2012.02.005.

[4] A. Mottet, C. de Haan, A. Falcucci, G. Tempio, C. Opio, and P. Gerber, “Livestock: on our plates or eating at our table? A new analysis of the feed/food debate,” *Global Food Security*, vol. 14, pp. 1–8, Sep. 2017, doi: 10.1016/j.gfs.2017.01.001.

[5] S. V. Shchitov *et al.*, “Using infrared radiation in preparation of feed for farm livestock,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 1, pp. 148–154, Jul. 2019, doi: 10.11591/ijeecs.v15.i1.pp148-154.

[6] F.-H. D. Witt and O. Einkamerer, “Factors affecting the voluntary feed intake of livestock,” *AFMA Matrix*, vol. 22, no. 1, pp. 29–33, 2013.

[7] D. Datta, “Indian fodder management towards 2030: a case of vision or Myopia,” *Xplore International Research Journal Consortium www.irjcjournals.org*, vol. 2, no. 2, p. 33, 2013, [Online]. Available: http://hayandforage.com/marketing/archive/0201-hay-.

[8] S. M. Sala, “An analysis of the factors influencing participation of pastoralists in commercial fodder value chain for livelihood resilience in Isiolo County, Kenya,” 2019. http://erepository.uonbi.ac.ke/handle/11295/107167.

[9] K. Ouédraogo, A. Zaré, G. Korbéogo, O. Ouédraogo, and A. Linstädter, “Resilience strategies of West African pastoralists in response to scarce forage resources,” *Pastoralism*, vol. 11, no. 1, p. 16, Dec. 2021, doi: 10.1186/s13570-021-00210-8.

[10] M. D. Abdi, “Assessment of feed resources and identification of coping strategies during drought by Pastoralists and Agro-Pastoralists At Gode District of Somali Region, Ethiopia,” *Haramaya University Library and Information Service*, 2018. http://localhost:8080/xmlui/handle/123456789/3151.

[11] G. L. Meena, S. Mishra, and P. Chand, “Assessment of demand and supply of livestock feed and fodder resources in Rajasthan,” *Indian Journal of Animal Nutrition*, vol. 35, no. 2, p. 166, 2018, doi: 10.5958/2231-6744.2018.00025.7.

[12] A. Cogato, F. Meggio, M. D. A. Migliorati, and F. Marinello, “Extreme weather events in agriculture: a systematic review,” *Sustainability (Switzerland)*, vol. 11, no. 9, p. 2547, May 2019, doi: 10.3390/su11092547.

[13] C. Godde *et al.*, “Climate change and variability impacts on grazing herds: Insights from a system dynamics approach for semi-arid Australian rangelands,” *Global Change Biology*, vol. 25, no. 9, pp. 3091–3109, Sep. 2019, doi: 10.1111/gcb.14669.

[14] M. P. Reynolds and H. J. Braun, *Wheat Improvement: Food Security in a Changing Climate*. Cham: Springer International Publishing, 2022.

[15] F. Gale, “China’s import potential for beef, corn, pork, and wheat,” *Washington DC: United States Department of Agriculture Economic Research Servic*, 2014, doi: 10.22004/ag.econ.329067.

[16] G. Aparicio and V. Pinilla, “International trade in wheat and other cereals and the collapse of the first wave of globalization, 1900-38,” *Journal of Global History*, vol. 14, no. 1, pp. 44–67, Mar. 2019, doi: 10.1017/S1740022818000347.

[17] J. Jung, M. Maeda, A. Chang, M. Bhandari, A. Ashapure, and J. Landivar-Bowles, “The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems,” *Current Opinion in Biotechnology*, vol. 70, pp. 15–22, Aug. 2021, doi: 10.1016/j.copbio.2020.09.003.

[18] A. Moussaid *et al.*, “Machine learning applied to tree crop yield prediction using field data and satellite imagery: a case study in a Citrus Orchard,” *Informatics*, vol. 9, no. 4, p. 80, Oct. 2022, doi: 10.3390/informatics9040080.

[19] S. Kohli, G. T. Godwin, and S. Urolagin, “Sales prediction using linear and KNN regression,” in *Singapore: Springer Singapore*, 2021, pp. 321–329.

[20] N. Mohamed-Amine, M. Abdellatif, and B. Belaid, “Artificial intelligence for forecasting sales of agricultural products: a case study of a moroccan agricultural company,” *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 1, p. 100189, Mar. 2024, doi: 10.1016/j.joitmc.2023.100189.

[21] J. Rincon-Patino, E. Lasso, and J. C. Corrales, “Estimating avocado sales using machine learning algorithms and weather data,” *Sustainability (Switzerland)*, vol. 10, no. 10, p. 3498, Sep. 2018, doi: 10.3390/su10103498.

[22] S. Bayona-Oré, R. Cerna, and E. T. Hinojoza, “Machine learning for price prediction for agricultural products,” *WSEAS Transactions on Business and Economics*, vol. 18, pp. 969–977, Jun. 2021, doi: 10.37394/23207.2021.18.92.

[23] R. K. Paul *et al.*, “Machine learning techniques for forecasting agricultural prices: a case of brinjal in Odisha, India,” *PLoS ONE*, vol. 17, no. 7 July, p. e0270553, Jul. 2022, doi: 10.1371/journal.pone.0270553.

[24] B. M. Pavlyshenko, “Machine-learning models for sales time series forecasting,” *Data*, vol. 4, no. 1, p. 15, Jan. 2019, doi: 10.3390/data4010015.

[25] J. F. Mgaya, “Application of ARIMA models in forecasting livestock products consumption in Tanzania,” *Cogent Food and Agriculture*, vol. 5, no. 1, p. 1607430, Jan. 2019, doi: 10.1080/23311932.2019.1607430.

[26] H. V. Nguyen, M. A. Naeem, N. Wichitaksorn, and R. Pears, “A smart system for short-term price prediction using time series models,” *Computers and Electrical Engineering*, vol. 76, pp. 339–352, Jun. 2019, doi: 10.1016/j.compeleceng.2019.04.013.

[27] Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang, and F. Y. Wang, “Forecasting horticultural products price using ARIMA model and neural network based on a large-scale data set collected by web crawler,” *IEEE Transactions on Computational Social Systems*, vol. 6, no. 3, pp. 547–553, Jun. 2019, doi: 10.1109/TCSS.2019.2914499.

[28] T. W. Yoo and I. S. Oh, “Time series forecasting of agricultural products’ sales volumes based on seasonal long short-term memory,” *Applied Sciences (Switzerland)*, vol. 10, no. 22, pp. 1–15, Nov. 2020, doi: 10.3390/app10228169.

[29] K. Sharifani and M. Amini, “Machine learning and deep learning: a review of methods and applications,” *World Information Technology and Engineering Journal*, vol. 10, no. 07, pp. 3897–3904, 2023, [Online]. Available: https://ssrn.com/abstract=4458723.

[30] A. R. S. Parmezan, V. M. A. Souza, and G. E. A. P. A. Batista, “Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model,” *Information Sciences*, vol. 484, pp. 302–337, May 2019, doi: 10.1016/j.ins.2019.01.076.

[31] S. Abdelmajid, A. Mukhtar, M. B. Baig, and M. R. Reed, “Climate change, agricultural policy and food security in Morocco,” in *Emerging Challenges to Food Production and Security in Asia, Middle East, and Africa: Climate Risks and Resource Scarcity*, Cham: Springer International Publishing, 2021, pp. 171–196.

[32] V. Boselli *et al.*, “System dynamics applied to terraced agroecosystems: the case study of assaragh (Anti-Atlas Mountains, Morocco),” *Water (Switzerland)*, vol. 12, no. 6, p. 1693, Jun. 2020, doi: 10.3390/W12061693.

[33] F. Ameur, H. Amichi, and C. Leauthaud, “Agroecology in North African irrigated plains? Mapping promising practices and characterizing farmers’ underlying logics,” *Regional Environmental Change*, vol. 20, no. 4, p. 133, Dec. 2020, doi: 10.1007/s10113-020-01719-1.

[34] I. Boujenane, “Socio-economic characteristics of farmers and constraints to livestock development in the Drâa-Tafilalet region of Morocco,” *Journal of Animal Science and Veterinary Medicine*, vol. 8, no. 5, pp. 179–189, Oct. 2023, doi: 10.31248/jasvm2023.393.

[35] I. Martinez, E. Viles, and I. G. Olaizola, “Data science methodologies: current challenges and future approaches,” *Big Data Research*, vol. 24, p. 100183, May 2021, doi: 10.1016/j.bdr.2020.100183.

[36] B. El Houssain, A. Fadlaoui, K. Allali, and R. Arrach, “Contract farming in the morocco cereal sector: contract clauses, ambiguity, and opportunism,” *International Journal of Agricultural Economics*, vol. 4, no. 5, p. 245, 2019, doi: 10.11648/j.ijae.20190405.17.

[37] S. K. Sharma, “Morocco: product specific support to wheat under WTO regime,” in *The WTO and Food Security*, Singapore: Springer Singapore, 2016, pp. 115–132.

[38] N. Rose and L. Dolega, “It’s the weather: quantifying the impact of weather on retail sales,” *Applied Spatial Analysis and Policy*, vol. 15, no. 1, pp. 189–214, Mar. 2022, doi: 10.1007/s12061-021-09397-0.

[39] J. M. Rickman, G. Balasubramanian, C. J. Marvel, H. M. Chan, and M.-T. Burton, “Machine learning strategies for high-entropy alloys,” *Journal of Applied Physics*, vol. 128, no. 22, Dec. 2020, doi: 10.1063/5.0030367.

[40] J. M. Addoum, D. T. Ng, and A. Ortiz-Bobea, "Temperature shocks and establishment sales," *Review of Financial Studies*, vol. 33, no. 3, pp. 1331–1366, Mar. 2020, doi: 10.1093/rfs/hhz126.
[41] D. Ni, Z. Xiao, and M. K. Lim, "A systematic review of the research trends of machine learning in supply chain management," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 7, pp. 1463–1482, Jul. 2020, doi: 10.1007/s13042-019-01050-0.
[42] N. Yeasmin, S. H. Amin, and B. M. Tosarkani, "Machine learning techniques for grocery sales forecasting by analyzing historical data," in *Learning and Analytics in Intelligent Systems*, vol. 25, 2022, pp. 21–36.
[43] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.
[44] B. KOCARIK GACAR and İ. DEVECİ KOCAKOÇ, "Regresyon Analizleri mi Karar Ağaçları mı?," *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, pp. 251–260, Dec. 2020, doi: 10.18026/cbayarsos.796172.
[45] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019*, May 2019, pp. 1255–1260, doi: 10.1109/ICCS45141.2019.9065747.
[46] F. Huettmann, "Boosting, bagging and ensembles in the real world: an overview, some explanations and a practical synthesis for holistic global wildlife conservation applications based on machine learning with decision trees," in *Machine Learning for Ecology and Sustainable Natural Resource Management*, Cham: Springer International Publishing, 2018, pp. 63–83.
[47] S. S. Azmi and S. Baliga, "An overview of boosting decision tree algorithms utilizing AdaBoost and XGBoost boosting strategies," *International Research Journal of Engineering and Technology*, no. May, pp. 6867–6870, 2020, [Online]. Available: www.irjet.net.
[48] A. Y. Hassan and M. H. Saleh, "Intelligence framework dust forecasting using regression algorithms models," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 1, pp. 177–184, Oct. 2023, doi: 10.11591/ijeecs.v32.i1.pp177-184.
[49] G. Behera and N. Nain, "Grid search optimization (GSO) based future sales prediction for big mart," in *Proceedings - 15th International Conference on Signal Image Technology and Internet Based Systems, SISITS 2019*, Nov. 2019, pp. 172–178, doi: 10.1109/SITIS.2019.00038.
[50] A. Qahtan, "Machine learning evaluation (cross-validation, metrics, importance scores)," in *Clinical Applications of Artificial Intelligence in Real-World Data*, Cham: Springer International Publishing, 2023, pp. 175–187.
[51] J. Dessain, "Machine learning models predicting returns: why most popular performance metrics are misleading and proposal for an efficient metric," *Expert Systems with Applications*, vol. 199, 2022, doi: 10.1016/j.eswa.2022.116970.

BIOGRAPHIES OF AUTHORS

Mohamed Amine Nebri holds an M.Sc. in Computer Systems Engineering, which he obtained in 2018 from the Polydisciplinary Faculty of Beni Mellal at Sultan Moulay Slimane University, Morocco. Currently, he is pursuing his Ph.D. studies within the same faculty. His research interests revolve around advanced technologies applied in agribusiness, including artificial intelligence, business intelligence, computer science, and the internet of things (IoT). For academic inquiries, he can be contacted at email at mohamedamine.nebrifpb@usms.ac.ma or through Gmail at aminetkdamine@gmail.com.

Dr. Moussaid Abdellatif researcher in Artificial Intelligence, Modelization, and Computer Science, holds a Ph.D. from the National School of Science and System Analysis, Mohammed 5 University in Rabat, Morocco. With an extensive publication record spanning various international journals and conferences, Dr. Moussaid has made significant strides in applying Artificial Intelligence across diverse domains including agriculture, healthcare, and industrial sectors. Alongside his academic pursuits, he brings good experience in teaching Artificial Intelligence and Computer Science, shaping the minds of future innovators. He can be contacted at email: abdellatif_moussaid@um5.ac.ma.

Dr. Belaid Bouikhalene professor of mathematics and information sciences received a Ph.D. Presently serving as the Vice-Dean of the Polydisciplinary Faculty at Beni Mellal, Sultan Moulay Slimane University in Morocco, Dr. Bouikhalene previously led the Department of Informatics and Mathematics Sciences at the same institution. His scholarly pursuits span a wide array of fields including optimization and operational research, heuristic methods, statistics, probability, reliability, game theory and mechanism design, simulation and modeling techniques, process robotics and automation, data analytics, big data, and scientific computing. Dr. Bouikhalene conducts his research within the esteemed LIMATI laboratory at Sultan Moulay Slimane University. He can be contacted at email: b.bouikhalene@usms.ma.