

Prediction of Agro Products Sales Using Regression Algorithm

Terungwa Simon Yange^{1,*}, Charity Ojochogwu Egbunu¹, Oluoha Onyekwere², Kater Amos Foga²

¹Department of Mathematics/Statistics/Computer Science, University of Agriculture, Makurdi, Nigeria

²Department of Computer Science, University of Nigeria, Nsukka, Nigeria

Email address:

lordesty2k7@gmail.com (S. T. Yange)

*Corresponding author

To cite this article:

Terungwa Simon Yange, Charity Ojochogwu Egbunu, Oluoha Onyekwere, Kater Amos Foga. Prediction of Agro Products Sales Using Regression Algorithm. *American Journal of Data Mining and Knowledge Discovery*. Vol. 5, No. 1, 2020, pp. 11-19.

doi: 10.11648/j.ajdmkd.20200501.12

Received: June 3, 2020; Accepted: June 17, 2020; Published: July 6, 2020

Abstract: This study aimed at developing a system using support vector machine (SVM) that will forecast sales of farm products for an agricultural farm so that managers can take strategic decisions timely to better market the excess farm products which some by nature are perishable. The sales prediction model used SVMs and Fuzzy Theory. The implementation was done using Python Programming Language. The system comprised of three (3) modules: web interface, flask and the SVM Framework. To evaluate the result of the SVM model, the RBF neural network was used as a benchmark. Data of previous sales records from University of Agriculture Makurdi (UAM) farm was used to train and test the system. After training the network with data which covered the time period from 21st January, 2017 to 30th June, 2019, the remaining data which covered from 1st July 2019 up to the 31st December, 2019 was used to test and validate the forecasting performance of the system. The Forecasting Precision (FP) value for the SVM was 96.75% and that of the RBF neural network forecasting value was 90.55%. Analysis from the results shows that the forecasting system with SVM had a greater precision in the sales of agricultural products.

Keywords: Prediction, Regression, Algorithm, Agricultural Products, Sales, SVM

1. Introduction

Nigeria is one of the leading countries that deals with agriculture [1]. Nigeria is known as the world's largest producer of yam (23million tonnes), cassava (54 million tonnes), plantain and several other agricultural products ranging from plants, vegetables, fruits and also poultry management involving rearing of goats, chickens, pigs, fish farming, eggs production and lots of others [2]. Agricultural products require lots of strategy in its management due to its quick nature of spoilage and difficulty in storage. Companies or organizations that must succeed in this business are left with no choice but to put different measures in check to reduce the loses as a result of spoilage. When there is excess of the produce and the production is not commensurable with sales i.e. low sales turn over and high supply, the tendency to suffer loss becomes inexpiable as the perishable nature of

farm produce remains a problem that requires considerate tactics to deal with. Post-harvest losses in Africa ranges from 4 to 80% [3] and Nigeria agriculture suffers losses as much as 20% from grains, 50% from agriculture and 30% from roots and tubers [1]. In 2017, the executive director of Nigerian Stored Products Research Institute (NSPRI) says Nigeria records N2. 7trillion deficit annually from post-harvest losses [2]. Losses from Postharvest apparently results in wastage of resources like fertilizer, water, land, seeds etc. as these resources could have been used for other purposes. Technology that is aimed at reducing post-harvest losses among farmers will add great value to the economy of a nation and improve efficiency in business. There is also a key need in focusing on reduction of food loss and waste at the early stage of supply chain; this will include efforts employed at the farm level where the impact seems the strongest. Forecasting is the process of making prediction of the future based on past and present data and analysis of

trends. Sales forecasting is the ability to forecast future sales base on the records from past months' sales. Ability to predict future market demands of agricultural products is very important in determining the decisions to be taken before farm production and in the development of marketing strategies. Business operations with perishable agricultural products will profit immensely from accurate forecasting of sales. losses [4].

This study aims at developing an algorithm based on support vector machine (SVM) that will forecast sales of agricultural farm produce for a farm that deals with perishable agricultural products. The model developed used data from previous sales from the farm to train and to test the model considering some factors that influences sales. This will guarantee generalization performance for the SVM imbedded with linear regression to effectively forecast sales of perishable farm products. The rest of the article is arranged in the following order. Section 2 gives the Literature Review while Section 3 presents the basic theory and model of SVM. Then the forecasting system framework based on the SVM and Linear Regression is explored in Section 4. In Section 5, the proposed model is presented, and numerical examples are used to investigate the forecasting performance of the model. As a conclusion, unique contributions of this article, limitations of the research and some future research directions are given in Section 6.

2. Review Related Works

Lots of researches have been carried out on agricultural forecasting systems. While some are interested in the forecasting of crop yield and hence developing systems that can forewarn possible outbreak of diseases in crops in order to take precautionary and timely steps to avoid them [5], some governments have shown interest in developing estimates and forecasts on their farm economy for adequate planning; they put into consideration the total value they can derive from farm production of goods and services, the net farm income, the net cash income, and also the cash earned from sales of farm products. Some researchers have also shown interest in price forecasting of agricultural products for the benefit of agribusiness industries, farmers and policy makers in proper management of food security [6][7]. Xu *et al.* [8] while trying to bridge the gap between demand and supply of agricultural products observed logistics as a factor valuable in the appropriate estimation of community demand and supply. Their study assessed the difficulty on demand forecasting of agricultural products logistics in a community in Beigin and reference data from agricultural outlets within the community were collected and a model was designed to satisfy consumers' demand and optimize profits of operators as well. Lawrence & Godwin [9] hinged their research on the change experienced in prices of fresh agricultural products like cabbage and celery from time to time. The fluctuation been attributed to seasons, supply and demand imbalance and of course agricultural market information on the fresh available products. Devising a way that the farmers' products

can be rightly aligned and channeled to demand customers so as to avoid over production and subsequently spoilage and be purchased at reasonable prices that farmers would be pleased with is very vital. Some agricultural products like dairy products have volatile demand pattern and are often influenced by rapid environmental changes; Frisvold & Murugesan [10] recommended an accurate forecast to predict market demands for such perishable dairy products in the milk processing industry and findings showed that high profits, good inventory control as well as high order fill rate were achieved. While it is important to look critically at the level of demands of certain agricultural products in certain areas, demands sometimes depends on the purchasing power and as a matter of fact, the purchasing power have a role to play in the amount or quantity of products to acquire; hence, the need to consider the sales factor and be able to forecast it considering different factors that can have an influence on it is very necessary and important so planning can be effective and losses greatly minimized especially for perishable items.

2.1. Factors That Influences Sales of Agricultural Products

Temperature (high, medium, low): There is a relationship between weather and business activities and this has guided human behavior in different fields [11]. the level of temperature on some farm produce sure affects sales. On hot weathers for instance, eggs tend to spoil quickly and so forecasting the amount of sales within such periods will influence farmers' or agribusiness owners' decision on the right quantity to avail for specific location based on the predicted sales rate and proper preservation techniques to apply. Also, Poultry managers and those into livestock business deals with losses especially for broilers during very hot season and extremely cold seasons as well. Knowledge of predicted sales will give an idea on the amount to keep for sales within such periods.

Break/Holiday: Sales are higher during weekends and special breaks for festivities. Most people who are busy during the weekdays go to the markets to buy food stuff during weekend. Also during special holidays, more sales are made compare to other days. Targeting periods for higher sales especially during festive seasons where excessive cooking of food stuffs like rice and Poultry is on the increase, sales would definitely be more than the usual.

Season (Winter, Spring, Summer): the availability of crops is more at certain seasons. There is a season for planting of crops; while for some, it is shortly before the rainy season; others it could be during the raining season (rice, maize etc.). Irrespective of the season for planting, the sales of most crops is at its peak around the period of harvest and as a matter of fact, it appears to be cheaper that period. People use such opportunity to make purchase in bulk and preserve them against the period of scarcity. While traders utilize this opportunity to make more profit later on. Agricultural products are usually more expensive in the period of scarcity most times just before the planting season of such crops. This factor sure has an effect on the level of sales of different farm products.

Weather (Sunny, Rainy, Windy). Irrigation farming is not common in Nigeria and so most crops are grown seasonally. Apart from the fact that weather plays an important role in agriculture production [12], it stills does in the sales of agricultural products after post-harvest. The prices of products like sorghum, millet, maize etc. is higher in some of the seasons due to their unavailability in those seasons. Producers use weather data for specific production and marketing decisions.

Location: To predict the success of a business, location is a very important factor to consider. According to Suttle [13], it is important that businesses be established in locations that generate the most customer traffic as the impact on the business would be so obvious. Kerani and Wanjohi [14] carried out a study to establish some factors that could influence the marketing of agricultural products among small scale farmers; 'access to information' and 'middle-men to sell these products' were factors that showed significant influences on sales. It is safe to say that location plays an important role in the sale of the products. If business is done in a vicinity where most of the people in the community farm these products, locating a strategic place where business would thrive would be the only way forward if they must make sales; this is the reason, farmers in remote villages would need to transport their goods to customer target regions else there would be waste from spoilage as its presently experienced in the interiors of several parts in Nigeria.

The demand for a product is generated by a complex interaction of many factors. If it were possible to understand the effect of these factors and how they interrelate, the job of sales forecasting would be relatively straight forward. All that would be done is to develop a mathematical model that could give a very accurate estimate of the future demand and the sales executive must consider some kind of sales forecast for effective decision and planning.

2.2. Different Models Used to Forecast Agricultural Systems

Time Series Techniques: This technique utilizes an index of sequential and successive equally spaced points in time. It uses discrete stamped data. Annadanapu and Ravi [15] applied the time series technique in combination with a regressive integrated moving average model to identify an underlining structure and the model was fitted on agriculture food production with R software. Ruekkasaem and Sasananan (2018) also used the time series to predict suitable crop rotation among some crops and rice cultivation planning in the midst of limited resources; the method was found to be more useful than the traditional method as it yielded higher profits.

Exponential Smoothing Model: Abid *et al.* [16] in their study to forecast area to grow and produce potato in Pakistan, using the best fitted model employed different forecasting models such as Linear trend model, Quadratic trend model, Exponential growth model, double exponential and the S-curve. The exponential model had the least value of

forecasting error and hence was found fit for the forecasting of potato area and production. Similar research was carried out to forecast the prices of chilli in Byadgi market where different exponential smoothing (single, double and triple) were considered but the single Exponential Smoothing was the best.

Principal Component Analysis: Research had surfaced on prediction of agricultural energy and Nikkhah *et al.* [17] identified that a problem of strong correlation from among the energy inputs in agricultural systems existed when the Artificial Neural Networks model was used and so they used the principal components as model input and not as raw data and the result showed that an improved ANN model prediction.

Linear Regression Analysis: This technique does a multivariate analysis that considers factors and group them into response and explanatory variables and aids in decision making. Sellam and Poovammal [18] used Regression Analysis to analyze some environmental parameters that can influence crop yield like Annual Rainfall, Food Price Index *etc.* and established a relationship among the parameters and the result showed that the factors 'annual rainfall', 'food price index' and 'area under construction influenced crop yield. Shastry *et al.* [19] in the bid to predict crop yield utilizing historic data from soil parameters, weather parameters and crop yield used regression techniques like quadratic, interactions, and polynomial in predicting the yields of maize, wheat and cotton and the best regression model with the least prediction error was selected. Several other models like Artificial Neural Networks, Clustering, SVMs and lots of other statistical analysis techniques have been used in different agricultural forecasting systems [20]. While using each forecasting model individually and observing success, combination of different techniques to counter the shortcomings of a particular model seems to be more successful. Emphasis on researches on agricultural forecasting systems has touched several areas however, the aspect of sales prediction has very little focus. Our study aims to develop an SVM model combined with Linear Regression to predict sales for the University of Agriculture farm.

3. Materials and Methods

This section shows the different methods that are used to achieve the aim of this research.

3.1. Support Vector Machine Algorithm

The Support Vector Machine (SVM) Algorithm is a non-linear generalisation of the generalised portrait algorithm developed in the 1960s, which is firmly grounded in the framework of the statistical learning theory. SVMs are linear learning machines, which mean that a linear function is always used to solve the regression problem. When dealing with non-linear regression, the input vector, x , is mapped into a high-dimensional feature space, z , via a non-linear mapping, and then conducting linear regression in this space

[21]. The derivation given below is adopted from Du *et al.* [21].

Given a set of data points $G = \{(x_i, d_i)\}_{i=1}^n$ (x_i is the input vector, d_i is the desired value and n is the total number of data patterns), SVMs approximate the function using the following:

$$y = f(x) = w\phi(x) + b \quad (1)$$

Where $\phi(x)$ is the high-dimensional feature space, which is non-linearly mapped from the input space x . Coefficients w and b are estimated by minimising risk function $R(C)$:

$$\text{Minimise } R(C) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^n L_\varepsilon(d_i, y_i) \quad (2)$$

$$\text{Where } L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon & |d - y| \geq \varepsilon \\ 0 & |d - y| \leq \varepsilon \end{cases} \quad (3)$$

and ε is a prescribed parameter. The term $L_\varepsilon(d, y)$ is the so-called ε -insensitive loss function. This loss function defines a flat region which takes the flatness $y = f(x)$ as the centre, the thickness of which is 2ε . When the data samples are in the flat region, the loss is equal to 0, if the discrepancy between the predicted and the observed values is less than ε . When the data samples are not in the flat region, linearity penalty is added to the function. The term $\frac{1}{2}\|w\|^2$ is used as a measure of function flatness. The constant C , which influences a trade-off between an approximation error and the weights vector norm $\|w\|$, is a design parameter chosen by the user.

To obtain the estimations of w and b , Equation (2) is transformed into Equation (4) by introducing positive slack variables ξ_i and ξ_i^* , as follows:

$$\text{Minimise } R(w, \xi^{(*)}) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^n (\xi_i, \xi_i^*) \quad (4)$$

$$\text{Subject to } \begin{cases} d_i - w\phi(x_i) - b \leq \varepsilon + \xi_i \\ w\phi(x_i) + b - d_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

By introducing Lagrange multipliers α_i, α_i^* ($\alpha_i^* \alpha_i^* = 0, \alpha_i, \alpha_i^* \geq 0, i = 1, \dots, n$) and exploiting the optimality constraints, we construct the Lagrange function according to the Lagrange dual theory.

Specifically,

$$\begin{aligned} L = & \frac{1}{2}\|w\|^2 + C \sum_{i=1}^n (\xi_i, \xi_i^*) \\ & - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - d_i + w\phi(x_i) + b) \\ & - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + d_i - w\phi(x_i) - b) \\ & - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (5)$$

In Equation (5), η_i, η_i^* are the dual variables, which satisfy $\eta_i, \eta_i^* \geq 0$. The search for an optimal saddle point ($w, b, \alpha_i, \alpha_i^*$) is necessary because Lagrangian L must be minimised with respect to w and b .

As the optimal solution, we have

$$\delta_b L = \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0$$

$$\delta_w L = w - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(x_i) = 0 \quad (6)$$

$$\delta_{\xi_i^{(*)}} L = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0$$

We obtain the dual problem by substituting (6) into (5). Specifically, the dual problem is as follows:

$$\begin{aligned} \max_{R(\alpha_i, \alpha_i^*)} : & -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \phi(x_i), \phi(x_j) \rangle \\ & - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n d_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (7)$$

$$\text{Subject to } \begin{cases} \sum_{i=1}^n (\alpha_i + \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad i=1, \dots, n$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in the feature space. Lagrange multipliers α_i, α_i^* are obtained by maximising function (7). Based on the nature of quadratic programming, only coefficients α_i, α_i^* will be assumed as non-zero, and the data points associated with them could be referred to as support vectors.

One basic idea in designing non-linear SVMs is to map input vectors x into vectors z of a higher dimensional feature space ($z = \phi(x)$, where ϕ represents a mapping). An input space (x -space) is spanned by components x_i of an input vector x , and a feature space (z -space) is spanned by components $\phi(x)$ of a vector z . By performing such a mapping, it is expected that the learning algorithm will be able to linearly separate images of x by applying the linear SVM formulation in a z -space. This approach is also expected to lead to the solution of a quadratic optimisation problem with inequality constraints in z -space. There are two basic problems in taking this approach: the first one is the choice of mapping $\phi_i(x)$; and the second problem is connected with a phenomenon called the 'curse of dimensionality'. This explosion in dimensionality can be avoided by noticing that training data appear only in the form of inner products $z_i^T z_j$, which are placed by inner products

$$z^T z_i = [\phi_1(x), \phi_2(x), \dots, \phi_n(x)] [\phi_1(x_i), \phi_2(x_i), \dots, \phi_n(x_i)]^T$$

in a feature space; the latter is expressed by using the function:

$$K(x_i, x_j) = z_i^T z_j = \phi^T(x_i) \phi(x_j)$$

where $K(x_i, x_j)$ is named as the kernel function, which is a function in the input space. The basic advantage of using a kernel function lies in avoiding having to perform a mapping $\phi(x)$. Instead, the required inner products in a feature space are calculated directly by computing kernels for given training data vectors in an input space. Thus, using the chosen kernel, a SVM that operates in an infinite-dimensional space can be constructed. In addition, by applying kernels, one does not even have to know what the actual mapping $\phi(x)$ is (Kecman 2001; Campbell 2002).

The value of the kernel function is equal to the inner product of two vectors x_i and x_j , in feature spaces $\phi(x_i)$ and $\phi(x_j)$. That is,

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = (\phi(x_i) \cdot \phi(x_j))$$

Any function that satisfies the Mercer's condition can be used as the kernel function. It should be pointed out that training SVMs is equivalent to optimising Lagrange multipliers α_i, α_i^* with constraints based on function (7).

Then, the regression function given by Equation (1) has the following explicit form:

$$y = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (\phi(x), \phi(x_i)) + b \quad (8)$$

$$y = f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b$$

3.2. Conceptual Representation of the Model

Accuracy in prediction is highly needed in making decision of any kind. As a matter of facts, using regression models for prediction is usually based on the assumption that the prospective value of a variable is linked to its past values [21]. The aim of using regression for prediction is to uncover patterns in historical data and used same to forecast into the future. Nonetheless, the future values of some items are affected by future values of one or more factors. For instance, the sale of agricultural products also obeys these rules and in a similar vein, they are affected by factors such as weather condition, temperature, season, break (holidays) and so on. Therefore, the single value regression prediction is not effective for forecasting sales of agro products. Also, previous methods of prediction were mostly applied to formulate various models centred on the original data and offered little to deal with the original data, which may contain some noisy data and void information [21]. Having considered this, we have formulated a model whose schematic view is shown in Figure 1. This model first preprocesses the historical sales data which involves smoothing and normalisation. Secondly, the model processes dynamic data such as weather data, temperature data, week data, location, season, breaks/holidays etc. corresponding to the previous sales data, by fuzzy methods. Then, the training samples inputted to SVM model are trained and learnt for adjusting the parameters to optimal values. With this, the model forecasts future sales in terms of quantity sold or the proceeds to be realised after the machine completes learning. Besides, information relevant to forecasting are processed by fuzzy methods and the results is channeled into the forecasting phase where the SVM algorithm is applied to carry out the prediction of future sales. Finally, forecasting is performed and the values are obtained after forecasting samples are inputted to the trained SVM model. With regard to the uniqueness of this model which considered the effects of factors such as weather conditions, temperature data and holidays data, etc., it shows that the result will be greatly improved with higher accuracy. Summarily, the phases of

Figure 1 are discussed below.

Data Sources: The model comprised of two data sources – past data (historical data) which is used in the training phase of the model, and real time data which is used in the testing phase of the model.

Data Preprocessing: The data collected into the model undergo denoising, imputation, smoothing and normalization before it used for the forecasting. The model uses fuzzy set theory for preprocessing of data.

Sales Forecasting: The prediction of future sales is done at this phase. This is achieved using support vector machine algorithm.

Decision Making: The output from the system is used for decision making by farmers and business men. This output is presented in the form of graphs/charts.

3.3. Model Formulation

Predicting the future sales for agricultural products is a complicated task, which could be treated as regression function $y = f(x) = w\phi(x) + b$. The output of this function is either the quantity sold or the total sales as the case may be, represented by y, and x is the input values for the function which include the factors the influence sales, such as historical sales, temperature, weather, breaks/holidays, season, location etc.

The aim of this model is to find a mapping from inputs x to the sales y, which has a better generalisation performance. Using the factors above, the inputs for the model are shown as:

$$X = [Q_{(k-1)}, Q_k, TS_{(k-1)}, TS_k, S_{(k-1)}, S_k, L_{(k-1)}, L_k, T_{(k-1)}, T_k, C_{(k-1)}, C_k, B_{(k-1)}, B_k, D_{(k-1)}, D_k]$$

K: This is the time under consideration and its values include day, week, month and year

k-1: This refers to the previous time under consideration. For instance, if 'day' is considered, k-1 means the previous day and k means the current day, if 'week' is considered, k-1 means the previous week and k means the current week, and so on.

$Q_{(k-1)}$: Previous quantity sold

Q_k : Quantity currently been sold

$TS_{(k-1)}$: Total sales for the previous time

TS_k : Total sales for the present time

$S_{(k-1)}$: Previous season (values for season include: Harmattan (dry season) and Rain (rainy season))

S_k : Present season

$L_{(k-1)}$: Previous location

L_k : Present location

$T_{(k-1)}$: Previous temperature (values include: high, mild, cold)

T_k : Present temperature

$C_{(k-1)}$: Previous weather condition (values include: cloudy, rainy/overcast and sunny)

C_k : Present weather condition

$B_{(k-1)}$: previous break/holiday (that is whether the break falls on weekday or weekend)

B_k : Present break/holiday

$D_{(k-1)}$: previous day of the week (weekday or weekend)

D_k : Present day of the week

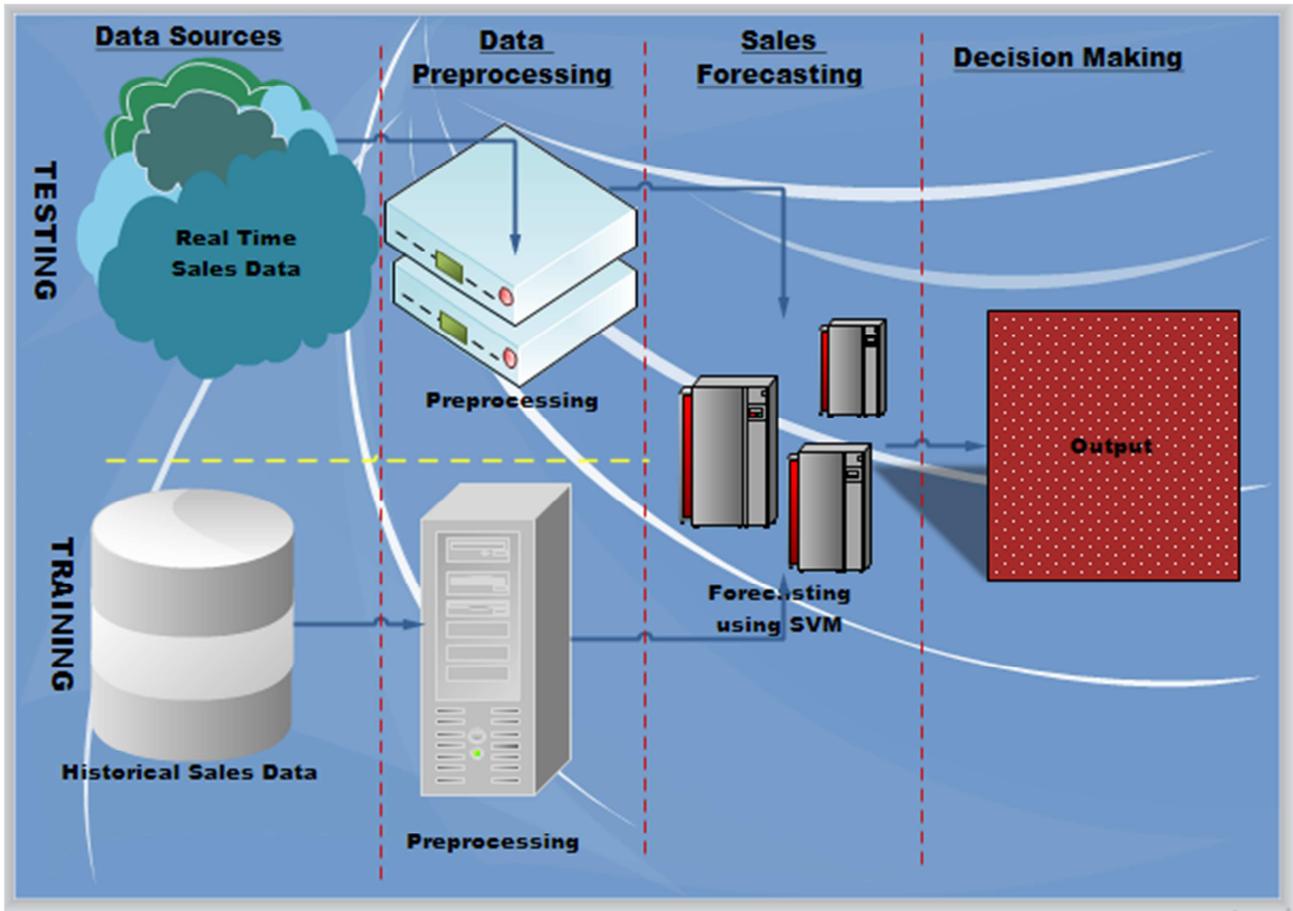


Figure 1. Schematic View of the Forecasting Model.

3.4. Processing of Data Using Fuzzy Theory

Fuzzy methods describe how to deal with vague statements or uncertain observations. It is supposed that the universe of discourse u is equal to set $\{x\}$, $u_A(x)$ is the mapping between u and closed interval $[0, 1]$ and certain vague subset of u is made by $u \rightarrow [0, 1]$, $x \rightarrow u_A(x)$. According to Fuzzy Theory, if A is a vague set, and the mapping is named as membership function of A . $u_A(x)$ is named as membership which reflects the membership degree of $x \rightarrow A$.

Being used to predict model sales of agricultural products, effect factors such as day of the week (working or weekend), temperature, weather, break, location, season etc., will be processed by the fuzzy theory. These factors are transformed into fuzzy values by the membership function, which will affect forecasting precision (FP) and robustness of the model.

Membership functions for Day of the week is of the following form:

$$u_d = \begin{cases} 0 & \text{Saturday, Sunday} \\ 1 & \text{Mon, Tues, Wed, Thur, Fri} \end{cases}$$

Membership functions for Break/holiday is of the following form:

$$u_b = \begin{cases} 0 & \text{Weekend (Saturday, Sunday)} \\ 1 & \text{Working day (Mon, Tues, Wed, Thur, Fri)} \end{cases}$$

Membership functions for weather condition is of the following form:

$$u_c = \begin{cases} -1 & \text{Rainy} \\ 0 & \text{Cloudy/Overcast} \\ 1 & \text{Sunny} \end{cases}$$

Membership functions for temperature is of the following form:

$$u_t = \begin{cases} -1 & \text{Cold } 0 \leq t \leq 10 \\ 0 & \text{Mild } 11 \leq t \leq 25 \\ 1 & \text{High } 26 \leq t \leq 40 \end{cases}$$

Membership functions for season is of the following form:

$$u_s = \begin{cases} 0 & \text{Harmattan Nov, Dec, Jan, Feb, March, Apr} \\ 1 & \text{Rain May, Jun, Jul, Aug, Sept, Oct} \end{cases}$$

4. Results

This study adopts the review of existing record method for obtaining daily sales data. This was applied to the Federal University of Agriculture Makurdi Farm for a period of three

(3) years: 2017, 2018 and 2019. The attributes of the dataset include: address, temperature, day of the week, break/holiday, season, quantity, weather, units sold, unit price, unit cost, total revenue, total cost, date, total profit of the goods (given in Naira) for a given Item or goods, with over 500000 instances (observations). This data is available as a CSV (Comma separated value) file that can easily be loaded into the system for training the model. The data was collected in the raw form (not scaled) and was partitioned into two: training data points which covers the time period from 21st January, 2017 to 30th June, 2019, and testing data points which covers from 1st July 2019 to the 31st December, 2019.

The implementation is done using Python Programming Language. The system comprised of three (3) modules: web interface, flask and the SVM Framework. The input is passed on to the Flask server. Flask is a micro web application framework written in python that includes a web server that can be used for testing and development. The supplied feature is restructured into a numpy (a specialized form of array for scientific computing in Python) array and passed on to the Support Vector Machine model which is encapsulated in the Flask server.

The SVM framework is complex and its construction

depends on the number of support vectors, rather than on the dimensionality of the feature space (Du *et al.*, 2013). The input vector is

$$X=[Q_{(k-1)}, Q_k, Ts_{(k-1)}, Ts_k, S_{(k-1)}, S_k, L_{(k-1)}, L_k, T_{(k-1)}, T_k, C_{(k-1)}, C_k, B_{(k-1)}, B_k, D_{(k-1)}, D_k]$$

the output vector is $y = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b$ where n is the number of the support vector.

The trained model makes prediction based on this input and passes the value back to the Flask server which makes it available on the web page. The exact value of the sales and the predicted sales returned to the user via the flask server using the convention adopted by Sales Manager as output on the same page. Different experiments were conducted to and the results are shown in the graphs in Figure 2.

The result from the system was evaluated by comparing its Forecasting Precision with that obtained using Radial Basis Function (RBF).

$$FP = \left[1 - \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{A(i)-F(i)}{A(i)} \right)^2} \right] X100.$$

Where A(i) and F(i) represent actual sales and the forecasting values, respectively.

5. Discussion

Table 1. Comparison of SVM and RBF.

	Training Data		Test Data		
	SVM	RBF	SVM	RBF	RBF
FP	93.881	91.913	96.75	90.55	
Average solution values	Exec. time: 10.2s Support vectors: 896		Exec. time: 7.4s	Exec. time: 0.09s Support vectors: 205	Exec. time: 0.03s

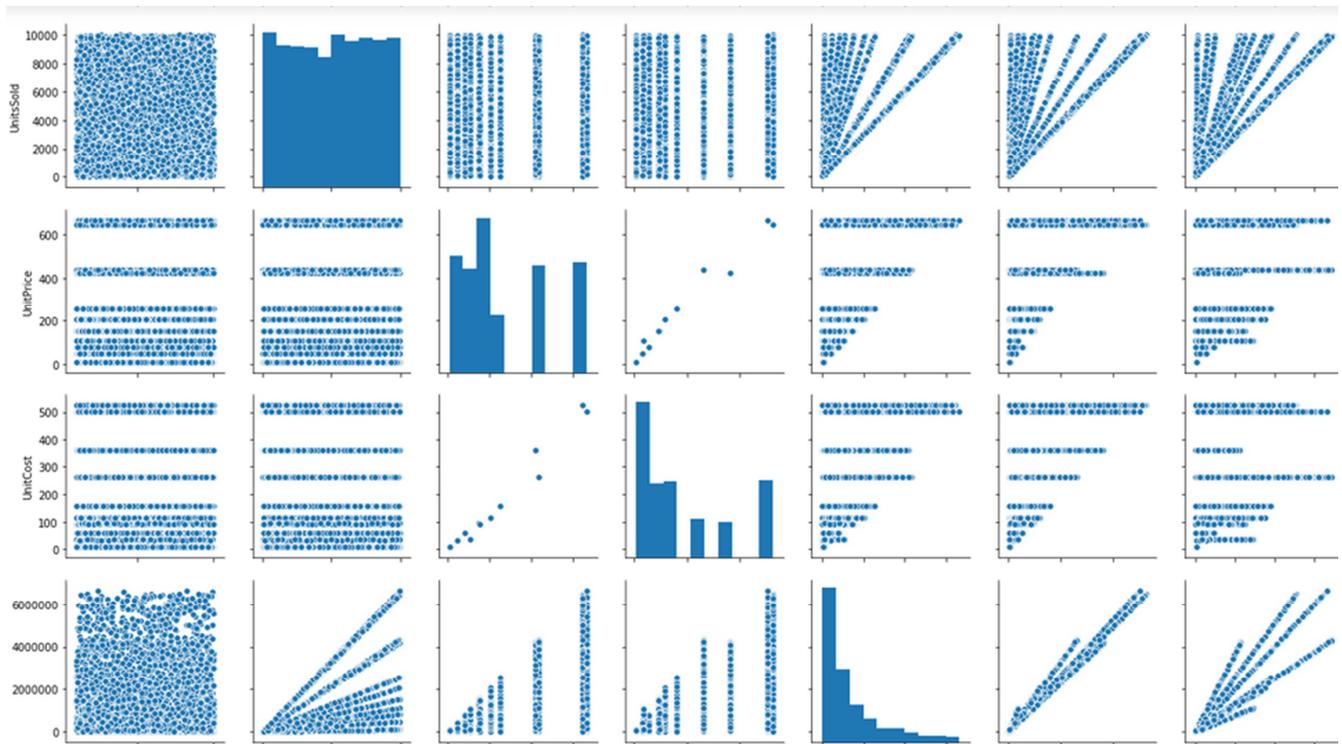


Figure 2. Simple plots to check out the data.

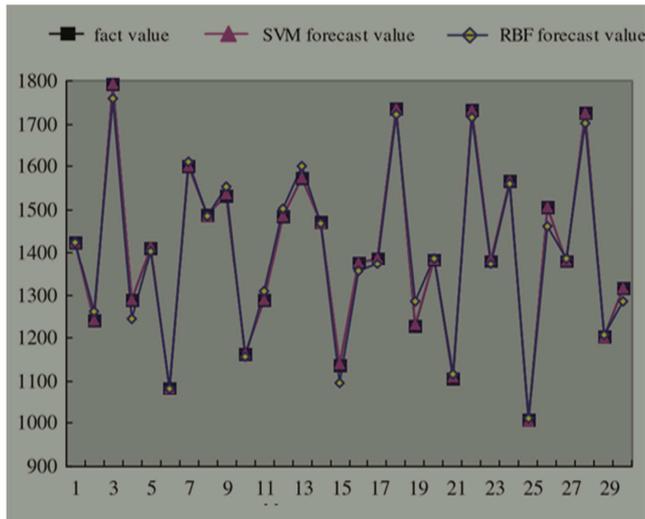


Figure 3. Comparison of actual sales, SVM and RBF.

This research developed a model for the prediction of agricultural product sales using SVMs and Fuzzy Theory. To evaluate the result of this model, the RBF neural network was used as a benchmark. After training the network which covers the time period from 21st January, 2017 to 30th June, 2019, the remaining data which covers from 1st July 2019 up to the 31st December, 2019 was used to test and validate the forecasting performance of the model. Figure 3 shows the relation between actual sales, SVMs forecasting values and RBF neural network forecasting values. This shows that the values obtained from the model are closer to the actual values as compared to those obtained from the RBF. With this, it is obvious the model performs better than the RBF. The sensitivity analysis of the model was done using FP as the performance criteria and the result is shown in Table 1. With the values in Table 1, the model used in this research outperforms the neural network methods used for the same purpose.

Experiments were conducted considering the effects of the factors outlined earlier. When the irrespective of the location, the sale was higher during weekends and holidays; in fact on days preceding holidays of festive periods like Easter, Salah, Christmas, New Year, the sale was also higher in all the locations. When temperatures are higher, most people do not go to markets and the sales drops. When the temperature is low and the weather is cloudy or rainy, people do not go to market and this affect the sales negatively. Using the FP values, the precision of the model was higher than that of the RBF, this shows clear that SVM was better than RBF in terms of performance.

6. Conclusion

In this research, a sales forecasting system was developed using SVM. Fuzzy logic was used in preprocessing the data which was collected from the University of Agriculture Makurdi Farm to cater for an effective sales forecasting process. With the results from this system, business managers can forecast sales so as to re-organize marketing strategies on where to channel excess produce timely to reduce spoilage

and increase sales which implies more profit at the end. The SVM model proves to be quite accurate in sales prediction with 96.75% accuracy beating the 90.55% accuracy from RBF neural networks. Other forecasting models may be considered for future works in the bid to get a better performance than the model proposed in this study. Future research may target how to forecast and improve sales of each farm product restricting the study to individual farm products and researchers may also want to find out the impact of certain environmental factors as peculiar to the sale of each farm product.

References

- [1] Essien J. A., Essien J. B. & Bello R. S. Product Quality Preservation and Agricultural Transformation in Nigeria. *Science Journal of Business and Management*. Vol 3, No. 5-1, 2015, pp. 35-40.
- [2] Agronigeria. ng (2017). NSPRI says Nigeria records N2.7trillion deficit annually from post-harvest losses. Available at URL: <https://agronigeria.ng/2017/10/23/nspri-says-nigeria-records-n2-7trillion-deficit-annually-post-harvest-losses/>.
- [3] Azeez, B. (2019). Nigeria loses over \$50b annually to postharvest waste. *Nigerian Stored Products Research Institute available at URL: https://tribuneonline.ng/nigeria-loses-over-50b-annually-to-postharvest-waste/*.
- [4] Alsem, K. J., Leeflang, P. S. & Reuyl, J. C. (1989). The Forecasting Accuracy of Market Share Models using Predicted Values of Competitive Marketing Behaviour. *International Journal of Research in Marketing*, 6: 183-98.
- [5] Ramasunbramanian V. (2013). Forecasting Techniques in Agriculture. Available at URL: http://cabgrid.res.in/cabin/publication/smfa/Module%20IV/5.%20Applications%20of%20Technology%20Forecasting%20methods%20in%20Agriculture_Rama.pdf.
- [6] Jha, G. K. & Sinha, K. (2013). Agricultural Price Forecasting Using Neural Network Model: An innovative information delivery system. *Agricultural economics research*. 26 (2). 229-239.
- [7] Ruekasaem & Sasananan (2018) *International Journal of Mechanical Engineering and Technology (IJMET)* Volume 9, Issue 7, July 2018, pp. 957-971, Available online at <http://www.iaeme.com/ijmet/issues.asp?IType=IJMET&VType=9&IType=7> ISSN Print: 0976-6340 and ISSN Online: 0976-6359.
- [8] Xu, G., Piao S. & Song, Z. (2015). Demand Forecasting of Agricultural Products Logistics in Community. *American Journal of Industrial and Business Management*, 5: 507-517.
- [9] Lawrencea, M. & Godwin, P. (2006) Judgemental Forecasting: A Review of Progress Over the Last 25 Years. *International Journal of Forecasting*. 22: 493-518.
- [10] Frisvold, G. & Murugesan, A. (2013). Use of Weather Information for Agricultural Decision Making. *Weather, Climate, and Society*. 5. 55-69. 10.1175/WCAS-D-12-00022.1.

- [11] Bahng, Y. & Kincade, D. (2012). The relationship between temperature and sales: Sales data analysis of a retailer of branded women's business wear. *International Journal of Retail & Distribution Management*, 40.
- [12] Kdhananjayamurthy B. V., Krishnamurthy K. N. & Murthy K. B (2018) Price forecasting of Chilli -by an Exponential Smoothing. *Journal of Current Agricultural Sciences*. Vol. 8, Issue, 10 (A), pp. 318-322.
- [13] Suttle, R. (2019). The Impact of Location on Business Success. Available at URL: <https://bizfluent.com/facts-5859788-impact-location-business-success.html>.
- [14] Karani D. K. & Wanjohi J. (2017). Factors Influencing Marketing of Agricultural Produce Among Small-Scale Farmers: A Case of Sorghum In Giaki Location, Meru County Kenya. *International Journal of Economics, Commerce and Management* United Kingdom, 5 (8).
- [15] Annadanapu, P. K. & Ravi, B. (2017). Time Series Data Analysis on Agriculture Food Production. 520-525. 10.14257/astl.2017.147.73.
- [16] Abid, S., N. Jamal, M. Z. Anwar & Zahid, S. (2018). Exponential growth model for forecasting of area and production of potato crop in Pakistan. *Pakistan Journal of Agricultural Research*, 31 (1): 24-28.
- [17] Nikkhah A., Rohani A., Rosentrater K. A., Assad M. E. & Ghnimi S. (2019). Integration of principal component analysis and artificial neural networks to more effectively predict agricultural energy flows. *Environmental progress and sustainable Energy*. 38 (4).
- [18] Sellam, V. & Poovammal, E., (2016). Prediction of Crop Yield using Regression Analysis. *Indian Journal of Science and Technology*. 9. 10.17485/ijst/2016/v9i38/91714.
- [19] Shastry, A., Sanjay, H. A., Bhanusree (2017). Prediction of Crop Yield Using Regression Techniques. *International Journal of Soft Computing*. Vol 12 (2): 96–102.
- [20] Zotteria, G. & Kalchschmidt, M. (2007). A Model for Selecting the Appropriate Level of Aggregation in Forecasting. *International Journal of Production Economics Processes*, 108: 74–80.
- [21] Du, X., F., Leung, S. C. H., Zhang, J. L. & Lai, K. K. (2013). Demand Forecasting of Perishable Farm Products using Support Vector Machine. *International Journal of Systems Science*, 44 (3): 556-567.