

# Churn and Customer Segmentation Analyses with Data Mining Techniques for a Bookstore Company

Ozlem Odabas, Mustafa Cem Kasapbasi

**Abstract:** Data mining, through piles of very large data is the process of obtaining meaningful data. Nowadays, rapidly developing technique. In this technique; data are grouped, classified according to the relationship, the model is created. In the last stage; the generated models reviewed. Impacts of data mining are widely used, one of the areas allocated to the customer analysis and segmentation of customers. In this study, bookstore customer groups and customer of segment showing the tendency to leave are analyzing; campaigns and marketing strategies that are appropriate to the groups identified. Classification techniques are used for Churn Analysis, clustering techniques are used for Customer Segmentation, and then the appropriate model was created. WEKA software was used to determine the model to be created.

**Keywords:** Data Mining, Churn Analysis, Customer Segmentation, Classification, Clustering

## I. INTRODUCTION

We will define the data mining, data that is outbound process information. Large, complex data from the stack, the next step is to provide meaningful estimation of rules or steps and analyzing of the information, whether with a variety of software. Nowadays, with the advancement of technology, competition between companies has increased. Technology, living conditions for reasons such as customers' buying behavior, trends, expectations has changed dramatically. Companies; customer acquisition, customer loyalty and improve the company's profitability as well as thoughts can get with a good customer relationship management. For this, the companies should pay attention to the analysis of churn. Churn, subscriber-based service that provides company within a certain period, the ratio among all subscribers cancel their subscriptions of the ones who called. [5] Company owners in a competitive environment the customer obtain, to respond to their expectations, customer behaviour in order to comply with the concept of Customer Relationship Management to better understand and interpret. This concept, both market share and customer share contained in it should be known. A customer in more than one ways to sell products, gain customer loyalty by establishing mutual good relationships and customer is to try to activate. For this purpose, attract the attention of customers and marketing methods should be determined. Because a customer win, obviously it will cost more than the customer retention. Therefore, the customers at hand is always valuable, should work to retain them. Customer dissatisfaction, increased competition, lower prices and/or better quality that offer the service in the presence of competitors, substitution products, the effects of regulation, are factors that accelerate churn.

**Revised Version Manuscript Received on January 09, 2016.**

**Ozlem Odabas**, Department of Computer Engineering, Istanbul Commerce University, Turkey.

**Mustafa Cem Kasapbasi**, Department of Computer Engineering, Istanbul Commerce University, Turkey.

These situations must be avoided.[5] With the developing technology, now it is possible that the data can be kept in larger databases. When you are prompted to obtain meaningful data from big data enters into the data mining method. Methods and tools in data mining with the data, using appropriately converted into useful information. Depending on the development of technology and for the needs of the age, a rapid increase in the size of the data. Data mining nowadays, the decision-making process that hosts applied in many areas. Used in the fields of data mining; in particular marketing, banking, insurance, medical, telecommunications, shipping and transportation, tourism, stock market, etc. widely used in the industry.

## 1.1. Literature Review

Churn analysis of the topic, let's talk about other work having taken from:

"Data mining in non-structural data analysis of the text and Web Mining" when I examined the article; Web mining, non-structural data and converts it into structural shape analysis of the data that is associated with the web. Text mining, analysis of very large documents. This method is based on churn customer analysis in the telecom sector are reviewed. For this method the decision tree algorithm C5.0 has been used. According to this algorithm, the separation of the tendency of the reason the customer is roaming. As a result, text and web mining of unstructured data, structured as a great success in converting, analyzing become unbendable. Classification and clustering algorithms for customer analysis, was utilized. "Customer analysis and customer segmentation with data mining techniques, a cosmetic brand reserved" when I examined an article about cosmetics for the customers of the firm by analyzing customer group that tends to stay appropriate strategies have been developed. Decision tree algorithm J48, has the highest rate of accuracy. In this study, was applied to a cosmetic company in Turkey for the first time. The results will be an important resource for production engineers. Association analysis in Data Mining, consumer behavior Modeling" in the study, the data were analyzed according to Association rules belonging to a large grocery chain. Customers which product, which they received with the product. According to the purchasing behavior of customers, the market the purpose of increasing the profit margin; shelf arrangements, relationships with each other products, some of its discount campaigns and the work begins. "Customer churn analysis in the telecommunications industry" in the study, a company that operates in the telecom sector, tends to leave the customers that has been identified. For this, logistic regression and a decision tree algorithm for classification was utilized. According to the results, the company to retain customers, aims to develop specific marketing strategies for them. "Artificial Neural

networks and Social Network Analysis with the help of the Turkish Telecommunications Market analysis of Customer Loss" study, loss of customers in the telecommunications industry has been estimated with Artificial Neural Networks. Customer communications network, the risk of loss of customers analysed with the help of social network analysis on the network locations and the effects were investigated. [4]

### II. APPLICATION

#### 1.2. Problem Definition

The purpose of the study, customers are grouping by receive the product, applied service and qualifications; to develop marketing strategies that will appeal to each customer group. Then, tends to buy more and more profitable customers, achieved is to prepare a strategic plan to keep. Bookstores, sends free books to customers purchases. At discounts and special days customers via e-mail are informed. Also bookstores, advertising on Facebook and websites.

#### 1.3. Data Mining Process

Firstly, customers are saved in the database. All data is collected and useless data is cleared. Blank fields in data by assigning the average value meaningful. With a meaningful dataset, campaign management and strategic plans are made to suit customer groups. New marketing strategies are developed for churn customers. The next step, data is cleared from the database that will not impact on the results of the analysis. Bookstore data set, when there is not missing value in the character field; numeric field is located the missing values in the field SPENDING AMOUNT. These missing values by determining the average spending amount, this record by assigning missing values to blank fields was destroyed.

- SELECT SUM(SHOPPING\_AMOUNT)/100 AS AVERAGE\_AMOUNT FROM BOOKSTORE
- UPDATE BOOKSTORE SET SHOPPING\_AMOUNT=AVERAGE\_AMOUNT WHERE SHOPPING\_AMOUNT IS NULL;

After data cleaning and removing missing data steps, data conversions are made. First, date of birth column is converted to age with SQL commands.

- UPDATE BOOKSTORE SET DATE\_OF\_BIRTH = SUBSTR (SYSDATE, 7,11 ) - SUBSTR (AGE, 7,11 )

The next step in creating the apparent age group; new age groups values were set to the colon.

- UPDATE BOOKSTORE SET AGE= '24-35' WHERE AGE <=35 AND AGE >24
- UPDATE BOOKSTORE SET AGE = '36-45' WHERE AGE <=45 AND AGE >35
- UPDATE BOOKSTORE SET AGE = '45+' WHERE AGE >45

Customers' residence regions, for the purpose of more effective and accurate classification is divided into 3 groups. Accordingly; residing in the Anatolian side 1, residing in the European side 2, as non-resident in İstanbul are numbered 3.

- UPDATE BOOKSTORE SET CITY='1' WHERE CITY='Anatolian Continent';

- UPDATE BOOKSTORE SET CITY = '2' WHERE CITY = 'European Continent';
- UPDATE BOOKSTORE SET CITY = '3' WHERE CITY = 'Out-of-Province';

Customers job states are lettered as if ' YES ' is ' Y ', if ' NO ' is ' N '

- UPDATE BOOKSTORE SET JOB\_STATUS='Y' WHERE JOB\_STATUS = 'YES'
- UPDATE BOOKSTORE SET JOB\_STATUS='N' WHERE JOB\_STATUS = 'NO'

Customers churn states are lettered as if ' YES ' is ' Y ', if ' NO ' is ' N '

- UPDATE BOOKSTORE SET CHURN = 'Y' WHERE CHURN = 'YES'
- UPDATE BOOKSTORE SET CHURN = 'N' WHERE CHURN = 'NO'

The customer is done shopping, the latest from today's date was determined by subtracting shopping periods. After these periods, ' 0-6 ', ' 7-12 ' and ' 12 + ' are grouped into ' dividing in the periods.

- UPDATE BOOKSTORE SET PERIOD='0-6' WHERE PERIOD <=6 AND PERIOD >0
- UPDATE BOOKSTORE SET PERIOD = '7-12' WHERE PERIOD <=12 AND PERIOD >6
- UPDATE BOOKSTORE SET PERIOD = '12+' WHERE PERIOD >12

Customers are numbered according to their membership type.

- UPDATE BOOKSTORE SET MEMBERSHIP\_TYPE = '1' WHERE MEMBERSHIP\_TYPE='GOLDEN'
- UPDATE BOOKSTORE SET MEMBERSHIP\_TYPE = '2' WHERE MEMBERSHIP\_TYPE='PLATINUM'
- UPDATE BOOKSTORE SET MEMBERSHIP\_TYPE = '3' WHERE MEMBERSHIP\_TYPE='SILVER'

Customers, receive free books according to the state, lettered this area.

- UPDATE BOOKSTORE SET FREE\_BOOK = 'Y' WHERE FREE\_BOOK = 'YES'
- UPDATE BOOKSTORE SET FREE\_BOOK = 'N' WHERE FREE\_BOOK = 'NO'
- UPDATE BOOKSTORE SET FREE\_BOOK = 'S' WHERE FREE\_BOOK = 'SOMETIMES'

The types of books that have been received by customers categorized by numbered .

- UPDATE BOOKSTORE SET BOOK\_TYPE = '1' WHERE BOOK\_TYPE='Scientific'
- UPDATE BOOKSTORE SET BOOK\_TYPE = '2' WHERE BOOK\_TYPE='Computer'
- UPDATE BOOKSTORE SET BOOK\_TYPE = '3' WHERE BOOK\_TYPE='Medicine'
- UPDATE BOOKSTORE SET BOOK\_TYPE = '4' WHERE BOOK\_TYPE='Pharmacy'
- UPDATE BOOKSTORE SET BOOK\_TYPE = '5' WHERE BOOK\_TYPE='Engineering'
- UPDATE BOOKSTORE SET BOOK\_TYPE = '6'

WHERE BOOK\_TYPE ='Architecture'

Customers shopping in general average expenditure by looking at which they made, are grouped into 3 different according to the value in the range. These values are '100-250', '251-400', '400+'.

- UPDATE BOOKSTORE SET SHOPPING\_AMOUNT='100-250' WHERE SHOPPING\_AMOUNT<=250 AND SHOPPING\_AMOUNT >100
- UPDATE BOOKSTORE SET SHOPPING\_AMOUNT='251-400' WHERE SHOPPING\_AMOUNT<=400 AND SHOPPING\_AMOUNT>250
- UPDATE BOOKSTORE SET SHOPPING\_AMOUNT='400+' WHERE SHOPPING\_AMOUNT >400

ATTRIBUTES	
BEFORE DATA TRANSFORMATION	AFTER DATA TRANSFORMATION
Date of birth	Age (24-35, 36-45, 45+ )
City ( Asian continent, European continent, out-of-province )	City (1,2,3)
Job Status	Job Status (Yes, No)
Churn Status	Churn (Yes, No)
Shopping History	Shopping Period (0-6 , 6-12, 12+ months)
Membership Type(Gold, Platinum, Silver)	Üyelik Tipi (1, 2, 3)
Free Book (Yes, No, Sometimes)	Free Book (Y, N, S)
Book Type (Scientific, Computer, Medicine, Pharmacy, Engineering, Architecture)	Book Type (1,2,3,4,5,6)
Shopping Amount	Shopping Amount (100-250,251-400,400+)

Table1.Qualification Names

### 1.3.1. Identification of Dataset

Let's examine attributes after the data cleaning and data transformation process. Fields, after the data transformation process, has been categorized and grouped. If I were to describe briefly; the date of birth for all to customers, the resulting age information is divided into 3 different categories as 24-35, 36-45, 45 years. All customers that they are seated regions of; the Asian continent 1, the European Continent 2, Outside the City 3, in 3 different groups that we are categorized. Our costumers' job status, if they work 'Y', if they do not work 'N' shaped two category. All costumers' churn status is categorized two different group that if they registered this bookstore 'Yes', not 'No'. No mean, it still means the bookstore customer. Shopping History, according to their latest shopping time-6 months, 7-12 months, 12 from the longer form is divided into three different groups. Type of membership, Gold is 1, Platinum is 2, Silver is 3 are listed. Free book status, regardless of their shopping from their customers, receive free promotional book is Yes as Y, No as N, Sometimes as S are listed. Types of book, Scientific Books as 1, Computer as 2, Medicine as 3, Pharmacy as 4, Engineering as 5, Architecture as 6 are numbered. Shopping amount is categorized as 100-250, 251-400, 400+, according to the amount of the total shopping.

### 1.3.2. Modeling

Redundant data is cleaned, brought into the next phase is associated with each other. At this stage, different models are tried on the data set and the margin for error minimal model is selected. In this application, to establish the appropriate model, called the 'WEKA', written in Java package program was used. WEKA is a software that can run on different operating systems. In the first step of the modeling; to give up being a bookstore customer and turning to another bookstore that identify for analysis churn costumers. In the second step, to develop and implement marketing strategies in order to attract more customers and different customers, on customer segmentation was studied.

### 1.4. Churn Customer Analysis

The aim of Churn Customer Analysis is define leave customers tendency and those customers get to keep. For that, marketing strategies, plans and programs are develop. You can define the properties of the customers and showing a tendency to leave those customers get to keep marketing strategies, plans, programs to develop. Churn status is identify for main variable. In order to identify a group of churn customers, firstly the classification technique was used. In Weka program, different classification algorithms are tested. At the same time, the reliability of these algorithms were compared. The choice of algorithm is important for high accuracy of the model setup. To test the accuracy of the model, a data set of 100 people on 10 different groups are created. First group is test group, other groups are a learning group. 10-cross-validation test was applied on dataset as one is test group and other one is learning group. Finally, every step of the way by taking the average of the resulting error rates, created for each model error rate is calculated. Algorithms that are used for the purpose of testing validity and error rates of these algorithms, sharpness, F – values of the comparison criteria as shown in Table.

ALGORITHM	NAİVE BAYES	LOGISTIC REGRESSION	ID.3	J.48	One R	Zero R
Correctly Classified as Sample Percentage	80	87	89	90	79	75
Incorrectly classified as Sample Percentage	20	13	7	10	21	25
Kappa Statistic	0.5062	0.675	0.803	0.7403	0.3636	0
Mean Absolute Error	0.1933	0.1551	0.1022	0.1267	0.21	0.3783
Square Root Of The Mean Error	0.3499	0.3658	0.2887	0.3019	0.4583	0.4337

Absolute Error	51.0913	40.999	28.9431	33.5077	55.5172	100
----------------	---------	--------	---------	---------	---------	-----



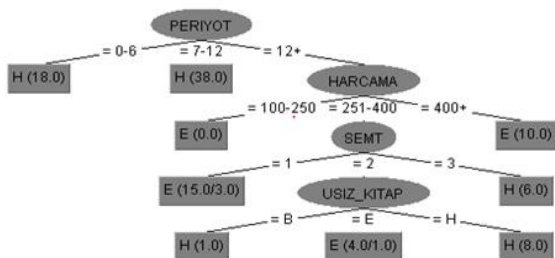
Relative Error Kashyap	80.66 99	84.33 56	69.86 49	69.61 67	105.66 04	100
TP Rate	0.8	0.87	0.927	0.9	0.79	0.75
FP Rate	0.253	0.15	0.086	0.14	0.47	0.75
Sharpness	0.819	0.882	0.932	0.903	0.772	0.563
Sensitivity	0.8	0.87	0.927	0.9	0.79	0.75
F – values	0.807	0.874	0.929	0.901	0.773	0.643
ROC Area	0.915	0.85	0.891	0.915	0.66	0.433

**Table 2. Classification Algorithms And Accuracy**

The values in the table were obtained with the program WEKA. WEKA is an open source software. [2] Licensed. The software on the computer should have Java 1.4 or the latest version installed for it to work. [2] Weka use ARFF (Attribute-Relation File format ) for the variable called that allows you to define ASCII text files. [2] In the first part of the ARFF file, variables, the relationships between each of these variables, and after you have defined the type of each variable is close to the value of the variable. At the beginning of the data related to @DATA. Variables, relationships and data defining a shape of the arff file is as follows:

```
@relation Bookstore
@attribute AGE{24-35,36-45,45+}
@attribute CITY{1,2,3}
@attribute JOB{E,H}
@attribute CHURN{E,H}
@attribute PERIOD{0-6,7-12,12+}
@attribute MEMBERSHIP_TYPE{1,2,3}
@attribute FREE_BOOK {B,E,H}
@attribute BOOK_TYPE {1,2,3,4,5,6}
@attribute AMOUNT{100-250,251-400,400+}
@data
24-35,3,Y,N,0-6,1,N,1,100-250
24-35,3,Y,N,0-6,2,N,3,100-250
```

In WEKA program, Naive Bayes, Logistic Regression, ID3, J.48, One R, Zero R algorithms are applied on dataset. And results are wrote table. As can be seen from the table, the highest accuracy rate (90) classification algorithm is J.48 algorithm. Thus, we have decided that we are going to use algorithm model. J.48 algorithm shall be based on for analysis on churn customer. Each row is a node in the decision tree; the bottom line shows the child nodes. The first of the number specified with parentheses, under nodes, how many had been classified correctly of the event, while the second shows the number of misclassified event.



**Figure 1. J.48 Decision Tree**

According to decision tree, our stages as folloes: shopping period, amount, city and free book. According to the results

of J.48, decision tree algorithm, we can evaluate our data in case of the following ingredients:

1. When we look to the first branch, showing a tendency to leave the customer group, the most recent period of making purchases more than 12 months.
2. Looking to the second branch, showing a tendency to leave customer group, those who spend in the range of \$ 100-250 and the guys who are spending over \$ 400. For not lose customers and customers ' do not prefer to rivals, campaignand strategies are implement. These are:

- Customer birthday, Christmas and special occasions, the customer is the most purchased book from the type of the book can be a gift.
- A gift certificate can be given to the customer to have made by spending.
- During certain periods of the year can be made in beli discount books. This information can be given to customers via e-mail or telephone.
- To increase customer satisfaction according to the customer's spend, stationery should give as a gift to customers.
- To increase customer satisfaction according to the customer's spend, the vision of the film in the DVD should give as a gift to customers.
- To increase customer satisfaction according to the customer's spend, office supplies should give as a gift to customers.
- To increase customer satisfaction according to the customer's spend, customers earn points for every purchase and then spending those points on other purchases by what they have done in a discount or free books, other get free products to use.

3. According to the third branch, spend 250-400 TL is in the range of customers to be effective where they reside as a reason for leaving. "Asia" group of customers residing in the hold should be applied to strategies may include:

- Regular ads and promotions, on the continent of Asia is that despite some school or workplace in front of them or something most people can see public transport is used in places where billboards or posters can be hung using the bookstore of the Clipboard.
- Sitting in the Asian continent customer group, and the appropriate discounts and promotional campaigns.
- Sitting on the Anatolian Side interested in customers who order books, according to the amount of their shopping at discount or free shipping no shipping costs.
- Customers in the Anatolian continent, orders can be delivered more quickly, and the order is compared to a small gift (pen) may be given.

4. According to the last branch 2 numbered with the European continent show a tendency to leave the customer the reason of free books take or not. But, according to the results tend to leave, despite the fact they get free book showing customers are also available. According to these results to be developed strategies are as follows:



- Should be tried in different campaigns for clients that tends to stay despite receiving free books. In addition to free books, stationery, bookmarks, small gifts such as a calendar can be given.
- The customer never sent free book, gift book can be sent once a year according to the state spending.
- Despite receiving free books, the reasons for separation should be examined for churn customers and precautions should be taken serious.

### 1.5. Customer Segmentation Analysis

For Customer Segmentation Analysis, Simple K mean Clustering algorithm was used. In this algorithm, there is not a learning method. Model was learned unsupervised in clustering algorithm. In Simple K Mean algorithm, kvalue, shows the number of identified clusters. [2] We assign a value to the value of k, according to obtain different sets. Every different error rates in clustering.SSE (Sum of Squared Error-SSE) refers to the sum of the squares of the error.The smallest SSE value is selected. Simple K P algorithm can be obtained by using 6 different customer group is located in Table 6.

Simple K Mean	k=2	k=3	k=4	k=5	k=6
SSE	256.0	212.0	207.0	194.0	186.0

Table3. SSE Values

ATTRIBUTES	GROUP1	GROUP2	GROUP3	GROUP4	GROUP5	GROUP6
AGE	36-45	36-45	24-35	24-35	36-45	45+
CITY	1	1	1	1	2	2
JOB	N	Y	Y	Y	Y	Y
CHURN	N	N	N	N	N	Y
PERIOD	0-6	12+	0-6	7-12	7-12	12+
MEMBERSHIP_TYPE	1	1	1	1	1	2
FREE_BOOK	N	N	N	S	S	S
BOOK_TYPE	4	1	1	2	2	2
AMOUNT	100-250	251-400	100-250	100-250	251-400	400+
THE RATIO OF THE TOTAL CUSTOMER	%4	%43	%26	%9	%8	%10

Table 4. Customer Groups

The simple K-Mean algorithm according to the results of the assessment can be made in the following manner:

1. Although first group of customers shopping periods is often, expenditures is low, interest in pharmacy, are not working, and did not avail himself of the opportunity to free book. This group has higher consistent shopping can be applied to the following strategies:
  - 36-45 years is in the range of this group in order to increase their spending; In addition to the book according to the amount of spending on Pharmaceuticals-related magazine. It also sent a free book in a certain period can be increased spending.
  - Gain a method to increase the amount of cross-selling. In this method, the aim is to deliver products that are associated with each other in one package and this package is to provide the customer to receive the discount. By delivering products to customers to make

more purchases in this manner is provided, and in this way, the gain of the bookstore is increased. Pharmacy with, Medical-related book if you take the amount you spend by offering a discount package at a certain rate can be increased.

2. A significant portion of the bookstore profits, offers from 2. group of customers. 6. group's shopping period is long, but it is more than the amount you spend. To retain customers in this group;
  - Customers of 2. and 6. Group are not work. They may be not use Internet often. The interest of this group is to raise even higher and increase the gains; bookstore catalogs, specials products, campaigns, sent via email free of charge customers in certain time zones.
  - This group may not be charge shipping from some shopping.
  - According to this group of customers in the period, free book may be sent as a gift.
3. Third group have a high rate of customers make a purchase, the amount of spending is at intermediate level. The store is the second largest portion of the proceeds from this group.This group has a higher amount of customers to shop will be determined to ensure that the steps are:
  - The average age of customers in this group 24-35. For customers who spend over a certain amount , cinema or a concert ticket can be given. The discount coupons which are valid in certain restaurants and cafes may be granted. This is expected to increase the amount of campaign spending.
  - The cross-selling method is suggestions to increase the amount of the expenditure, we can also apply this group. If informatics and computer books taken with together, a certain amount of spending can be increased by applying a discount rate.
  - The free book can be given this group, per shopping.
4. The fourth group of customers making purchases rate is low. At the same time spending rates are also low. This group of customers, allowing you to shop for higher amounts to be determined in these steps:
  - 24-35 age range of the customers in this group. Young and middle age group covers. Customers who make purchases over a certain amount to concert tickets, movie tickets can be given. The discount coupons which are valid in certain restaurants and cafes may be granted. This campaign is thought to be an effective method for increasing the expenditure.
  - The cross-selling suggestions is increase the amount of the expenditure method, we can also apply to this group. If information books and computer books are taken together, the rate of spending can be increased by applying a certain percentage discount.
5. The fifth group customers shopping periods although not often, the expenditures are high, they work, they sometimes take a free book and they are interested in the computer field. The higher the consistent of this group do the shopping, to ensure that these strategies can be applied:
  - The working group and 36-45 age range, for



increasing spending the amount of shopping and shopping periods make often, according the spending amount, also send magazines about computer with books.

- Computer related seminars, fairs, etc. can be informed with this customer group via e-mail.
6. Store a large portion of its income from the group, the second and sixth. But showing a tendency to leave the sixth group of customers spending high, shopping periods are short. This customer group shopping for durations can be monitored frequently to make certain strategies. These are:
- Applying a discount to all products at certain times, this group the frequency of making purchases is increased.
  - According to the amount of spending for the group over the age of 45, classical-style music CDs can be sent.
  - Free books can be send per shopping.

### III. EVALUATION

Data mining, hidden, noteworthy, previously unknown data to analyze is the technique to produce meaningful results. In other words, the validity of the data and to obtain data that can be applied in dynamic process. In this technique, in addition to the numerical data, non-numeric data to reveal meaningful information with analyses. In this process, artificial neural networks, statistical methods, memory-based methods, and these methods such as decision trees and the methods used and these methods are evolving quickly. Many programs have been developed that support data mining algorithms, WEKA is one of them. In my work, using data mining methods, were grouped customers who subscribe to the bookstore and who prefer to be a member of another company and leave the bookstore, customers or membership allocated from the behaviour of the characteristic features reveal the models that have been developed. Retain customers, to increase the earnings of the bookstore campaign and marketing strategies have been developed. Before you create a certain model data set using SQL commands is clear of unnecessary data. clean data with data conversion, can be used in the data mining process. After this step, different algorithms are tried. Then, this data is modeled by several algorithms. Logistic Regression, Naïve Bayes, ID3, J.48, One and Zero R algorithms are applied in dataset in Weka program. The used algorithms compare according to the accuracy rate. As a result of this comparison, the model with a decision tree algorithm J. 48 that has a high rate of accuracy has been created. In this way, the separated customer behavior is modeled. strategies to retain this customer group have been identified. Secondly, simple K-Mean algorithm is applied on dataset and divided into certain groups of customers. To increase sales, campaign and marketing strategies are improved separately for each different customer group. The rapid development of technology, to the development of mining has prepared the ground and revealed the necessity. [3] The studies with data mining approach, were applied on bookstore customers and the results are noted. Customer acquisition, customer loyalty, customer retention, and the company's earnings is important significantly. This study shed light on other bookstores.

### REFERENCE

1. Rygielski C., Wang J., Yen D., "Data Mining Techniques for Customer Relationship Management", *Technology in Society*, 24:483-502, 2002.
2. Berkay M, Çamur E, Korum, Özkan Ö ve Rzayeva Z, "Veri Kümelerinden Bilgi Keşfi : Veri Madenciliği" Yüksek Lisans Tezi
3. <http://www.gurunlu.com/>
4. <http://ceng.gazi.edu.tr/~ozdemir/teaching/dm/slides/10.DM.TWM.pdf>
5. <http://www.iszekam.net/>