T.C.

BAHCESEHIR UNIVERSITY

NEWSPAPER VENDOR SALES PREDICTION BY USING DATA MINING TECHNIQUES

Master Thesis

Barış AKBAŞ

İSTANBUL, 2011

T.C.

BAHCESEHIR UNIVERSITY
The Graduate School of Natural and Applied Sciences
Computer Engineering Graduate Program

NEWSPAPER VENDOR SALES PREDICTION BY USING DATA MINING TECHNIQUES

Master Thesis

Barış AKBAŞ

Supervisor: Assoc. Prof. Dr. Adem KARAHOCA

İSTANBUL, 2011
T.C.

BAHCESEHIR UNIVERSITY
The Graduate School of Natural and Applied Sciences

Computer Engineering Graduate Program

Name of the thesis: NEWSPAPER VENDOR SALES PREDICTION BY USING
DATA MINING TECHNIQUES

Name/Last Name of Student : Barış Akbaş

Date of Thesis Defense : 15.08.2011

The thesis has been approved by The Graduate School of Natural and Applied
Sciences.

Signature

Assoc. Prof. Dr. F. Tunç BOZBURA

Active Director

This is to certify that we have read this thesis and that we find it fully adequate in
scope, quality and content, as a thesis for the degree of Master of Science.

Examining Committee Members

Assoc. Prof. Dr. Adem Karahoca (Supervisor)   :

Asst. Prof. Dr. M. Alper Tunga   :

Asst. Prof. Dr. Yücel Batu Salman   :

ABSTRACT

NEWSPAPER VENDOR SALES PREDICTION BY USING DATA MINING
TECHNIQUES

Akbaş Barış

The Graduate School of Natural and Applied Sciences, Computer Engineering
Graduate Program

Supervisor: Assoc. Prof. Dr. Adem Karahoca

August 2011, 55 pages

This study focuses on the newspaper sales prediction problem. This is a
Significant problem especially for newspaper companies, which always try to
reduce the number of sell outs and the number of return in the delivered
newspapers. This is crucial for decreasing the cost and increasing the profit.
It is basically a time-series problem, so it is attempted to be solved with the Holt-
Winter's, Linear Regression and FCM+ANFIS methods. The relationship
between the inputs and the outputs are investigated and calculated by these three
methods. The results are presented in the findings section, where they are also
compared. The best method for each outlet is selected according to forecast
evaluation measures.

Keywords: Newspaper sales prediction; Time-Series Prediction; Holt-Winter's;
Linear Regression; ANFIS

ÖZET

VERİ MADENCİLİĞİ YÖNTEMLERİ KULLANILARAK GAZETE SATIŞLARININ TAHMİN EDİLMESİ

Akbaş Barış

Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Yüksek Lisans Programı

Danışman: Doç. Dr. Adem Karahoca

Ağustos 2011, 55 sayfa

Bu çalışmada gazete satışlarının tahmin edilmesi problemine odaklanılmıştır.Satış tahminlerine bağlı olarak iade edilen ürün sayısını azaltmaya çalışmak, özellikle gazette şirketlerinin zararlarını azaltıp kârlarını artırabilmeleri için çözülmesi gereken önemli bir problemdir. Bu aynı zamanda bir zaman-serisi tahmin problemidir. Bu problem, Holt-Winters, Doğrusal Regresyon ve FCM+ANFIS yöntemleri ile çözlülmeye çalışılmıştır. Bu üç yöntem ile çözüm uygulanmış ve sonuçlar, bulgular bölümünde karşılaştırılmıştır. Her bir satış mekanı için kullanılacak en iyi yöntem, tahmin değerlendirme kıstaslarına gore seçilmiştir.

Anahtar Kelimeler: Gazete satş tahmini, Zaman-serisi tahmini, Holt-Winters, Doğrusal Regresyon, ANFIS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Adaptive Neuro Fuzzy Inference System	:	ANFIS
Autoregressive Integrated Moving Average	:	ARIMA
Area Under Curve	:	AUC
Confusion Matrix	:	CM
False Negatives	:	FN
False Positives	:	FP
False Positive Rate	:	FPR
Fuzzy Inference System	:	FIS
Mean Absolute Error	:	MAE
Mean Absolute Percentage Error	:	MAPE
Matrix Laboratory	:	MATLAB
Receiver Operating Characteristic	:	ROC
Root Mean Square Error	:	RMSE
True Negative Rate	:	TNR
True Negatives	:	TN
True Positive Rate	:	TPR
True Positives	:	TP
Vector Autoregressive Moving Average	:	VARMA

1. INTRODUCTION

This thesis focuses on the problem of forecasting the sales of a newspaper, using several computational methods. Automating this daily decision task and using computational methods for forecasting is essential for reducing the cost and increasing the profit.

Newspapers distribution is an everyday task that requires a lot of planning and work. The number of newspapers that are delivered to an outlet each morning is not determined randomly. Experts use their knowledge and prediction skills, in order to find the right numbers of newspapers to send to the outlets every day. Estimating the right number of delivery is crucial. If too many newspapers are delivered, then they will not be sold and will be returned to the company. The more newspapers are returned, the more is the cost of this to the newspaper company. If the delivery is too little, then they will be sold out early in the day, so the potential customers in the rest of the day will not have chance to buy the newspaper, and this will decrease the profit of the company. The aim of sales prediction is keeping the returned newspapers at minimum while increasing the sales.

As the number of unsold newspapers (return) increase, the cost for the company increases. Therefore it is essential to decrease this number. For this reason, a newspaper company needs a successful program for predicting the future sales numbers. A sell-out is not something beneficial for the newspaper company, either. Because it indicates that more newspapers could have been sold, and more profit could have been made. So the number of newspapers delivered to an outlet should be an optimized number, and it should be estimated with the least possible error.

Forecasting the sales of the day is done considering some data about that day in addition to experience from the previous days. The factors about the day can be if

the topics of the agenda are interesting, or whether the weather will be sunny or rainy, etc. The experts put their forecast in Equation 1.1, together with a safety term, while determining how many newspapers should be delivered to each outlet:

$$\text{Delivery} = \text{Forecast} + \text{Safety Coverage} \qquad (1.1)$$

Obviously, this decision requires experience, if a human being is responsible for it. The traditional method is to use an expert's experience to decide how many newspapers to deliver to each outlet, everyday. The limited simulation environments have been forcing people to use the traditional method. However, with the development of technology, and with the computers high processor speeds, data mining methods have developed enormously. Like many other problems, the newspaper sales forecasting is also attempted to be solved with computational methods. The expert's memory and decision making capabilities in such problems are slowly being replaced with analytical and computational skills.

In HaberTürk newspaper distribution, this forecasting has been done manually, as it is in many other newspapers today. It means that the number of newspapers that will be delivered to each outlet is determined by an experienced person, only by using his experience. This might be an efficient enough technique up to now. But with the increasing computational power available, and with the developing data mining methods and algorithms, this specific problem has become a hot topic, as the newspaper sales prediction problem. In this study, this challenging problem is attempted with computational methods.

The two main methods used are the Linear Regression and Holt-Winters, which are applied in the R-Project statistical environment. These methods are based the idea that this is a time-seires prediction problem (Chatfield 2001). The demand prediction problem is attempted to be solved with ANFIS algorithm in MATLAB as the third method, and the results of these three methods are presented in the results part of this thesis.

The aim of this study is to automate the sales prediction process, and solve it computationally rather than manually. The authors believe that this will reduce the error made in predicting the numbers, and result in less cost to HaberTürk newspaper company.

The number of HaberTürk newspapers that will be delivered to 25.000 outlets has to be decided every day. This is a challenging problem because first of all newspaper sales are not static in Turkey. Secondly, the sales vary depending on the day of the week, the weather condition and vacations. Also, the promotions or breaking news usually have a large impact on the sales.

2. LITERATURE REVIEW

There is a large amount of work done in literature, about the newspaper sales and sales prediction problems, as well as the data mining techniques. (McCleneghan 1997) discusses the decreasing newspaper sales and how to increase the sales in the newspaper business. (McCarthy 2009) gives valuable information on demand analysis. In (Rafenberg 1998), the importance of the problem of predicting the newspaper sales is presented in detail, giving examples about the newspaper Le Monde. These studies prove the importance of the newspaper sales prediction problem and lead us to investigate the computational methods for solving this problem.

As for the computational solution approaches to solve this problem, first of all (Porteus 2008) explains and formulizes the problem very well. As explained in (Porteus 2008), this problem is defined as the newsvendor problem because of its nature: the newspaper distributer has to estimate the daily demand for newspaper beforehand, so as to deliver the optimum number of newspapers to the outlets. Delivering too little or too much are both disadvantageous for the company. An important aspect of this problem is that the number of delivery needed is highly random.

Newspapers in the world have supported research in this area. In (Ragg 2002), a study on the newsvendor problem is done for BILD, the well-known newspaper in Germany. In this study, a prediction system is developed using Bayesian learning based on neural networks. Another similar system called JED (Just Enough Delivery) is presented in (Heskes 2003) and this system is used by De Telegraaf, a major Dutch newspaper company. As explained in this article, the solution approaches to this problem are classified in two groups: the rule based systems, and the adaptive systems. The rule-based approaches contain user-specified simple rules of the type ''if the outlet has been sold out the last two weeks, increase the delivery with two copies." The adaptive systems use the

explanatory variables like "recent sales figures, sales figures from the last year, seasonal variables, information on competitors, holidays, weather, news content and so on" (Heskes 2003). However the authors of the article warn us about the high risk of over-fitting in adaptive systems.

In (Lauer 2003), another solution approach to the same problem is presented. In this article, as opposed to (Heskes 2003), the seasonal variable does not exist and a shorter history of sales is used for prediction. A mathematical model and its implementation are deduced and a data cleaning approach is presented. The model developed in this study is compared with a neural prediction system, again applied on the sales data of the German newspaper BILD.

The newspaper demand prediction is performed in (Cardoso 2007), which uses fuzzy clustering for database exploration and fuzzy rules for prediction. In this article it is concluded that this method works better than neural-network based predictors and autoregressive forecasters. In (Fakharudin 2009) the same problem is attempted with artificial neural networks. (Ali 2009) focuses on a similar problem, a grocery sales prediction problem. The authors claim that the "simple time series techniques perform very well for periods without promotions. However, for periods with promotions, regression trees with explicit features improve accuracy substantially."

In Table 2.1, the previous studies on similar problems are listed. This table consists of the studies on mainly the sales prediction problem, also referred to as demand prediction problem. There are also studies on irregular time-series, since it is a more general case.

Table 2.1: Previous studies on sales prediction problem and irregular time-series

Authors	Year	Methodology
Ali, et al.	2009	Benchmark method
		Linear Regression
		Suport Vector Regression
Cardoso, et al.	2007	Fuzzy prediction method
Cipra, et al.	2008	Exponential Smoothing based on ARIMA (Autoregressive integrated moving average)
Cooray, et al.	2007	Exponential Smoothing (Holt Winters, Double Seasonal)
		VARMA (Vector Auto Regressive Moving Average)
Fakharudin, et al.	2009	Neural Networks
Hanzak, et al.	2008	Imporved Holt Method
Heskes, et al.	2003	Neural Networks with Bayesian Learning
Kalekar, et al.	2004	Exponential Smoothing (Holt Winters)
Lauer, et al.	2003	Neural Networks with Bayesian Learning
Ragg, et al.	2002	Neural Networks with Bayesian Learning
Şen, et al.	2009	Holt-Winters - ARIMA (Autoregressive integrated moving average)

In this chapter the state of the art in this field is presented. In the following chapter, the methods used in the prediction programs and algorithms will be introduced. The main elements will be the Holt-Winters method, Linear Regression Method, ANFIS method and Fuzzy C-Means clustering method. The implementation of these methods and algorithms in our specific problem will also be explained. The problem will be formulized and the solution steps will be explained in detail. In the 4th chapter the prediction results, and the errors obtained by these methods are presented and compared. The 5th chapter is the concluding chapter which presents the summary, the lessons learned, and the discussion of the results.

3. MATERIALS AND METHODS

As mentioned in the previous chapter, researches have been investigating the methods for the sales prediction problem for years. In this chapter, the sales prediction problem and its solutions are explained in detail. The assumptions are stated, and the solution approach is explained. For the specific problem presented in this thesis, these three main methods are focused on: Holt-Winters, Linear Regression and FCM+ANFIS. The information given at the beginning of this chapter about the methods is essential for understanding the application of these methods to this problem.

3.1 DATA SET

For sales forecasting, a computer program requires the recent sales history. This helps the program to analyze the range in which the numbers are varying, how often they make unexpected peaks or falls, etc. If there is a periodicity in the sales history, then it is easier to make reasonable guesses. For example, the periodicity may be due to the weeks of the day.

Since the sales is the single output of the system for training an ANFIS model, first of all a history of sales is needed. HaberTürk newspapers' sales information is used as the database for this problem. The data set at hand covers the sales data for 378 days (from 01-04-2010 to 13-04-2011), and for 378 outlets all over Istanbul. So the total amount of data lines is 378 x 378 = 142884. Since there are too many lines to fit in this document, a representative part of the data was shown in Table 3.1

In Table 3.1, the first column shows the date, the second column shows the outlet code, the third column shows the number of newspapers delivered to that outlet that day, the fourth column shows the sales of the corresponding day and the outlet, and the fifth column shows the number of newspapers unsold and returned.

Table 3.1: The sales history of the outlets from 01-04-2010 to 13-04-2011

Date	Outlet	Delivery	Sales	Return
2010-03-01	3517	7	7	0
2010-03-01	3559	50	35	15
2010-03-01	3622	18	12	6
2010-03-01	3638	30	22	8
2010-03-01	3649	12	11	1
2010-03-01	3673	8	8	0
2010-03-01	3682	25	25	0
...				
2010-03-02	3517	7	7	0
2010-03-02	3559	40	32	8
2010-03-02	3622	20	18	2
2010-03-02	3638	32	24	8
2010-03-02	3649	15	11	4
2010-03-02	3673	15	9	6
2010-03-02	3682	35	25	10
...				
2010-03-03	3517	7	7	0
2010-03-03	3559	40	30	10
2010-03-03	3622	17	11	6
2010-03-03	3638	30	26	4
2010-03-03	3649	17	9	8
2010-03-03	3673	14	6	8
2010-03-03	3682	33	23	10
...				
2011-04-13	3517	7	7	0
2011-04-13	3559	55	45	10
2011-04-13	3622	18	12	6
2011-04-13	3638	26	26	0
2011-04-13	3649	17	13	4
2011-04-13	3673	23	19	4
2011-04-13	3682	37	28	9
…				

The first two methods use only the sales history information, however the estimation process is supported by additional information like weather condition history and promotion history with FCM+ANFIS. A list of the features used in addition to the sales data is given in Table 3.2

Table 3.2: A list of additional features

Abbreviation	Meaning
Weather	Weather Condition (Sun, Cloud, Rain)
Promotion	Additional Pages/Magazines given with the newspaper
Day	Day of the week (Monday - Sunday)
Holiday	Whether that day is a national holiday or not
Cluster No	The cluster number of the outlet after applying FCM
Output	Number of newspapers sold

Considering all the effects depending on the date and the outlet, a software algorithm can be used to train on the existing data and make predictions about the future sales data, using the ANFIS method.

Weather condition is an important parameter for newspaper sales. It is observed that in dry days the sales are higher. If the weather is rainy or stormy, low sales are not surprising. If the weather condition history is also available in addition to the sales history, then the software can also analyze the effect of the weather condition on sales. Naturally, the weather forecast will be an additional clue for forecasting the sales of a future day.

Promotions are very effective in selling newspapers. If there are unexpected peaks in the sales history, it is reasonable to check which promotions were distributed that day. The promotions for the children and women are especially important. Another effective parameter is the day of the week; whether it is a weekday or weekend. The holidays are also important. These are the days which the regular paths of people change. So the sales are affected.

The weather condition, the promotion, the day of the week and the holiday information are the parameters that depend on the specific day, or the date. There are also parameters which depend on the sales place.

One such parameter can be the general behavior of that outlet. Does it have a high sales potential, or just a few customers every day? This is an important question. The places can be divided into several clusters before the estimation process, by using Fuzzy C-Means (FCM) algorithm, for example. One cluster includes the small shops, and another includes the places with high sales rates. Depending on the total cluster number, there might be clusters in between these two, with moderate sales rates. The cluster that an outlet belongs gives an idea about the average sales rates of that outlet.

Besides the general expectation from an outlet, the weekly affects may also be included. For this purpose the last week's sales average is calculated for each day and each outlet. How many newspapers that outlet has sold in the last 7 days in average, gives an important clue about what to expect from there these days.

The last two were the effective parameters which depend on the outlet.

The inputs of the ANFIS method, as mentioned before, depend on the day and the place. The inputs depending on the day are listed below:

- The weather condition of the day
- Whether that day is a national holiday or not
- The promotion information (if/which promotion was given with the newspaper that day)
- The day of the week

The properties of each day are given in a matrix format in Table 3.3. In the actual matrix, there are 378 lines, each line corresponding to a day. But here, because of the space limitations, only a part of the matrix is presented, showing a limited

number of days. This much of the matrix is sufficient to understand the structure of it.

Table 3.3: The parameters of the data set, depending on the day

2010-04-01	3	0	2	4
2010-04-02	2	0	6	5
2010-04-03	1	0	2	6
2010-04-04	1	0	2	7
2010-04-05	1	0	2	1
2010-04-06	3	0	3	2
...				
2011-04-13	1	0	1	3

Each row of Table 3.3 corresponds to a day. The columns, their meanings and the values that they can take are explained in   Table 3.4.

Table 3.4: The day-related inputs and their numeric representatives

COLUMN NUMBER	MEANING	VALUES
1	Date	
2	Weather condition	1: Snow/storm
		2: Rain
		3: Sunny/couldy
3	Whether it is a national holiday	1: Yes
		0: No
4	Type of promotion	1: No promotion
		2: Other promotions
		3: Football magazine (12-16 page)
		4: CD for kids/Lottery Results
		5: Women's magazines/Card games
		6: Football magazine (32 page)
5	The day of the week	1: Monday
		2: Tuesday
		3: Wednesday
		4: Thursday
		5: Friday
		6: Saturday
		7: Sunday

Another type of inputs that affect the sales data is the outlet-related inputs. These are:

The cluster number of the outlet(which was explained in the previous subsection)

The average sales of the past 7 days, for that outlet.

The cluster number is already a numeric value and it can take the values of 1, 2 or 3. The average sales number for the past week is also a numeric value and it helps ANFIS consider the recent sales conditions for the relevant outlet.

So, we can conclude that a total of 7 inputs lead to 1 output (the number of newspapers sold) for each day and for each outlet. After analyzing the system as such, it is easy to turn these inputs into numbers and feed them to ANFIS, using MATLAB.

3.2 METHODS

The methods used for forecasting in this study are the Holt-Winters, Linear Regression and ANFIS. In this section these methods will be explained. Also, the FCM algorithm will be presented because it is used for clustering before applying ANFIS. The forecast evaluation methods will be presented at the end of this chapter, for comparing the performances of the forecasting methods.

3.2.1 Holt-Winters

Holt-Winter's Method is a popular method for time-series forecasting problems, like the one subject to this study. It is based on exponential-smoothing approach (Winters 1960, Goodwin 2010, Kalekar 2004). The input of the method is an array of periodically collected samples of data. The output is the prediction for a future sample. The algorithm is presented in Figure 3.1. (The figure is taken from (Dahyot 2010).)

Init:

$$L_s = \frac{1}{s}\sum_{i=1}^{s} y_i$$

$$b_s = \frac{1}{s}\left[\frac{y_{s+1}-y_1}{s} + \frac{y_{s+2}-y_2}{s} + ... + \frac{y_{2s}-y_s}{s}\right]$$

$$S_i = y_i - L_s, \; i = 1,... s$$

and choose 0<α<1 and 0<β<1

Compute:

$$(Level)\; L_t = \alpha * (Y_t - S_{t-s}) + (1-\alpha)*(L_{t-1} + b_{t-1})$$

$$(Trend)\; b_t = \beta * (L_t - L_{t-1}) + (1-\beta)*b_{t-1}$$

$$(Seasonal)\; S_t = \gamma * (Y_t - L_t) + (1-\gamma)*S_{t-s}$$

$$(Forecast\;\; for\;\; period\;\; m)\; F_t + m = L_t + m*b_t + S_{t+m}$$

until no more observation y_t are available

and subsequent forecasts: $F_{t+k} = L_t + k\,b_t + S_{t+k-s}$

Figure 3.1: Holt-Winters algorithm

As seen in the algorithm, the future samples of the time-series data are dependent on the past samples. The functions relating the future to the past is given in the algorithm. The coefficients used in these functions are calculated within the algorithm, and they are dependent on the past data, as well. The method analyzes the change in data through time or the chosen coordinate, in order to predict the future.

The coefficients and the variables used in the algorithm are listed in Table 3.5

Table 3.5: The coefficients and the variables used in Holt-Winters algorithm

α	Alpha
β	Beta
γ	Gamma
s	Number of periods ahead to forecasts – length of the seasonality
L_t	Level of the series at time t
b_t	Trend of the series at time t
S_t	Seasonal component at time t

An important feature of this method is that it takes the seasonal effects into account. That is, the prediction is done with an assumption that the data has a periodic ingredient in it. And this period has to be determined beforehand. For example if the data are taken once a month, and an annual periodicity is expected, then the period of the data (s) should be taken as 12.

The algorithm can be applied with any convenient software program. In this study, the method is applied with R-Project statistical environment.

3.2.2 Linear Regression

This method is also a popular method for time-series prediction problems (Kalu 2010). The aim is to predict a future instant of the data, given the history of it. So the data at hand is only an array of samples, which are taken equidistantly. A general formula for the linear regression method can be as follows:

$$Y = b_0 + b_1 * X1 + b_2 * X2, \tag{3.1}$$

where Y is the prediction and the X's are the past instances of the data array. Suppose that the data at hand is a daily data. As can be seen from the formula, the prediction of tomorrow can be a function of yesterday and/or several days ago. This can be chosen according to the nature of the problem, by the user.

Since the X values are known, the unknowns are the b_0, b_1 and b_2 coefficients in this formula. The role of the linear regression is to find out these coefficients so that a prediction can be made. These coefficients are found by the algorithm of linear regression, by analyzing the past data.

3.2.3 Fuzzy C-Means Clustering

Fuzzy C-Means is one of the most well-known techniques for clustering a set of data points (Bezdek 1981). The process of clustering data points can be considered as dividing the data set into convenient groups, so that the points in the same group have similar properties. The properties mentioned here are represented by the dimensions of the data vector.

The coordinates of the data points and the number of desired clusters are input to the Fuzzy C-Means (FCM) algorithm. The algorithm eventually finds the cluster centers, and the percentages of the data points showing with what percentage they belong to each cluster.

The flow chart for the FCM algorithm is presented in Figure 3.2. Here, "e" is the value of the specified threshold as a stopping criterion for the algorithm. (The figure is taken from (Jaiswal 2007)).
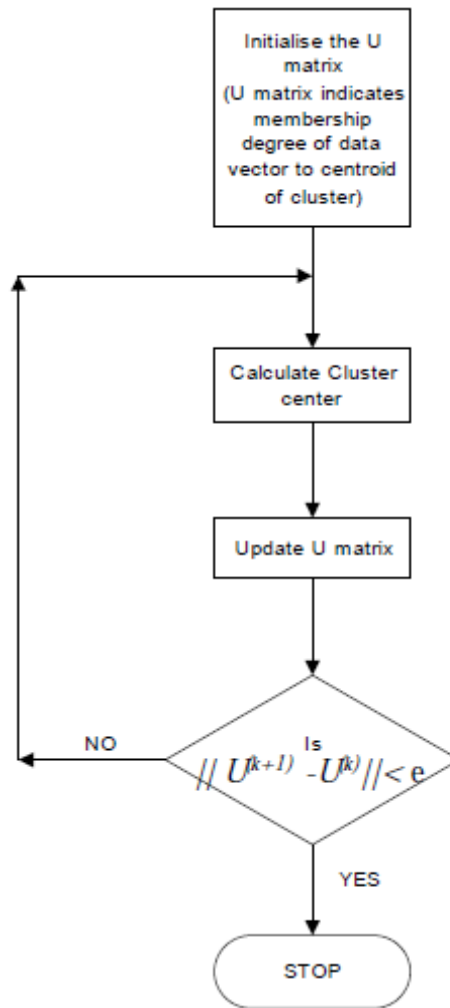
Figure 3.2: The basic flow chart for the Fuzzy C-Means algorithm

A data point is defined by its coordinates. If the data set is one dimensional, then each data point has only one numeric value. If the data set is n-dimensional, then one has to fix the point in an n-dimensional space by specifying one numeric value for each dimension.

Suppose that there is a 3-dimensional data set and this set will be divided into 4 clusters. At first step, the percentages that each data point belongs to each cluster center (membership degrees: u_{ik}) are randomly initialized according to Equation 3.2

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,...,n$$

(3.2)

where c is the number of clusters, n is the number of data points, i is the index of clusters and j is the index of data points.

Then the centers' coordinates c_i are calculated for each cluster, using the Equation 3.3.

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}{}^{m} x_j}{\sum_{j=1}^{n} u_{ij}{}^{m}}$$

(3.3)

where x_j are the coordinates of the data points. According to these center coordinates, the updated membership degrees are calculated, and then the new centers, and so on. This iterative process continues until either a given number of iterations are reached, or until the change in the dissimilarity function is narrowed down to a specified threshold range. The dissimilarity function calculated at each step of the algorithm is given by Equation 3.4.

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{k, x_k \in G_i} d(x_k - c_i)$$

(3.4)

where $d(x_k - c_i)$ is the Euclidian distance between the data point x_k and the cluster center c_i.

When the algorithm stops, each data point has 4 membership percentages for each cluster, summing up to 1. The algorithm is based on the approach to minimize the sum of the Euclidean distances between the data points and their cluster centers.

3.2.4 ANFIS

ANFIS aims to find out the rules between the inputs and the outputs of a data set, so that the unknown outputs can be estimated by giving the inputs to the system. In other words, it is used for obtaining a non-linear mapping from an input space to an output space, by the use of fuzzy sets (Jang 1993). The name "ANFIS" stands for "Adaptive-Network-based Fuzzy Inference System". This system can offer a solution for types of problems where there is a certain set of numerical inputs leading to a certain set of outputs. The non-linear relation between them is tried to be estimated by the procedure explained below.

The procedure is illustrated in Figure 3.3. First, the input characteristics are mapped into input membership functions. Then the input membership functions are used to find out rules, which will lead to a set of output characteristics. The output characteristics are used to derive the output membership functions, and they will be used to obtain a single values output. The adaptive network that is built in this procedure is similar to neural networks. The network, its inputs and its output are clearly illustrated in Figure 3.3. (The figure is taken from (Jang 1993).)
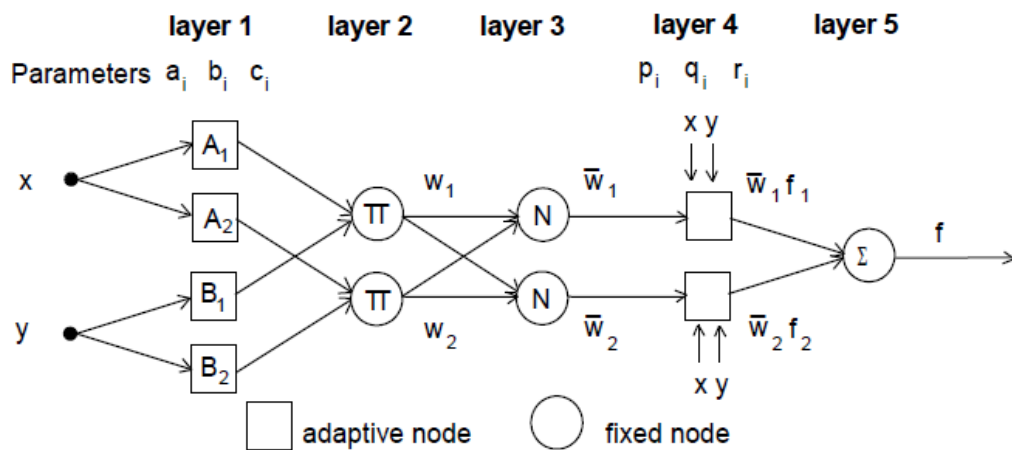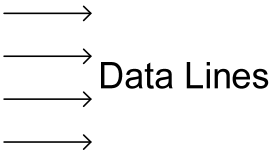

Figure 3.3: ANFIS structure, x and y being the inputs and f being the output

What ANFIS does is basically take the inputs and outputs, and analyze the relationship between them. From a simpler point of view, suppose that we have a system at hand with 3 inputs and 1 output. And we have 1000 instances from this system. That is to say, we know that when [1 2 3] are input to the system [4] is the output, when [1 3 5] are input [7] is the output, and there are 1000 such examples known beforehand. This database will be used to train the ANFIS algorithm to learn the relationship between the inputs and the output. This data is formed into a matrix, so the example data set turns into the form as shown in Figure 3.4 The inputs and outputs are given to the ANFIS training algorithm in that matrix format.

Input-Output Matrix:

Input Columns			Output Column
1	2	3	4
1	3	5	7
...	...	...	...

Data Lines

Figure 3.4: The input-output matrix for ANFIS, for the example data set.

ANFIS builds a network composed of layers of nodes, and trains this network according to the given 1000 examples. At the end of the training, the system learns which inputs might lead to which outputs. Then if the system is given an input, that is a triple like [1 3 4], it estimates an output like [6] for this input set, based on the experience gained during training. It gives similar outputs for similar input sets.

The data set (composed of inputs and outputs) used for the training is called the "training data". The data set composed of only inputs, and which is fed into the system after training is called the "test data". The test data does not include the

output column seen in Figure 3.4, because the outputs are asked from the trained ANFIS model.

The training is done within the ANFIS algorithm, step by step. As the training steps proceed, the system understands the input-output relationship better. At each step, the error between the estimated outputs and the actual output is calculated. And the coefficients of the network are updated so as to reduce this error, iteratively. This iteration can proceed until a specified error is reached or until a given number of training steps (epoch number) is reached. The number of steps of training can be selected by the user.

One might think that the more training steps are taken, the better the relationship is learned. However, it might easily be the case that the number of steps taken for training is too much, so the system is trained for the training data very well but it is not capable of estimating the unknown outputs for the test data any more. So the training should be enough, but not too much. Otherwise the system is "over-trained" and unsuccessful in predicting the outputs for any data other than the training data.

To overcome this problem, the learning algorithms like this one need another data set called the "checking data" for supervising the training process. Checking data is similar to training data. At each training step, the error of the checking data is also calculated, although it is not used for updating the network coefficients. As the system learns the input-output relationship, both the training error and the checking error decrease, but when the system begins over-training, the checking data begins to increase. This step is chosen by the algorithm to be the optimum step to fix the parameters of the network.

Once the parameters of the network are determined, then the test data are given to the system and the outputs are asked.

MATLAB is used for applying the ANFIS method to the specific problem, which is the sales prediction, in this study. In MATLAB, there are the necessary functions for building an ANFIS model (MATLAB 2008). But for using these functions, first of all the inputs and the outputs of the problem has to be stated clearly. The inputs and outputs of the system are fed into the necessary functions in the form of matrices composed of numbers only. The adaptive network and the membership functions are built by MATLAB, depending on several parameters selected by the user. Among these parameters are the number of nodes in the first layer, the maximum number of iterations, etc. The application of the ANFIS method with MATLAB to the problem presented in this study is explained in detail in the next chapter.

3.2.5 Forecast Evaluation Criteria

Root Mean Square Error (RMSE)

If there is a real time series and a prediction series as in the case of this study, then the error between the two series (signals) should be calculated in some way. One of the most common error measures is the "Root Mean Square Error (RMSE)". This is the square root of the sum of the squares of error of each prediction in a time series. It is give by the formula in Equation 3.5:

$$RMSE = \sqrt{mean((Y_t - F_t)^2)}$$

(3.5)

where Y_t is the real time series and F_t is the prediction at instant t.

Mean Absolute Error (MAE)

Another forecast evaluation measure is the "Mean Absolute Error (MAE)" which is the average of the absolute errors between the real series and the prediction and given by Equation 3.6:

$$MAE = mean(|Y_t - F_t|)$$

(3.6)

Both RMSE and MAE have to eliminate the negative sign in the errors, while calculating the average error. RMSE does this by taking the square of the errors, where MAE takes the absolute value of the errors.

Mean Absolute Percentage Error (MAPE)

A very popular forecast evaluation measure is the "Mean Absolute Percentage Error (MAPE)" which is the average of the absolute of percentage errors between the real series and the prediction and given by Equation 3.7:

$$MAPE = mean(|100(Y_t - F_t)/Y_t|)$$

(3.7)

This method differs from the first two, in the way it measures the errors. The first two measures takes the absolute errors into account, independent of the signal value. However MAPE takes the percentage error into account, which is more useful in most cases. The absolute error might not make any sense, when the real value of the signal is not known or when it varies a lot. MAPE satisfies the need of comparing the error with the real value, by taking the percentage error.

These forecast evaluation criteria give important clues about the accuracy of the prediction series. (De Gooijer 2005) It is important to note that RMSE and MAE measures deal with the absolute error, that is, the difference between the

prediction and the real value. However, MAPE deals with the percentage error. If there is a wide range of numbers in the series, then MAPE is usually preferred.

Receiver Operating Characteristics (ROC) Plots

ROC Plots are useful for evaluating the performance of a binary classifier, or a binary prediction (Fawcett 2004, ROC Graph). In binary problems, first of all an event should be defined. The only possible cases are, the event occurs or the event does not occur. As for the predictions, there are again two cases: predicting that the event will occur or predicting that the event will not occur.

According to which combination is valid for a single data point, a 2x2 chart can be naturally formed as shown in Table 3.6.

Table 3.6: The confusion matrix for a binary problem

		Predicted	
		Positive	Negative
Actual	Positive	**TP**	**FN**
	Negative	**FP**	**TN**

The chart in Table 3.6 is called the confusion matrix, and it represents the number of percentage of the correctly predicted events (TP), correctly predicted non-events (TN), falsely predicted events (FP) and falsely predicted non-events (FN).

The event mentioned here can easily be defined as a quantity to be below a limit value. If the predicted value is below the limit value, then the algorithm predicts an event. If the predicted value is above limit value, then the algorithm predicts non-event. Then these predictions are compared with the real event/non-event cases and the confusion matrix is formed. From the confusion matrices, the ROC plots are formed.

The ROC plot is basically the true positive rate (TPR) versus the false positive rate (FPR). These ratios are calculated from the confusion matrix as follows:

$$TPR = TP / (TP + FN) \hspace{6cm} (3.8)$$

$$FPR = FP / (FP + TN) \hspace{6cm} (3.9)$$

In the ROC plot, it is desirable to have the points as close to the point (0,1) as possible. This point is where all the predictions are true positives and there are no false positives. This is a desirable case, so the ROC plot which is closer to the (0,1) point is more successful. It is important to remind that being close to this point also means that the area under the curve has its largest value.

3.3 APPLICATION OF THE METHODS TO SALES PREDICTION PROBLEM

3.3.1 Sales Prediction with Holt-Winters

As mentioned in the previous sections, Holt-Winters is a very convenient method to use with the problems of time-series prediction. The sales prediction of newspapers is also this type of problem therefore in this subsection the application of Holt-Winters Method to sales prediction problem is presented.

The only input this method requires is the sales history of an outlet, in order to predict a future day's sales. The outlets are treated separately, in this method. Therefore the program is run as many times as the number of outlets. For a single outlet, the program works as follows:

First of all, the sales history of an outlet, which is basically a number array, is input to the R-Project statistical software environment. The software calculates three coefficients: α, β and γ for each day. These three coefficients are necessary for calculating the functions give in Equations 3.10 – 3.13.

$$(Level)\ L_t = \alpha * (Y_t - S_{t-s}) + (1 - \alpha) * (L_{t-1} + b_{t-1}) \tag{3.10}$$

$$(Trend)\ b_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * b_{t-1} \tag{3.11}$$

$$(Seasonal)\ S_t = \gamma * (Y_t - L_t) + (1 - \gamma) * S_{t-s} \tag{3.12}$$

$$(Forecast\ for\ period\ m)\ F_t + m = L_t + m * b_t + S_{t+m} \tag{3.13}$$

Here t is the index of the day that is focused on. y_t is the sales for that day. s is the length of the seasonal cycle and is taken as 7 in this study. The seasonal cycle is taken as 7 because a weekly periodicity is expected, due to the nature of the sales data. The functions L_t, b_t and S_t which are written in the first three lines are calculated in order to get a forecast value, from the formula in the fourth line. The R-Project is used to calculate the necessary coefficients α, β and γ for each day, and also the functions L_t, b_t and S_t. Out of these six important values, forecasting is done by using any software environment, using the F_{t+m} function.

It is important to note that the functions L_t, b_t and S_t are functions of the coefficients α, β and γ in addition to previous values of their own. For example, in order to calculate L_t, the software needs to have calculated S_{t-s}, L_{t-1} and b_{t-1}, and similarly for b_t and S_t. The R-Project makes the necessary initializations when starting the calculations of these three functions.

For each day, the functions L, b and S are calculated from the beginning. This is because of the fact that the data changes from day to day. And using the old values, that is the values of L, b and S from the previous days might cause cumulative errors in the end. As the days proceed, the new data are added to the array and for predicting the next day one more data has to be considered. That is to say, for example L53 calculated for the day 54 does not have the same value as L53 calculated for the day 55. So the functions L b and S have to be updated everyday.

The same necessity is valid for the coefficients α, β and γ as well. However, the calculation of these coefficients has quite high computationally complexity. So updating them for each day takes more time than an acceptable amount, with the available computational power. The solution to this problem is found by updating these coefficients once a week. This way, time is gained in return for an acceptable amount of inaccuracy.

3.3.2 Sales Prediction with Linear Regression

Another method that can be successfully applied to sales prediction problem is the linear regression method. For the application of this method, R-Project statistical program is used. The necessary coefficients for forecasting are obtained from this program, given the sales history.

This method, like Holt-Winters method, also considers each outlet separately. So, the sales history given to the program as an input is an array of past sales numbers, until the day that is forecast. The output is the coefficients in the forecast formula shown by Equation (3.11):

$$Y(t+2) = b_0 + b_1 * Y(t+2-7) + b_2 * Y(t) \qquad (3.11)$$

This formula can be interpreted as: today is the day $(t+1)$, the forecast day is tomorrow, so it is day $(t+2)$. The two sales data that is used in this formula belong to yesterday (t) and one week before the forecast day $(t+2-7)$. The b coefficients come from the R-Project.

R-Project analyzes the sales history and calculates the coefficients b_0, b_1, b_2, for each day in a week. This is because of the expected periodicity of 7 days, in the sales data. So the output of the program is a set of three coefficients, for seven days, which makes 21 coefficients in total. As the days proceed, new sales data is added to the input array. So the coefficients have to be updated, and they are updated one a week.

Once the coefficients are obtained from the linear regression analysis, the forecast formula can be applied in any environment that is able to do simple math. As mentioned before, this procedure is applied to each outlet separately so the program is run as many times as the number of outlets, with the relevant sales data.

3.3.3 Sales Prediction with FCM + ANFIS

Sales prediction problem is a very convenient one for using the ANFIS method. Unlike the other methods, all the outlets are treated at the same time, with the ANFIS method. So the sales history and the other observed parameters for all the outlets form a very big and complicated matrix, therefore the computational complexity of the algorithm is high.

As mentioned in the previous chapter, while using the ANFIS method, determination of the inputs and the output is very important. In this section, the inputs and outputs of the system will be explained in detail, and then the application of the method will be presented.

However, before these, the Fuzzy C-Means clustering for the outlets have to be presented. This is because of the fact that the results of the clustering process will be used as inputs in ANFIS.

Therefore, this section begins with presenting the clustering process, then introducing the data set, which is composed of the inputs and the outputs. The section will conclude with the application of the method to the data at hand. The results will be presented and discussed in the following chapter.

Clustering the outlets with Fuzzy C-Means

Clustering the outlets before the application of ANFIS is necessary because all outlets do not behave in the same way. Some of them have a high potential of sales, some of them are smaller. There are 378 outlets that observed in this study,

and among these some of them behave similarly to each other. The aim of clustering is to find out the outlets that behave similarly, so that this information will be useful in predicting the future sales of these outlets.

What is known prior to clustering is the sales data of 378 outlets for 378 days (from 01.04.2010 to 13.04.2010). This data is formed into a matrix as shown in Table 3.7. (The actual size of this matrix is 378 x 378 however, due to space limitations only a part of it is shown here.)

Table 3.7: The outlet vs date matrix which is given to MATLAB for Fuzzy C-Means clustering.

DATE

OUTLET	1-Apr-2010	2-Apr-2010	3-Apr-2010	4-Apr-2010	5-Apr-2010	6-Apr-2010	7-Apr-2010	8-Apr-2010	9-Apr-2010	10-Apr-2010	…
2	174	205	175	185	156	200	185	189	207	172	
3	85	99	75	50	100	100	69	87	107	48	
4	43	48	40	27	49	42	44	40	49	45	
5	44	45	72	110	50	39	44	39	45	86	
7	37	41	21	28	20	31	27	21	27	30	
10	23	26	25	35	20	24	21	19	21	24	
12	22	25	28	34	21	22	19	24	23	27	
13	30	30	42	42	32	31	35	24	33	31	
14	21	25	30	24	21	29	28	25	25	29	
15	30	26	27	21	25	18	17	22	22	28	
17	32	30	14	21	35	27	28	27	26	12	
24	17	20	18	18	14	15	15	18	16	15	
26	33	32	44	46	32	34	31	36	34	40	
…											...

The outlet vs date sales matrix is input to Fuzzy C-Means function (fcm.m) in MATLAB. The algorithm considers each outlet is a data point with 378 dimensions, since there are 378 sales data in total for each outlet. The desired number of clusters is also input to the algorithm and chosen to be 3 for this

problem. The output of this algorithm is 3 clusters of outlets, each having similar sales rates. The cluster number will be an important parameter for an outlet, in the prediction phase. The results of clustering will be used as input in the ANFIS algorithm. The cluster numbers of the outlets are shown in Table 3.8. In this table, the first column shows the outlet code, the second column is the cluster number given by the FCM algorithm and the third column is the average sales of that outlet. The average sales data for each outlet is observed to be coherent when the cluster numbers are considered. (Although there are 378 clusters in total, only a part of them are shown here.)

Table 3.8:   The outlets and their cluster numbers after Fuzzy C-Means clustering.

OUTLET	CLUSTER NUMBER	OUTLET SALES AVERAGE
2	1	180.775132
3	2	81.7857143
4	2	38.9365079
35	3	10.0026455
36	2	29.1560847
125	3	7.22486772
127	2	44.7460317
369	2	89.7751323
503	3	3.7010582
751	2	56.7328042
3517	3	7
3559	2	33.9206349
4018	3	6.21957672
8891	2	63.7566138
8930	1	116.380952
8966	2	34.1375661
8983	1	223.320106
9009	2	83.9920635
9026	3	24.2936508
9042	1	257.082011
...		

Cluster 1 is composed of 6 outlets, whose average sales are between the interval [116.3810 - 257.0820]. Cluster 2 is composed of 80 outlets whose average sales are between the interval [28.6508 - 111.0079]. Cluster 3 is composed of 292

outlets whose average sales are between the interval [1.1614 - 28.8942]. This data is illustrated in Table 3.9.

Table 3.9: The properties of clusters resulting from FCM

	Number of Outlets	Number of Records	Min. Average Sales Value	Max. Average Sales Value
Cluster 1	6	6x378 = 2268	116.3810	257.0820
Cluster 2	80	80x378 = 30240	28.6508	111.0079
Cluster 3	292	292x378 = 11036	1.1614	28.8942

Inputs and Outputs of ANFIS

As explained in the dataset section, the sales depend on many factors like the weather conditions, the promotion type, the day of the week, whether it is holiday or not, and also on the outlet itself. There are many inputs to the system, but only one output and it is the sales of the day. The input-output matrix prepared for the sales prediction problem for this study is given in Table 3.10.

Table 3.10: The input-output matrix of the sales prediction problem for ANFIS

Columns: 1	2	3	4	5	6	7	8	9
08-Apr-10	3517	1	0	2	4	3	7	7
08-Apr-10	3559	1	0	2	4	2	29.57143	32
08-Apr-10	3622	1	0	2	4	3	14.71429	14
08-Apr-10	3638	1	0	2	4	3	21.42857	15

In Table 3.10 the columns that represent the inputs of the algorithm are:

3: weather condition,

4: national holiday or not,

5: promotion information,

6: day of the week,

7: cluster number of the outlet,

8: average sales for the past week.

The output is the 9th column; it is the sales data for the corresponding day and the outlet.

Table 3.10 can be interpreted as "for each data line, when the first 7 values are the input, the 9th value is the output". What ANFIS uses for training is this information, as many times as the number of data lines. The more data lines, the better the algorithm learns the relation between the inputs and the output.

Running the ANFIS Algorithm with MATLAB

The input-output matrix was explained previously. Forming this matrix is essential for the success of the algorithm. Once the input-output matrix is formed, the first thing to do is to determine the training data, the checking data and the testing data.

The training data is used to train the network structure that is formed within the ANFIS algorithm, for the data set at hand. If the checking data accompanies the training process, the results are better. From the input-output matrix, the lines which will be a part of the training data and those which will be a part of the checking data are selected randomly. This selection is done randomly because a homogeneous training data is required. If training data is composed of only a group of data lines that are very similar to each other, then the training process is not expected to be very successful for various testing data sets chosen from the same input-output matrix.

When the training data and the checking data are selected, the rest of the input-output matrix is used as testing data. The most important issue with the testing data is that it does not include the output column, which is the last column of the input-output matrix.

Table 3.11: ANFIS parameters

Range of Influence	0.5
Number of Epochs	100
Training Error Goal	0
Initial Step Size	0.01
Step Size Decrease Rate	0.9
Step Size Increase Rate	1.1

For training and testing, the MATLAB functions designed for the steps of ANFIS algorithm are used. Among these functions are genfis2, anfis and evalfis. The first one analyzes the training data, the second one combines this information with the checking data set and the last one is used for getting the predictions for the test data. During this process, the algorithm parameters are selected as listed in Table 3.11.

The ANFIS algorithm eventually calculates the membership functions for each input, and also the training and checking errors for the training process. The membership functions are used for testing, in which the output column is not given to the algorithm, but asked from it as the prediction. The results of the algorithm will be presented in the next chapter.

3.3.4 Correcting the Predictions According to Standard Deviation

After the predictions are obtained, it is useful to check the predictions, to see if they are reasonable or not. For example, if they contain a negative value, this is not a realistic prediction, since the minimum sales can be zero. In this case, replacing the negative values with zero can be a useful strategy.

In order to understand if the predictions are reasonable, the standard deviations are used in this study. The standard deviation is a statistical measure that implies how far the data points can go away from the mean value. This value can be calculated easily using any software. It is given by Equation 3.14 as follows, where Equation 3.15 shows the calculation of the mean.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

(3.14)

where,

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

(3.15)

If a data set has normal distribution, then nearly % 95 of its members are within a range determined by the mean value and the standard deviation. This range can be formulated as follows:

$(\mu - 2* \sigma) <$ a selected data point $< (\mu + 2* \sigma)$

(3.16)

where μ is the mean and σ is the standard deviation. It is statistically correct to assume that this formula holds for 95% of the data. So the predictions should not exceed $(\mu + 2* \sigma)$, and they should not be smaller than $(\mu - 2* \sigma)$.

All the sales predictions obtained for this study, with all three methods are checked and corrected if necessary, according to this principle.

3.3.5 Comparison of the Predictions with Different Methods

When comparing the performance of different methods, the forecast evaluation criteria are applied. They measures used in this study are RMSE, MAE, MAPE and ROC plots. The calculation of RMSE, MAE and MAPE were previously presented. They are statistical formulas and their application is rather simple: the numerical cumulative error values for each method are compared. However for obtaining the ROC plots for the data at hand, the problem needs to be re-defined in order to fit in the ROC plot calculations.

Binary Sales Problem for the ROC Plot

Predicting the sales is important for many reasons which were discussed in the introduction chapter. Now it is time to emphasize another aspect of predicting the sales. The sales prediction problem can also be considered as predicting if the sales will be above or below a limit for a given day and an outlet. This is also an important problem, since precautions can be taken if a low sales number is expected for a future day.

In this subsection, the sales prediction problem will be defined as a binary problem, and the ROC plots regarding this problem will be presented in the findings section. In binary problems, first of all an event should be defined.

For this case, the event is defined as going below the average sales value for an outlet. Therefore, the average sales data is the limit that will be compared with the sales predictions. If the predicted sales number is below the sales average, then the algorithm predicts an event. If the predicted sales number is above the average sales, then the algorithm predicts non-event. Then these predictions are compared with the real event/non-event cases and the confusion matrix is formed for each outlet, separately. From the confusion matrices, the ROC plots are formed and presented in the findings section.

4. FINDINGS

In this chapter, the results of the application of the methods mentioned in the previous chapter are illustrated with comparative plots. Since the data set is large, presenting all the prediction data for all outlets is not possible. The prediction with the three methods and the actual sales data will be presented for the selected outlets.

4.1 ANFIS OUTPUTS

The ANFIS algorithm was run once with all the data, including all days and outlets. The training process gives two very important outputs: The training error and the checking error, which are presented in Figure 4.1 and Figure 4.2.
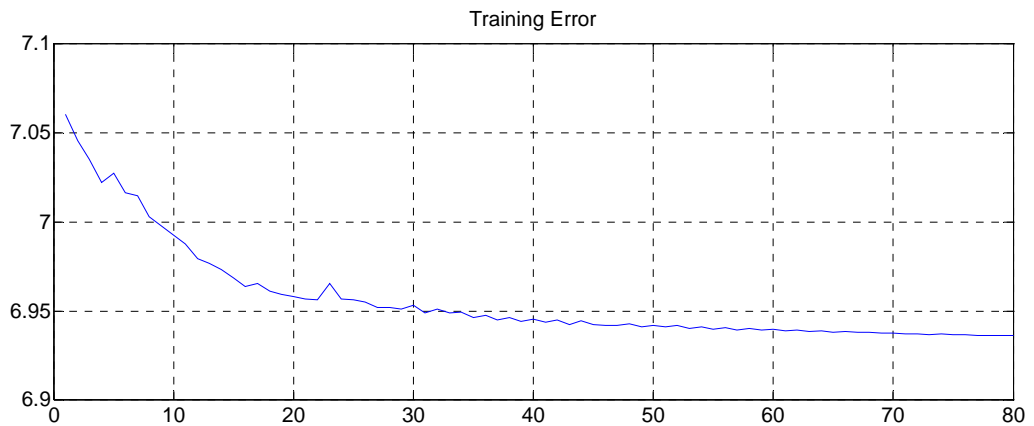

**Figure 4.1: Training error from the ANFIS algorithm.**
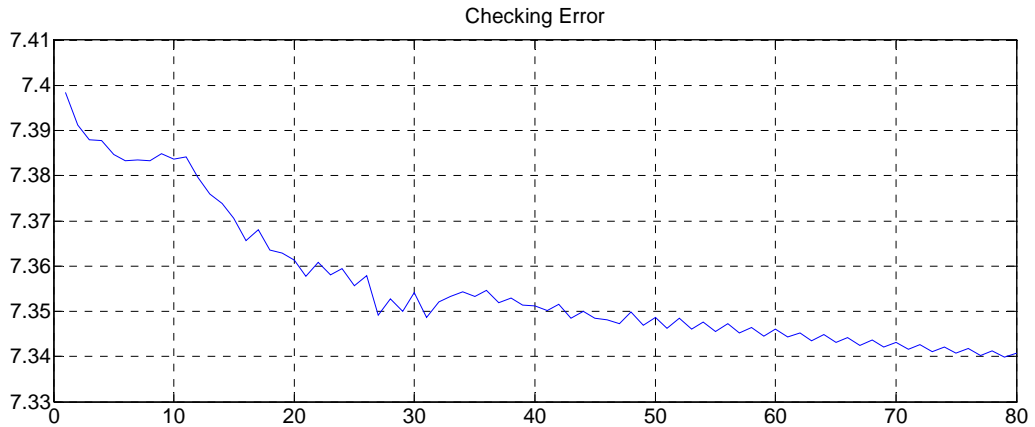
Figure 4.2: Checking error from the ANFIS algorithm.

The horizontal axes for the training and checking error plots is the epoch number, which is the iteration counter for the ANFIS algorithm. At each iteration the coefficients are updated and an error is calculated. These plots show the change in those errors. The iteration, which gives the smallest checking error (in order to prevent over-training), is chosen to be the best one, and the predictions are done using the coefficients at this iteration.

The membership functions of the ANFIS algorithm is given in Figure 4.3 and Figure 4.4.
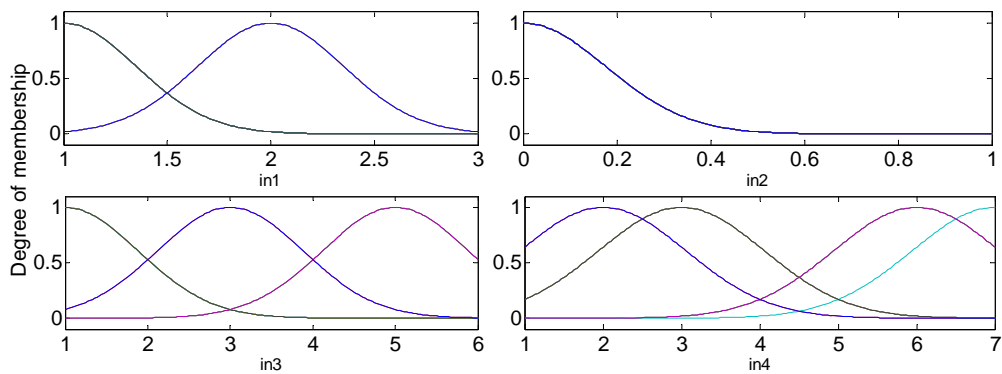
**Figure 4.3: The input membership functions from ANFIS algorithm.**
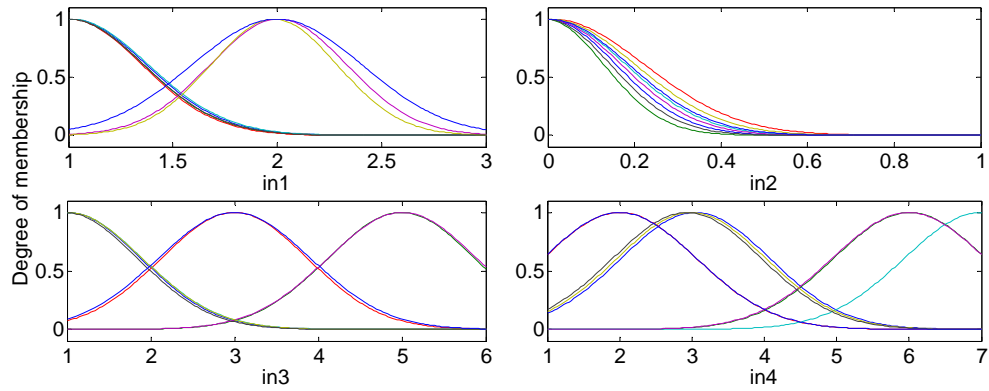
Figure 4.4: The output membership functions from ANFIS algorithm.

4.2 COMPARISON OF THE RESULTS

As mentioned in the precious chapter, the test data for the ANFIS method was chosen from the input-output matrix. Since it was chosen randomly, there are the predictions for every outlet, but not necessarily for consecutive days. In order to visualize the results, the predictions concerning an outlet vs. time are plotted. For example,

Figure 4.5: Comparison of the predictions of the three methods and the actual sales data for Outlet 2, from cluster2.
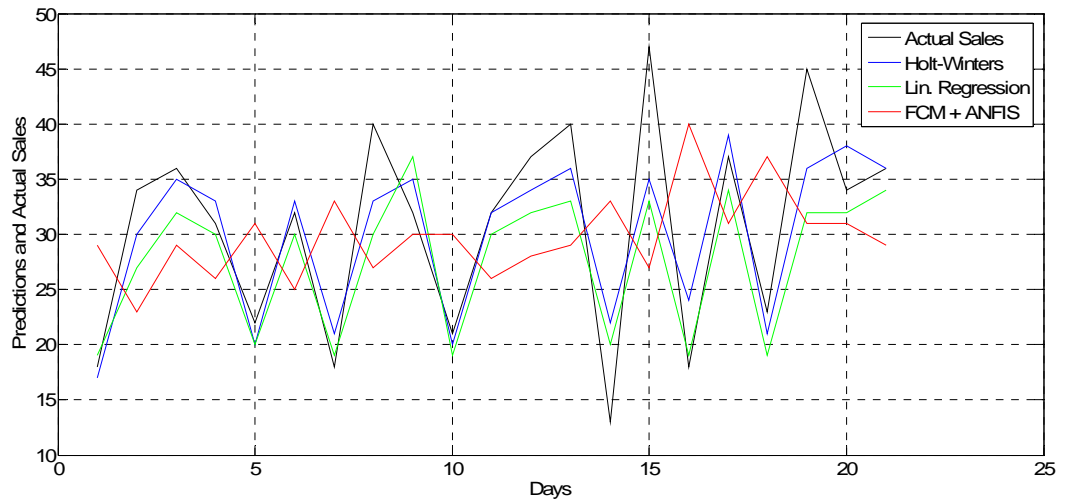
Figure 4.6 and Figure 4.7 show the predictions for the Outlets 2, 13 and 7 from clusters 1, 2 and 3, respectively.
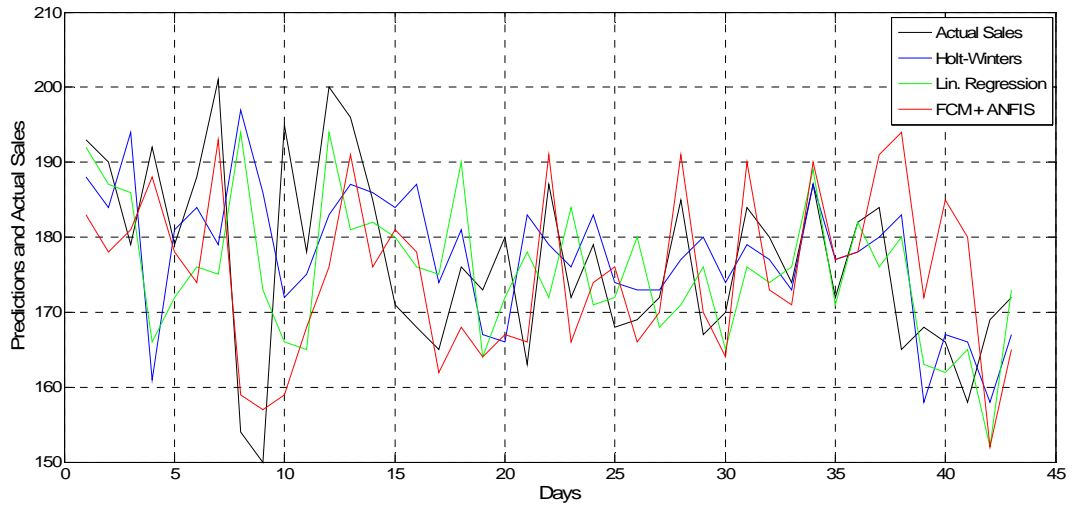
Figure 4.5: Comparison of the predictions of the three methods and the actual sales data for Outlet 2, from cluster2.
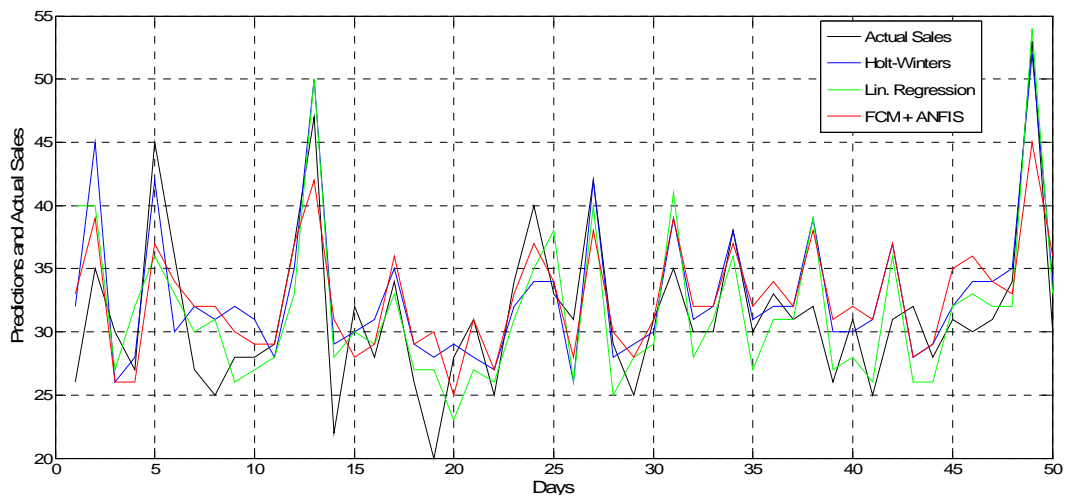
Figure 4.6: Comparison of the predictions of the three methods and the actual sales data for Outlet 13, from cluster 2.

38

Figure 4.7: Comparison of the predictions of the three methods and the actual sales data for Outlet 7, from cluster 3.

4.3 CHOOSING THE BEST METHOD

In this section, the predictions obtained by three different methods are compared using the forecast evaluation criteria. These are the RMSE, MAE and MAPE values and the ROC plots. For each outlet, one method is better among the three. The best method for each outlet will eventually be selected according to the MAPE values.

According to statistical analyses among all outlets in all three clusters, the following results are obtained. First of all, it is useful to remind that the 378 outlets were divided into 3 clusters by FCM, the first cluster having the highest potential customers, and the third the lowest.

Since there are three prediction values for each outlet, a statistical measure to compare these three is necessary. In statistics, the root-mean-square-error (RMSE) values are important indicators for measuring the similarity of two signals. While comparing the actual sales and the prediction, RMSE calculates the square root of the mean of the squared errors of each corresponding pair. The

39

RMSE values for each of the methods for the three example outlets are shown in Table 4.1.

Table 4.1: RMSE values for the predictions of the three methods, for the example outlets.

	Outlet 2 (Cluster 1)	Outlet 13 (Cluster 2)	Outlet 7 (Cluster 3)
Holt-Winters	13.4624	4.0820	5.3379
Lin. Regression	14.0464	4.3402	5.7629
ANFIS	13.5031	4.9232	9.1605

The RMSE values of each method are calculated for each cluster. This data is illustrated in Figure 4.8.
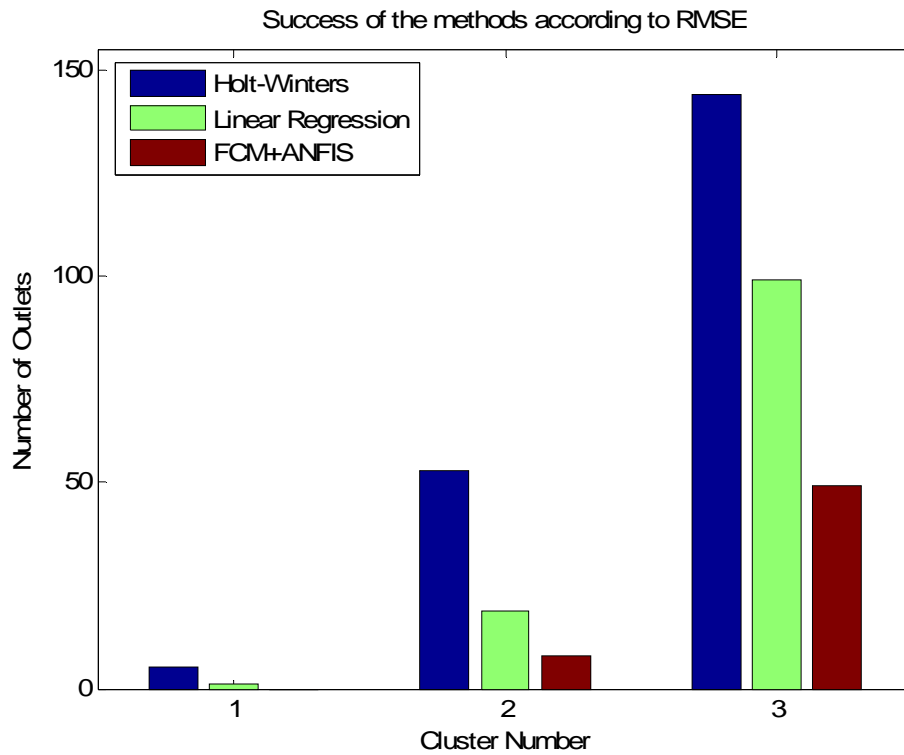
Figure 4.8: Comparison of the outlet numbers for the three methods according to RMSE measure, in all three clusters.

The data illustrated in Figure 4.8 is tabulated with amount and percentages in Table 4.2.

Table 4.2: **Comparison of the outlet numbers for the three methods according to RMSE measure, in all three clusters.**

	Cluster 1		Cluster 2		Cluster 3		Total	
	Amount	Percentage	Amount	Percentage	Amount	Percentage	Amount	Percentage
Holt-Win.	5	83,3%	53	66,3%	144	49,3%	202	53,4%
Lin. Reg.	1	16,7%	19	23,8%	99	33,9%	119	31,5%
FCM+ANFIS	0	0,0%	8	10,0%	49	16,8%	57	15,1%
Total	6	100,0%	80	100,0%	292	100,0%	378	100,0%

Similarly, the same comparison is made with the other statistical measures and their results are illustrated in Figure 4.9 and Table 4.2, Figure 4.10 and Table 4.3
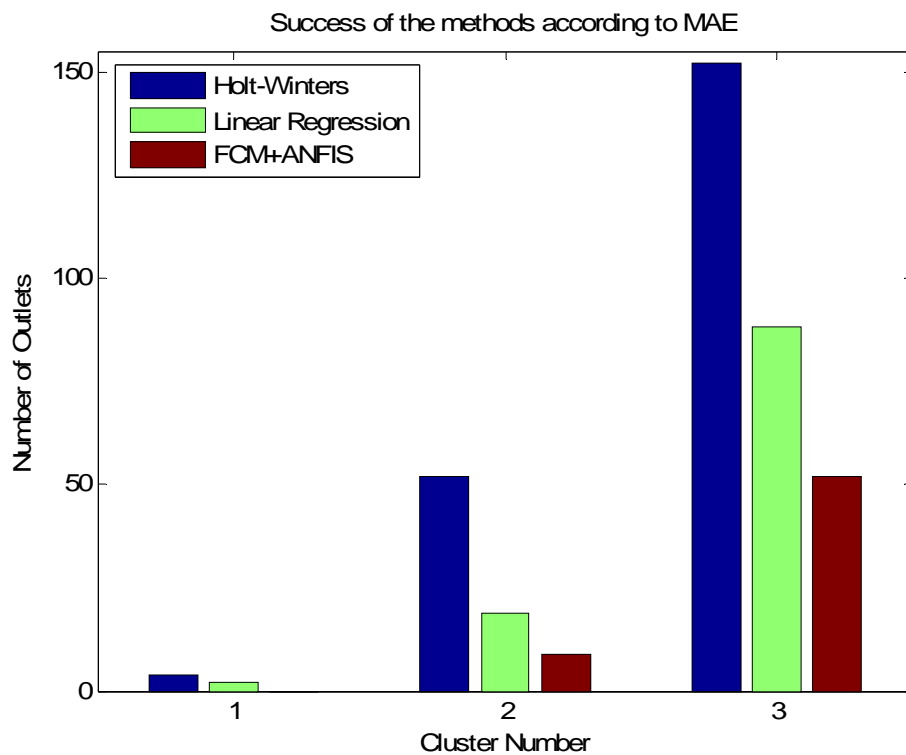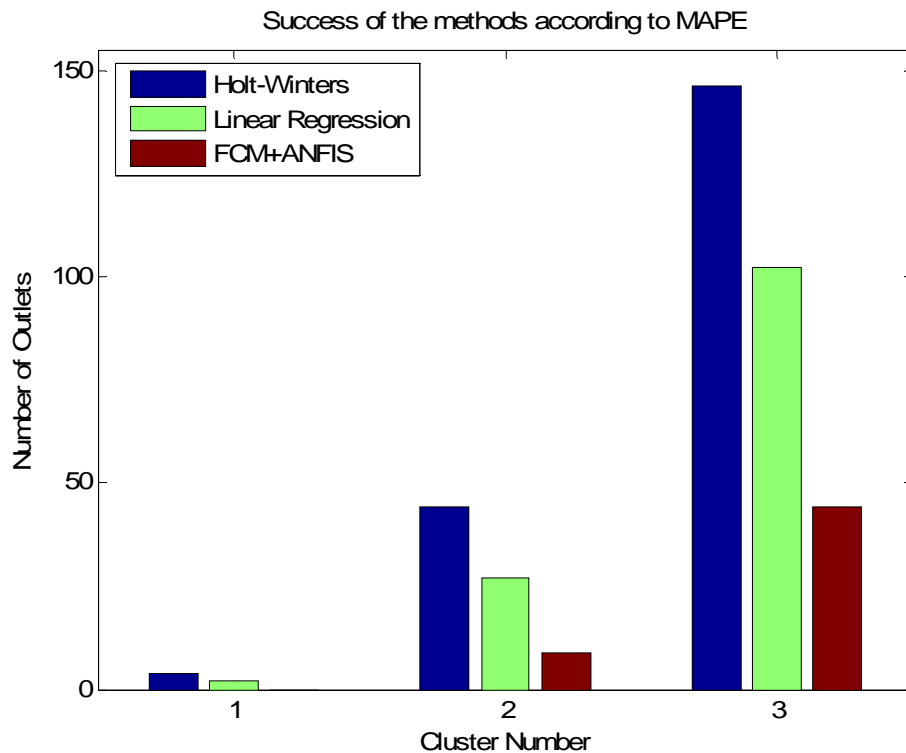

Figure 4.9: **Comparison of the outlet numbers for the three methods according to MAE measure, in all three clusters.**

The data illustrated in Figure 4.9 is tabulated with amount and percentages in Table 4.3.

Table 4.3:   Comparison of the outlet numbers for the three methods according to MAE measure, in all three clusters

	Cluster 1		Cluster 2		Cluster 3		Total	
	Amount	Percentage	Amount	Percentage	Amount	Percentage	Amount	Percentage
Holt-Win.	4	66,7%	52	65,0%	152	52,1%	208	55,0%
Lin. Reg.	2	33,3%	19	23,8%	88	30,1%	109	28,8%
FCM+ANFIS	0	0,0%	9	11,3%	52	17,8%	61	16,1%
Total	6	100,0%	80	100,0%	292	100,0%	378	100,0%

When the clusters are observed, there seems to be one best method: the Holt-Winters method. However, it is important to note that his is only a general view. When the outlets are observed separately, it is seen that different methods are better for different outlets. This is the reason why the forecast evaluation measures are necessary.


Figure 4.10: Comparison of the outlet numbers for the three methods according to MAPE measure, in all three clusters

The data illustrated in Figure 4.10 is tabulated with amount and percentages in Table 4.4.

Table 4.4:   Comparison of the outlet numbers for the three methods according to MAPE measure, in all three clusters

	Cluster 1		Cluster 2		Cluster 3		Total	
	Amount	Percentage	Amount	Percentage	Amount	Percentage	Amount	Percentage
Holt-Win.	5	83,3%	53	66,3%	144	49,3%	202	53,4%
Lin. Reg.	1	16,7%	19	23,8%	99	33,9%	119	31,5%
FCM+ANFIS	0	0,0%	8	10,0%	49	16,8%	57	15,1%
Total	6	100,0%	80	100,0%	292	100,0%	378	100,0%

Also the ROC plots for the three different methods for three example outlets are produced, and they are presented below.
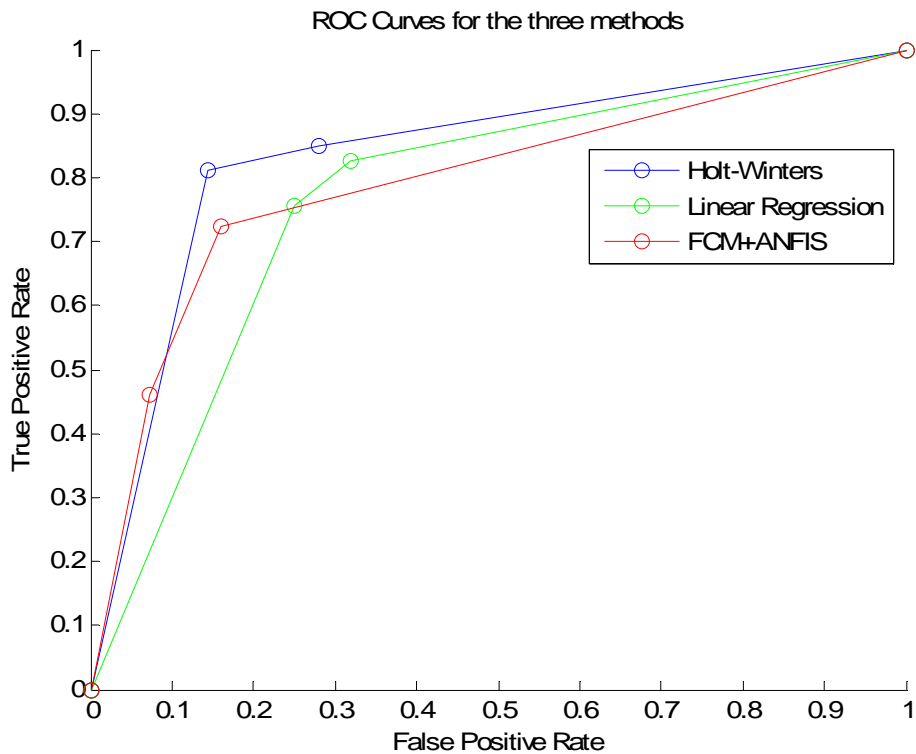
Figure 4.11: The ROC plots for the three methods, for the Outlet 5249.

43

Since there are three different methods used for each cluster, the ROC graphs include three different plots, one for each method. An example ROC graph is shown for an outlet in in Figure 4.11. The figure shows that this outlet is best modeled by the Holt-Winters method.
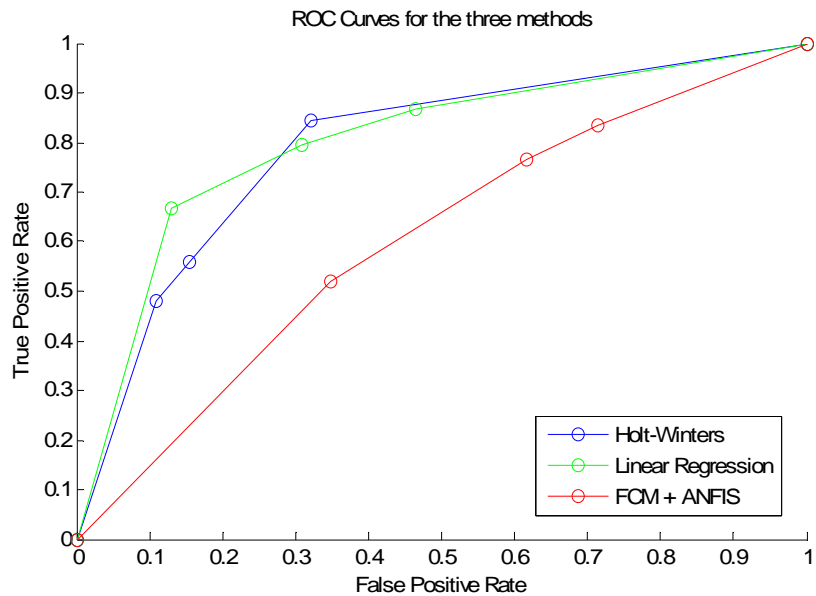
Figure 4.12: The ROC plots for the three methods, for the Outlet 4957.

Similarly, the ROC plots for two other outlets are presented in Figure 4.12 and Figure 4.13.
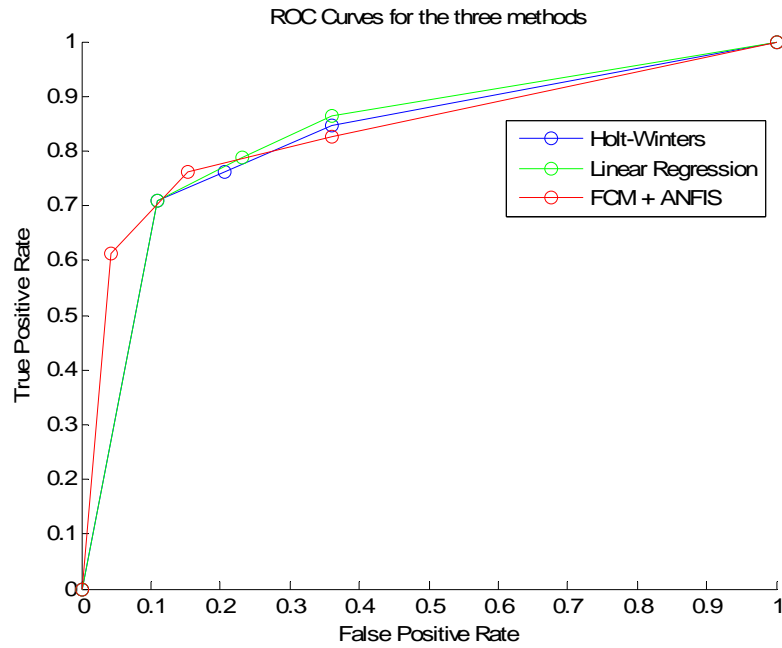
Figure 4.13: The ROC plots for the three methods, for the Outlet 109.

As mentioned before, there are 378 outlets subject to this study. The ROC plot for each outlet is generated and the successes of the methods are compared according to the ROC plots. A comparison similar to the ones presented previously for the other statistical performance measures is given below in Figure 4.14 and in Table 4.5, for the ROC plots.

Success of the methods according to ROC
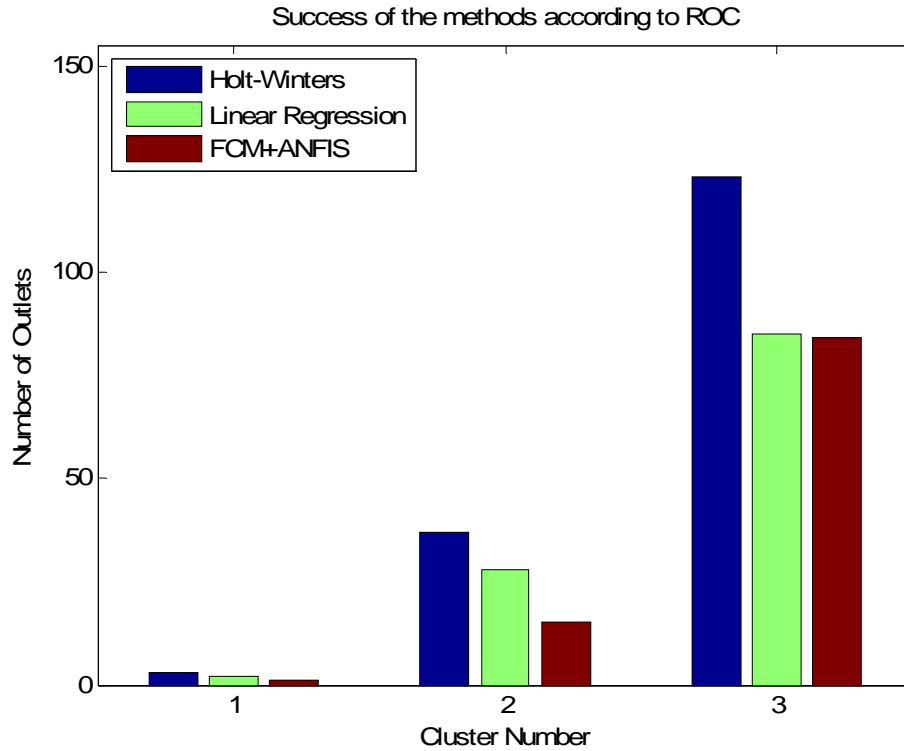
Figure 4.14: **Comparison of the outlet numbers for the three methods according to ROC plots, in all three clusters**

The data illustrated in Figure 4.15 is tabulated with amount and percentages in Table 4.5

Table 4.5: **Comparison of the outlet numbers for the three methods according to ROC plots, in all three clusters**

	Cluster 1		Cluster 2		Cluster 3		Total	
	Amount	Percentage	Amount	Percentage	Amount	Percentage	Amount	Percentage
Holt-Win.	3	50,0%	37	46,3%	123	42,1%	163	43,1%
Lin. Reg.	2	33,3%	28	35,0%	85	29,1%	115	30,4%
FCM+ANFIS	1	16,7%	15	18,8%	84	28,8%	100	26,5%
Total	6	100,0%	80	100,0%	292	100,0%	378	100,0%

For each outlet, the best method is selected by looking at the MAPE measures. MAPE is chosen to be a more reliable measure in this study, since it considers the percentage error. All the outlets have different range of minimum and maximum sales. An error of 3 newspapers may not change much for an outlet whose average sales is 180. However the same 3 newspaper difference is quite a large

number, when the subject is an outlet with an average of 7 newspapers a day. This is why the MAPE measure is chosen to be the critical measure for choosing the best method for each outlet.

MAPE calculations for the Outlets, for the three methods are presented in Table 4.6. In the last column of this table, 1 represents Holt-Winters, 2 represents Linear Regression, and 3 represents the ANFIS methods.

Table 4.6: MAPE values for the outlets, for the three methods.

Outlet Code	Cluster No	MAPE with Holt-Win.	MAPE with Lin. Regression.	MAPE with ANFIS.	Best method
2	1	0.05521532	0.05834955	0.05859926	1
3	2	0.21183901	0.17695941	0.40752615	2
4	2	0.25023368	0.23146276	0.50003892	2
5	2	0.08935867	0.08576665	0.26082078	2
7	3	0.18054032	0.17852632	0.34143797	2
10	3	0.11169673	0.1018921	0.11020088	2
12	3	0.1081267	0.11451641	0.14779433	1
13	2	0.11813921	0.11794806	0.12630137	2
14	3	0.10561709	0.10657316	0.10095575	3
15	3	0.20670439	0.19036789	0.19667304	2
17	3	0.18083318	0.19242803	0.28027858	1
24	3	0.2046096	0.19926718	0.18132068	3
26	2	0.0779608	0.08712871	0.09464922	1
27	3	0.22282555	0.20083365	0.2466849	2
30	3	0.4827556	0.55579482	0.69572829	1
32	2	0.09366335	0.0853241	0.08690855	2
33	2	0.14663002	0.15046759	0.25800588	1
35	3	0.11848719	0.13387471	0.13660757	1
36	2	0.0979819	0.11392234	0.11126439	1
...	...	...	...	...	...

5. DISCUSSION AND CONCLUSIONS

One of the most important problems in newspaper distribution is predicting the number of newspapers that will be sold. Increasing the sales is the ultimate goal; however, forecasting the sales is crucial as well. This is necessary for decreasing the unsold newspapers, and also preventing the sell-outs.

In this study, sales prediction problem is focused on. It is an everyday problem for the newspaper companies. It is very important and the solution can lead to saving big amounts for the newspaper companies.

Sales prediction is basically a time-series prediction problem, and it has been attempted with several methods in the literature. This study applies three of the convenient methods: Holt-Winters, Linear Regression and Adaptive-Network Fuzzy Inference System with Fuzzy C-Means Clustering (FCM + ANFIS).

The given data is the sales history of 378 outlets for 378 days. The desired output is the future sales predictions. The three mentioned methods were investigated and adapted to this problem in this study. The results of these methods were presented and compared in the findings chapter.

It is observed that certain methods worked better for some outlets. The Holt-Winters and the Linear Regression methods generally give similar results, since their approach is more similar to each other, than ANFIS. These two methods are simple and successful. However, ANFIS can make better predictions than the other two, depending on the date and the outlet.

Holt-Winters and the Linear Regression are simple methods that are easy to implement, however ANFIS allows more inputs and parameters to be taken into account.

The first two methods were rather simple, compared to ANFIS, because ANFIS works with the adaptive networks, with a lot more coefficients and functions to optimize. Another important reason why ANFIS was more complicated was that more inputs were used with the ANFIS method. The Holt-Winters and Linear Regression only took the sales history as input, but ANFIS also took the weather information, the promotion information, the day of the week… etc. This is why the input-output matrix was also much larger and complex.

Holt-Winters and Linear Regression methods only take the sales history as input, so they are better for predicting more stable outlets, in terms of sales. If the customers of an outlet are more or less the same every day, and if the sales are not affected by many parameters and the sales are generally stable, then these two methods generate more accurate prediction results compared to FCM+ANFIS. However, the outlets which are severely affected by the additional parameters can be better predicted with ANFIS. For example, an outlet may be in a ferry-boat and the ferry-boat's schedule is highly dependent on weather conditions. So sunny days, this outlet might sell much more than a stormy day in which it might sell zero newspapers. ANFIS is better at predicting this type of unstable sales graphs, since it takes many additional effects into account.

In an outlet where the sales are stable, the Holt-Winters and Linear Regression are simple and successful enough. For this type of outlets, the ANFIS needs more computational resources than Holt-Winters and Linear Regression.

It is learned that ANFIS is a powerful method for working with large data matrices. If there are more than 20 input parameters ANFIS method needs high computational infrastructure to complete modeling and prediction processes. To prevent this, the parameters have to be selected carefully at the beginning of ANFIS method. Too much perfection can cost a lot of time.

In the sales prediction problem, a combination of the three methods can be used with an appropriate algorithm to select which prediction is better. There are

forecast evaluation measures that can be applied to time-series predictions and they are also presented in this study. Performance of multiple methods can be compared and evaluated by using these measures. MAPE (Mean Absolute percentage Error) is chosen to be the best measure because it considers the percentage error. Since the sales numbers are varying in a wide range, the comparing the percentage errors rather than the absolute ones is highly preferable.

According to MAPE, some outlets give better results with one of the three methods. With a simple algorithm, the MAPE results of all outlets are calculated and compared, this way the method which is the optimum for each outlet is determined.

REFERENCES

Books

CHATFIELD, C., 2001. Time-Series Forecasting, Chapman and Hall/CRC Press.

MATLAB, 2008. Fuzzy logic toolbox user's guide, The MathWorks, Inc..

Periodicals

ALİ, Ö., Sayın, S., et al., 2009. SKU demand forecasting in the presence of promotions, Expert Systems with Applications, vol.36, no.10, pp.12340 – 12348.

BEZDEK, J, 1981. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, NY, USA.

CARDOSO, G. and GOMIDE, F. A. C., 2007. Newspaper demand prediction and replacement model based on fuzzy clustering and rules., presented at Inf. Sci., pp.4799-4809.

CIPRA, T. and HANZAK, T., 2008. Exponential Smoothing for Irregular Time Series, Kybernetica, 44 (3), pp.385 – 399.

COORAY, T.M.J.A., Analysis and forecasting of multiple seasonal time series models, in Proceeding of the 13th Annual Symposium, pp.12-15, 2007.

DE GOOIJER, Jan G. and HYNDMAN, Rob J., 2005. 25 Years of IIF Time Series Forecasting: A Selective Review.

FAKHARUDIN, A. S., MOHAMAD, M. A., et al., 2009. Newspaper vendor sales prediction using artificial neural networks, International Conference on Education Technology and Computer, pp.339-343.

FAWCETT, T., 2004. ROC Graphs: Notes and Practical Considerations for Researchers, Machine Learning, Kluwer Academic Publishers.

GOODWIN, P., 2010. The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong, Foresight.

HANZAK, T., 2008. Imporved Holt method for irregular time series, Proceedings of WDS, Part I, pp.62 – 67.

HESKES, T., SPANJERS, J., et al, 2003. Optimising newspaper sales using Neural-Bayesian technology. Neural Computing and Applications vol.12, no.3-4, pp.212-219.

JANG, J.-S. R., 1993. ANFIS: Adaptive-network-based fuzzy inference system, IEEE Transactions on Systems Man and Cybernetics, vol.23, no.3, pp.665–685.

KALEKAR, P. S., 2004. Time series Forecasting using Holt-Winters Exponential Smoothing, in Kanwal Rekhi School of Information Technology, Tech. Rep.

LAUER, M., RIEDMILLER, T. et al., 2003. The smaller the better: Comparison of two approaches for sales rate prediction. Advences in Intelligent Data Analysis V. pp.451-461.

McCARTHY, P., LEI, L., 2009. Regional demands for pulp and paper products. Journal of Forest Economics, vol.16 no.2, pp.127-144.

McCLENEGHAN, J. S., 1997. Searching for 'Analysis' in the Southwest: AM vs. PM Newspapers in Four 'Fire Zones', Journal of The Social Science, vol.34 no.1, pp.21-33.

PORTEUS, 2008. The Newsvendor Problem, Building intuition : insights from basic operations management models and principles, edited by Dilip Chhajed and Timothy Lowe, Springer, New York ; London.

RAGG, T., MENZEL, W., et al., 2002. Bayseian learning for sales rate prediction for thousands of retailers, Neurocomputing, vol.43, no.1-4, pp.127-144.

ŞEN, A. B. and KABA, G., 2009. The effects of using leading economic indicatiors in increasing forecast accuracy, Marmara University İİBF Journal, vol.XXVII, no.II, pp.397-411.

WINTERS, P. R., 1960. Forecasting sales by exponentially weighted moving averages, Management Science, vol.6, pp.324-342.

Others

Dahyot, R., 2010. Lecture notes on Time Series,
      http://www.scribd.com/doc/46481836/Dahyot-Course-Time-Series-in-R-2010.

JAISWAL, M. P., 2007. Clustering Blog Information, Master's Projects. Paper
      36. http://scholarworks.sjsu.edu/etd_projects/36.

KALU, E., KAW, A., NGUYEN, C., 2010. Linear Regression,
      http://numericalmethods.eng.usf.edu/topics/linear_regression.html.

CURRICULUM VITAE

Name / Last Name: Barış AKBAŞ

Address: Gültepe Mah. Reyhan Cad.Sekman Sok. No:8/3 K.Çekmece /IST

Place/Date of Birth: İstanbul – 1975

Foreign Languages: English

High School: Kabataş Erkek Lisesi - 1992

University: Kocaeli Uni. - Electronics and Communication Engineering - 1998

Professional Experience:

Ciner Media Group, IT Infrastructure Manager,		2008 – Present
Ciner Media Group, Network & System Engineer,		2005 - 2008
HAVAŞ A.Ş., Network & System Engineer,		2003 – 2005
VİTEL A.Ş:., Network Engineer,		2000 - 2003