

Food sales prediction model using machine learning techniques

Hussam Mezher Merdas, Ayad Hameed Mousa

Department of Computer Science, College of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq

Article Info

Article history:

Received Jan 9, 2023

Revised Mar 15, 2023

Accepted Apr 7, 2023

Keywords:

Bagging regressor

Decision tree

Food sales prediction

Gradient boosting

Least absolute shrinkage and selection operator

Random forest regressor

Support vector machines

ABSTRACT

Food sales prediction means how to obtain future results of sales of companies. The purpose of this step is to increase the profits of these companies by avoiding spoilage of food products and avoiding buying more quantities than the needs of these companies, which means the accumulation of these products in the warehouses without selling them. Stocked and expired products require a model that guesses the actual future need for these products. In this study, a model for food sales prediction using machine learning algorithms is proposed to achieve two objectives, first: make a comparison between two datasets, one dataset with a high correlation between its features, and another dataset has a low correlation between its features. The second objective is to use several machine learning algorithms for prediction and comparing between these algorithms to find the best three algorithms that give the best prediction. By using the most important metrics such as root mean square error (RMSE) and mean square error (MSE) found the best three algorithms by using the first dataset are support vector machines (SVMs), least absolute shrinkage and selection operator (LASSO), and bagging regressor) and the best three algorithms by using the second dataset are (gradient boosting, random forest regressor, and decision tree).

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Hussam Mezher Merdas

Department of Computer Science, College of Computer Science and Information Technology, University of Kerbala

Karbala, Iraq

Email: hussam.m@s.uokerbala.edu.iq

1. INTRODUCTION

Food sales are an inspiring topic for researchers today due to their importance for companies, merchants, and entrepreneurs [1]. However, food sales prediction is an important and critical step and at the same time useful for these companies to know the expected quantities of food will need in the future and ensure that food is not damaged as a result of its expiration date [2]. Accurate prediction of food sales is a difficult task due to the dynamic change of external parameters. In general Time series data control food sales prediction [1]. The huge amounts of data collected and the large number of features affect each other making data impossible to be processed manually. Here comes the role of artificial intelligence to solve this problem and automatically predict the future of sales. It is the responsibility of artificial intelligence algorithms to process the data and give the appropriate prediction. In this regard, many algorithms differ from each other in the results they provide to give a good prediction.

Many researchers have addressed the topic of food sales prediction, and many algorithms have been used. But most researchers did not use only one, two, or three algorithms. Also, they often use one dataset. As previously mentioned, this study, used more than one dataset and more than one algorithm. So, the related works can be listed below. According to the study mentioned in source [3], the researchers designed two prediction models: the main prediction model and the weather prediction model to prove the effect of weather

factors on the sales prediction of fresh products. The algorithms used are random forest (RF), ridge regression (RR), and support vector machine (SVM). The results of this study by using the relative reduction rate of the using root mean square error (RMSE) for the three algorithms were: ridge regression 68.90%, RF 23.66%, SVM 59.52%, and the results of this study by using relative reduction rate of the mean absolute percentage error was: ridge regression 66.2%, RF 34.99%, and SVM 61.13%.

According to Prabhakar *et al.* [4], predicted the future sales of one of the leading competitor's grocery store chains in the market. This study explains the algorithms used to predict the unit sales of products sold by this grocery store. Predictive modeling is used to solve the analytics problem that came from daily trading, continuous selling, and the relationship of products to each other. The results of this study by using RMSE: linear regression 11.97, gradient boosting 9.5, neural network (NN) 9.82, SVM still running for 5 hours without result.

According to Arunraj *et al.* [5], this study aims to build a model which tries to take into consideration all the effects of the on-demand factors, to predict the daily sales of spoiled food in a store. So, using a prediction system that Calculates uncertainty and accuracy which are important measures to prevent food waste. The results of this study by using RMSE is 0.613.

According to Posch *et al.* [6], a designed model for predicting future sales of items in restaurants. Which use two Bayesian generalized additive models. The first model assumes a normally distributed for future sales, while the second one depended on the negative binomial distribution. They used two datasets collected from point of sale (POS) systems in a restaurant and a staff canteen. The results showed that the used approach often provided the most robust point predictions and the best. The negative binomial distribution of the model is more accurate than the ones obtained by the other.

According to Meulstee and Pechenizkiy [7], an ensemble learning approach for sales prediction has been presented. Dynamic integration of classifiers employs to guarantee good handling of fluctuations and seasonal changes in consumer demands. So, they focus their study on how the food wholesale forecasting is currently operated in the company, and how can improve product prediction. They evaluate their approach by using real data collected by the retailing company. The results prove that the ensemble learning approach is better than the currently used baseline, and if the feature set for a target product is complemented with features, the ensemble learning approach can handle seasonal changes. And if information about holidays and the weather is presented explicitly in a feature set, an ensemble can become more accurate.

Finally, according to Schmidt *et al.* [8], each day is broken up into three sales periods because of predicted partitions of sales days, so they focused on the middle timeslot, where sales predicting should extend for one week. They use a dataset generated by three years of sales between 2016 to 2019 from a restaurant for researching sales prediction methods. They make preprocess and convert raw data to a workable dataset. So, they tested many machine learning (ML) algorithms on the dataset, such as recurrent neural network (RNN) and ridge model. As a result, the Ridge regression model was the best model for one-day forecasting with a mean absolute error (MAE) of 214, and the temporal fusion transformer model was the best for one-week forecasting with an MAE of 216.

In this paper different datasets are used, one with a strong correlation, this dataset is obtained from "publicly available Alibaba's Tianchi platform data", this dataset consists of 15 attributes and 1,000 objects containing sales data of different cities and branches between its features, and the other dataset with little correlation includes sales data from ten stores in different cities sourced from Kaggle, consisting of 12 attributes and 8,523 objects, to study the results of both. The two datasets were examined to detect the missing values, as they were compensated using appropriate mathematical operations, such as taking the mean for the rest of the values. As well as processing the repeated values, and spelling errors in the entries were also processed, so these datasets were filtered and prepared in the appropriate way to present them to the algorithms that will use. These algorithms will be ten algorithms. Finally, several metrics will be used to evaluate these algorithms to select the best three. To take it into account as the appropriate algorithm for such datasets. The main objective of this process is to help companies to know the future sales of their products and the consequent reduction of losses resulting from food spoilage and achieving more profits.

2. THE PROPOSED MODEL

The proposed model is based on two key steps as shown in Figure 1. The first step is to use two datasets with different content and different correlation between their features. The goal of this step is to make a practical comparison between these two datasets by doing the necessary processing on them and inserting them into several algorithms and observing the results to benefit from this step in subsequent studies or for companies that have similar databases. The second step is to use ten different ML algorithms after making preprocessing data that utilize. Then evaluate these algorithms using several different metrics. And by using these metrics, the best three algorithms have been determined and recommended for future utilization in the field of predicting food sales.

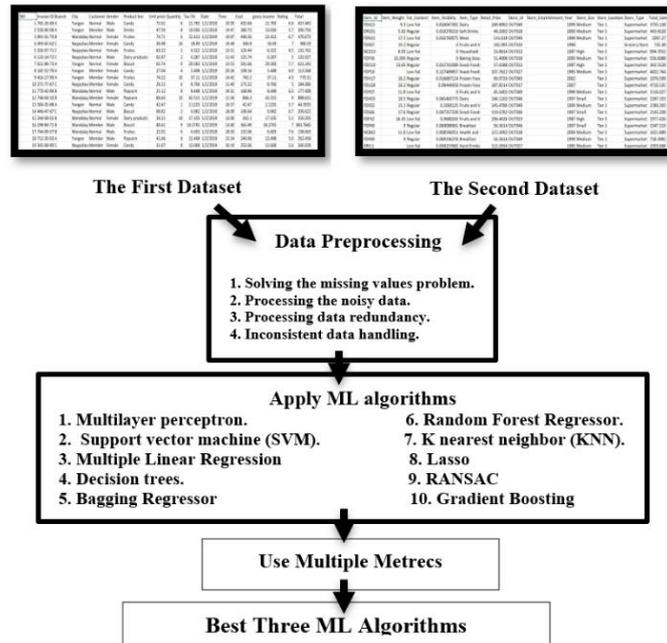


Figure 1. The proposed model

3. METHOD

3.1. General steps

The main steps in the proposed model start with loading the two datasets, then performing the visualization and pre-processing operations, this step is followed by measuring the degree of correlation between the features of each dataset separately. After that the ten ML algorithms are applied to each dataset, then the results are evaluated using accuracy measures. Finally, the three best algorithms that gave the highest accuracy are obtained with each dataset separately as shown in Figure 2.

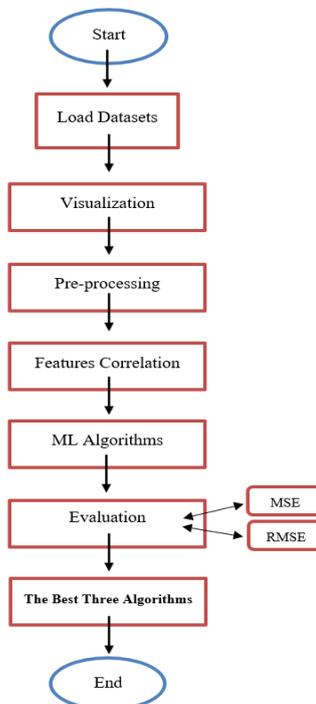


Figure 2. General steps

3.2. The datasets correlation

To find out the degree of correlation between the features of the datasets, the “heatmap correlation matrix” was used. It is a visualization tool provided by “The Seaborn library in Python”. It is a matrix of colors used to describe the correlation between attributes in each selected dataset. As clearly shown in Figures 3(a) and 3(b), the higher intensity of the color between two attributes indicates that one attribute depends on the other attribute. While lower intensity is Indicate the weakness of the attribute’s dependence on the corresponding attribute. Therefore, it can be concluded that the correlation between the dataset 1 attributes is higher than the dataset 2 attributes correlation.

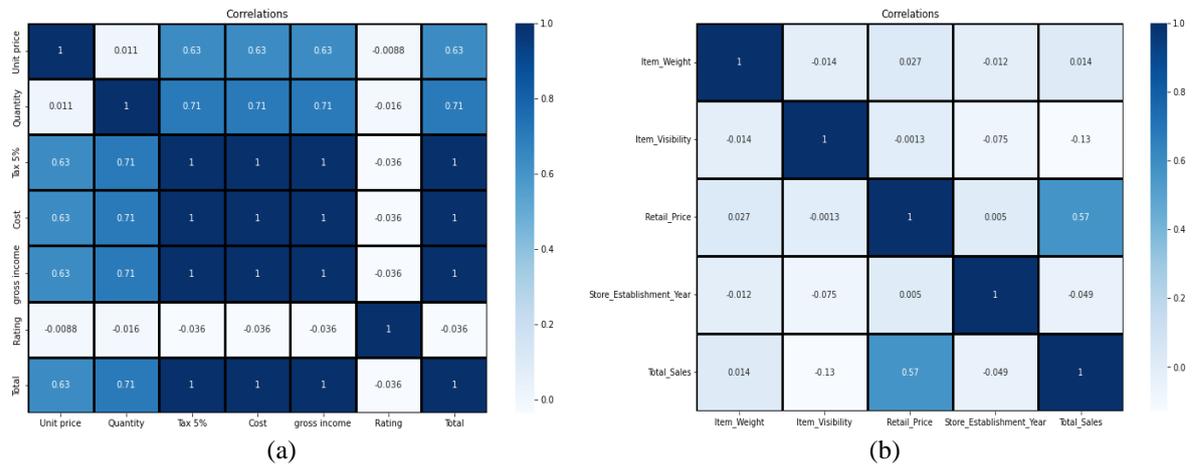


Figure 3. The correlation between features (a) dataset selected 1 and (b) dataset selected 2

3.3. ML algorithms used

In the context of this study, ten supervised ML algorithms (multilayer perceptron, SVM, multiple linear regression, decision trees, RF regressor, K-nearest neighbor (KNN), bagging regressor, least absolute shrinkage and selection operator (LASSO), random sample consensus (RANSAC), and gradient boosting) were used. The goal of that is to achieve a comprehensive and effective comparison to obtain satisfactory results. The following sub-paragraphs highlighted the selected ML algorithms that were conducted.

3.3.1. Multilayer perceptron

The multilayer perceptron (MLP) is a type of neural network this type is the most frequently and most known used [9]. It is a mathematical system derived from the human nervous system and the structure of the brain [10]. The MLP has three basic layers; the input layer, the hidden layer, and the output layer. As a human brain, the information transmitting in one direction from the input layer to the output layer in multilayer feed-forward networks. The Boolean function had been used in feed-forward networks [11]. The ability of the network to perform complex functions depends on the number of hidden layers and the number of nodes in each hidden layer [11]. However, the complexity of the increased number of hidden layers and the number of nodes does not necessarily mean performs better.

3.3.2. SVM

SVMs is a supervised learning method used as a classification and regression prediction tool [12]. SVM based on ML theory to improve predictively. SVM uses hypothesis space of linear functions in a high dimensional feature space, by using a learning algorithm from optimization theory [12]. It is being used for many applications in the regression field. Later, support vector regression (SVR) was derived from SVM which had high performance in the nonlinear regression problems [13]. The idea was: based on the information of given training samples, input variables are mapped from the primal space to higher dimensional feature space built a linear regression function [14].

3.3.3. Multiple linear regression

Linear regression is a method that estimates the value of a variable by using another variable importance [15]. In machine learning, linear regression is a popular and widely-used method. The simplicity of linear regression makes it so popular [16]. By using past data, it tries to predict values. Linear regression has two types: simple and multiple linear regression. Multiple linear regression is the most popular type of

predictive analysis [17]. Multiple linear regression determined a relationship among multi-random variables. So, it shows how the relationship is located between multiple independent variables on the one hand and a dependent variable on the other hand [18]. Its equation can be written as:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 \cdots + \beta_kX_k + \varepsilon$$

where Y_i is predicted (dependent), β is a parameter, X is an independent variable, and ε is the random error.

3.3.4. Decision tree

From the concept of a real tree, the concept of a decision tree emerged in the field of artificial intelligence [19]. A decision tree is used to represent decisions and decisions making. The decision tree algorithm is a supervised learning algorithm used for regression and classification [20]. A decision tree can be used to create a training model that can use to predict the value of the target variable by using decision rules from history data (training data) [21]. When using decision trees in prediction we compare the root attribute values with the record's attribute values. The depth of the decision tree controls its efficiency [22].

3.3.5. Random forest regressor

RF collects multiple decision trees, its response is to give efficient results by using of bagging mechanism [23]. Bagging is one of the most common ensemble techniques [24]. Ensemble methods use several learning models to give the best prediction rather than using any machine learning algorithm alone [25]. RF was used for classification and regression. For regression, the algorithm returns the average prediction or mean of the individual trees [25]. Controlling the depth of the RF allows for optimizing and giving the best prediction result [26]. Each decision tree makes a number of splits and by using these splits we find max depth [23].

3.3.6. K-nearest neighbors

The KNN algorithm is a supervised learning approach used for regression and classification [27]. Based on the similarity between the target and other available cases, KNN calculates the prediction value by using the distance measure [28]. The Euclidian distance is the most popular approach for finding the distance [28]. KNN works by finding the K most similar neighbors (instances) to the testing point from our datasets [29]. By using the Euclidean distance formula $\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$, KNN finds the mathematical distance between points [30]. Selection best value of K is difficult. Noise will affect the result if K was small and if K was a large value which means computationally expensive [30]. To select the best K value for our data we used different values of K and implement the algorithm several times.

3.3.7. Bagging regressor

Bootstrap Aggregation or as it is called Bagging is one of the ensemble machine learning algorithms that combine many decision trees to make a prediction [24]. Bagging works as a basis for several algorithms. From the idea of the "bootstrap" sample, the bootstrap aggregation had come [24]. That means using a sample drawn from the dataset with replacement, so it is selected multiple times as a new sample. A population statistic had been found from a small data sample by using the bootstrap sampling technique. This gets through drawing multiple bootstrap samples then calculating the statistic on each, and taking the mean statistic for all samples. That gives an accurate approach to estimating statistical quantities compared [24]. Can create an ensemble of decision tree models by using the same approach, by using multiple bootstrap samples drawing from the training dataset then fitting a decision tree on each. Finally, combining predictions of the decision trees gives a more robust prediction than a single decision tree [31].

3.3.8. LASSO regression

LASSO is a method used for regression [32]. LASSO works as a regularization technique. Regularization is used to avoid overfitting the data to enhance the prediction accuracy by using it over regression methods [33]. LASSO shrinks data values and takes a central point as the mean by using an operation called shrinkage [34]. Ridge regression and LASSO regression are the main regularization techniques [32]. In this study, we will focus on LASSO regression. LASSO regression uses the L1 regularization technique [32].

3.3.9. RANSAC regression

The RANSAC algorithm is an approach that works as an iterative method that estimates the parameters of a model from a set of data inputs to cope with outliers [35]. This algorithm was proposed by Fischler and Bolles [36]. RANSAC had been used for regression in our study because it works very well with outliers [37].

So, it works by selecting a random subset of data samples and then using this subset to estimate model parameters [38]. In the next step, it finds the samples that are within model error tolerance. In this step, this algorithm considered these samples as a consensus set. So, the data samples in the consensus are called inliers and the rest are called outliers [39]. RANSAC trains the model by using samples in the consensus if it has a high count. It returns the model that has the smallest error by repeating these steps for some iterations.

3.3.10. Gradient boosting

Gradient boosting is a popular machine learning technique used for classification and regression [40]. It works well with missing values and outliers and it can discover nonlinear relationships between features [41]. Gradient boosting is one of the ensemble methods which combines multiple weak models to give better performance [42]. It is a boosting algorithm that minimized the bias error of the model [43]. Mean square error (MSE) had been used as a cost function for regression tasks [44]. Gradient boosting is a very powerful technique for prediction [41]. It optimizes the prediction accuracy of multiple functions and finally gives a high-accuracy prediction.

4. RESULTS AND DISCUSSION

The proposed model in this study adopted important metrics to compare the accuracy obtained from each algorithm. These metrics are RMSE and MSE. The results are shown in Tables 1 and 2. And by using two types of datasets and ten different algorithms the following was achieved:

- The algorithms give excellent outputs when applied to the first dataset, which had a high correlation between its features, unlike the other dataset, which had a low correlation between its features.
- The best three algorithms applied on dataset 1 that have the highest accuracy and lowest (MSE and RMSE) are (SVR, LASSO, and bagging regressor). And the best three algorithms applied to dataset 2 are (gradient boosting, random forest regressor, and decision tree).

Table 1. Results of the dataset 1

Algorithm	RMSE	MSE	Accuracy
Multilayer perceptron	1.8867	3.5598	99.99
SVR	0.0510	0.0027	100.00
Multiple linear regression	2.6009	6.7647	100.00
Decision tree	2.6937	7.2561	99.99
Random forest regressor	2.0922	4.3775	99.99
KNN	22.6160	511.4860	99.02
Bagging regressor	1.4915	2.2246	100.00
LASSO	0.0210	0.0004	100.00
RANSAC	1.7274	2.9842	100.00
Gradient boosting	2.1722	4.7184	99.99

Table 2. Results of the dataset 2

Algorithm	RMSE	MSE	Accuracy
Multilayer perceptron	1251.1710	1565428.9678	46.52
SVR	1238.1194	1532939.8127	47.63
Multiple linear regression	1203.8843	1449337.6418	50.48
Decision tree	1105.0661	1221171.0931	58.28
Random forest regressor	1094.0445	1196933.5811	59.11
KNN	1545.5220	2388638.2604	18.39
Bagging regressor	1212.0442	1469051.3589	49.81
LASSO	1203.9653	1449532.5095	50.47
RANSAC	1611.3576	2596473.3284	11.29
Gradient boosting	1091.6557	1191712.2726	59.28

5. CONCLUSION

The correlation between datasets features has a significant impact on the results of the applied algorithms. When using datasets that have a high correlation between their features, it is preferable to use (SVR, LASSO, and bagging regressor) algorithms. And when using datasets that have a low correlation between their features, (gradient boosting, random forest regressor, and decision trees) are preferable. This study sets a clear roadmap for future studies and saves them time to use the appropriate algorithm for datasets with certain features, as well as saves the effort for companies by putting in their hand appropriate algorithms to apply to their datasets.

REFERENCES

- [1] H. Bousqaoui, I. Slimani, and S. Achchab, "Comparative analysis of short-term demand predicting models using ARIMA and deep learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3319–3328, 2021, doi: 10.11591/ijece.v11i4.pp3319-3328.
- [2] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 1st ed. Australia: OTexts, 2013.
- [3] X. Wang, D. Lin, W. Fan, and T. Wang, "Research on sales forecast of fresh produce considering weather factors," in *Proceedings of the International Conference on Electronic Business (ICEB)*, 2018, pp. 541–548.
- [4] V. Prabhakar, D. Sayiner, U. Chakraborty, T. Nguyen, and M. A. Lanham, "Demand forecasting for a large grocery chain in Ecuador," *GüHub*. pp. 1–18, 2018.
- [5] N. S. Arunraj, D. Ahrens, M. Fernandes, and M. Müller, "Time series sales forecasting to reduce food waste in retail industry," *The 34th International Symposium on Forecasting*, 2014. doi: 10.13140/RG.2.1.4829.1607.
- [6] K. Posch, C. Truden, P. Hungerländer, and J. Pilz, "A Bayesian approach for predicting food and beverage sales in staff canteens and restaurants," *International Journal of Forecasting*, vol. 38, no. 1, pp. 321–338, Jan. 2022, doi: 10.1016/j.ijforecast.2021.06.001.
- [7] P. Meulstee and M. Pechenizkiy, "Food sales prediction: 'if only it knew what we know,'" in *2008 IEEE International Conference on Data Mining Workshops*, 2008, pp. 134–143. doi: 10.1109/ICDMW.2008.128.
- [8] A. Schmidt, M. W. U. Kabir, and M. T. Hoque, "Machine learning based restaurant sales forecasting," *Machine Learning and Knowledge Extraction*, vol. 4, no. 1, pp. 105–130, Jan. 2022, doi: 10.3390/make4010006.
- [9] M.-C. Popescu, E. B. Valentina, P.-P. Liliana, and M. Nikos, "Multilayer perceptron and neural networks," *WSEAS Transactions on Circuits and Systems*, vol. 8, no. 7, pp. 579–588, 2009, doi: 10.5555/1639537.1639542.
- [10] W. Wiharto, H. Kusananto, and H. Herianto, "Hybrid system of tiered multivariate analysis and artificial neural network for coronary heart disease diagnosis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 2, pp. 1023–1031, Apr. 2017, doi: 10.11591/ijece.v7i2.pp1023-1031.
- [11] W. Castro, J. Oblitas, R. Santa-Cruz, and H. Avila-George, "Multilayer perceptron architecture optimization using parallel computing techniques," *PLOS ONE*, vol. 12, no. 12, Dec. 2017, doi: 10.1371/journal.pone.0189369.
- [12] V. Jakkula, "Tutorial on support vector machine (SVM)," *Northeastern University*. Accessed: Apr. 20, 2022. [Online]. Available: <http://www.ccs.neu.edu/course/cs5100f11/resources/jakkula.pdf>
- [13] R. M. Balabin and E. I. Lomakina, "Support vector machine regression (SVR/LS-SVM) - An alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data," *Analyst*, vol. 136, no. 8, pp. 1703–1712, 2011, doi: 10.1039/c0an00387e.
- [14] A. Shmilovici, "Support vector machines," in *Data Mining and Knowledge Discovery Handbook*, New York: Springer-Verlag, pp. 257–276. doi: 10.1007/0-387-25465-X_12.
- [15] D. R. Thomas, P. C. Zhu, and Y. J. Decady, "Point estimates and confidence intervals for variable importance in multiple linear regression," *Journal of Educational and Behavioral Statistics*, vol. 32, no. 1, pp. 61–91, Mar. 2007, doi: 10.3102/1076998606298037.
- [16] S. Weisberg, *Applied linear regression*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005. doi: 10.1002/0471704091.
- [17] G. K. Uyanik and N. Güler, "A study on multiple linear regression analysis," *Procedia - Social and Behavioral Sciences*, vol. 106, pp. 234–240, 2013, doi: 10.1016/j.sbspro.2013.12.027.
- [18] K. F. Nimon and F. L. Oswald, "Understanding the results of multiple linear regression: beyond standardized regression coefficients," *Organizational Research Methods*, vol. 16, no. 4, pp. 650–674, Jul. 2013, doi: 10.1177/1094428113493929.
- [19] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics*, vol. 18, no. 6, pp. 275–285, Jun. 2004, doi: 10.1002/cem.873.
- [20] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhama, "Analysis of various decision tree algorithms for classification in data mining," *International Journal of Computer Applications*, vol. 163, no. 8, pp. 15–19, Apr. 2017, doi: 10.5120/ijca20171913660.
- [21] E. J. Shim *et al.*, "An MRI-based decision tree to distinguish lipomas and lipoma variants from well-differentiated liposarcoma of the extremity and superficial trunk: classification and regression tree (CART) analysis," *European Journal of Radiology*, vol. 127, Jun. 2020, doi: 10.1016/j.ejrad.2020.109012.
- [22] Y. Liu and S. Yang, "Application of decision tree-based classification algorithm on content marketing," *Journal of Mathematics*, pp. 1–10, Mar. 2022, doi: 10.1155/2022/6469054.
- [23] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu, and L. Zhu, "A missing power data filling method based on improved random forest algorithm," *Chinese Journal of Electrical Engineering*, vol. 5, no. 4, pp. 33–39, 2019, doi: 10.23919/CJEE.2019.000025.
- [24] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine learning paradigms in hydrology: a review," *Journal of Hydrology*, vol. 598, Jul. 2021, doi: 10.1016/j.jhydrol.2021.126266.
- [25] D. Opitz and R. Maclin, "Popular ensemble methods: an empirical study," *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, 1999, doi: 10.1613/jair.614.
- [26] E. Scornet, "Tuning parameters in random forests," *ESAIM: Proceedings and Surveys*, vol. 60, pp. 144–162, 2017, doi: 10.1051/proc/201760144.
- [27] A. Pandey and A. Jain, "Comparative analysis of KNN algorithm using various normalization techniques," *International Journal of Computer Network and Information Security*, vol. 9, no. 11, pp. 36–42, Nov. 2017, doi: 10.5815/ijcnis.2017.11.04.
- [28] S. Zhang, D. Cheng, Z. Deng, M. Zong, and X. Deng, "A novel KNN algorithm with data-driven k parameter computation," *Pattern Recognition Letters*, vol. 109, pp. 44–54, Jul. 2018, doi: 10.1016/j.patrec.2017.09.036.
- [29] S. Karthick, "Semi supervised hierarchy forest clustering and Knn based metric learning technique for machine learning system," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 9, pp. 2679–2690, 2017.
- [30] A. R. Lubis, M. Lubis, and A.- Khowarizmi, "Optimization of distance formula in K-Nearest Neighbor method," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 1, pp. 326–338, Feb. 2020, doi: 10.11591/eei.v9i1.1464.
- [31] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug. 1996, doi: 10.1007/BF00058655.
- [32] J. Ranstam and J. A. Cook, "LASSO regression," *British Journal of Surgery*, vol. 105, no. 10, p. 1348, 2018, doi: 10.1002/bjs.10895.
- [33] F. Santosa and W. W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 4, pp. 1307–1330, 1986, doi: 10.1137/0907087.
- [34] F. X. Diebold and M. Shin, "Machine learning for regularized survey forecast combination: partially-egalitarian LASSO and its derivatives," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1679–1691, 2019, doi: 10.1016/j.ijforecast.2018.09.006.
- [35] R. Subbarao and P. Meer, "Beyond RANSAC: user independent robust regression," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, 2006, doi: 10.1109/CVPRW.2006.43.

- [36] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981, doi: 10.1145/358669.358692.
- [37] T. Strutz, "Data fitting and uncertainty: a practical introduction to weighted least squares and beyond," *Vieweg + Teubner*, Wiesbaden, pp. 1-244, 2011.
- [38] M. S. Barkhordari and M. S. Es-Haghi, "Straightforward prediction for responses of the concrete shear wall buildings subject to ground motions using machine learning algorithms," *International Journal of Engineering, Transactions A: Basics*, vol. 34, no. 7, pp. 1586–1601, Jul. 2021, doi: 10.5829/IJE.2021.34.07A.04.
- [39] V. Jarunakarint and S. Utama, "Detection of risky riding patterns of motorcyclists based on deep learning and linear regression," *International Journal of Science*, vol. 18, no. 1, pp. 136–151, 2021.
- [40] O. Wisesa, A. Andriansyah, and O. Ibrahim Khalaf, "Prediction analysis for business to business (B2B) sales of telecommunication services using machine learning techniques," *Majlesi Journal of Electrical Engineering*, vol. 14, no. 4, pp. 145–153, 2020, doi: 10.29252/mjee.14.4.145.
- [41] S. M. Piryonesi and T. E. El-Diraby, "Data analytics in asset management: cost-effective prediction of the pavement condition index," *Journal of Infrastructure Systems*, vol. 26, no. 1, Mar. 2020, doi: 10.1061/(ASCE)IS.1943-555X.0000512.
- [42] P. Kumari and D. Toshniwal, "Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance," *Journal of Cleaner Production*, vol. 279, Jan. 2021, doi: 10.1016/j.jclepro.2020.123285.
- [43] S. Islam and S. H. Amin, "Prediction of probable backorder scenarios in the supply chain using distributed random forest and gradient boosting machine learning techniques," *Journal of Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00345-2.
- [44] A. Subasi, M. F. El-Amin, T. Darwich, and M. Dossary, "Permeability prediction of petroleum reservoirs using stochastic gradient boosting regression," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 7, pp. 3555–3564, Apr. 2022, doi: 10.1007/s12652-020-01986-0.

BIOGRAPHIES OF AUTHORS



Hussam Mezher Merdas     obtained his Bachelor's degree in Computer Science from University of Baghdad, Iraq, in 2016. Works as a lecturer at the College of Computer Science and Information Technology, University of Kerbala, Information Technology Department, Iraq. He can be contacted at email hussam.m@s.uokerbala.edu.iq.



Ayad Hameed Mousa     received his Ph.D. in computer science in 2017 and is an assistant professor at the College of Computer Science and Information Technology, University of Kerbala, department of computer science. His research is situated in the field of Technology and Innovation, with a special focus on software engineering, business intelligence modeling, data mining, data warehouse, and e-learning modeling. He is actively engaged in several scientific projects (bilateral cooperation, national projects). A university lecturer from 2005 and ongoing. He is the author or co-author of more than 34 papers in international journals and international conference contributions. He can be contacted at email: ayad.h@uokerbala.edu.iq.