



## FORECASTING MODELLING AND IT'S APPLICATION

**Lecturer Engin OĞUZAY**

Department of Computer Programming, İstanbul  
Vocation School, Maltepe University

**Associate Professor Erdem UÇAR**

Department of Computer Engineering, Edirne  
Faculty of Engineering, Trakya University

**Email:** [engin\\_oguzay@maltepe.edu.tr](mailto:engin_oguzay@maltepe.edu.tr), [erdemucar@trakya.edu.tr](mailto:erdemucar@trakya.edu.tr)

### Abstract

Predictive data mining models (predictive) model developed by the application of the results with the known data and established the development of a model by moving the results of using this model is to estimate the unknown data sets, the values for the results.

Regression Analysis with the implementation of the existing data, the formulas used in making the estimates. Forecasting, modeling and application; explained by the method of machine learning and regression analysis, the results of present study are summarized and evaluated.

**Keywords:** Forecasting modeling, machine learning, regression analysis.

### 1. REGRESSION ANALYSIS

Regression analysis of predictive variables referred to as the independent variable, dependent variable, including the so-called Weightings determine the value of the variable to be estimated is a combination of independent variables.

Regression used to assess the relationship between the two variables. Keeping up with the regression curve is usually a set of points in predicting future values based on past values, a technique referred to. Independent variables in the regression technique is used to solve classification problems, database properties, estimated to be in the dependent variable refers to the class label.

Regression Analysis is used to predict continuous values. Regression is similar to classification. The most important difference can be estimated regression assessment of the sorting variable is a continuous number[Tantuğ, 2002]. Delivery methods, depending on the air temperature, wind speed and humidity are used in estimating regression analysis.

Modeled data according to the data available in different regression models. Linear regression, multiple regression, nonlinear regression techniques used for modeling continuous variables, logistic regression can be used to model discrete or categorical data[Akpınar, 2002].

Regression Analysis with the implementation of the existing data, the formulas used in making the estimates.

Using the techniques of linear or logistic regression, the function is obtained from the data available. Using known values of the variables used in estimation of other variables. Regression terminology, the variable to be estimated "dependent variable" variable or variables used to predict the dependent variable in the "independent variable" is called.

Regression relationship between the independent variables and the dependent variable is applied to obtain the best statistical technique that identifies the function[Berson, 2000]. Data mining applications in the database tables features independent variables on the dependent variable is referred to as the target property or the class label. As in other supervised learning techniques for implementing data mining is known as a result of using the data in the regression analysis generated a model that defines the relationship.

Logistic regression and the result of the variable double, triple and multiple categories of cause-and-effect relationship between explanatory variables and a method of viewing. Logistic Regression in recent times, due to the ease of use outside the numerical data to be interpreted in a comfortable came to the fore and has become a widely used method. Widely used in medicine, biology, economics, used more frequently in areas such as.

The purpose of logistic regression, with the outcome variable of one or more of the argument is to establish a model[Borgonovo, 2004]. When the difference of the values of the other regression methods, a continuous outcome variable, logistic regression, discrete variable, the



result is used in cases where two or more values. In this method, the effects of explanatory variables on the probability of the dependent variables are calculated as the determination of the probability of risk factors is provided.

Mathematical representation of the logistic regression equation can be done as follows:

$$y_i = \sum_{k=0}^p \beta_k X_{ik} + u_i$$

here;

A: The dependent variable,

$\beta$ : Regression coefficients,

X: the independent variables,

u: error term.

Logistic regression, the effects of the independent variables on the probability of outcome variables, and calculates the probability of these risk factors allows the identification.

There are 2 main reason for us to use logistic distribution. This is one of the reasons the mathematical point of view, considering the logistics distribution is a function that is very flexible and easy to use. Second, the logistics of the distribution itself is an expression of biologically sensible. Calculate the value of the dependent variable, rather than the possibility of getting a value of the dependent variable is calculated.

## 2. REFRIGERATOR CONTROL SYSTEM

### MODELLING OF APPLIED FORECASTING

Refrigerator control system, machine learning method of finding the critical level in the system aims to calculate and forecast consumption period. Interpretation of the data and to display the results (simulated application) were performed. This section details are as follows:

Product entry and exit movements were transferred to a refrigerator and a refrigerator product movement history's database already installed on the system. This information is transferred to the refrigerator and new products will operate separately, various analyzes. This analysis;

1. Predictive analysis. Here refrigerator for the past inputs and consumption of the product will be analyzed, and therefore need to be running out of products that will be estimated.
2. Refrigerated products, the current state of a percentage. 10% of the products at the bottom boundary representation

For the carrying out of this system, the following steps were taken:

- Editing the database infrastructure
- Web service software

- Transferred to the system that the border products
- Exhaustion likely that products are transferred to the system

PC software

- Historic products to enter
- New product entry
- Based on the input that each new
  - According to the shape of the consumption of the product before consumption, to estimate the termination time (predictive analysis)
  - Sub-border listing of the remaining products

Interpretation of the data and the results displayed in the web environment (simulated application) form part of the main page the following Figure 2-1 is output to the screen is as follows:

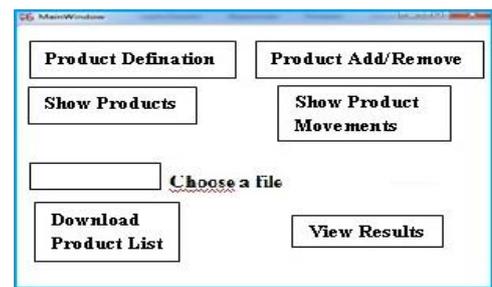


Figure 2-1. Displayed in a Web environment (simulated practice) portion of the screen is the main form page

Interpretation of the data and the results displayed in the web environment (simulated practice) portion of the end products of the critical level and recently the following Figure 2-2, which results in a display screen output is as follows:

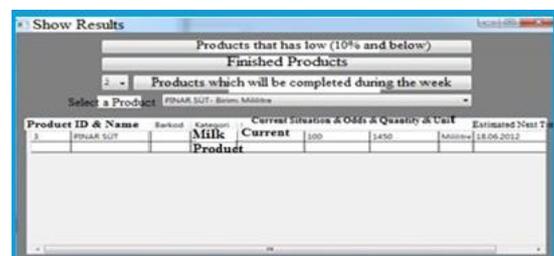


Figure 2-2. Simulated the application part of the critical level and recently the results of end products to display

Interpretation of the data and the results displayed in the web environment (simulated application) part of the



product, which is related to the add-drop is part of the output to the screen in Figure 2-3 are as follows:



Figure 2-3. Removing the add-on product, which is part of the simulated application screen

Interpretation of the data and the results displayed in the web environment (simulated application) which is part of the product movements on the part of the output to the screen shown in Figure 2-4 is as follows:

Figure 2-6. The screen shot on the new product identification procedures

Here, the simulated application predictive forecasting, analysis, and product entries on the refrigerator with the past, and hence the need to run out by the analysis of consumption goods will be given an application for estimating, machine learning outcome for the graphics (the last consumption period, the weighted average daily use of the product the final value, the total consumption time last case) are shown:

Date	16.01.2012	19.01.2012	20.01.2012	21.02.2012	23.01.2012	28.01.2012	29.01.2012	03.02.2012
Current Amount	200	0	400	200	800	600	100	500
Learning								
Last time consuming			4		3			11
The amount of final consumption			200		200			700
Daily Rate			50		66.66666667			63.63636364
New Quantity			200		400			400
Total Depletion time			4		7			18
Total Amount of Depletion			200		400			1300
Estimated Depletion time			8					
Estimated Time to Purchase Next			28.01.2012		20.01.2012			23.01.2012
Previous Date of Purchase			16.01.2012					00
Previous Purchase Amount			200		400			60.16161617
Mean Daily Use			50		58.33333333			61.11111111
Weighted Average Daily Use			50		57.14285714			

Figure 2-7. Forecasting analysis of the simulated application

Interpretation of the data and the results displayed in the web environment (simulated application) which is part of the product on the part of the output to the screen shown in Figure 2-5 are as follows:

Product ID & Name	Barcode	Category	Current Situation & Odds & Quality & Unit	Estimated Next
1 Capsoy 200 ml		fruit juice	Current 100 200	Minute
2 KEOLA		Cola	0	Minute
3 PINAR SUT		Milk Prod	Current 100 1450	Minute 18.06.2012

Figure 2-5. Part of the product display screen associated with the application of simulated

In addition, new product identification procedures related to the following Figure 2-6 is display output is as follows:

Every time a new product is added to the amount or value of the product is removed from the product table is updated. The new items are added, updated in a previous purchase amount and the purchase date and the amount of final consumption, consumption of time is calculated. In addition, the new total consumption and total consumption amount of time will be updated. However, the weighted average is calculated daily use, update performed, a new weighted average and the new estimated current consumption of the product by the amount of time will be updated the next time of purchase.

Final consumption and final consumption quantities of the application period values in Chart 2-1 are obtained from the analysis of:



PRODUCT ID	Last Depletion Amount	Last Depletion time
2	200	3
2	700	11
2	600	15

Chart 2-1. Product\_id number 2 final consumption amounts and final consumption of the product

Final consumption quantities and values of final consumption in the application period results have been obtained from the analysis of Figure 2-8 :

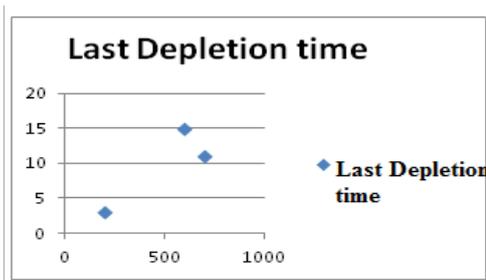


Figure 2-8. End of the application to consume simulated time

Practical products of daily use weighted average values are the last values are obtained from the analysis of Chart 2-2:

PRODUCT ID	PRODUCT PROCESS ID	PRODUCTS FOR DAILY USE LAST VALUE WEIGHTED AVERAGE
2	1	66
2	2	64
2	3	51

Chart 2-2. Product\_id number 2 last values weighted average daily use of the product

As a result of analysis of the final values weighted average daily use of the application; Figure 2-9, results are obtained:

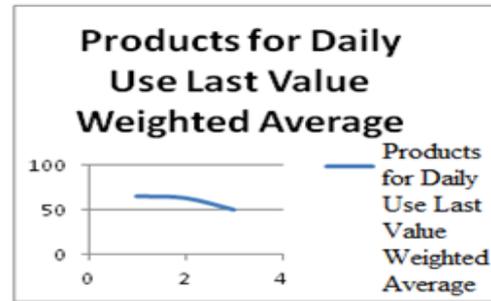


Figure 2-9. Weighted average values of simulated daily use of the application

In addition, the application developed simulated shown in Chart 2-3 and Figure 2-10, the latest state of the product is obtained from the total consumption time:

PRODUCT ID	Depletion Total Amount of Current	Total Depletion Time Last Status
2	200	3
2	900	14
2	1500	29

Chart 2-3. The status of a total duration of consumption of the product id number 2

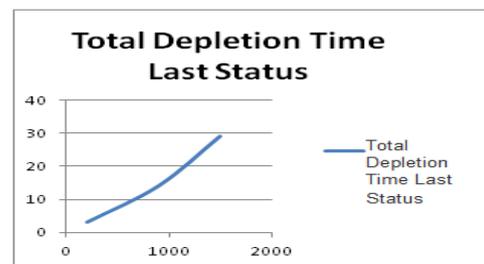


Figure 2-10. Total consumption of simulated time the status of the application

The display for refrigerator products are defined in the application of simulated product availability are shown in Figure 2-11 below:

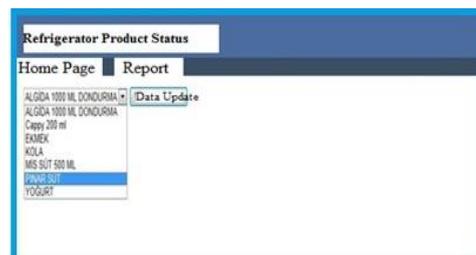




Figure 2-11. Simulated on the display refrigerator products are defined in the application for product status

The application of simulated product-based product, daily use of products with an average refrigerator, expiry time, according to the criteria of the latest situation of total consumption time of the calculation of the estimated time of consumption are shown in Figure 2-12:

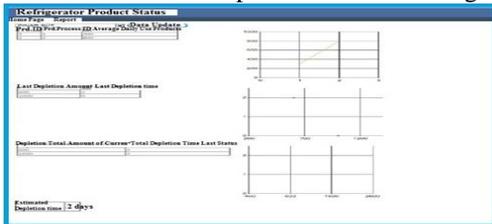


Figure 2-12. Product-based product, the average daily use, expiry time, according to the criteria of the latest situation of total consumption time of the calculation of the estimated time of consumption

Simulated refrigerator is running out of application products with product availability (10% and below) on the basis of the week and time of your next purchase finished products is calculated on the estimated values, the following results are shown in Figure 2-13:

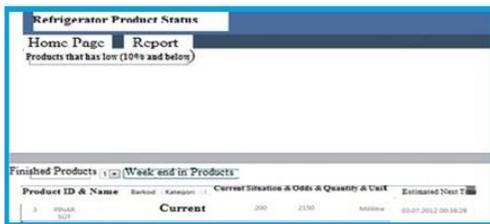


Figure 2-13. Running out of the products (10% and below) and the week ending about products on the basis of the calculation of the estimated time of your next purchase

However, the simulated application, flow, and the current situation in the calculation schemes are shown in Figure 2-14 :

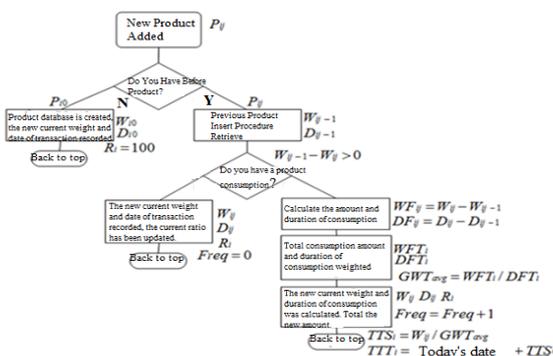


Figure 2-14. Flowchart simulated application

Simulated application, a new product is added to the system, at first, before the system before any product is any product not analyzed. If you do not have a product defined earlier in the product database by creating a new current weight and the date of transaction is recorded and updated in the current rate (Wi0, D, i0). There is a product already on the system from the previous product is the process of adding information (Wij-1, D ij-1) and status of consumer products is checked (Wij-1-Wij > 0).

Product consumption is not available, the most recent new addition to the amount of the previous insertion process the amount of the difference between the current crop of process is checked and the new current weight, date of transaction is saved, the current rate is updated (Wij, Dig, Freq = 0). In the case of product consumption, consumed amounts and duration of consumption is calculated (WFij, DFij). The amount of total consumption, total consumption of the weighted average amount of time and is updated daily consumption (WFTi, DFTi, GWTavg).

As a result of all these operations, the new current weight and the date of transaction is recorded. And the current rate of consumption is calculated according to the amount of new time will be updated.

Here, the weighted average value of daily intake (GWTavg), (2.1) is provided by the number of equations.

$$GWT_{avg} = \frac{(W_{ij+1} - W_{ij}) + (W_{ij} - W_{ij-1}) + (W_{ij-1} - W_{ij-2}) + \dots}{(D_{ij+1} - D_{ij}) + (D_{ij} - D_{ij-1}) + (D_{ij-1} - D_{ij-2}) + \dots} \quad (2.1)$$

In addition, the expected consumption time (days) basis (2.2) is the number of equations.

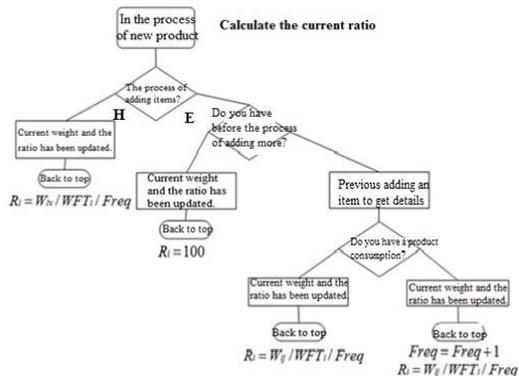
$$TTS_i = W_{ij} / GWT_{avg} \quad (2.2)$$

Estimated date of consumption (estimated after the date of purchase) with respect to ta (2.3) was carried out using equation.

$$TTT_i = \text{Today's Date} + TTS_i \quad (2.3)$$

Estimated date of purchase after the previous transaction history, with the addition of the estimated consumption is time to day basis.

The following Figure 2.15, the current rate of simulated calculation flow chart of the application are drawn:



**Figure 2-15.** Current Odds Calculator Flow Chart of simulated application

The calculation of the current rate, the system when the new product introduction process, the product has been made the process of adding the existing weight and the current rate is updated. If you made the process of adding items to the system, see if there is an earlier insert operations.

More items to add pre-made, only the current weight and the current rate is updated. Product made in the process of adding the previous product is the process of adding information. The product is also available depending on the weight and the current rates of consumption will be updated.

All calculations in the application of both the process of adding realized. Records are kept during the removal process, but calculated from the process of adding a calculation is made without time consuming.

## CONCLUSION

Simulated and the results of application software developed in light of predictive models method, regression analysis was established with the help of this model is intended to predict the results of unknown data sets of values for the results.

Established in accordance with this data model, the system used to estimate the critical level and the estimated duration of consumption.

## REFERENCE

- [1] **Tantuğ A.C., (2002).** Data Mining and Clustering, Istanbul Technical University, Institute of Science and Master of Science Thesis, Istanbul, p.1.
- [2] **Akpınar, H. (2000).** "Knowledge discovery in databases and data mining", Journal of Istanbul University, Faculty of Business Administration, 29: 1-22.
- [3] **Berson, A.; Smith, S.; Thearling, K. (2000).** "Building Data Mining Applications for

CRM", McGraw-Hill Professional Publishing, New York, USA.

[4] **F. Borgonovo, A. Capone, M. Cesana, L. Fratta. (2004).** ADHOC MAC: a new MAC architecture for ad hoc networks providing efficient and reliable node densities. point-to-point and broadcast services, Wireless Networks, vol. 10, no. 4, pp. 359-366.

[5] **Buruncuk G., (2006).** Data Mining For Customer Segmentation And Profiling: A Case Study For A Fast Moving Consumer Goods (Fmcg) Company, Bogazici University, Institute of Management Information Systems Master of Science Thesis, Istanbul, p.7.

[6] **H.Uğuz vd., (2000).** "Web Usage Mining Method Using Apriori Algorithm Implementation of Web Log Records", IJCI Proceeding of International Conference on Signal Processing, C.1, P.2 s: 499-501.

[7] **I.F. Akyidiz, W. Su, Y. Sankarasubramaniam, and E.Cayirci. (2005).** Wireless sensor networks: A survey. Computer Networks, 3:393-422.

[8] **Margaret H. Dunham, (2003).** Data mining Introductory and Advanced topics, Prentice Hall, s 13.

[9] **Santi, P. (2005).** Topology control in wireless ad hoc and sensor networks. ACM Computing Surveys 37, 164–194

[10] **Altıntop U, (2006).** Application of Data Mining Techniques for Internet-Based Education, Kocaeli University, Institute of Science and Master of Science Thesis, Kocaeli, p.15.

[11] **ZhaoHui Tang ve Jamie MacLennan, (2005).** Data Mining with SQL Server 2005, Wiley Publishing, Indianapolis, , s.13-17.