

Article

Development and Validation of Innovative Machine Learning Models for Predicting Date Palm Mite Infestation on Fruits

Maged Mohammed ^{1,2,*} , Hamadttu El-Shafie ^{1,3}  and Muhammad Munir ¹ 

¹ Date Palm Research Center of Excellence, King Faisal University, Al-Ahsa 31982, Saudi Arabia

² Agricultural and Biosystems Engineering Department, Faculty of Agriculture, Menoufia University, Shebin El Koum 32514, Egypt

³ Department of Crop Protection, Faculty of Agriculture, University of Khartoum, Shambat 13314, Sudan

* Correspondence: memohammed@kfu.edu.sa

Abstract: The date palm mite (DPM), *Oligonychus afrasiaticus* (McGregor), is a key pest of unripe date fruits. The detection of this mite depends largely on the visual observations of the webs it produces on the green fruits. One of the most important problems of DPM control is the lack of an accurate decision-making approach for monitoring and predicting infestation on date fruits. Therefore, this study aimed to develop, evaluate, and validate prediction models for DPM infestation on fruits based on meteorological variables (temperature, relative humidity, wind speed, and solar radiation) and the physicochemical properties of date fruits (weight, firmness, moisture content, total soluble solids, total sugar, and tannin content) using two machine learning (ML) algorithms, i.e., linear regression (LR) and decision forest regression (DFR). The meteorological variables data in the study area were acquired using an IoT-based weather station. The physicochemical properties of two popular date palm cultivars, i.e., *Khalas* and *Barhee*, were analyzed at different fruit development stages. The development and performance of the LR and DFR prediction models were implemented using Microsoft Azure ML. The evaluation of the developed models indicated that the DFR was more accurate than the LR model in predicting the DPM based on the input variables, i.e., meteorological variables ($R^2 = 0.842$), physicochemical properties variables ($R^2 = 0.895$), and the combination of both meteorological and the physicochemical properties variables ($R^2 = 0.921$). Accordingly, the developed DFR model was deployed as a fully functional prediction web service into the Azure cloud platform and the Excel add-ins. The validation of the deployed DFR model showed that it was able to predict the DPM count on date palm fruits based on the combination of meteorological and physicochemical properties variables ($R^2 = 0.918$). The deployed DFR model by the web service of Azure ML studio enhanced the prediction of the DPM count on the date fruits as a fast and easy-to-use approach. These findings demonstrated that the DFR model using Azure ML Studio integrated into the Azure platform can be a powerful tool in integrated DPM management.

Keywords: decision forest regression (DFR); integrated pest management (IPM); linear regression (LR); machine learning dataset (MLD); meteorological variables; physicochemical properties; predictive models; web service



Citation: Mohammed, M.; El-Shafie, H.; Munir, M. Development and Validation of Innovative Machine Learning Models for Predicting Date Palm Mite Infestation on Fruits. *Agronomy* **2023**, *13*, 494. <https://doi.org/10.3390/agronomy13020494>

Academic Editors: Gniewko Niedbała, Magdalena Piekutowska, Tomasz Wojciechowski and Mohsen Niazian

Received: 26 December 2022

Revised: 5 February 2023

Accepted: 7 February 2023

Published: 8 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The date palm mite (DPM), *Oligonychus afrasiaticus* (McGregor) (Acari: Tetranychidae) is an important and key pest of immature date palm fruits in the Middle East and Africa [1]. It can tremendously reduce the yield and quality of date fruit, causing economic losses [2–6]. This is because the DPM attacks the developing fruits, representing the final yield component. Such direct damage cannot be compensated, unlike damage on frond pinnae (leaflets) [7].

Date fruit development is categorized into hababook, kimri, khalal, biser, rutab, and tamr based on developmental status, color, shape, and chemical properties [8]. Infestation

by the date mite may start after fruit set or it may come late at the kimri stage when the fruits are green with high moisture content. Fruit chemistry greatly influences the DPM multiplication. DPM reproduces profoundly on developing or unripe fruits (kimri stage), characterized by low sugar, high acidity, and high moisture content [9,10]. Mite populations decline during the biser stage, when fruit color changes to yellow or red depending on the date palm cultivar [11–13]. During the rutab stage, when fruits ripen, mite populations decrease and continue to decline further during the tamr or fully ripen stage [14].

The world-famous date palm varieties, such as *Medjool*, *Deglet Noor*, *Barhee*, and *Khalas*, are highly susceptible to mite infestation, and damage estimation may range between 30–80% [9]. Mite infestation usually starts at the calyx end of the fruit and then progresses toward the fruit tip. The fruits shrivel, harden, stop development, and produce gum-like exudates, and their content of water and sugars are largely reduced [14]. Heavily infested fruits become covered with fine dust and sand particles, which render them unfit for human consumption [9,15].

One of the most important problems of DPM control is the lack of an accurate decision-making approach for monitoring and predicting infestation on date fruits [6]. This problem led to the heavy use of chemical acaricides in an attempt to control the DPM. However, the heavy use of chemical acaricides created many problems, including environmental pollution and the development of resistance among the mite populations [2,3]. Therefore, monitoring and sampling DPM are considered the backbone of any integrated management program. However, it is extremely difficult to sample mites through a direct count of individuals due to the small size of the mite relative to the size of the date palm and date fruit [16].

Moreover, it is also challenging to detect mite infestation at an early stage, and it is possible only after the appearance of silken webs, and at that time, the mite colonies would have already established themselves and started inflicting damage on fruits [3,7]. To date, no sampling plan is available for the DPM, which can be used in management programs [11]. Due to these difficulties in sampling and monitoring, scientists have studied some modern methods of controlling DPM during the last decade. For example, simple forecasting models based on meteorological variables, dispersion index pests, and geostatistics have been used to forecast DPM infestation on date fruits [17].

Modern environmental and precision agriculture research has recently merged with machine learning (ML) techniques to find out innovative solutions for agricultural issues [18–20]. The most common regression methods in Azure ML are linear regression (LR), neural network regression (NNR), and decision forest regression (DFR) [21]. The artificial neural network (ANN) has been used to predict the number of insects collected in various traps [22]. The recent success of the DFR algorithm in classification and regression problems is largely responsible for its adoption in ML. The random forests algorithm are a more versatile technique that may solve various problems in regression, classification, density estimation, semi-supervised learning, manifold learning, and active learning [23].

The user can utilize Azure ML to deploy the ML predictive model in production for commercial usage after developing and testing it using any ML regression algorithm. Azure ML's simplicity of deployment in production is a key distinction. The prediction model can be developed and made available as a web service. Once the model is deployed, it may be accessed as a web service from various devices, including tablets, computers, and even smartphones [21]. For predicting the risk that pests may cause plant damage, ML techniques are regarded as useful resources [18]. The impact of weather on pest development in rice in the context of ML-based applications revealed a significant correlation between the number of insect catches and temperature compared to other factors [24]. The risk of a mountain pine beetle outbreak of up to seven years was predicted using nine ML models, and although the prediction performance of various models changes with the history length of the covariates, the NN and naive Bayes models predicted more accurately, followed by the generalized boosted regression trees model [25]. The damage caused by the insects *Spodoptera frugiperda* and *Dichelops melacanthus* was predicted using random forest, support vector machine, extreme gradient boost, and neural network models based on the spectrum

response in maize, and the random forest-based framework algorithm was recommended as the best one [26]. The appearance of *Helicoverpa armigera* insects was successfully predicted using temperature and RH environmental variables in the ML model [27].

Accordingly, the ML prediction models could be an alternative tool for predicting the mite counts on the date palm fruits and providing valuable data for initiating management measures against this severe pest. To our knowledge, no research has been conducted previously for predicting DPM count on date fruits based on the combination of the meteorological variables data and physicochemical properties of date fruits using machine learning models. Therefore, the main objective of this study was to develop, evaluate, and validate machine-learning-based prediction models to predict DPM infestation on date fruits using Azure ML studio based on the meteorological variables data and the physicochemical properties of date fruits during development. To achieve this objective, different steps were followed according to the following sequence:

- Sampling and counting of DPM on developing fruits from the first of April (one month after the fruit set) to the end of August on a weekly basis;
- Analysis of the physicochemical properties of date fruits during development as sampled in the first step;
- Recording of the meteorological variables data in the study area using an IoT-based weather station;
- Development and evaluation of prediction models based on the meteorological variables data and the physicochemical properties of the date fruits;
- Validation and deployment of the developed models.

2. Materials and Methods

2.1. Study Area and Meteorological Data Collection

The present study was performed in an arid region at the experimental farm of Date Palm Research Center of Excellence (latitude: 25°16′04.3″ N, longitude: 49°42′30.1″ E) at the Agricultural Training and Research Station, King Faisal University (KFU), Saudi Arabia. The experiment was performed during two successive seasons (2021 and 2022). The study began on 1 April (one month after fruit set) and lasted until the end of August in each experimental season. The meteorological variables data and fruit properties at various development stages, i.e., hababook, kimri, khalal, rutab, and tamr, of two fully grown date palm cultivars (*Khalas* and *Barhee*), were collected from the study area for training and testing the machine learning regression models for predicting DPM infestation on date palm fruits. In addition, the main meteorological variables data, i.e., the maximum temperature, minimum temperature, average temperature, maximum relative humidity, minimum relative humidity, average relative humidity, average wind speed, max solar radiation, and average daily solar radiation, were collected by an IoT-based weather station installed at the study area by Mohammed et al. [28].

2.2. Date Palm Mite Sampling and Counting Data

In most cases, field mite infestation is clumped, i.e., it is not evenly distributed among fruit bunches on the date palm. Therefore, a cluster sampling approach [29] was followed in the mite sample collection. According to this method, infested bunches were categorized into different strata depending on the level of infestation (low, medium, and high) before fruit samples were randomly selected from each stratum. Ten palm trees from each cultivar (*Khalas* and *Barhee*) were selected randomly from the date palm farm in the study area. Four fruit bunches from the four cardinal directions were randomly selected from each date palm tree. Twenty five fruits were randomly taken from each bunch (a total of 100 fruits) every week. They were put into an ice box to immobilize mites before being brought to the laboratory for counting and estimating the rate of infestation.

Due to the small size of the DPM (0.2–0.5 mm), its high density, mobility, and ability to hide beneath the silken webs, it is extremely difficult to make a direct visual count in the field [30]. Therefore, indirect counting followed. Mites were extracted from infested fruits

using the funnel technique, where the Tullgren funnels 6 banks (Tullgren Funnel, Burkard Manufacturing Co. Ltd., Rickmansworth, Hertfordshire, UK) were used. Further extraction of the remaining mites was performed using a mite-extracting machine (brushing machine, Leedom Enterprises, Mi-Wuk Village, CA, USA). This machine has two electric motors, one for rotating the turntable glass palate and the other for rotating the brushes that dislodge the mite from the infested fruits. Brushing was performed for 5 s for each fruit.

Additionally, a fine camel hairbrush was gently moved on the fruit skin to ensure all mites on the infested fruits had been removed. Brushed mites were deposited on a rotating glass plate (12.5 cm diameter) after coating it with an adhesive substance consisting of 55% corn syrup, 44% glycerol, and 1% liquid detergent [30]. The glass plate was placed over a counting grid (pie grid) with 12 wedges. After the brushing machine was switched off, the glass plate with the counting grid was removed, and mite specimens in all wedges were enumerated under a dissecting microscope. Figure 1 shows the extraction of DPM from infested date fruit.

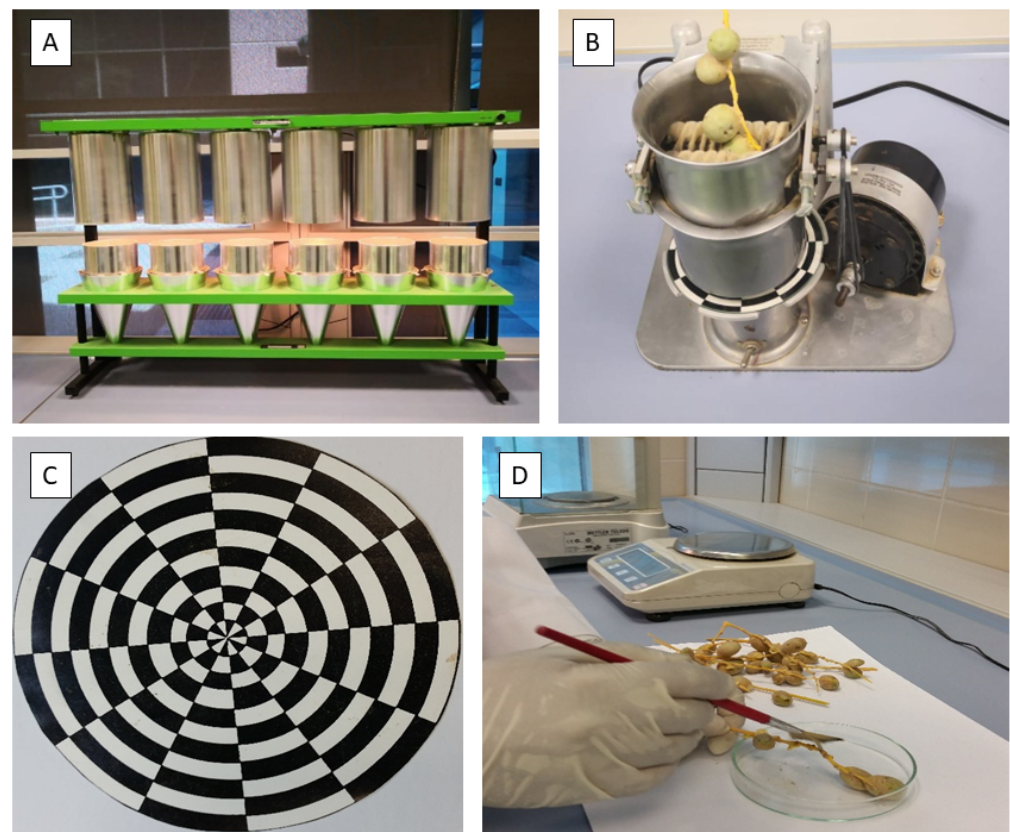


Figure 1. Extraction of date palm mite (DPM) from infested date fruit. Tullgren funnel (A); Mite-extracting machine (B); Mite grid (C); and hair brushing (D).

2.3. Physicochemical Properties of the Fruit

To collect the data of the physicochemical properties for training the models, the samples of two date fruit cultivars (*Khalas* and *Barhee*) were picked during fruit development at five stages, i.e., kimri, khalal, biser, rutab, and tamr. The fruit weights were measured using an electronic balance (Sartorius Lab Instruments GmbH and Co., Göttingen, Lower Saxony, Germany). Fruit pH was determined using a pH meter (S400, Mettler-Toledo LLC, Columbus, OH, USA). Fruit firmness was measured using a Koehler penetrometer (Thomas Scientific, Swedesboro, NJ, USA). The fruit moisture content was determined by drying a date fruit sample of 20 g of each cultivar under a vacuum at 70 °C using a vacuum-drying oven. The sample weight was measured at regular intervals after 48 h in the oven until it was constant [31]. Total soluble solids (TSS) were measured with a digital laboratory

refractometer (Model: RFM 840, Richmond Scientific Ltd. Unit 9, Lancashire, UK). The anthrone–sulfuric acid colorimetry method was used to determine the total sugar [32,33]. About 100 µL of each cultivar's date palm fruit extract was taken, and 900 µL of distilled water was added. Additionally, 1 mL of the anthrone reagent was added, and the mixture was feted for 8 min before cooling to room temperature. The absorbance was measured at 630 nm wavelength using a spectrophotometer (Genesys 20, Thermo Scientific, Waltham, MA, USA). The amount of total sugar in the sample was determined using a standard graph constructed by plotting standard concentration on the x -axis against absorbance on the y -axis. To determine the tannin content in date palm fruit, 5 g fruit pulp was homogenized with 25 mL of 80% methanol and then centrifuged. The supernatant was collected, and the precipitant was re-extracted with 80% methanol and centrifuged. The supernatant was taken, and 100 mL of distilled water was added. Then, a 1 mL sample of the solution was mixed with 6 mL of distilled water and 0.5 mL of Folin–Ciocalteu reagent and shaken. After 3 min, 1 mL of saturated sodium carbonate was added. To make 10 mL of the total solution, 1.5 mL of distilled water was added, and left for one hour at ambient temperature, before measuring absorbance at 750 nm wavelength using a spectrophotometer (Genesys 20, Thermo Scientific, Waltham, MA, USA). The tannin content was calculated using a calibration curve obtained by measuring the absorbance of known concentrations of gallic acid [34].

2.4. Machine Learning Algorithms

In this study, two regression algorithms, i.e., linear regression (LR) and decision forest regression (DFR), were developed, evaluated, and validated using Microsoft Azure Machine Learning (ML) Studio to determine the best model for predicting the DPM on date palm fruits based on three input variables: (1) The meteorological variables of the study area, i.e., the maximum temperature (T_{Max}), minimum temperature (T_{Min}), average temperature (T_{Avg}), maximum relative humidity (RH_{Max}), minimum relative humidity (RH_{Min}), average relative humidity (RH_{Avg}), average wind speed (WS_{Avg}), max solar radiation (SR_{Max}), average daily solar radiation (DSR_{Avg}); (2) The physicochemical properties data of the date fruits during the development stages, i.e., fruit weight (FW), fruit firmness (FF), fruit moisture content (FMC), total soluble solids (TSS), total sugar (TS), and tannin content (TC); and (3) The combination data of meteorological variables and physicochemical properties. The following is a description of the regression algorithms used within Azure ML in this study.

2.4.1. Linear Regression

Linear regression (LR) analyses describe the linear relationship between a response variable (outcome) and the explanatory variables (input). The purpose of LR is to fit a linear model between the response and independent variables to predict the outcome given a set of observed independent variables. The LR equation is as follows:

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_n X_{ni} + \varepsilon \quad (1)$$

where i is n observations, Y is the response variable, X_1 , X_2 , X_3 , and X_n are the independent variables used to predict the outcome, β_0 is the intercept of the regression line, β_1 , β_2 , β_3 , and β_n are the coefficients of the independent variables, and ε is the model's errors

2.4.2. Decision Forest Regression

Decision forest regression (DFR) regression is an ensemble ML technique that involves the creation of numerous decision tree models. These models score new data by evaluating each decision tree and then determining the outcome based on the best result from all decision trees. Combining several ways of deciding can result in a more accurate prediction or result. Random forest models can be used for regression and classification based on problems that include most ML systems and solutions. The predicted value in the regression problems is numeric and on some scale or range. In the DFR, each decision tree will predict a numeric value, and the algorithm will compute the average of these values for the final predicted outcome. Although decision trees using DFR are a famous ML algorithm, they

tend to overfit models, and this overfit can lead to higher-than-expected errors in predicting unrecognized data.

Generally, the boosted decision tree and decision forest regression algorithms build an ensemble of decision trees and use them for predictions. The critical difference between the boosted decision tree and decision forest regression algorithms is that in the boosted decision tree algorithms, numerous decision trees are grown in series such that the output of one tree is delivered as input to the next tree.

During the training phase for the DFR, all trees are independently trained, while the testing and predictions are made through weighted voting on the most confident prediction. The decision trees with higher prediction confidence will have a greater weight in the final prediction decision of the ensemble. A simple averaging operation can aggregate voting according to Equation (2). Figure 2 shows a simple example of a decision forest regression consisting of three trees.

$$p(c|v) = \frac{1}{T} \sum_{t=1}^T P_t(c|v) \quad (2)$$

where $p_t(c|v)$ represents the posterior distribution obtained by the t tree, c is the discrete or continuous labels, and v is the data point.

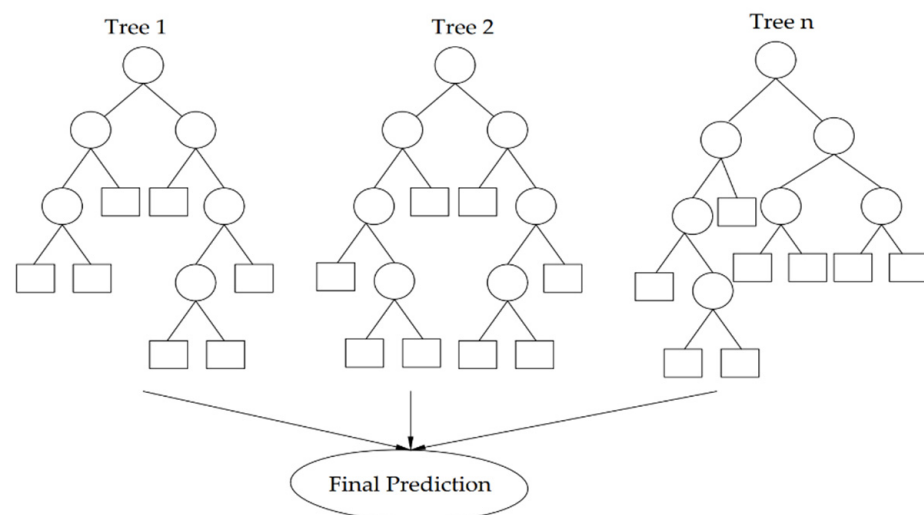


Figure 2. Simple structure for decision forest regression (DFR).

2.5. Building the Azure ML Predictive Models

Figure 3 shows a complete experiment to predict DPM infestation on date palm fruits based on three input variables, i.e., the meteorological variables data of the study area (T_{Max} , T_{Min} , T_{Avg} , RH_{Max} , RH_{Min} , RH_{Avg} , WS_{Avg} , SR_{Max} , and DSR_{Avg}), the physicochemical properties data of the date fruits during the development stages (FW, FF, FMC, TSS, TS, and TC), and the combination data of meteorological variables and physicochemical properties. In addition, the experiment included data processing and building two predictive models, i.e., LR and DFR, using Microsoft Azure ML Studio.

The following describes the components of the predictive experiment.

2.5.1. Data Acquisition

After collecting the DPM infestation counts on the date fruits, meteorological variables data on the study area, and physicochemical properties of the fruits of the two selected cultivars, the dataset was uploaded among other Microsoft Azure ML Studio datasets (Dataset.scv module in Figure 3) and it applied the normalization, if needed, after selecting only the relevant columns.

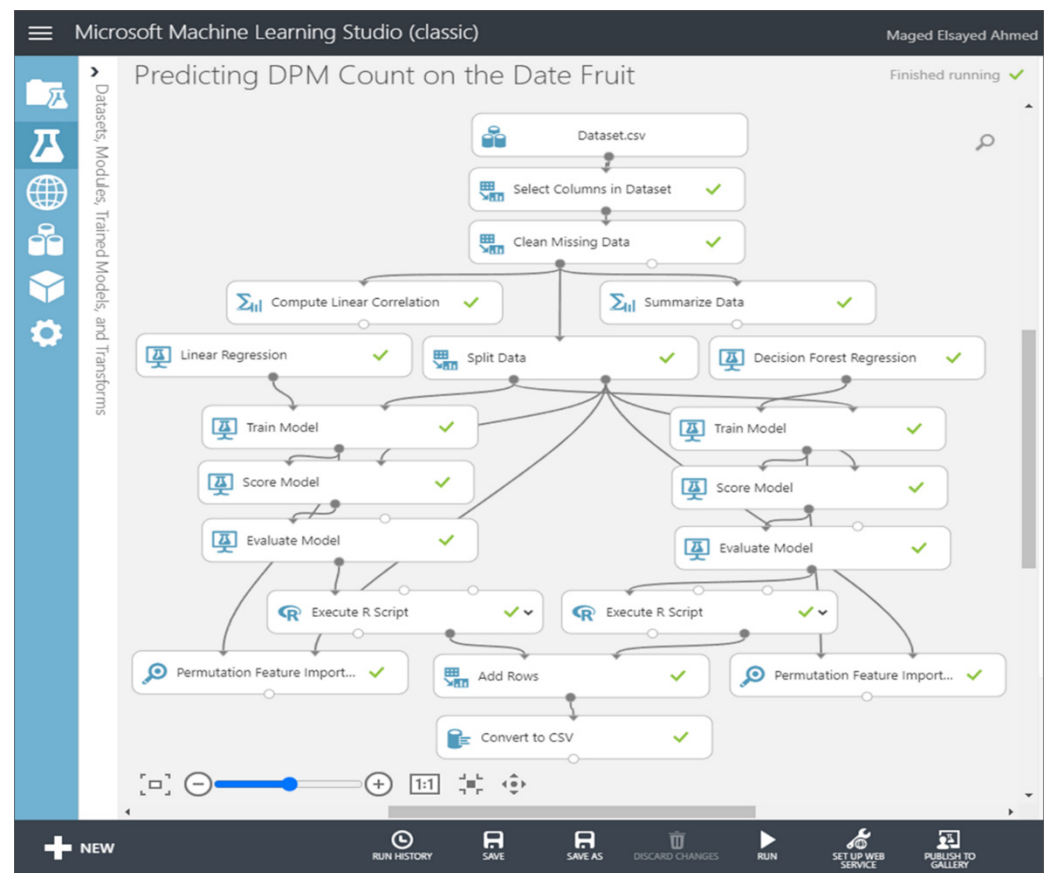


Figure 3. A screenshot for an experiment of the data analysis and building the predictive models for predicting DPM infestation on the date fruit using Microsoft Azure Machine Learning.

2.5.2. Data Analysis

The top section of Figure 3 shows the experiment's first part covering data preparation and analysis. First, the select column in the dataset module was used to determine essential variables for prediction. Next, the data was loaded and cleaned for missing data using the clean missing data module. Next, the clean missing data module was used in the experiment workspace to use measures of central tendency for missing value imputation. Next, the summary statistics of the data were obtained using summarize data module. Finally, the compute linear correlation module was used to determine the linear correlation between the input features.

2.5.3. Training the Models

After performing the data preprocessing, the next step was to train the LR and DFR models based on the variables input for predicting DPM infestation. Up to the split data module, the top half of the experiment performed the data preprocessing step. The split data module split the data into two categories, a training group comprising 70% and a test category with the remaining 30% of the initial dataset. Each one of the selected algorithms was trained with the same training data and tested with the same test data using the Train Model modules.

2.5.4. Performance of the Models

The Evaluate Model module was used to evaluate the performance accuracy of the trained LR and DFR models. The model evaluation took two datasets as inputs, the first was a testing dataset to score the model, and the second was an optional dataset for comparison. In this study, we used the first input only. After experimenting, we studied the developed model's performance using the evaluation metrics, i.e., negative log likelihood (NLL),

mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), relative squared error (RSE), and coefficient of determination (R^2). The values of these evaluation variables were converted to datasets from Azure ML into a CSV format that can be downloaded and opened using Microsoft Excel, exported, or shared with R modules using three modules, i.e., execute R script, add rows, and convert to CSV. The execute R script module was used to run R code in the Azure ML pipeline to transform the evaluation metrics data of the prediction models into an excel file. The permutation feature importance scores of the variables provided a trained model and a test dataset category using the permutation feature importance modules. The permutation feature importance worked by randomly varying the values of the feature column (one column at a time) and evaluating the model.

2.5.5. Deployment of the Model

After we created and evaluated a predictive model suitable for predicting DPM infestations, we deployed it using Azure ML to make it easier to use as a predictive Azure ML web service on the Azure cloud platform to make it easier to use as a web service. The predictive experiment was automatically obtained after selecting the train model module for the best-developed model. Figure 4 shows the predictive experiment of the web service that was created automatically for the prediction in the Azure ML studio platform web service.

In addition, Azure ML web services were added as Excel add-ins using the obtained URL and the application programming interface (API) key. Figure 5 shows the method used to download the workbook directly from the Azure ML web service.

2.6. Statistical Analysis

The analysis of variance (ANOVA) of the data regarding physicochemical properties of the date fruit recorded at five fruit development stages of the two date palm cvs. and the meteorological variables in the study area were conducted using GenStat data analysis software, Version 11 (VSNi International Ltd., Hemel Hempstead, UK). The least significant difference test (LSD) was used for multiple comparisons of means for each fruit property and meteorological variable.

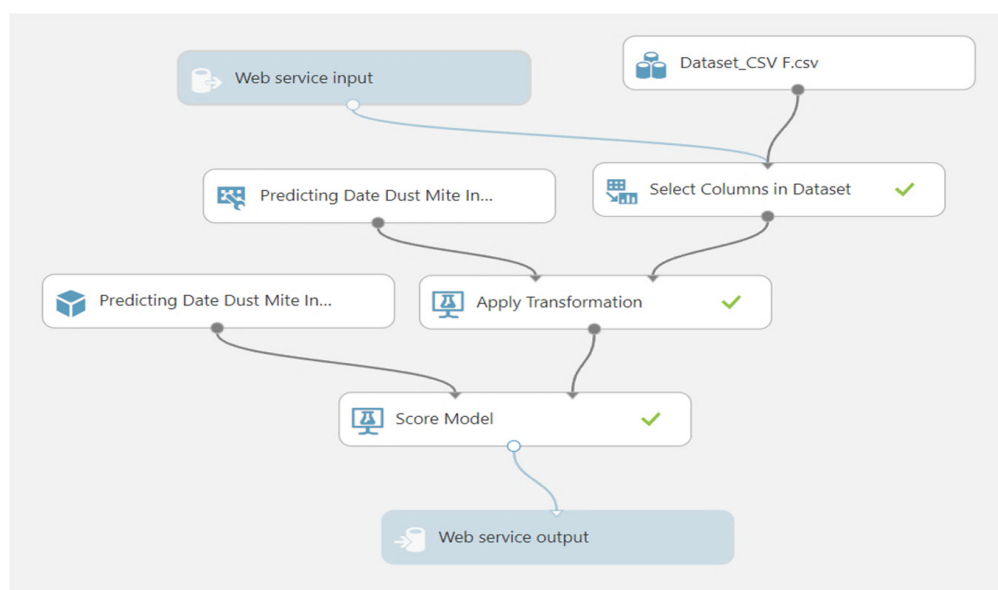


Figure 4. A screenshot for the experiment of the web service that was created for the prediction model in Azure Machine Learning.

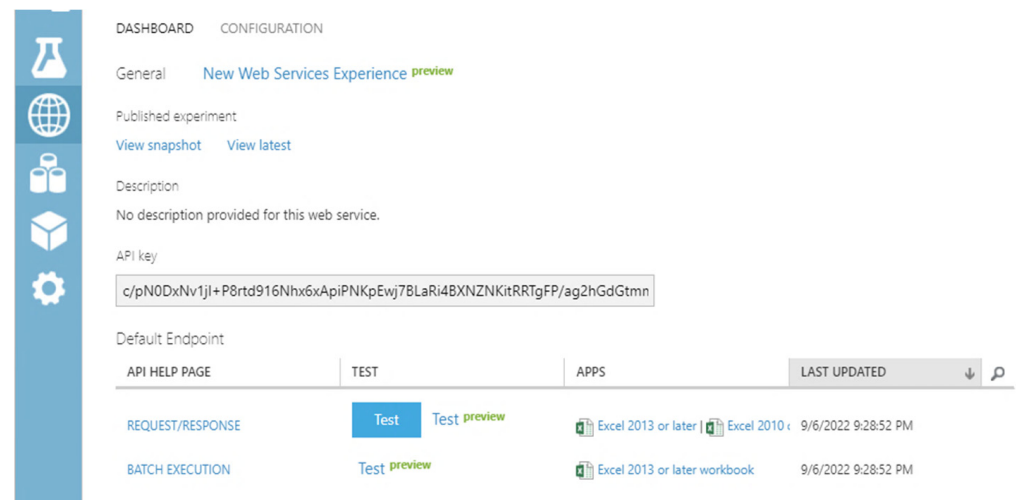


Figure 5. A screenshot of the web service that was created for prediction in Azure Machine Learning.

3. Results

3.1. Meteorological Variables

Table 1 shows the observed monthly mean values $\pm$ standard deviation of the meteorological variables in the study area. The highest temperature was during July, while the lowest was during April. The highest RH was during August, while the lowest RH was during June. The highest wind speed was during May, followed by April, with a slight difference. There was no significant difference in the max. solar radiation and average daily solar radiation intensity between the months in the study period.

Table 1. Meteorological variables in the study area during 2021–22 seasons. The T_{Max} , T_{Min} , T_{Avg} , RH_{Max} , RH_{Min} , RH_{Avg} , WS_{Avg} , SR_{Max} , DSR_{Avg} in the table represent the maximum temperature, minimum temperature, average temperature, maximum relative humidity, minimum relative humidity, average relative humidity, average wind speed, max. solar radiation, and average daily solar radiation, respectively.

Meteorological Variables	Months				
	April	May	June	July	August
T_{Max} (°C)	39.66 ± 8.09^c	46.15 ± 2.76^a	45.79 ± 8.31^b	47.07 ± 1.09^a	46.57 ± 1.53^a
T_{Min} (°C)	23.87 ± 5.57^c	28.06 ± 3.17^b	29.56 ± 5.27^a	30.47 ± 2.22^a	28.08 ± 1.31^b
T_{Avg} (°C)	31.65 ± 6.63^d	37.23 ± 2.58^b	37.77 ± 6.71^b	38.76 ± 1.24^a	36.9 ± 1.17^c
RH_{Max} (%)	47.15 ± 17.36^b	36.09 ± 12.78^c	33.37 ± 15.27^d	38.01 ± 18.23^c	55.43 ± 19.88^a
RH_{Min} (%)	10.68 ± 4.02^a	6.71 ± 1.81^c	6.13 ± 2.55^c	7.1 ± 2.5^b	7.49 ± 2.31^b
RH_{Avg} (%)	34.81 ± 4.59^b	27.06 ± 11.17^c	23.55 ± 13.83^d	28.41 ± 16.29^c	39.7 ± 18.82^a
WS_{Avg} (km/day)	26.06 ± 16.28^b	27.07 ± 19.12^a	18.63 ± 14.98^c	10.38 ± 3.76^d	8.22 ± 3.78^e
SR_{Max} (kW/h)	1.29 ± 0.25^a	1.26 ± 0.07^a	1.18 ± 0.21^a	1.22 ± 0.05^a	1.21 ± 0.05^a
DSR_{Avg} (kW/h)	0.48 ± 0.10^a	0.53 ± 0.01^a	0.5 ± 0.09^a	0.49 ± 0.02^a	0.48 ± 0.01^a

Data are the mean of days of the month, and $\pm$ represents the standard deviation within days. Different letters within each row indicate significant mean differences as compared to the LSD ($p < 0.05$).

3.2. Date Palm Mite Count Data

Figure 6 shows the observed daily average values of the DPM count per fruit versus the day of the year in the experimental area. The average mite population was 0.92 ± 1.05 in April (days 91–120) and increased gradually during May (days 121–151) when it reached 4.53 ± 3.52 . The maximum count of mites of 25.63 ± 3.58 was recorded in July (days 182–212) before it declined to 10.6 ± 4.64 in August (days 213–243). The highest mite count coincided with an average ambient temperature of 38.76 and RH of 28.41 (Table 1).

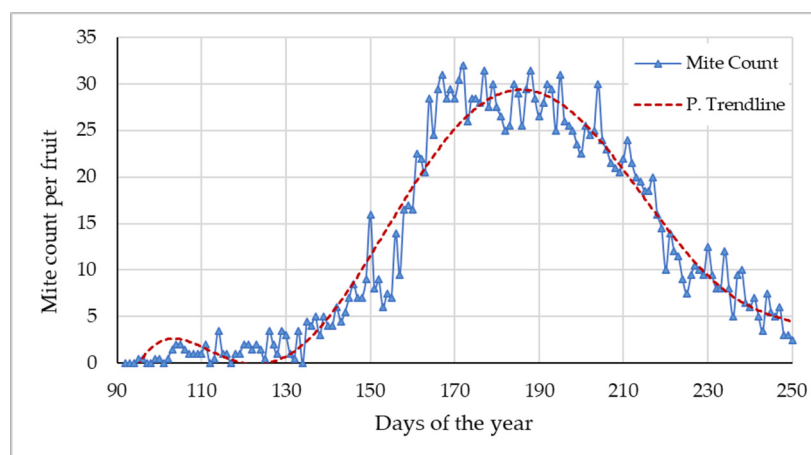


Figure 6. The average count of the date palm mite (mite count) vs. day of the year and the polynomial trendline (P. trendline) from April to the end of August in the study area during the 2021–22 seasons.

3.3. Physicochemical Properties of Date Fruit

Table 2 shows the physicochemical properties of date palm fruit cvs. *Khalas* at different fruit development stages, i.e., kimri (1-April to 21-May), khalal (22-May to 24-June), biser (25-June to 18-July), rutab (19-July to 6-August), and tamr (7-August to 31-August). Furthermore, this table shows the physicochemical properties of date palm fruit cvs. *Barhee* at different fruit development stages, i.e., kimri (1-April to 29-May), khalal (30-May to 27-June), biser (28-June to 30-July), rutab 31 July to 16-August), and tamr (17-August to 31-August). At each of the five fruit development stages, all fruit quality indicators revealed a statistically significant difference ($p < 0.05$). The fruit weight of both cultivars linearly increased from the kimri stage to the biser stage, which was decreased afterward at the rutab and tamr stages. Similarly, both cultivars showed statistically similar but highest fruit pH at the biser, rutab, and tamr stages of fruit development, which was statistically at par at these three stages. The fruits of both cultivars were harder at the kimri stage, and gradually became softer until the tamr stage. The highest fruit moisture content was recorded at the kimri stage of development in both cultivars, which significantly decreased in later stages. Fruit TSS and total sugar showed a similar trend in both cultivars. Both fruit quality attributes were lowest at the kimri stage, which linearly increased up to tamr stage. However, maximum tannin content was determined at the khalal stage of fruit development, which progressively decreased and was lowest at the tamr stage.

3.4. The Correlation between the Variables

Table 3 displays the coefficients of the linear correlation between the meteorological variables data and the mite count per fruit. Table 3 indicates a significant positive correlation between the count of DPM per fruit and maximum, minimum, and average temperature, and average daily solar radiation, which means that the count of DPM per fruit will increase with their increase. On the contrary, it is observed that there are negative significant correlation coefficients between the count of DPM per fruit and maximum, minimum, and average relative humidity, maximum solar radiation, and wind speed, and this means that by increasing their values, the number of DPM per fruit will decrease. There was no significant correlation between the maximum solar radiation and the number of DPM per fruit; therefore, this variable was excluded from the training models used in this study.

Table 4 displays the linear correlation coefficients between the variables of the physicochemical properties of the fruit data and the DPM count per fruit. For example, Table 4 indicates significant positive correlation coefficients between the count of DPM per fruit and fruit weight, pH, TSS, and tannin content, which means that the count of DPM per fruit will increase with the increase in their values. On the contrary, it is observed that there are negative significant correlation coefficients between the count of DPM per fruit and fruit

firmness, moisture content, and total sugar, which means that by increasing their values, the number of DPM per fruit will decrease.

Table 2. Physicochemical properties of date palm cvs. *Khalas* and *Barhee* at different fruit development stages, i.e., kimri, khalal, biser, rutab, and tamr. The FW, FF, FMC, TSS, TS, and TC in the table represent the fruit development stages, fruit weight, fruit firmness, fruit moisture content, total soluble solids, total sugar, and tannin content, respectively.

Cultivars	Fruit Properties	Fruit Development Stages				
		Kimri	Khalal	Biser	Rutab	Tamr
<i>Khalas</i>	FW (g)	8.20 ± 0.14 ^d	12.73 ± 0.53 ^b	14.25 ± 0.48 ^a	12.08 ± 0.35 ^b	10.45 ± 0.24 ^c
	Fruit pH	5.08 ± 0.35 ^b	5.44 ± 0.09 ^b	6.50 ± 0.16 ^a	6.36 ± 0.20 ^a	6.64 ± 0.19 ^a
	FF (kg)	9.23 ± 0.31 ^a	9.28 ± 0.15 ^a	7.83 ± 0.31 ^b	6.43 ± 0.35 ^c	3.47 ± 0.15 ^d
	FMC (%)	79.95 ± 0.49 ^a	70.67 ± 2.01 ^b	61.39 ± 2.75 ^c	46.76 ± 2.21 ^d	16.28 ± 1.01 ^e
	TSS (Brix)	15.27 ± 1.99 ^e	25.49 ± 2.57 ^d	38.38 ± 0.83 ^c	51.49 ± 2.32 ^b	60.58 ± 4.37 ^a
	TS (%)	12.53 ± 2.83 ^e	26.54 ± 2.17 ^d	38.24 ± 1.83 ^c	52.87 ± 2.31 ^b	63.35 ± 1.68 ^a
	TC (%)	4.33 ± 0.75 ^b	6.39 ± 0.22 ^a	1.84 ± 0.08 ^c	1.03 ± 0.07 ^d	0.30 ± 0.03 ^e
<i>Barhee</i>	FW (g)	7.23 ± 0.16 ^e	9.59 ± 0.09 ^b	10.46 ± 0.23 ^a	9.11 ± 0.07 ^c	7.83 ± 0.18 ^d
	Fruit pH	5.04 ± 0.14 ^c	6.34 ± 0.17 ^b	6.82 ± 0.07 ^a	6.66 ± 0.07 ^a	6.73 ± 0.19 ^a
	FF (kg)	9.07 ± 0.38 ^a	8.73 ± 0.55 ^a	7.27 ± 0.21 ^b	6.57 ± 0.15 ^c	3.23 ± 0.12 ^d
	FMC (%)	80.06 ± 0.97 ^a	71.18 ± 4.17 ^b	63.28 ± 2.05 ^c	50.98 ± 0.62 ^d	21.47 ± 1.89 ^e
	TSS (Brix)	18.35 ± 2.81 ^d	29.86 ± 3.77 ^c	51.37 ± 1.74 ^b	54.35 ± 2.66 ^b	59.19 ± 1.05 ^a
	TS (%)	13.69 ± 1.51 ^e	32.17 ± 1.44 ^d	56.15 ± 4.01 ^c	63.46 ± 1.99 ^b	69.81 ± 1.06 ^a
	TC (%)	3.55 ± 0.16 ^b	4.59 ± 0.46 ^a	0.95 ± 0.04 ^c	0.59 ± 0.02 ^{cd}	0.22 ± 0.01 ^d

Data are the mean of three independent replicates, and $\pm$ represents the standard deviation within replicates. The statistical analysis is based on a single-factor completely randomized design. The treatment means were compared by the least significant difference (LSD) test. Different letters within each row indicate significant mean differences as compared to the LSD ($p < 0.05$).

Table 3. The correlation coefficients between the meteorological data and the mite count per fruit.

	T _{Max}	T _{Min}	T _{Avg}	RH _{Max}	RH _{Min}	RH _{Avg}	WS _{Avg}	SR _{Max}	DSR _{Avg}	MC
T _{Max}	1	0.739 **	0.938 **	−0.408 **	−0.678 **	−0.405 **	−0.172 **	−0.045	0.356 **	0.534 **
T _{Min}	0.739 **	1	0.902 **	−0.484 **	−0.379 **	−0.422 **	−0.132 **	−0.049 *	0.161 **	0.608 **
T _{Avg}	0.938 **	0.902 **	1	−0.520 **	−0.586 **	−0.488 **	−0.134 **	−0.019	0.335 **	0.630 **
RH _{Max}	−0.408 **	−0.484 **	−0.520 **	1	0.513 **	0.802 **	−0.134 **	−0.395 **	−0.425 **	−0.34 **
RH _{Min}	−0.678 **	−0.379 **	−0.586 **	0.513 **	1	0.509 **	−0.030	−0.214 **	−0.412 **	−0.314 **
RH _{Avg}	−0.405 **	−0.422 **	−0.488 **	0.802 **	0.509 **	1	−0.137 **	−0.351 **	−0.396 **	−0.301 **
WS _{Avg}	−0.172 **	−0.132 **	−0.134 **	−0.134 **	−0.030	−0.137 **	1	0.311 **	0.288 **	−0.241 **
SR _{Max}	−0.045	−0.049 *	−0.019	−0.395 **	−0.214 **	−0.351 **	0.311 **	1	0.381 **	−0.057
DSR _{Avg}	0.356 **	0.161 **	0.335 **	−0.425 **	−0.412 **	−0.396 **	0.288 **	0.381 **	1	0.228 **
MC	0.534 **	0.608 **	0.630 **	−0.341 **	−0.314 **	−0.301 **	−0.241 **	−0.057	0.228 **	1

The T_{Max}, T_{Min}, T_{Avg}, RH_{Max}, RH_{Min}, RH_{Avg}, WS_{Avg}, SR_{Max}, DSR_{Avg}, and MC in the table represent the maximum temperature, minimum temperature, average temperature, maximum relative humidity, minimum relative humidity, average relative humidity, average wind speed, max. solar radiation, average daily solar radiation, and mite count per fruit, respectively. ** The correlation is significant at the 0.01 level. * The correlation is significant at the 0.05 level.

Table 4. The correlation coefficients between the meteorological data and the date palm mite count per fruit.

	FW	pH	FF	FMC	TSS	TC	TS	MC
FW	1 **	−0.006	0.261 **	0.148 **	−0.087 *	−0.168 **	0.182 **	0.135 **
pH	−0.006	1 **	−0.708 **	−0.679 **	0.885 **	0.898 **	−0.826 **	0.587 **
FF	0.261 **	−0.708 **	1 **	0.979 **	−0.895 **	−0.892 **	0.856 **	−0.449 **
FMC	0.148 **	−0.679 **	0.979 **	1 **	−0.890 **	−0.865 **	0.816 **	−0.451 **
TSS	−0.087 *	0.885 **	−0.895 **	−0.890 **	1 **	0.972 **	−0.925 **	0.652 **
TS	−0.168 **	0.898 **	−0.892 **	−0.865 **	0.972 **	1 **	−0.912 **	0.643 **
TC	0.182 **	−0.826 **	0.856 **	0.816 **	−0.925 **	−0.912 **	1 **	−0.688 **
MC	0.135 **	0.587 **	−0.449 **	−0.451 **	0.652 **	0.643 **	−0.688 **	1 **

The FW, FF, FMC, TSS, TC, TS, and MC in the table represent the fruit development stages, fruit weight, fruit firmness, fruit moisture content, total soluble solids, tannin content, total sugar, and mite count per fruit, respectively. ** The correlation is significant at the 0.01 level. * The correlation is significant at the 0.05 level.

3.5. Evaluation of the Prediction Models

Table 5 shows the performance evaluation results of two predictive models trained on the same three input combinations: (1) Meteorological variables (MV); (2) Physicochemical properties variables (PPV); and (3) The combination of meteorological and physicochemical properties' variables (MPPV) to predict the DPM count on the date fruit. The meteorological variables included: T_{Max} , T_{Min} , T_{Avg} , RH_{Max} , RH_{Min} , RH_{Avg} , WS_{Avg} , and DSR_{Avg} . The physicochemical properties included: FW, FF, FMC, TSS, TC, and TS.

Table 5. Performance of the ML predictive models in predicting date dust mites count per fruit based on the meteorological variable (MV) in the study area (test data: 2021–22 seasons), physicochemical properties variables (PPV), and meteorological and physicochemical properties' variables (MPPV).

Input Variables	Metrics	ML Models	
		LR	DFR
MV	MAE	6.553	3.126
	RMSE	7.993	4.336
	RAE	0.664	0.317
	RSE	0.536	0.158
	R ²	0.464	0.842
PPV	MAE	5.046	2.622
	RMSE	6.273	3.536
	RAE	0.511	0.266
	RSE	0.330	0.105
	R ²	0.670	0.895
MPPV	MAE	5.778	2.009
	RMSE	7.338	3.091
	RAE	0.580	0.202
	RSE	0.446	0.079
	R ²	0.554	0.921

The LR and DFR represent the linear regression and decision forest regression, respectively. The MAE, RMSE, RAE, RSE, and R² represent the metrics of the mean absolute error, root mean squared error, relative absolute error, relative squared error, and coefficient of determination, respectively.

The results of the performance metrics, i.e., MAE, RMSE, RAE, RSE, and R² in Table 5 indicated that the decision forest regression (DFR) is better than the linear regression (LR) for predicting the DPM infestation based on the three input variables of MV, PPV, and MPPV. The performance of the DFR model for the prediction based on the combination of meteorological and physicochemical properties' variables followed by physicochemical property variables is better than the prediction based on the meteorological variables only.

Figure 7 displays the scatter plots of the observed and the predicted count of the DPM on the date palm fruits by the LR (Figure 7A) and DFR (Figure 7B) models in the evaluation phase based on the meteorological variables only. These figures indicated that DFR was the best model, and on the contrary, LR overpredicts when observations are low and underpredicts at the highest end of observations, and there is a lot of variability in the residuals overall, whereas, the DFR model is generally close to the 1:1 line, although there is still a much less pronounced tendency to overpredict at lower abundances and vice versa.

Figure 8 displays the scatter plots of the observed and the predicted count of the DPM on the date palm fruits by the LR (Figure 8A) and DFR (Figure 8B) models in the evaluation phase based on the physicochemical properties' variables. This figure indicated that the DFR model was the best for predicting DPM infestation on fruits. On the contrary, the LR models had a deficient predictive performance based on the physicochemical property variables of the two date palm cultivars, i.e., *Khalas* and *Barhee*.

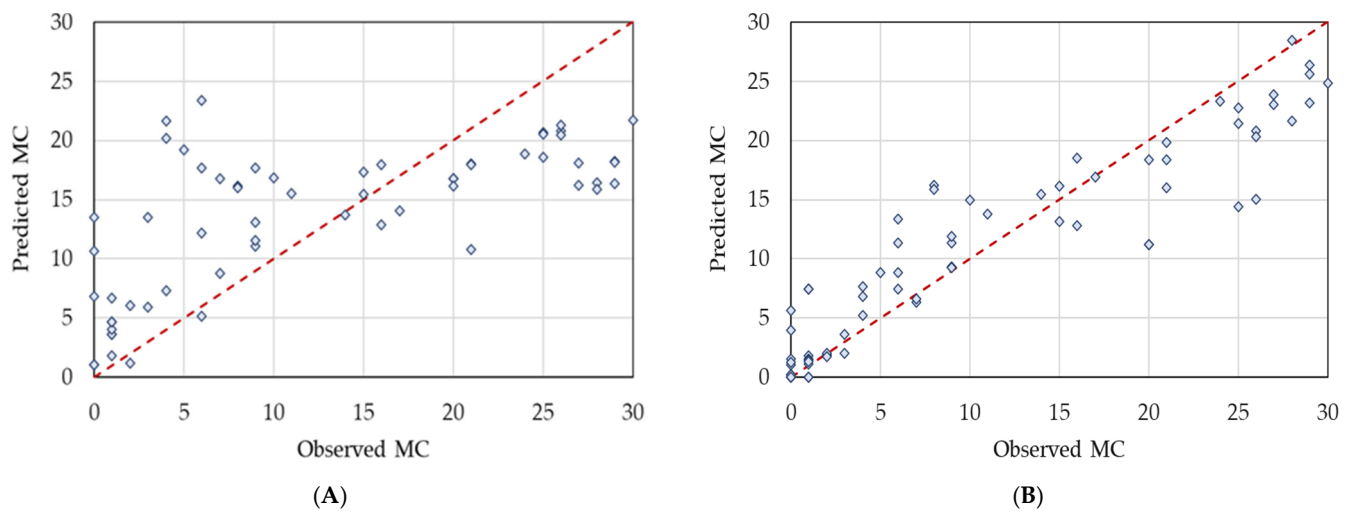


Figure 7. Scatter plots of the observed and the predicted count of the date palm mite (MC) on the date palm fruits using two Azure ML algorithms, i.e., linear regression (A) and decision forest regression (B) based on the meteorological data in the study area.

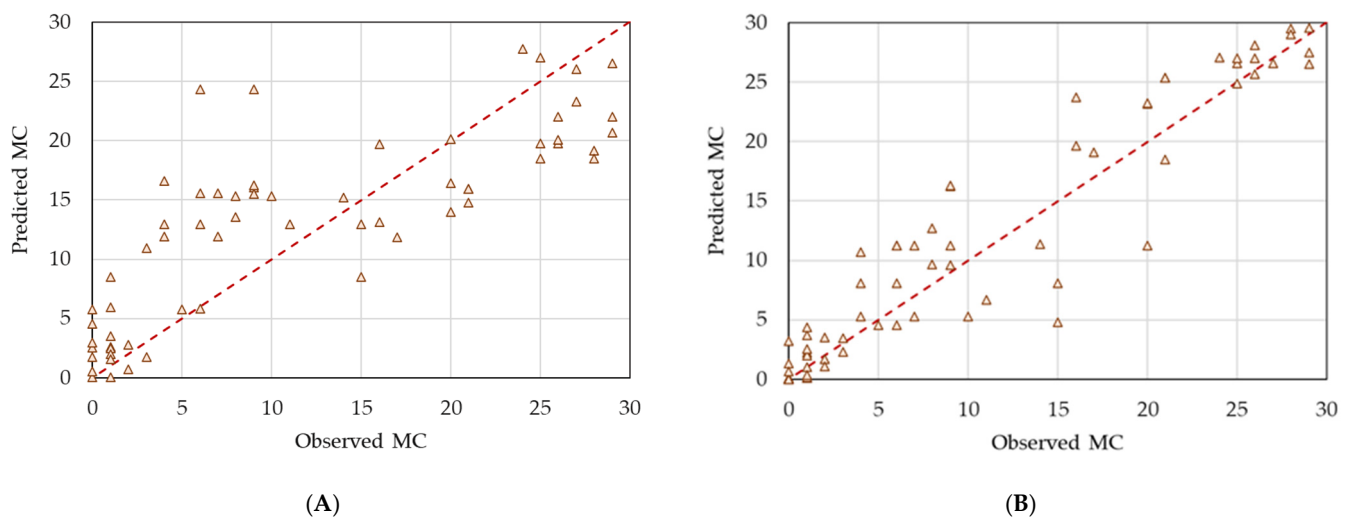


Figure 8. Scatter plots of the observed and the predicted count of the date palm mite (MC) on the date palm fruits using two Azure ML algorithms, i.e., linear regression (A) and decision forest regression (B) based on the physicochemical properties of two date palm cultivars, i.e., *Khalas* and *Barhee*.

Figure 9 displays the scatter plots of the observed and the predicted count of the DPM on the fruits by the LR (Figure 9A) and DFR (Figure 9B) models in the evaluation phase based on the combination of the meteorological and the physicochemical properties' variables. This figure indicated that the DFR was the best model, followed by the DFR model for predicting the DPM infestation on date palm fruit. The regression line between the observed and the predicted values for the DFR model nearly overlapped the 1:1 line ($y = x + 0$) with $R^2 = 0.92$. On the contrary, the LR model had a deficient predictive performance based on the meteorological variables.

Generally, the evaluation results of the developed models based on physicochemical property variables only or based on meteorological variables only showed lower performance than the developed models based on the combination of the meteorological and physicochemical properties' variables.

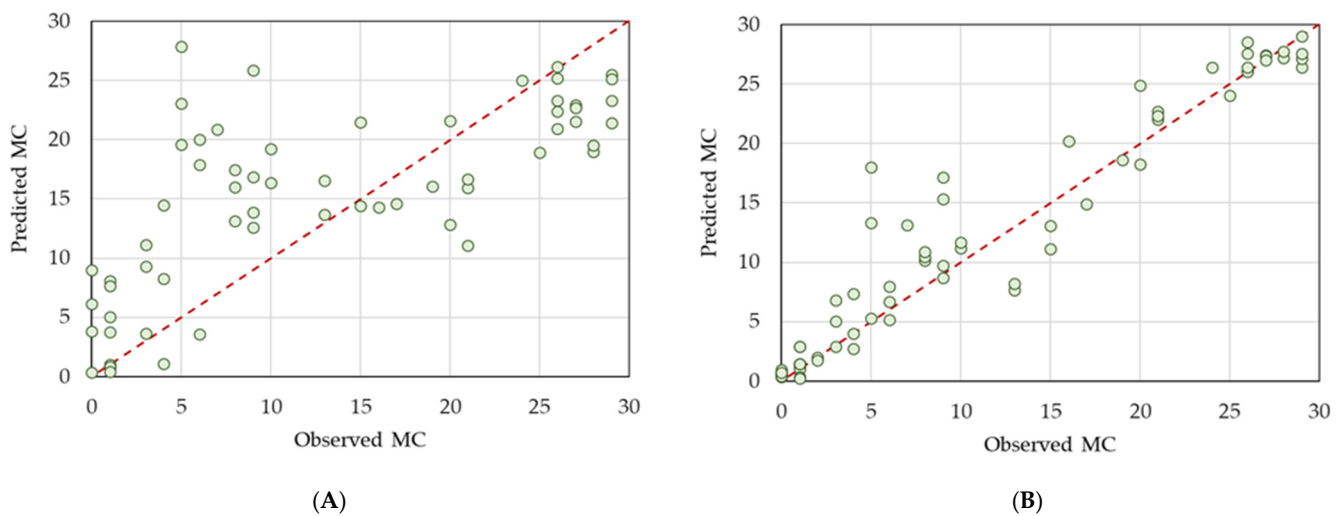


Figure 9. Scatter plots of the observed and the predicted count of the date palm mite (MC) on the date palm fruits using two Azure ML algorithms, i.e., linear regression (A) and decision forest regression (B) based on the combination of the meteorological and the physicochemical properties' variables.

3.6. Variables Importances

3.6.1. Meteorological Variables

Figure 10 shows the input variable's importance scores for the meteorological variables. The scores are used to determine the best input variables in the models. The input variable's importance scores were computed using the permutation feature importance (PFI) module in Azure ML Studio. The computed importance scores, in this case, are only relative to ranking the feature variables of a dataset in the trained model from the most important to the least important in order of permutation importance scores. The idea of the permutation importance algorithm is like the feature randomization process used in the random forests algorithm. The PFI computes importance scores for the input variables by selecting the model's sensitivity to random permutations of the values of the input variables.

Accordingly, the average, maximum, and minimum temperatures were the most important inputs for the LR model (Figure 10A). The average and maximum temperatures and the daily average solar radiation were the most important inputs for the DFR model (Figure 10B).

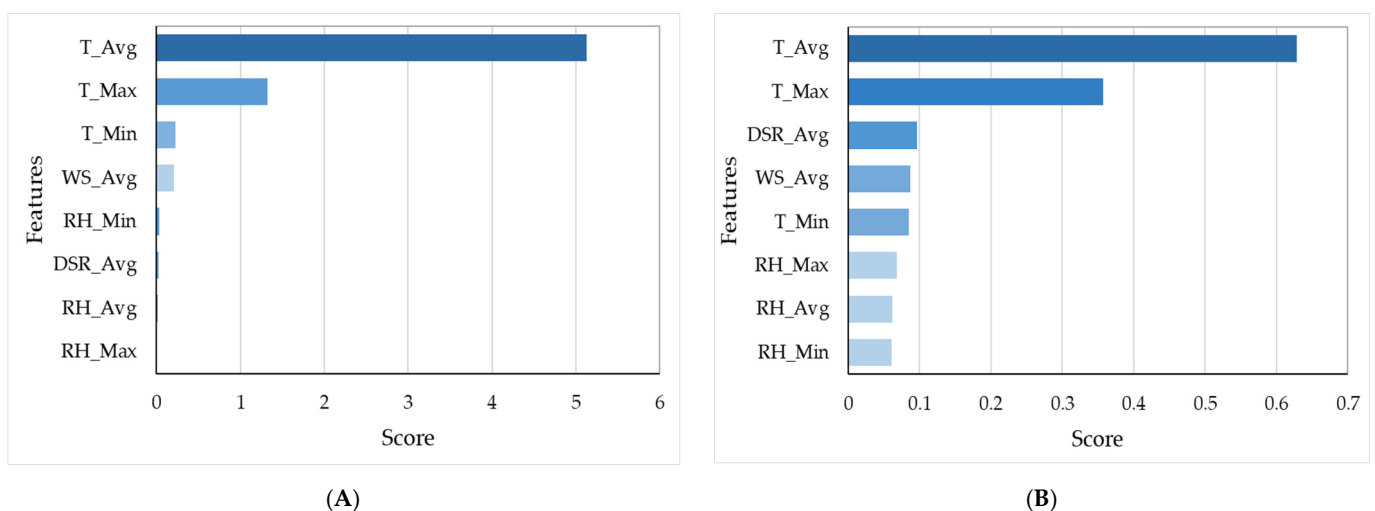


Figure 10. The importance of the meteorological variables input in the developed models of the linear regression (A) and the decision forest regression (B).

3.6.2. Physicochemical Properties

Figure 11 shows the input variable's importance scores for the physicochemical properties' variables. The total sugar, fruit firmness, tannin content, and fruit pH were the most important inputs for the LR model (Figure 11A). The fruit moisture content, tannin content, fruit firmness, and TSS were the most important inputs for the DFR model (Figure 11B).

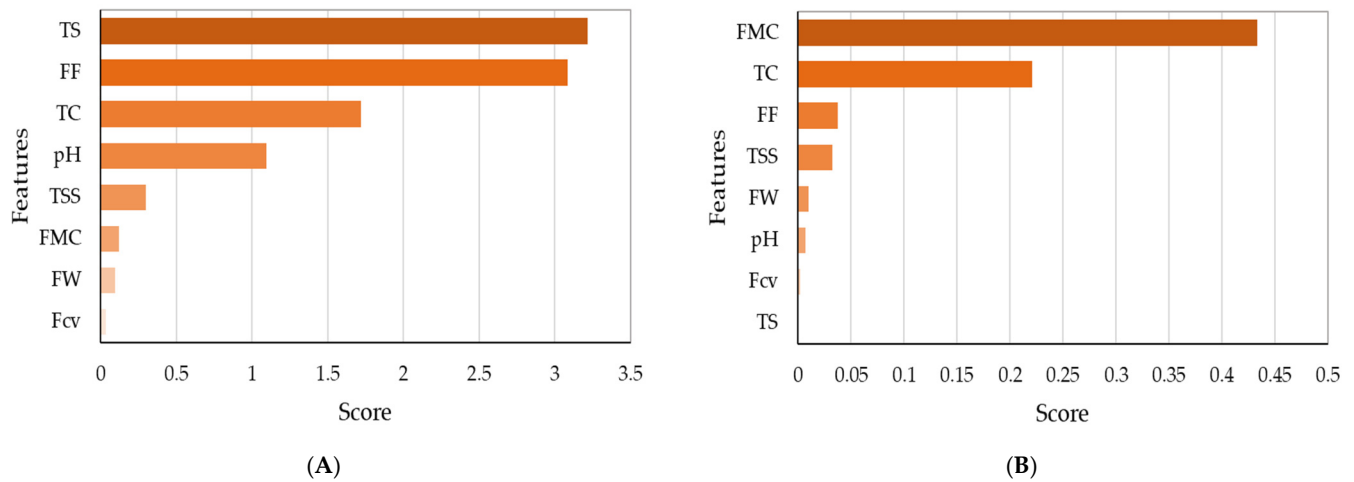


Figure 11. The importance of the physicochemical properties' variables input in the developed models of linear regression (A) and decision forest regression (B).

3.7. Model Deployment

After building and testing the ML predictive models to predict the DPM infestation on fruits, the best model (DFR) was deployed using Microsoft Azure ML studio to facilitate the use of developed models as a web service. The predictive model is a scalable cloud web service, which is readily available over the Internet by any web browser. Once a model is deployed, it can be easily used in the web service tab in the Azure ML Studio platform, as shown in Figure 12. In addition, the web service also displays the APIs to add as Excel add-ins using the obtained URL and the API key.

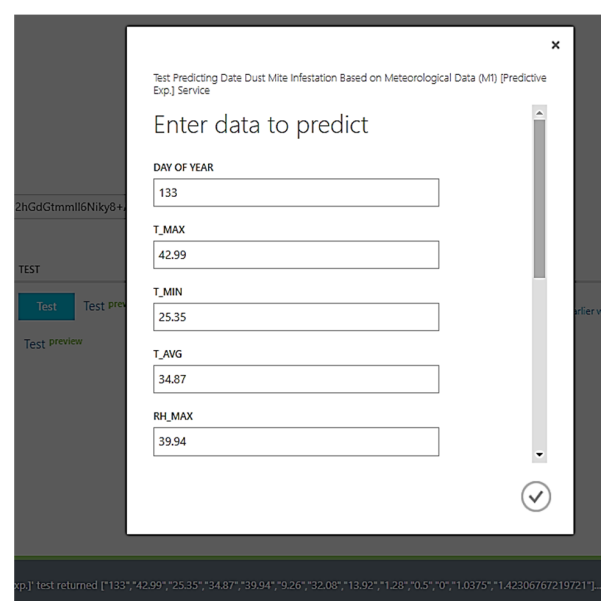


Figure 12. A web service tab in the Azure ML Studio platform with an input field for test predicting date palm mite infestation based on the input variables.

The Excel add-in facilitated the use of web services published by Microsoft Azure ML for evaluating the prediction models. When the Excel workbook is saved, the web services are also saved, so Azure ML web services can share the workbook with other users and enable them to use the deployed web service. The defined functionalities in the user interface were not complex. The user interfaces successfully allowed inputting the meteorological and physicochemical properties' variables with numerical values. After inputting the required variables, the results of the DPM count were predicted efficiently.

3.8. Validation of the DFR Prediction Model

Based on the performance evaluation of the developed models, the DFR model was used to predict the DPM count on date fruit based on the combination of the meteorological and physicochemical variables. This validation is crucial to ensure that the DFR model can accurately predict the DPM count based on the input variables. Figure 13 compares the actual and predicted DPM counts on date palm fruit using the developed DFR prediction model, which is based on variables related to meteorological and physicochemical variables. The predicted count and actual count are similar for the majority of DPM counts throughout the year ($R^2 = 0.918$). For example, Figure 13 shows that the predicted DPM increased exponentially from Day 90 to Day 150. After that, there was not a significant increase until Day 201, after which there was a decline.

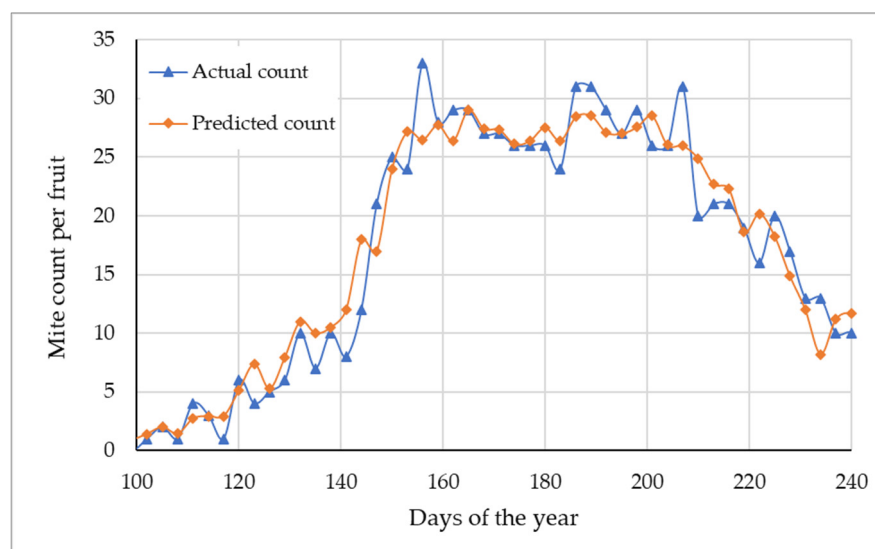


Figure 13. Actual count vs. predicted count of date dust mites on the date palm fruits using the developed decision forest regression (DFR) predictive model based on the combination of the meteorological and the physicochemical variables.

4. Discussion

The prediction of pest populations helps specify pest management strategies to reduce pesticide use and is an integral part of the successful application of integrated pest management (IPM) [35]. The successful management of DPM depends largely on detection, proper sampling, and monitoring, which are essential for decision-making regarding the timing of initiation of control measures. However, sampling DPM based on counting individual mites is difficult and impractical because the mite is so small, and the palm is so large [11]. The population dynamics of DPM in date palm groves are affected by the climate, the chemistry of date fruits, on-farm cultural practices, predators, and the chemical treatment for mites and other date palm pests [3,6,9]. The forecasting and prediction of DPM are largely based on the interaction with the host and the environment [17].

In this study, the average mite population per fruit was 0.92 ± 1.05 in April and increased gradually to reach the maximum count of 25.63 ± 3.58 in July, which corresponds to days 182–212 of the year. This highest mite count coincided with a temperature of 38.76°C

and RH of 28.41%. These results are consistent with the findings of other authors [5,10,11]. These results reflect the impact of meteorological variables on the development of DPM. Perring et al. [36] used meteorological models to predict the percentage of bunch webbing on date fruits caused by the Banks grass mite, *Oligonychus pratensis*. They reported that a bunch infested with a single female mite required 1520-degree days (DD) to reach a web rating of 1 or 7% covering the exterior of the fruit bunch [36]. The overall effect of temperature on the life cycle parameters of the DPM includes the net reproductive rate, intrinsic population growth rate, and shortening of the mite doubling time [9]. The findings of the present study show that the meteorological variables recorded using the cloud-based IoT platform coincide with previous studies carried out in the present study area [28,37–39]. Several previous studies confirmed that for pest prediction models, meteorological variables, i.e., temperature, RH, and sunshine duration, are often used as abiotic predictors in developing prediction models [24,40–43]. The finding in this study indicated that the average temperature was the most important input for the two developed models, compared with other meteorological variables.

Furthermore, regarding the physicochemical properties of date palm fruit, the previous studies' results confirmed the present study's findings. For example, Al-Shahib compared the fruit weight and pH of different date palm cultivars and observed that both traits were lowest at the kimri stage compared to the khalal stage [44]. They also reported that moisture content was highest at the kimri stage, whereas total sugars were highest at the tamr stage. At the kimri and khalal stages of fruit development, tannins—astrigent, bitter-tasting polyphenolic compounds—are often more abundant, and degrade in further stages and are minimal at the fully ripe tamr stage [45,46]. At the kimri, khalal, and tamr stages, total soluble solids and total sugars gradually rise [47]. The decrease in moisture content of dates from the kimri stage to the tamr stage is associated with an increase in sugar content [44], reducing sugars rise at the kimri stage, while sucrose concentration rises at the khalal stage. Due to moisture loss, the fruit weight drops at the rutab stage, and sucrose is converted into reducing sugars [45]. The decrease in fruit water content and elevated sugar content retard date palm mite development, and thus, may confer resistance in some date palm cultivars to injury by DPM [14]. The water content of fruits (86%) also encouraged the establishment of the mite. The population of DPM on date palm fruit declined as the fruit moisture content decreased to below 75% and the TSS to 15% [14]. During fruit development, sugars reach their peak at the tamr stage, primarily by reducing sugars [47]. Our results indicate that there are negative significant correlation coefficients (-0.912) between the count of DPM per fruit and fruit total sugar, which means that by increasing total sugar values, the number of DPM per fruit will decrease. In this respect, Palevsky et al. stated that the population of DPM on date fruit declined as total sugars increased [14]. When a date fruit is in the khalal and early rutab phases of ripening, sugar makes up around 60% of its dry weight [48]. This clearly explains the sharp decline in DPM count during these fruit developmental stages.

Models based on ML algorithms have not been used before to predict DPM infestation, especially based on the meteorological and physicochemical variables. However, a practical application of the developed ML model for predicting insect infestation based on meteorological data was found. In previous studies, ML models have been used to predict the insect population density or the number of insects. The four example ML models have been used to predict the appearance of other pests such as the bollworm *Helicoverpa armigera*, using meteorological data with an accuracy of 76.5% based on the air temperature and RH on a daily basis [27]. Sagar et al. [49] proposed and developed a prediction model, based on meteorological data, for *H. armigera*, which describes the count of insects weekly. The air temperature, RH, and sunshine hours were singled out as weather factors that affect the appearance of *H. armigera*. Blum et al. [50] developed an ML model based on satellite surface temperature data to predict the insects' appearance and estimate population dynamics. Among the environmental factors, temperature plays an essential role in the population build-ups of mites, as well as their mortality [51,52]. To identify apples with

codling moth infestations, physicochemical attributes (such as sugar content, firmness, pH, and moisture content) were recorded and processed through ML, explaining that the higher spectral absorbance by the infested apples can be elucidated by a combination of chemical and textural changes of the fruit due to the infestation [53]. In order to develop better models for tart cherries, Xing et al. [54] suggested that TSS and firmness could be complementary parameters for elucidating the distinction between the insect-infested and intact tart cherries. This demonstrates how examining the correlations between internal quality characteristics related to insect infestation, such as firmness and TSS, can improve the models' accuracy. Similarly, Jamshidi et al. [55] studied the feasibility of detecting pomegranate fruits with interior infestations caused by carob moth (*Ectomyelois ceratoniae*) larvae using pattern recognition models.

Tan et al., [35] applied LR and gradient-boosting decision tree models in the dynamic prediction of the striped rice pest population using meteorological variables and time series of related pests to facilitate the designation of a pest-control strategy. They mentioned that the gradient-boosting decision tree model produced more accurate pest predictions than the LR model. The DFR algorithm is one of the most promising techniques in ML and computer vision. The flexibility of the DFR framework further extends to tasks such as manifold learning, density estimation, and semi-supervised learning. The suitable forest framework allows the user to implement and optimize the underlying algorithm only once and then quickly adjust it to individual applications with moderately small changes [23]. Therefore, the results of the current study showed that the prediction of the DPM count on date palm fruits using the DFR model produced more accurate predictions than the LR model. The formal statistical tests of linear regression make assumptions regarding the distribution of the data, which cannot always be satisfied [56]. The efficient prediction of DFR is due to the assembly of the predictions of its decision trees. The performance of this assembly depends on the algorithm employed to train the decision forest. The utilization of the DFR is due to its efficiency during both training and testing. A key advantage of DFR is that the associated inference algorithms can be executed and optimized once. Yet, relatively minor changes to the DFR model, depending on the application, enable the user to solve many diverse tasks. The DFR can be applied to supervised, semi-supervised, and unsupervised tasks. The accuracy of the DFR model in the prediction is due to its ability to improve the learning process by simplifying the objective and reducing the number of iterations to arrive to an optimal solution. The LR model has few parameters and DFR has a lot more. That means that DFR will overfit more easily than LR. The DFR models have repeatedly proven themselves in various evaluations grading accuracy and efficiency, making them a fundamental component in the prediction approaches. Each iteration of the DFR involves adjusting the values of the coefficients, weights, and biases utilized to each input variable to predict the target value and minimize the difference between the predicted and actual target values. Therefore, the concept of a DFR is not to reject events that fail a criterion right away, but instead to check whether other criteria may help to predict these events properly. Therefore, the developed DFR prediction models were more accurate than the LR model in predicting the DPM infestation on the date fruits based on the input of meteorological variables, date fruits physicochemical properties' variables, and the combination of both meteorological and the physicochemical properties' variables.

5. Conclusions

In this study, a novel approach for predicting the date palm mite infestation on date fruit was found based on meteorological and physicochemical properties' variables using two popular algorithms, i.e., linear regression (LR) and decision forest regression (DFR). The best prediction model for the prediction of date palm mite count on the fruits was the DFR. Generally, the performance evaluation results of the developed DFR and LR models based on physicochemical properties or meteorological variables alone showed lower performance than those based on the combination of the meteorological and physicochemical properties' variables. The DFR model was deployed as a predictive web service

of the Azure cloud platform and the Excel add-ins. The deployed DFR model enhanced the prediction of date palm mite infestation on the fruit through a fast and easy approach. However, this study was performed only on two cultivars (*Khalas* and *Barhee*), and the parameters identified could be different for other resistant or early maturing cultivars under different meteorological conditions. Therefore, further study is needed to develop other predictive models that consider other variables not addressed in this study. Future research will focus on more field evaluation and model development, which could also be adapted for predicting other mite species having agricultural and economic importance.

Author Contributions: Conceptualization, M.M. (Maged Mohammed), H.E.-S. and M.M. (Muhammad Munir); methodology, M.M. (Maged Mohammed), and H.E.-S.; software, M.M.; validation, M.M. (Maged Mohammed) and H.E.-S.; formal analysis, M.M. (Maged Mohammed); investigation, M.M. (Maged Mohammed) and H.E.-S.; resources, M.M. (Maged Mohammed) and H.E.-S.; data curation, M.M. (Maged Mohammed) and H.E.-S.; writing—original draft preparation, M.M. (Maged Mohammed), H.E.-S. and M.M. (Muhammad Munir); writing—review and editing, M.M. (Maged Mohammed), H.E.-S. and M.M. (Muhammad Munir); visualization, M.M. (Maged Mohammed); project administration, M.M. (Maged Mohammed); funding acquisition, M.M. (Maged Mohammed) and H.E.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. GRANT1151].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available upon request from corresponding author.

Acknowledgments: The logistic and technical support provided by the Date Palm Research Center of Excellence, King Faisal University, is appreciated and sincerely acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jeppson, L.R.; Keifer, H.H.; Baker, E.W. *Mites Injurious to Economic Plants*; University of California Press: Oakland, CA, USA, 1975; ISBN 9780520335431.
2. Arbabi, M.; Latifian, M.; Askari, M.; Fassihi, M.T.; Damghani, M.R.; Khiaban, N.G.Z.; Rezai, H. Evaluation of Different Treatments in Control of *Oligonychus afasiaticus* in Date Palm Orchards of Iran. *Persian J. Acarol.* **2017**, *6*, 125–135. [\[CrossRef\]](#)
3. El-Shafie, H.A.F. An Upsurge of the Old World Date Mite (*Oligonychus afasiaticus*) in Date Palm Plantations: Possible Causes and Management Options. *Outlooks Pest Manag.* **2019**, *30*, 13–17. [\[CrossRef\]](#)
4. Alatawi, F.J.; Mirza, J.H.; Alsahwan, K.A.; Kamran, M. Field Population Sex Ratio of the Date Palm Mite, *Oligonychus afasiaticus* (McGregor). *Afr. Entomol.* **2019**, *27*, 336–343. [\[CrossRef\]](#)
5. Mirza, J.H.; Kamran, M.; Alatawi, F.J. Phenology and Abundance of Date Palm Mite *Oligonychus afasiaticus* (McGregor) (Acari: Tetranychidae) in Riyadh, Saudi Arabia. *Saudi J. Biol. Sci.* **2021**, *28*, 4348–4357. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Latifian, M.; Assari, M.J.; Modarresi-Najafabadi, S.S.; Amani, M.; Basavand, F.; Fasihi, M.T.; Zohdi, H.; Bagheri, A. Economic Injury Level of Date Spider Mite, *Oligonychus afasiaticus* (Acari: Tetranychidae) on Six Commercial Date Cultivars. *Persian J. Acarol.* **2021**, *10*, 451–466. [\[CrossRef\]](#)
7. El-Shafie, H.A.F.; Abdel-Banat, B.M.A.; Mohammed, M.E.A.; Al-Hajhoj, M.R. Monitoring Tools and Sampling Methods for Major Date Palm Pests. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* **2019**, *14*, 1–11. [\[CrossRef\]](#)
8. Ali-Dinar, H.; Mohammed, M.; Munir, M. Effects of Pollination Interventions, Plant Age and Source on Hormonal Patterns and Fruit Set of Date Palm (*Phoenix dactylifera* L.). *Horticulturae* **2021**, *7*, 427. [\[CrossRef\]](#)
9. Ben Chaaban, S.; Chermiti, B.; Kreiter, S. Effects of Host Plants on Distribution, Abundance, Developmental Time and Life Table Parameters of *Oligonychus afasiaticus* (McGregor) (Acari: Tetranychidae). *Pap. Avulsos Zool.* **2012**, *52*, 121–132. [\[CrossRef\]](#)
10. Ben Chaabane, S.; Chermiti, B. Characteristics of Date Fruit and Its Influence on Population Dynamics of *Oligonychus afasiaticus* McGregor (Acari: Tetranychidae) in the Southern Tunisia. *Acarologia* **2009**, *49*, 29–37.
11. Negm, M.W.; De Moraes, G.J.; Perring, T.M. *Mite Pests of Date Palms*; Wakil, W., Romeno Faleiro, J., Miller, T.A., Eds.; Springer International Publishing Switzerland: Cham, Switzerland, 2015; ISBN 978-3-319-24395-5.
12. Ben Chaaban, S.; Chermiti, B.; Kreiter, S. *Oligonychus afasiaticus* and Phytoseiid Predators' Seasonal Occurrence on Date Palm Phoenix Dactylifera (Deglet Noor Cultivar) in Tunisian Oases. *Bull. Insectology* **2011**, *64*, 15–21.
13. Yousof, D.E.; Mahmoud, M.E.E. Distribution of Date Palm Dust Mite *Oligonychus afasiaticus* Meg., (Acari: Tetranychidae) in Northern State in Sudan and Its Impact on Productivity of Fruits of Date. *Persian Gulf Crop Prot.* **2013**, *2*, 54–59.

14. Palevsky, E.; Borochoy-Neori, H.; Gerson, U. Population Dynamics of *Oligonychus afrasiaticus* in the Southern Arava Valley of Israel in Relation to Date Fruit Characteristics and Climatic Conditions. *Agric. For. Entomol.* **2005**, *7*, 283–290. [\[CrossRef\]](#)
15. Latifian, M.; Rahnema, A.A.; Amani, M. The Effects of Cultural Management on the Date Spider Mite (*Oligonychus afrasiaticus* McG) Infestation. *Int. J. Farming Allied Sci.* **2014**, *3*, 1009–1014.
16. El-Shafie, H.A.F. The Old World Date Palm Mite *Oligonychus afrasiaticus* (McGregor 1939) (Acari: Tetranychidae), a Major Fruit Pest: Biology, Ecology, and Management. *CABI Rev.* **2022**, *20*. [\[CrossRef\]](#)
17. Latifian, M. Date Palm Spider Mite (*Oligonychus afrasiaticus* McGregor) Forecasting and Monitoring System. *WALIA J.* **2014**, *30*, 79–85.
18. Mohammed, M.; Munir, M.; Aljabr, A. Prediction of Date Fruit Quality Attributes during Cold Storage Based on Their Electrical Properties Using Artificial Neural Networks Models. *Foods* **2022**, *11*, 1666. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Kashyap, P.K.; Kumar, S.; Jaiswal, A.; Prasad, M.; Gandomi, A.H. Towards Precision Agriculture: IoT-Enabled Intelligent Irrigation Systems Using Deep Learning Neural Network. *IEEE Sens. J.* **2021**, *XX*, 1–11. [\[CrossRef\]](#)
20. Mohammed, M.; El-Shafie, H.; Alqahtani, N. Design and Validation of Computerized Flight-Testing Systems with Controlled Atmosphere for Studying Flight Behavior of Red Palm Weevil, *Rhynchophorus ferrugineus* (Olivier). *Sensors* **2021**, *21*, 2112. [\[CrossRef\]](#)
21. Barga, R.; Fontana, V.; Tok, W.H. *Predictive Analytics with Microsoft Azure Machine Learning*; Apress: Berkeley, CA, USA, 2015; ISBN 978-1-4842-1201-1.
22. Tonnang, H.E.Z.; Nedorezov, L.V.; Owino, J.O.; Ochanda, H.; Löhr, B. Host-Parasitoid Population Density Prediction Using Artificial Neural Networks: Diamondback Moth and Its Natural Enemies. *Agric. For. Entomol.* **2010**, *12*, 233–242. [\[CrossRef\]](#)
23. Criminisi, A.; Shotton, J.; Konukoglu, E. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. In *Foundations and Trends in Computer Graphics and Vision*; NOW Publishers: Norwell, MA, USA, 2011; Volume 7, pp. 81–227.
24. Skawsang, Nagai; Tripathi; Soni Predicting Rice Pest Population Occurrence with Satellite-Derived Crop Phenology, Ground Meteorological Observation, and Machine Learning: A Case Study for the Central Plain of Thailand. *Appl. Sci.* **2019**, *9*, 4846. [\[CrossRef\]](#)
25. Ramazi, P.; Kunegel-Lion, M.; Greiner, R.; Lewis, M.A. Predicting Insect Outbreaks Using Machine Learning: A Mountain Pine Beetle Case Study. *Ecol. Evol.* **2021**, *11*, 13014–13028. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Garcia Furuya, D.E.; Ma, L.; Faita Pinheiro, M.M.; Georges Gomes, F.D.; Gonçalves, W.N.; Junior, J.M.; de Castro Rodrigues, D.; Blassioli-Moraes, M.C.; Furtado Michereff, M.F.; Borges, M.; et al. Prediction of Insect-Herbivory-Damage and Insect-Type Attack in Maize Plants Using Hyperspectral Data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102608. [\[CrossRef\]](#)
27. Marković, D.; Vujičić, D.; Tanasković, S.; Đorđević, B.; Randić, S.; Stamenković, Z. Prediction of Pest Insect Appearance Using Sensors and Machine Learning. *Sensors* **2021**, *21*, 4846. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Ahmed Mohammed, M.E.; Refdan Alhajhoj, M.; Ali-Dinar, H.M.; Munir, M. Impact of a Novel Water-Saving Subsurface Irrigation System on Water Productivity, Photosynthetic Characteristics, Yield, and Fruit Quality of Date Palm under Arid Conditions. *Agronomy* **2020**, *10*, 1265. [\[CrossRef\]](#)
29. Taherdoost, H. Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *SSRN Electron. J.* **2016**, *5*, 18–27. [\[CrossRef\]](#)
30. Macmillan, C.D.; Costello, M.J. Evaluation of a Brushing Machine for Estimating Density of Spider Mites on Grape Leaves. *Exp. Appl. Acarol.* **2015**, *67*, 583–594. [\[CrossRef\]](#)
31. AOAC Association of Official Analytical Chemists. *Association of Official Analytical Chemists Gaithersburg (Maryland): AOAC International*, 19th ed.; AOAC: Washington, DC, USA, 2012.
32. Ahmad, A.; Naqvi, S.A.; Jaskani, M.J.; Waseem, M.; Ali, E.; Khan, I.A.; Faisal Manzoor, M.; Siddeeg, A.; Aadil, R.M. Efficient Utilization of Date Palm Waste for the Bioethanol Production through *Saccharomyces Cerevisiae* Strain. *Food Sci. Nutr.* **2021**, *9*, 2066–2074. [\[CrossRef\]](#)
33. Mohammed, M.; Sallam, A.; Alqahtani, N.; Munir, M. The Combined Effects of Precision-Controlled Temperature and Relative Humidity on Artificial Ripening and Quality of Date Fruit. *Foods* **2021**, *10*, 2636. [\[CrossRef\]](#)
34. Taira, S. Astringency in Persimmon. In *Modern Methods of Plant Analysis*; Linskens, H.F., Paech, K., Sanwal, B.D., Tracey, M.V., Eds.; Springer: Berlin/Heidelberg, Germany, 1996; pp. 97–110.
35. Tan, S.; Liang, Y.; Zheng, R.; Yuan, H.; Zhang, Z.; Long, C. Dynamic Prediction of Chilo Suppressalis Occurrence in Rice Based on Deep Learning. *Processes* **2021**, *9*, 2166. [\[CrossRef\]](#)
36. Perring, T.M.; Holtzer, T.O.; Kalisch, J.A.; Norman, J.M. Temperature and Humidity Effects on Ovipositional Rates, Fecundity, and Longevity of Adult Female Banks Grass Mites (Acari: Tetranychidae). *Ann. Entomol. Soc. Am.* **1984**, *77*, 581–586. [\[CrossRef\]](#)
37. Mohammed, M.E.A.; El-Shafie, H.A.; Sallam, A.A.A. A Solar-Powered Heat System for Management of Almond Moth, *Cadra cautella* (Lepidoptera: Pyralidae) in Stored Dates. *Postharvest Biol. Technol.* **2019**, *154*, 121–128. [\[CrossRef\]](#)
38. Sagheer, A.; Mohammed, M.; Riad, K.; Alhajhoj, M. A Cloud-Based IoT Platform for Precision Control of Soilless Greenhouse Cultivation. *Sensors* **2021**, *21*, 223. [\[CrossRef\]](#)
39. Mohammed, M.; Riad, K.; Alqahtani, N. Efficient Iot-Based Control for a Smart Subsurface Irrigation System to Enhance Irrigation Management of Date Palm. *Sensors* **2021**, *21*, 3942. [\[CrossRef\]](#) [\[PubMed\]](#)

40. de Oliveira Aparecido, L.E.; de Souza Rolim, G.; da Silva Cabral De Moraes, J.R.; Costa, C.T.S.; de Souza, P.S. Machine Learning Algorithms for Forecasting the Incidence of Coffea Arabica Pests and Diseases. *Int. J. Biometeorol.* **2020**, *64*, 671–688. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Holloway, P.; Kudenko, D.; Bell, J.R. Dynamic Selection of Environmental Variables to Improve the Prediction of Aphid Phenology: A Machine Learning Approach. *Ecol. Indic.* **2018**, *88*, 512–521. [\[CrossRef\]](#)
42. Poggi, S.; Le Cointe, R.; Riou, J.B.; Larroudé, P.; Thibord, J.B.; Plantegenest, M. Relative Influence of Climate and Agroenvironmental Factors on Wireworm Damage Risk in Maize Crops. *J. Pest Sci.* **2018**, *91*, 585–599. [\[CrossRef\]](#)
43. Gu, Y.H.; Yoo, S.J.; Park, C.J.; Kim, Y.H.; Park, S.K.; Kim, J.S.; Lim, J.H. BLITE-SVR: New Forecasting Model for Late Blight on Potato Using Support-Vector Regression. *Comput. Electron. Agric.* **2016**, *130*, 169–176. [\[CrossRef\]](#)
44. Al-Shahib, W.; Marshall, R.J. The Fruit of the Date Palm: Its Possible Use as the Best Food for the Future? *Int. J. Food Sci. Nutr.* **2003**, *54*, 247–259. [\[CrossRef\]](#)
45. Tafti, A.G.; Fooladi, M.H. Changes in Physical and Chemical Characteristic of Mozafati Date Fruit During Development. *J. Biol. Sci.* **2005**, *5*, 319–322. [\[CrossRef\]](#)
46. Nadeem, M.; Anjum, F.M.; Zahoor, T.; Saeed, F.; Ahmad, A. Anti-Nutritional Factors in Some Date Palm (*Phoenix dactylifera* L.) Varieties Grown in Pakistan. *Internet J. Food Saf.* **2011**, *13*, 386–390.
47. Bacha, M.A.; Shaheen, M.A.; Nasr, T.A. Changes in Physical and Chemical Characteristics of the Fruits of Four Date Palm Cultivars. *Saudi Biol.* **1987**, *10*, 285–294.
48. Samarawira, I. Date Palm, Potential Source for Refined Sugar. *Econ. Bot.* **1983**, *37*, 181–186. [\[CrossRef\]](#)
49. Sagar, D.; Nebapure, S.M.; Chander, S. Development and Validation of Weather Based Prediction Model for Helicoverpa Armigera in Chickpea. *J. Agrometeorol.* **2017**, *19*, 328–333. [\[CrossRef\]](#)
50. Blum, M.; Nestel, D.; Cohen, Y.; Goldshtein, E.; Helman, D.; Lensky, I.M. Predicting Heliothis (Helicoverpa Armigera) Pest Population Dynamics with an Age-Structured Insect Population Model Driven by Satellite Data. *Ecol. Modell.* **2018**, *369*, 1–12. [\[CrossRef\]](#)
51. Jarošík, V.; Honěk, A.; Magarey, R.D.; Skuhrovec, J. Developmental Database for Phenology Models: Related Insect and Mite Species Have Similar Thermal Requirements. *J. Econ. Entomol.* **2011**, *104*, 1870–1876. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Alatawi, F.J. Field Studies on Occurrence, Alternate Hosts and Mortality Factors of Date Palm Mite, *Oligonychus afrasiaticus* (McGregor) (Acari: Tetranychidae). *J. Saudi Soc. Agric. Sci.* **2020**, *19*, 146–150. [\[CrossRef\]](#)
53. Ekramirad, N.; Khaled, A.Y.; Doyle, L.E.; Loeb, J.R.; Donohue, K.D.; Villanueva, R.T.; Adedeji, A.A. Nondestructive Detection of Codling Moth Infestation in Apples Using Pixel-Based Nir Hyperspectral Imaging with Machine Learning and Feature Selection. *Foods* **2022**, *11*, 8. [\[CrossRef\]](#)
54. Xing, J.; Guyer, D.; Ariana, D.; Lu, R. Determining Optimal Wavebands Using Genetic Algorithm for Detection of Internal Insect Infestation in Tart Cherry. *Sens. Instrum. Food Qual. Saf.* **2008**, *2*, 161–167. [\[CrossRef\]](#)
55. Jamshidi, B.; Mohajerani, E.; Farazmand, H.; Mahmoudi, A.; Hemmati, A. Pattern Recognition-Based Optical Technique for Non-Destructive Detection of Ectomyelois Ceratoniae Infestation in Pomegranates during Hidden Activity of the Larvae. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *206*, 552–557. [\[CrossRef\]](#)
56. Goldstein, R. Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures Models. *Technometrics* **2006**, *48*, 149–150. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.