



Universidad de Jaén
Escuela Politécnica Superior de Jaén
Departamento de Informática

Don Pedro José Sánchez Sánchez, tutor del Proyecto Fin de Carrera titulado: “Online Market Customer Analyze on the Social Diffusion Using with C4.5 Algorithm and Decision Tree”, que presenta Aytac Altintas, autoriza su presentación para defensa y evaluación en la Escuela Politécnica Superior de Jaén.

Jaén, MES de 2016

El alumno:

Los tutores:

AYTAC ALTINTAS

PEDRO JOSÉ SÁNCHEZ SÁNCHEZ

Escuela Politécnica Superior de Jaén

PREFACE

I am going to express my gratitude to my supervisor on this project Asst. Assoc. Dr. Esra ŞATIR, for helps me to work on this project. I have learned a great deal during process of the project.

I am going to thank Pedro José Sánchez Sánchez who is the director of the department of computer science of University of Jaén for sacrificing his time and mentoring all in process.

Also for guiding me throughout my whole life education, supporting, most importantly, giving me an awareness of how important my education and work discipline, I would like to express my thanks to my family with my best wishes.

[AYTAC ALTINTAS]
[aa000039]
[110101003014]

INDEX

PREFACE	II
INDEX	III
LIST OF FIGURES	V
LIST OF TABLES	VI
LIST OF ABBREVIATIONS.....	VI
ABSTRACT	VII
1. INTRODUCTION.....	1
1.1 MOTIVATION	1
1.2 OBJECTIVE	2
2. MAIN SECTIONS.....	3
2.1 MACHINE LEARNING.....	3
2.1.1 Decision Tree	4
2.1.2 An Overview of the Implementation of Decision Tree	4
2.1.3 Classification With Decision Tree	5
2.1.4 Elements of Decision Tree	6
2.1.5 Gain Information with Decision Tree.....	6
2.2 DATA MINING.....	7
2.2.1 C4.5 Algorithm	8
2.2.2 Why Used C4.5 Algorithm ?	9
2.3 SOCIAL DIFFUSION	10
2.3.1 What is the Social Diffusion ?	10
2.3.2 Relations between Online Marketing and Social Diffusion	11
2.3.3 Social Diffusion As An Result Attribute	13
2.4 WHAT ARE THE MAIN CHALLENGES ?	13
2.4.1 Online Market Research Tools.....	14
2.4.2 Research Tools and Techniques	15
2.4.3 Major Categories	15
3. USED TOOLS AND METHODS	17
3.1 ANALYZE.....	17
3.1.1 General Entropy & Relative Entropy	17
3.1.2 Data Fetch	17

3.2	DESIGN.....	18
3.2.2	Treeview Usage.....	18
3.3	USED PLATFORM AND PROGRAMMING LANGUAGE.....	20
3.3.1	ASP.NET.....	21
3.3.2	Visual Studio 2012.....	22
3.3.3	SQL Server Management Studio.....	22
3.4	DATASET	23
3.5	IMPLEMENTATION PROCESS.....	25
4.	RESULTS AND DISCUSSION.....	35
	BIBLIOGRAPHY	41

LIST OF FIGURES

Figure 0.1: Annual Growth January 2016 (KEMP, 2016)	VII
Figure 0.2: Internet Use: Regional Overview (KEMP, 2016).....	VIII
Figure 2.1: Entropy Formula	6
Figure 2.2: Relative Entropy Formula.....	7
Figure 2.3: Gain Calculation Formula.....	7
Figure 2.4: C4.5 Formula	8
Figure 3.1: Show Tree and Button	20
Figure 3.2: Dataset With Some Several Rows	24
Figure 3.3: Test Decision Tree in RapidMiner.....	26
Figure 3.4: Database Connection with SQL Server 2012	27
Figure 3.5: Baglan.cs.....	27
Figure 3.6: Some of Stored Procedure in Database.....	28
Figure 3.7: Stored Procedure Automatically Created by Method	28
Figure 3.8: Methods For All Calculatation Process	29
Figure 3.9: Formula of Entropy Method Code Implementation.....	29
Figure 3.10: Control Structure of Inovator Value of Smartphone Sales Attribute.....	29
Figure 3.11: Probability Calculus of Values of Smartphone Sales Attribute.....	30
Figure 3.12: Entropy and GenelEntropy Process in GenelEntropy Method	30
Figure 3.13: Counting of Website Activity Values For Relative Calculation.....	31
Figure 3.14: Relative Calculation of Website Activity	32
Figure 3.15: Create Object From Smartphone Class.....	33
Figure 3.16: Controls of Determine of Root	33
Figure 3.17: Controls of Determine of Frequent Branch.....	34
Figure 4.1: Decision Tree with Treeview.....	36
Figure 4.2: Other Representation of Decision Tree of Work	37

LIST OF TABLES

Table 2.1: Sample Application of C4.5 on Age Attribute	9
Table 2.2: Sample Application of C4.5 on New Age Attribute	10
Table 3.1: Used Programming Language and Platform	20
Table 4.1: Result Values of Smartphone Sale	38

LIST OF ABBREVIATIONS

OMCA	:Online Market Customer Analyze
Sub Value	:Each categorical value of an attribute. For example for a weather attribute subvalues would be rainy, cloudy etc.

ABSTRACT

ONLINE MARKET CUSTOMER ANALYZE ON THE SOCIAL DIFFUSION USING WITH C4.5 ALGORITHM AND DECISION TREE

The "internet" is a gigantic library, as well as a world-wide message board, telephone network, and publishing medium. It is open 24 hours a day, and you can find anything you want there, and say anything you want. (Walker, 2007)

As well as current events and blogs, information about almost any subject is available in depth and up to date. This is incredibly valuable for every subject you can imagine. Almost every college and government research organization is on the web along with libraries, educational institutions, associations, and many commercial directories and sites. If you appreciate the richness of the Web and the Internet, and get the benefits yourself, then you will be better able to provide services on them. You don't have to know how to do those technical things yourself. (Walker, 2007)



Figure 0.1: Annual Growth January 2016 (KEMP, 2016)

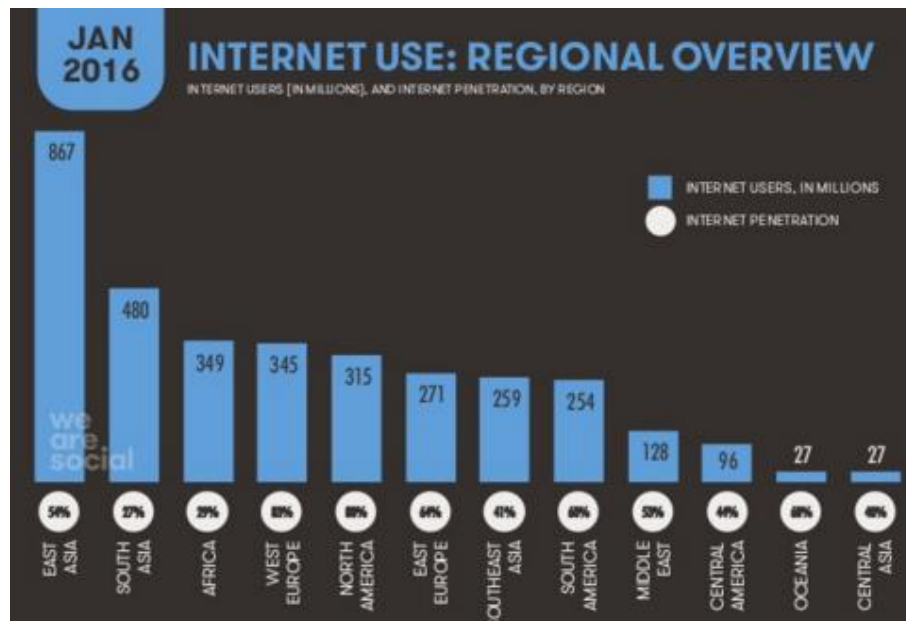


Figure 0.2: Internet Use: Regional Overview (**KEMP, 2016**)

Especially, shopping online is more satisfying for both of customer and vendor.

Because;

The seller store side benefits;

- They can have virtual store which they can their own products and which is accessible anytime, anywhere.
 - Virtual stores have lower cost than a real store.
 - They can provide information about the products in detail.
 - They can sell products according to the demand by keeping it in stock.
 - The possibility of finding new customers all over the world is not just based localization, it is in everywhere. It makes availability more facilitated.
- (Eticaretport, 2016)

Client side benefits;

- Users enables users to shop quickly without leaving the house.
 - With seeing more product types, selecting opportunity is possible.
 - Easier to find needed items.
 - Possible to get feedbacks about the product, thanks to comments by those who use the product will have clearer information.
- (Eticaretport, 2016)

Today, many e-commerce sites on behalf of individuals to carry out wholesale or retail has enabled us to open our own stores. These can be listed the advantages of this;

- Companies can offer customers a branch offices for information on the internet 7/24 opened. Possible to save time due to be transferred within seconds of data on electronic media.
 - More product options are available for customers.
 - Virtual store reduce your financial burden.
 - Because of being cost and stock information in the electronic environment, suppliers do not need to make stock. He/She knows that how much product is in stock and how much of them is sold.
 - It allows supplier to create a virtual shopping mall without need of a large office.
 - In a global environment like the Internet would not be easily to fluctuations, losses is not caused by inflation.
- (Eticaretpor, 2016)

One of the indispensable of Internet world included with many users is online shopping. Normally, amount and kind of sales, gain information from these customers who makes shopping are more important in this case.

1. INTRODUCTION

Today, online shopping has become very popular. For Customers, as much as the seller's performance is important, for seller sale performance is also important. In this case also customer analysis has importance for sellers. The next sales forecast of product makes factors such as market analysis, stock supply visible. This work has made seven different attributes of about six hundred and fifty customer of a technological online store. These attributes are gender, age, marital status, website activity, bought electronic in last 12 months, smartphone sale, method of payment, bought phone. The next sales forecasts are leaded to the conclusion with social diffusion that defines how the spread of knowledge in a community all customers. This analysis was carried out with decision tree which one of the machine learning method. One of the most important features of the decision tree is calculation of general entropy and relative entropy for determine the most distinctive attribute. The attribute which has highest gain is located top of the tree and attributes which have higher gain value are located higher up. In the data set some numerical values of some attributes may not be too critical, they can also have a wide range of values. For example, age value rather than the value of a customer, average age of all customers can be more meaningful. Therefore in some attributes entropy is calculated with using the C4.5 algorithm. In case of the conclusion, customers are examined in four different audiences in four different categories of social diffusion. They are innovator customer, early adopter customer, early majority customer, late majority customer. Implementation of the code is carried out with asp.net. Required information counts are retrieved from database with stored procedure. The tree display is made with TreeView.

1.1 MOTIVATION

In this project what motivates me; making a machine learning and data mining work with used customer in online marketing information. I have passion to realize gain information and resulting of this. Another motivation is being realized code implementation provided with gain information. Commenting on this getting results on my own project.

1.2 OBJECTIVE

The main objectives are ;

- Information extraction using with C4.5 algorithm and decision tree from data which is belong customer.
- Creating rules based on the information, commenting on it. These rules take place after creating attributes as an tree structure.
- Respectively determine some misclassifications in labels, the removal of outliers and noisy data, testing the suitability of the tree structure of the data set in natural language processing tool, namely the creation of the tree using the tool.
- Transferring the data set into SQL database management studio.
- Making connection between database and code.
- Implementation of the stored procedure to return the numeric value of the data required for calculations.
- Code implementation of decision tree calculations and tree structure exceptions to the controls.

2. MAIN SECTIONS

2.1 MACHINE LEARNING

Machine learning is a method of data analysis that automates analytical model building. Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look. The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. It's a science that's not new – but one that's gaining fresh momentum. Because of new computing technologies, machine learning today is not like machine learning of the past. While many machine learning algorithms have been around for a long time, the ability to automatically apply complex mathematical calculations to big data – over and over, faster and faster – is a recent development. (SAS)

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. The process of machine learning is similar to that of data mining. Both systems search through data to look for patterns. However, instead of extracting data for human comprehension -- as is the case in data mining applications -- machine learning uses that data to detect patterns in data and adjust program actions accordingly. Machine learning algorithms are often categorized as being supervised or unsupervised. Supervised algorithms can apply what has been learned in the past to new data. Unsupervised algorithms can draw inferences from datasets. (Rouse, 2016)

2.1.1 Decision Tree

Decision trees are helpful, not only because they are graphics that help you 'see' what you are thinking, but also because making a decision tree requires a systematic, documented thought process. Often, the biggest limitation of our decision making is that we can only select from the known alternatives. Decision trees help formalize the brainstorming process so we can identify more potential solutions. (Hawks)

It is one way to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning. (Wikipedia, 2016)

A decision tree is a graphical representation of possible solutions to a decision based on certain conditions. It's called a decision tree because it starts with a single box (or root), which then branches off into a number of solutions, just like a tree. The decision tree can be linearized into decision rules, where the outcome is the contents of the leaf node, and the conditions along the path form a conjunction in the if clause. In general, the rules have the form: If condition1 and condition2 and condition3 then outcome. Learning, come true with a set of sample. (Wikipedia, 2016)

2.1.2 An Overview of the Implementation of Decision Tree

In retrospect of decision tree, the made outstanding works are implemented with different algorithms. For instance, decision tree is implemented with ID3 algorithm, and fuzzy logic. With machine learning works in order to find the key factor in the business field studies were conducted, some of those; “Application of Decision-Tree Induction Techniques to Personalized Advertisements on Internet Storefronts” and “Using Decision Trees to Identify Key Factors of Keyword Advertisements”. Also datasets are enormous capacity on the current study. Continuously customers and information of customers are increasing on the internet, that's the reason why there is such a big throughput on datasets. Some other remarkable works as follows;

- “Japanese named entity recognition based on a simple rule generator and decision tree learning”,

- “Decision tree analysis of construction fall accidents involving roofers”,
- “Chinese Customers' Evaluation of Travel Website Quality: A Decision-Tree Analysis”,
- “Application of Improved Decision Tree Method based on Rough Set in Building Smart Medical Analysis CRM System”,
- “Predicting financial distress: A comparison of survival analysis and decision tree techniques”,
- “A Comparison of H1N1 and H3N2 Viruses Using Decision Tree and Apriori Algorithm”.

2.1.3 Classification With Decision Tree

Classification is a decomposition according to mutual features. Actually classification is a part of data mining. Also It is possible to applicate with decision tree. But to realize this, required thing is that each data line should be classified in the result attribute. Classification made in this project is implemented through social diffusion. Every customer in the data set is a part of a categorical value of social diffusion attribute – result attribute-.

Classification process has two step:

- Reveal the appropriate model for data sets
- Determine classification rules on the test data

Applicability condition for Decision Trees:

- The examples shown in the form of Attribute-value. A small number of discrete values easily applicable.
- Discrete Output value of the target function is there.
- Training data may contain errors. Both classifications of property in the value of both.
- Training data may contain missing attribute values.

2.1.4 Elements of Decision Tree

Decision tree is a structure which differentiate from other machine learning methods. In this tree structure prior thing is root. In the top of tree root is located. And rules are depended from root to branches and nodes. Elements of decision tree are listed below.

- *Node*: Each attribute is represented by "node".
- *Root*: The top structure , not income branches, 0 or more outcome branches.
- *Leaf*: Latest structure , income 1 branch, but not branching.
- *Internal Nodes*: Only one branch coming toward him before him and from which two or more branches outcome.
- *Branch* : Represents the outcome of the test and each leaf node represents a class label -decision taken after computing all attributes-

2.1.5 Gain Information with Decision Tree

Class qualifications that can be taken in a training set $\{C_1, C_2, \dots, C_k\}$ including divided into classes as “k” units. For class information set containing T the probability distribution of the classes PT:

$$P_T = \left(\frac{|C_1|}{|T|}, \frac{|C_2|}{|T|}, \dots, \frac{|C_k|}{|T|} \right)$$

$$Entropy(T) = H(T) = - \sum_{i=1}^n p_i \log_2 p_i$$

Figure 2.1: Entropy Formula

T representing the target attribute, class attribute value without depending on the nature of X, If divided into subsets like $T_1, T_2, \dots, T_n$; required information for determine the class of an element of T, It is a weighted average of the information necessary to determine the class of an element of T_i is on Figure 2.2 (Mitchell, 1997)

$$H(X, T) = \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i)$$

Figure 2.2: Relative Entropy Formula

In this case the gain criteria:

$$(X, T) = H(T) - H(X, T)$$

Figure 2.3: Gain Calculation Formula

Here provides the highest information gain X test is selected.

2.2 DATA MINING

With each passing day computer systems are getting cheaper and also their power grew. Processors are getting faster, disk capacity is increasing. No longer able to store larger quantities of data in computer and it can handle in less time. Besides, possible advances in computer network and also to reach quickly of these data from another computer. By cheapening of computers, digital technology is used more widely. Data is collected and stored directly in digital. As a result of that we can reach a detailed and accurate information. For example, the cash register used in the supermarket was merely a gathering machine. It was used to calculate the total of goods that have been bought by customer at that time. Today, thanks to point of sale terminals used in place of cash register, all the details of this movement can be kept. Thanks to movement information of these thousands of goods and thousands of customers which is stored, If movement in time of each goods and customer are codified with a customer number, possible to access and analyse data of a customer in time. In the supermarket example, making with

data analyse for each goods sales forecasts of next months are detectable. Customers can be grouped according to the goods which they buy; potential customers for a new product can be determined; with examining the movement of clients, also possible to make prediction about their behaviour. If thousands of goods or customer are expected to be, this analysis can not be done by eye and hand, required making automatically. Data mining is activated here: data mining large amounts of data that will allow us to make predictions about the future relations within the rules and is searchable using computer programs. The future, the near future at least, assuming not too much different from the past, rules issued from historical data will enable us to make accurate predictions for the future and will also be available in the future. Data mining has many fieldwork and methods but in retrospect when articles in which used C4.5 algorithm is examined as related data mining, due to its simplicity and rapidity, possible to decide that C4.5 algorithm is common instead of ID3. Some academic studies using C4.5 algorithm is as following: “Automatic rule generation for protein annotation with the C4. 5 data mining algorithm applied on SWISS-PROT”, “Efficient C4. 5 [classification algorithm]”, “A comparative analysis of methods for pruning decision trees”, “Applying the weak learning framework to understand and improve C4. 5”.

2.2.1 C4.5 Algorithm

The numerical value of some of the attributes in the data set without specifying any category can vary from person to person. In this case, to achieve a more lean content using C4.5 algorithm is possible for us. Also it enables us to provide a more simplified decision. The basic logic of algorithm C4.5 after collecting all the qualities of digital customer value divided by the number of customers.

$$t_i = \frac{v_i + v_{i+1}}{2}$$

Figure 2.4: C4.5 Formula

2.2.2 Why Used C4.5 Algorithm ?

Most of attributes contain categorical value in the dataset. Values which are not categorical are not concluded in terms of gain information and it causes occurring many branches in the implementation, it is not efficient. Some attributes vary person by person and they change. Because of this type attributes, I need to make it categorical using with C4.5 algorithm. Also by reducing the number of branches we can create a tree which has better interpretable capacity thanks to C4.5 algorithm. Sample implementation of “Age” attribute applied C4.5 algorithm is in “Table 2.1” and “Table 2.2”.

Table 2.1: Sample Application of C4.5 on Age Attribute

Age
61
27
29
49
29
60
53
63
46
25

Age Attribute = {61, 27, 29, 49, 29, 60, 53, 63, 46, 25}

$(61+27+29+49+29+60+53+63+46+25)/10=44.2$

If value is lower and equal to 44.2 , we will dedicate L.

If value is bigger than 44.2, we will dedicate B as a new value.

Table 2.2: Sample Application of C4.5 on New Age Attribute

Age Attribute	(New) Age Attribute
61	B
27	L
29	L
49	B
29	L
60	B
53	B
63	B
46	B
25	L

2.3 SOCIAL DIFFUSION

2.3.1 What is the Social Diffusion ?

Management and Technological Innovation, practical applicability is very high, is an interdisciplinary academic and industrial work sites. Adoption and Diffusion of Innovations adapt to context of technological innovation of Social Diffusion theory. Adoption and diffusion of innovations examine adoption of innovation (software, hardware, services, processes, paradigms, etc.) from target institutions or individuals adoption and the dynamics of the spread between target audiences. Particularly US, in many developed countries, Under of this main headline, The total number measured in tens of thousands, a huge volume of research done and numerous practical applications of each theory was realized by private and public institutions.

2.3.2 Relations between Online Marketing and Social Diffusion

Social diffusion offers one of the most robust theories for taking innovations in ideas, behaviors and practice to scale. Social marketing needs to focus on ‘pushing the curve’ of adoption of health practices among individuals, the adoption of effective interventions among practitioners, and the adoption of health-promoting and supportive policies among policy-makers. Adopting diffusion of innovations means that we understand how and why people adopt (or fail to adopt) healthier, environmentally conscious or socially beneficial behaviors. These characteristics include;

- The contextual factors that surround the adoption of new practices and policies,
- People’s perceptions of the innovation as a normative behavior among their reference group or peers (which studies suggest account for anywhere from 50-86% of the variance in the rate of diffusion of new behaviors),
- A risk-benefit analysis that substitutes certainty of outcomes and familiarity with the new behavior over costs and benefits of changing current ones.

Social marketers need to design their behavior, product and service offerings to answer the questions people have that are associated with adoption or termination:

- How is this better than what I currently do? ·
- How is it relevant to the way I go about my everyday life? ·
- Is it simple enough for me to do?·
- Can I try it first?
- Can I watch others and see what happens to them when they do it?

Social marketing for behavior change at scale also needs to explore more extensively the use of audience segments such as:

Innovators: who have a high tolerance of risk; are fascinated with novelty; usually viewed by others in their community as mavericks, not opinion leaders; and whose social networks transcend geographic boundaries.

Early Adopters: who are the community opinion leaders and well-connected socially and locally; have the resources and risk tolerance to try new things; and are the people who are watched by others – and they know it.

Early Majority: the people who are very engaged in local peer networks; rely on personal familiarity before adoption and continually ask the question ‘How does this new behavior, product or service help me?’

Late Majority: the group of people who are most sensitive to peer pressure and norms, yet very cautious about change of any kind; they seek to minimize uncertainty of outcomes and want to see the proof of relative advantage locally – not read about it on web sites or see it on television. (Lefebvre)

There are lessons we have learned over the years that can be used to re-conceptualize social marketing strategy to promote diffusion of product and service use as well as adoption of new behaviors. They are ;

- Find sound innovations/solutions. Ones that meet the criteria noted earlier and are relevant to people’s everyday lives.
- Provide opportunities for Innovators to discover them.
- Engage the curiosity of the Early Adopters.
- Target the weak ties between Innovators and Early Adopters through identifying the ‘boundary spanners’ who interact with multiple networks.
- Promote the work of Early Adopters.
- Create spanable social distances between groups through various means including using social media such as social network sites and blogs and convening meetings of the ‘unlike’ rather than the usual host of agents.
- Enhance the salience and attractiveness of the ‘positive deviants’ – the people who are already practicing the target behavior or using the product and service; put the practitioners of ‘new’ behaviors in contexts and situations that attracts imitation or modeling.
- ‘If you can’t imitate them, don’t copy them.’ Expect and encourage reinvention.
- Support time and energy for discovery, testing, networking, adapting, monitoring and preserving the past. Do not write off the traditionalists, especially when encouraging organizational change.
- It starts with yourselves and your partners.

(Lefebvre)

2.3.3 Social Diffusion As An Result Attribute

Social Diffusion, in the widest sense is a diffusion process of any ideas, attitudes and practices in a social or institutional context. This area is based on applied models of basic social psychology in supports of mathematical model, in context of Change.

Each branch or node provided from our data set is concluded with adoption of social diffusion. If a product is purchased when it is new we call it “inovator”, If a group sense benefit and adopt product it will be called with "Early adoption", a group of audience that occurs when a productivity gain of the product will be called with "Early majority", a group of customer who bought just when they need to take a product called with "Late majority".

2.4 WHAT ARE THE MAIN CHALLENGES ?

Today through use decision tree or other machine learning methods, many online market analyze is made but underlying thing is that make popular and fashionable market analyze because popular and fashinable things have sales force. For instance nowadays technological goods, smartphones has growing popularity and fashion. That is the reason why this work is compatible with the conditions of the day. On the other hand, social diffusion as a theory adds more power of interpretation in terms of mass approach according to online market analyze. Because its classification is depended result attribute. We can say this classification is result of each branch. In previous works but now works related social diffusion continues increasingly. Also for online marketing, know what would be the next sales and predict what kind of inverstment to do are important. This work is a demonstration of implementation of online market through database connection and without using a mining tool. In the past there is no similar work as my project in the internet. Another section is ease of use of mining tools, with less effort, without code implementation, a analyze of any dataset would be realized include decision tree. Considering thing that this work is a code implementation intended to improvement as a prototype.

2.4.1 Online Market Research Tools

The following techniques can be used to gather market information with the help of a few mouse clicks and keystrokes;

- **Keyword Search:** Take that a step farther by searching for "keywords" that people would use to find your type of products or services on the Internet. Keyword searches can also help remind you of product niches that you might not have considered. There are other reasons to conduct keyword searches. 'First, you're going to be reminded of product niches that you might not of thought of.' says Jennifer Laycock, editor-in-chief of Search Engine Guide, an online guide to search engines, portals and directories. 'Second, these services will also give you a guesstimate of how many existing sites already use that phrase,' Laycock continues. 'How many existing sites already offer that product.' WordTracker and Trellian's Keyword Discovery are popular keyword search engines.
- **Competitor Links:** A traditional search engine can also help you check out your competitors, their prices, and their offerings. Try typing 'link:www.[competitor's name].com' into Google to find out how many other sites link to your competitor's website. 'It is a great way to see a competitor's link development and PR campaigns,' says Shari Thurow, Web expert and author of the upcoming book Search Engine Visibility. 'Is the competitor promoting a product or service similar to your own? Maybe you can get publicity because you have a new or better product.'
- **Read Blogs:** Blogs are updated much more regularly than traditional websites and, therefore, they can be another gauge of public opinion. Search blogs by using blog-specific search engines, such as Technorati or Nielsen BuzzMetrics' Blogpulse. 'Blogs tend to move at a faster pace and be more informal in tone, so you're more likely to pick up conversation about a new product type or need on a blog than on a standard web site,' Laycock says.
- **Conduct Online Surveys:** Another way to gauge public opinion is through online surveys. While not as scientific as in-person or phone surveys that use a random sampling of the population, online surveys are a low-cost way to do market research about whether an idea or a product will be appealing to

consumers. Now many companies offer to conduct online research for you or give your company the tools to carry out your own surveying. Some online survey companies include EZquestionnaire, KeySurvey, and WebSurveyor. (BY INC. STAFF)

2.4.2 Research Tools and Techniques

There are a variety of types of market research tools - both offline and online - that are used by many large businesses and can be available to small and mid-sized businesses. When these techniques involve people, researchers use questionnaires administered in written form or person-to-person, either by personal or telephone interview, or increasingly online. Questionnaires may be closed-end or open-ended. The first type provides users choices to a question ("excellent," "good," "fair") whereas open-ended surveys solicit spontaneous reactions and capture these as given. Focus groups are a kind of opinion-solicitation but without a questionnaire; people interact with products, messages, or images and discuss them. Observers evaluate what they hear. (BY INC. STAFF)

2.4.3 Major Categories

- **Audience Research:** Audience research is aimed at discovering who is listening, watching, or reading radio, TV, and print media respectively. Such studies in part profile the audience and in part determine the popularity of the medium or portions of it.
- **Product Research:** Product tests, of course, directly relate to use of the product. Good examples are tasting tests used to pick the most popular flavors - and consumer tests of vehicle or device prototypes to uncover problematical features or designs.
- **Brand Analysis:** Brand research has similar profiling features ("Who uses this brand?") and also aims at identifying the reasons for brand loyalty or fickleness.
- **Psychological Profiling:** Psychological profiling aims at construction profiles of customers by temperament, lifestyle, income, and other factors and tying such types to consumption patterns and media patronage.

- **Scanner Research:** Scanner research uses checkout counter scans of transactions to develop patterns for all manner of end uses, including stocking, of course. From a marketing point of view, scans can also help users track the success of coupons and to establish linkages between products.
- **Database Research:** Also known as database "mining," this form of research attempts to exploit all kinds of data on hand on customers - which frequently have other revealing aspects. Purchase records, for example, can reveal the buying habits of different income groups - the income classification of accounts taking place by census tract matching. Data on average income by census tract can be obtained from the Bureau of the Census. Post-sale or Consumer Satisfaction Research. Post-consumer surveys are familiar to many consumers from telephone calls that follow having a car serviced or calling help-lines for computer - or Internet-related problems. In part such surveys are intended to determine if the customer was satisfied. In part this additional attention is intended also to build good will and word-of-mouth advertising for the service provider. (BY INC. STAFF)

3. USED TOOLS AND METHODS

3.1 ANALYZE

Actually main purpose of project making a customer analyze. This analyze is a work to calculate entropy and fetch some count provided information of customer from database. Analyze is concluded with decision tree and its obtained rules. Generate these rules depends pattern of decision tree. To generate decision tree, there are some circumstances are required to control with if structure. With control numerical value of gain information of each attribute, attribute which has greater value will be implemented as a current node in tree structure. For each attribute which will take place in the tree, recalculation of entropy is required. Because of this, implementing with functions and methods is better for code project. For instance stored procedures fetch numerical value of attribute for to use in method and method performs its calculations. In this analyze I have methods which are making general entropy, relative entropy calculations and its functions. These calculations is detailed in section 3.1.5 Implementation of Process.

3.1.1 General Entropy & Relative Entropy

General entropy is a part of decision tree analyze as needed calculate for each attribute. In this project, because of facility in code implementation, general entropy is executed by a method. For sample implementation please take a look "Figure 3.12".

Also relative entropy is required for decision tree analyze. In project implementation, while decision tree is in creation process, for each attribute, gain information is measured making with entropy calculation by method. Numerical value of general entropy is forever bigger than relative entropy. And in this case also gain information will never be negative value. Relative entropy calculation is executed by method.

3.1.2 Data Fetch

To be able to fulfill all calculations, fetch sub value and its count of each attributes for each calculation, from database is required. Stored procedure is a thing that provide

these fetch mission with a database connection. This fetch process has a key role for these calculations, it forms the backbone of the calculations.

3.2 DESIGN

In terms of design project has no visual richness, because functionality is prominent. Design requirements of project is representation of decision tree. With click a button decision tree is accessible. Also other options are possible for presentation of tree. In this work a have used TreeView which is compatible with code implementation platform. TreeView control is a powerful server-control for rendering TreeView UI. It supports a variety of programming models, from statically-defined trees, to dynamically constructed trees, to databound trees. The TreeView's rendering is fully customizable, allowing for a wide-range of look-and-feels for the control. It also supports rendering on a variety of browsers, and can take advantage of up-level capabilities, such as client-script on later desktop browser versions. (Microsoft .NET Framework SDK QuickStart Tutorials Version 2.0, 2005)

After code executed, In the localhost, show tree button will shown on the screen, with click the button decision tree will appear as treeview.

3.2.2 Treeview Usage

The topics in this section describe how to work with the ASP.NET TreeView Web control to display hierarchical data (such as a table of contents) in a tree structure in a Web page. The TreeView control is made up of TreeNode objects. The TreeView control can be bound to data. (Microsoft)

The TreeView Web server control is used to display hierarchical data, such as a table of contents or file directory, in a tree structure. It supports the following features:

- Automatic data binding, which allows the nodes of the control to be bound to hierarchical data, such as an XML document.
- Site navigation support through integration with the SiteMapDataSource control.
- Node text that can be displayed as either selectable text or hyperlinks.
- Customizable appearance through themes, user-defined images, and styles.

- Programmatic access to the TreeView object model, which allows you to dynamically create trees, populate nodes, set properties, and so on.
- Node population through client-side callbacks to the server (on supported browsers).
- The ability to display a check box next to each node.

The TreeView control is made up of one or more nodes. Each entry in the tree is called a node and is represented by a TreeNode object. The following table describes the three different node types. Root, is a node that has no parent node and one or more child nodes. Parent, is a node that has a parent node and one or more child nodes. Leaf is a node that has no child nodes. Although a typical tree has only one root node, the TreeView control allows you to add multiple root nodes to your tree structure. This is useful when you want to display item listings without displaying a single main root node, as in a list of product categories. Each node has a Text property and a Value property. The value of the Text property is displayed in the TreeView control, while the Value property is used to store any additional data about the node, such as data passed to the postback event that is associated with the node. (Microsoft)

See the used treeview shown in following “Figure 3.1”.



Figure 3.1: Show Tree and Button

3.3 USED PLATFORM AND PROGRAMMING LANGUAGE

ASP.net was better for platform choice in terms of code implementation connection with database used in project. ASP is suitable for sql management studio and visual studio.

Table 3.1: Used Programming Language and Platform

Programing Language	Code Platform	Tool	Database
ASP.net (C#)	Visual Studio 2012	RapidMiner 5	SQL Server Management Studio

3.3.1 ASP.NET

ASP.NET is a web development platform, which provides a programming model, a comprehensive software infrastructure and various services required to build up robust web applications for PC, as well as mobile devices. ASP.NET works on top of the HTTP protocol, and uses the HTTP commands and policies to set a browser-to-server bilateral communication and cooperation. ASP.NET is a part of Microsoft .Net platform. ASP.NET applications are compiled codes, written using the extensible and reusable components or objects present in .Net framework. These codes can use the entire hierarchy of classes in .Net framework. The ASP.NET application codes can be written in any of the following languages:

- C#
- Visual Basic.Net
- Jscript
- J#

ASP.NET is used to produce interactive, data-driven web applications over the internet. It consists of a large number of controls such as text boxes, buttons, and labels for assembling, configuring, and manipulating code to create HTML pages. ASP.NET web forms extend the event-driven model of interaction to the web applications. The browser submits a web form to the web server and the server returns a full markup page or HTML page in response. All client side user activities are forwarded to the server for stateful processing. The server processes the output of the client actions and triggers the reactions. Now, HTTP is a stateless protocol. ASP.NET framework helps in storing the information regarding the state of the application, which consists of:

- Page state
- Session state

The page state is the state of the client, i.e., the content of various input fields in the web form. The session state is the collective information obtained from various pages the user visited and worked with, i.e., the overall session state. To clear the concept, let us take an example of a shopping cart. User adds items to a shopping cart. Items are selected from a page, say the items page, and the total collected items and price are

shown on a different page, say the cart page. Only HTTP cannot keep track of all the information coming from various pages. ASP.NET session state and server side infrastructure keeps track of the information collected globally over a session. The ASP.NET runtime carries the page state to and from the server across page requests while generating ASP.NET runtime codes, and incorporates the state of the server side components in hidden fields. This way, the server becomes aware of the overall application state and operates in a two-tiered connected way. (Tutorialspoint)

3.3.2 Visual Studio 2012

Visual Studio Professional 2012 is a professional-quality integrated development environment that simplifies the tasks of creating, debugging, and deploying software for Windows, Microsoft Office and the web. To choose the download option that's best for you, read the descriptions later on this page.

Supported Operating System

- Windows 7 SP1 (x86 and x64)
- Windows 8 (x86 and x64)
- Windows Server 2008 R2 SP1 (x64)
- Windows Server 2012 (x64)

Hardware Requirements:

- 1.6 GHz or faster processor
- 1 GB of RAM (1.5 GB if running on a virtual machine)
- 10 GB of available hard disk space
- 5400 RPM hard disk drive
- DirectX 9-capable video card that runs at 1024 x 768 or higher display resolution (Microsoft)

3.3.3 SQL Server Management Studio

Microsoft® SQL Server® 2012 Express is a powerful and reliable free data management system that delivers a rich and reliable data store for lightweight Web Sites and desktop applications.

Supported Operating System; Windows 7, Windows Server 2008 R2, Windows Server 2008 Service Pack 2, Windows Vista Service Pack 2

- 32-bit systems
 - Computer with Intel or compatible 1GHz or faster processor (2 GHz or faster is recommended.)
- 64-bit systems
 - 1.4 GHz or faster processor
- Minimum of 512 MB of RAM (2 GB or more is recommended.)
- 2.2 GB of available hard disk space
(Microsoft)

3.4 DATASET

Data is worthless in itself. Information is wanted accordably with the objectives. Information is processed data for a purpose. Information is an analyzed data. Information can also be defined as that we extracted the data to give a answer a question. Data are not the only numbers or letters; also data are numbers and letters and their meanings. We call this data about data as “metadata”. Data set selection is very important, it should be considered to be a set of noiseless data which is efficient and providing information gain. There are 8 different characteristics of the customers in this data set. Seven of these attributes create one of the nodes of the decision tree. They are gender, age, marital status, website activity, bought electronic in 12 months, bought phone, payment method. Each attributes concluded in one value of the "smartphone sales" which is becoming outcome attribute .

- **Gender:** “M” signs male, “F” signs female.
- **Marital Status:** “S” signs single person, “M” also married one.
- **Age:** Age lists age of customer .It varies from user to user. Therefore I have applied C4.5 Algorithm. In the end, If age is bigger than 42, It will shown as “B”, else “K”.
- **Website Activity:** “Seldom” explains customer who has seldom website activity. “Regular” explains customer who has regular website activity. “Frequent” explains customer who has frequent website activity.

- **Bought Electronic in 12 months:** It shows whether there is electronic device that is bought in last 12 months. “Yes” signs that there is a device bought, “No” signs that there is not a device bought.
- **Bought Phone:** It shows whether there is a customer who bought smartphone before. “Yes” signs that there is a smartphone bought before, “No” signs that there is not a device bought before.

A	T	T	R	B	T	E	RESULT
Gender	Age	Marital Status	Website_Activity	Bought electronic in 12 months	Bought phone	Payment method	Smartphone sales
M	B	M	Seldom	Yes	No	Bank transfer	late majority
M	K	M	Regular	No	No	Bank transfer	Inovator
F	K	M	Seldom	Yes	Yes	Bank transfer	Early Adoption
M	B	M	Seldom	Yes	Yes	Website Account	Early Adoption
M	K	S	Seldom	Yes	No	Website Account	late majority
M	B	S	Seldom	Yes	Yes	Credit Card	Early Majority
F	B	S	Regular	No	Yes	Bank transfer	Early Adoption
F	B	S	Seldom	Yes	No	Bank transfer	late majority
M	B	M	Regular	Yes	Yes	Monthly Billing	Inovator
F	K	S	Seldom	Yes	No	Website Account	Early Majority
F	K	M	Regular	Yes	Yes	Website Account	Inovator
M	K	S	Seldom	Yes	Yes	Credit Card	Early Adoption
F	B	S	Seldom	Yes	No	Monthly Billing	late majority
M	K	M	Regular	Yes	Yes	Credit Card	Inovator
F	B	S	Regular	No	Yes	Bank transfer	Inovator
F	B	M	Seldom	Yes	No	Bank transfer	late majority
M	B	M	Seldom	Yes	Yes	Bank transfer	Early Adoption
F	K	M	Regular	Yes	Yes	Monthly Billing	Early Adoption
M	K	S	Seldom	Yes	No	Credit Card	Early Majority
M	B	M	Seldom	No	No	Credit Card	late majority
F	B	M	Seldom	No	Yes	Monthly Billing	Early Adoption
M	K	M	Seldom	Yes	No	Bank transfer	Early Majority
F	B	S	Seldom	No	No	Bank transfer	late majority
F	B	S	Seldom	Yes	No	Bank transfer	late majority
F	B	S	Seldom	Yes	No	Website Account	late majority
F	K	S	Regular	Yes	Yes	Bank transfer	Inovator
F	K	S	Seldom	No	Yes	Bank transfer	Early Adoption
F	K	M	Seldom	Yes	Yes	Bank transfer	Early Adoption

Figure 3.2: Dataset With Some Several Rows

- **Payment Method:** It shows which payment method is used in shopping by customer. “Bank Transfer” signs that payment is made with bank transfer. “Website Account” signs that payment is made with website account. “Credit Card” signs that payment is made with credit card. “Monthly Billing” signs that payment is made with monthly billing.

- **Smartphone Sales:** When a smartphone is turned out for sale, If there a sale in this first week, this type of sale is called with “Inovator”. “Early Adoption” is called for people who sense a benefit from product, adopt product. “Early Majority” is called for people who join shopping only when they have productivity gains, when product is required to buy. When they need to buy people are called with “Late majority”

Early Adoption : *Sales in 2.or 3. week*

Inovator : *Sales in 1. Week*

Early Majority : *Sales after 3 weeks and before 2 month*

Late majority : *Sales after first 2 months*

3.5 IMPLEMENTATION PROCESS

Before beginning application development, I had to be sure the dataset which I will handle it, I needed to go through the test tool. I ran this data as a decision tree in rapidminer of natural language processing program as seen Figure 3.3 . After results concluded as expected, I have started to excel works. It was necessary to remove noisy data.

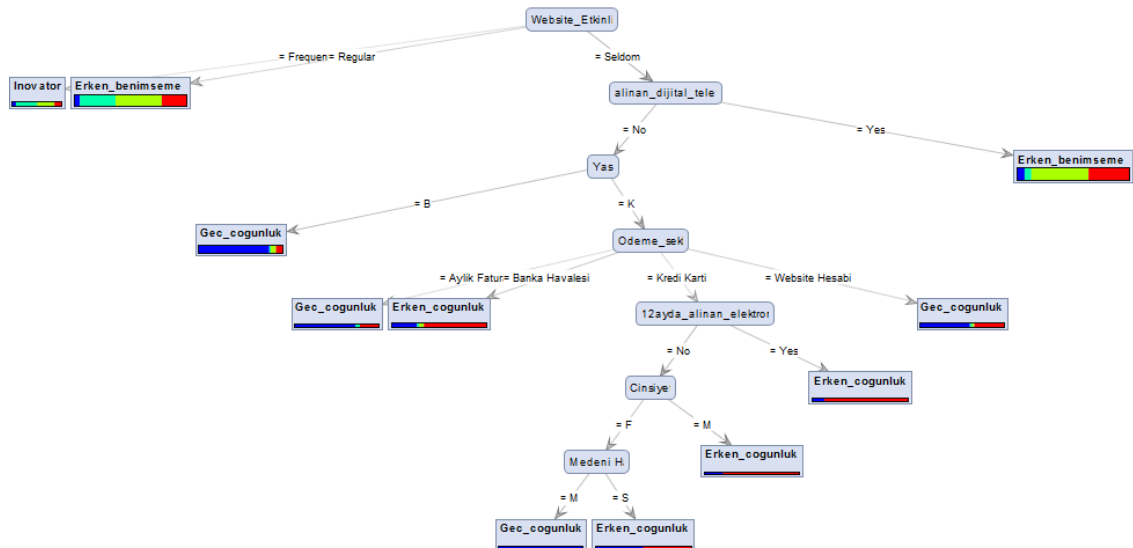


Figure 3.3: Test Decision Tree in RapidMiner

First, this work I have created column for each attribute on excel, then I have removed outliers. Additionally, before the testing on Rapidminer, because of rapidminer does not support to C4.5 algorithm, I have edited some numerical attributes as applied C4.5 algorithm. Then I have imported update data file.

Next step will be database connection. Aims are upload the data set to SQL Server Management Studio and then, create database connection and stored procedure. After creating a database on sql server, I have imported excel file to database. I needed to establish a connection between Microsoft Visual Studio' and SQL Server because of connectivity both of database and code platform. I have created the class called "baglanti.cs" for this connection and some method of this class I could make a connection between database and code. In addition, we describe here baglan.cs the data extraction process from the database as `tek_veri_getir`, `calistir`, `tablolar_getir` ile sağladık. `Tek_veri_getir` has brought both number and string. I could get data both and numerical and string and their counts.

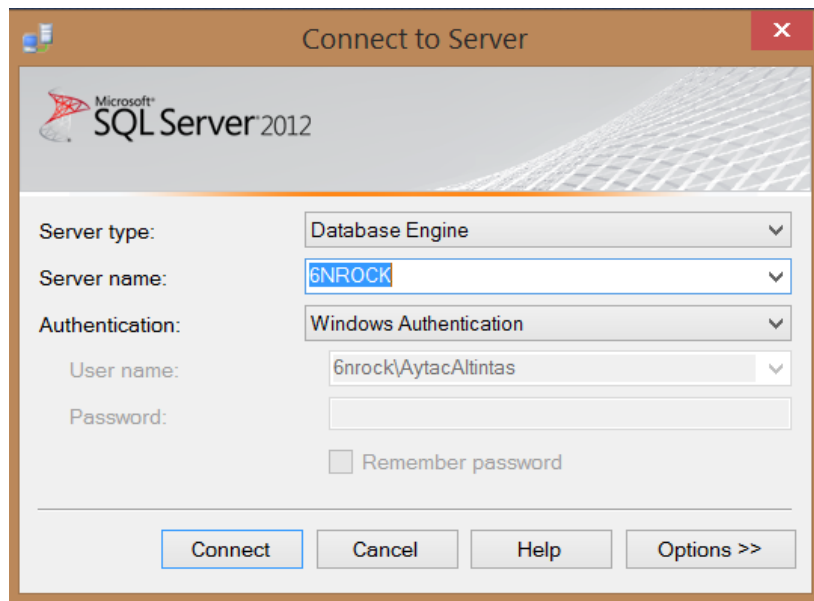


Figure 3.4: Database Connection with SQL Server 2012

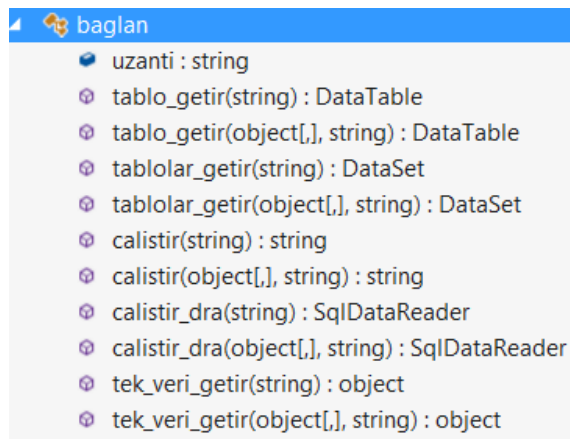


Figure 3.5: Baglan.cs

Another issue related to database operations are also stored procedures because calculations need some count of variable or string. In this point there was a class in ASP.net which facilitated my works. Otherwise for each calculations of attribute may required stored procedure each one. Returning to count on stored procedure is very important but the difficult thing is that, you need to create a stored procedure for each count of calculation. This was about 30-50 store procedure surrounding. The numerical value of them was able to get through GetCount automatically from the database without a store procedure command for each. It has provided me more convenient work

and also less time loss. Method for database connection which in the code as you see Figure 3.6 has been below.

```
public class SmartPhone
{

public double Say(string Cinsiyet, string Yas, string medeniHal,
string Website_Etkinlik, string ayda_alinan_elektronik,
string alinan_dijital_telefon, string Odeme_sekli, string Smartphone_alma)
{
    baglan bag = new baglan();
    object[,] icerik = new object[,] { { "@Cinsiyet", Cinsiyet },
    { "@Yas", Yas }, { "@medeniHal", medeniHal },
    { "@Website_Etkinlik", Website_Etkinlik }, { "@12ayda_alinan_elektronik", ayda_alinan_elektronik },
    { "@alanan_dijital_telefon", alinan_dijital_telefon }, { "@Odeme_sekli", Odeme_sekli },
    { "@Smartphone_alma", Smartphone_alma } };

    return Convert.ToInt32(bag.tek_veri_getir(icerik, "getCount"));
}
```

Figure 3.6: Some of Stored Procedure in Database

This method is within the Public Class smartphone as called "say". Object icerik which is belong this method as started “ `object[,] icerik = new object[,]`.” Then continued with {“@database variable “, `variable_name_in_code`}. Using with `tek_veri_getir` function from `baglan` class I have provided return of count of these database attributes. Current default case just makes connection, unless stated a particular value of the attribute, all these takes all the values of the variables are written burgundy color. "Say Class" automatically creates all database parameters of the stored procedure that we want to return and it fetchs with `tek_veri_getir`.

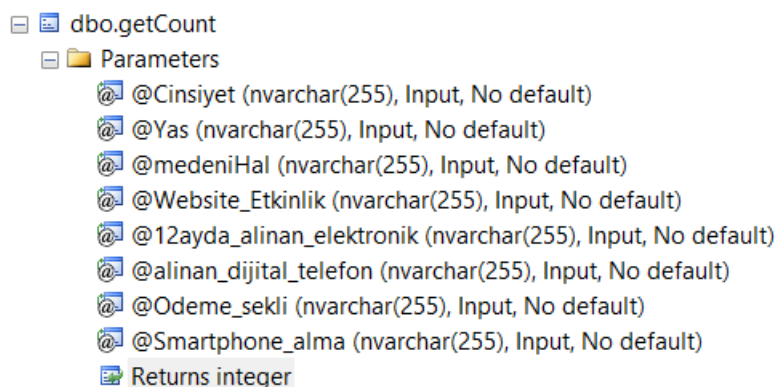


Figure 3.7: Stored Procedure Automatically Created by Method

I needed two different methods for the overall entropy calculation. Entropy method is a function of the entropy formula. Probability values derived from GenelEntropy method,

are is calculated in Entropy method with sending parameters. Later, In the GenelEntropy method, values of general entropy is returned.

```
public double Say(string Cinsiyet, string Yas, string medeniHal,
string Website_Etkinlik, string ayda_alinan_elektronik,
string alinan_dijital_telefon, string Odeme_sekli, string Smartphone_alma)...

public double Entropy(double x,double y,double z,double w)...
public double GenelEntropy(string a,string b,string c,string d,string e,string f,string g)...
public double WebSiteEtkinlikGoreceli(string a, string b, string c, string d, string e, string f)...
public double AlinanDijitalTelefonGoreceli(string a, string b, string c, string d, string e, string f) ...
public double YasGoreceli(string a, string b, string c, string d, string e, string f)...
public double OdemeSekliGoreceli(string a, string b, string c, string d, string e, string f)...
public double AydaAlinanElektronikGoreceli(string a, string b, string c, string d, string e, string f)...
public double CinsiyetGoreceli(string a, string b, string c, string d, string e, string f)...
public double MedeniHalGoreceli(string a, string b, string c, string d, string e, string f)...
```

Figure 3.8: Methods For All Calculation Process

```
public double Entropy(double x,double y,double z,double w)
{
    return -((x * Math.Log(x, 2)) + (y* Math.Log(y, 2)) + (z* Math.Log(z, 2)) + (w* Math.Log(w, 2)));
}
```

Figure 3.9: Formula of Entropy Method Code Implementation

Firstly, to calculate Genel Entropy, values of Smartphone Sale is calculated with say() as one by one before, some exceptional values like zero is checked with if structure. If here is a zero value, I have threw it a very small value like "0,0000001 ". This process is required for keep up the algorithm. Because the number zero will stop the algorithm. In the GenelEntropy, because of the probability of occurrence value of zero controls are made for each Early Adoption, Late Majority, Early Majority, Inovator. See Figure 3.10 for an example control structure.

```
double inovator;
if (Say(a, b, c, d, e, f, g, "Inovator") == 0)
    inovator = 0.000001;
else
    inovator = Say(a, b, c, d, e, f, g, "Inovator");
```

Figure 3.10: Control Structure of Inovator Value of Smartphone Sales Attribute

In the GenelEntropy method, Entropy calculation based on result attribute as required by general entropy process, calculation of the value of Smartphone Sales attribute is done. For this calculation, the probability values is required for the entropy formula. See to Figure 3.10 for the calculation of probability sample.

```
double PgecCogunlukolasiligi = (gecCogunluk) / (gecCogunluk + inovator + erkenBenimseme + erkenCogunluk);
double Pinovatorolasiligi = (inovator) / (gecCogunluk + inovator + erkenBenimseme + erkenCogunluk);
double PerkenBenimsemeolasiligi = (erkenBenimseme) / (gecCogunluk + inovator + erkenBenimseme + erkenCogunluk);
double PerkenCogunlukolasiligi = erkenCogunluk / (gecCogunluk + inovator + erkenBenimseme + erkenCogunluk);
```

Figure 3.11: Probability Calculus of Values of Smartphone Sales Attribute

Ultimately, after both controls are done for possibility of fetch zero value all and probability calculations is made. Just remains to return the General Entropy value. Last line of code in the GenelEntropy method is below, Figure 3.12.

```
double Genelentropi = Entropy(PgecCogunlukolasiligi, Pinovatorolasiligi,
                              PerkenBenimsemeolasiligi, PerkenCogunlukolasiligi);
return Genelentropi;
```

Figure 3.12: Entropy and GenelEntropy Process in GenelEntropy Method

After obtaining a result with entropy calculation, I have made relative entropy calculation for root of decision tree. It is required to make relative entropy calculation for each attributes one by one which will take place in the decision tree. In the relative entropy calculation, probability values are counted for each values of attributes just as general entropy calculation but in the relative calculation also I have counted entropy based on value of result attribute for each calculated attribute values. When entropy calculation is made, it is required to count probability values depending on the outcome. We can examine the relative attribute the calculation of the Website Activity from Figure 3.13 and Figure 3.14 In the website activity method firstly I have counted count of these webregular, webseldom, webfrequent which are category of this attribute with say() method from database. Then with this obtained the total values, I have calculated the probability value of each of these categorical values - web regular, web seldom, web frequent-.

```

public double WebSiteEtkinlikGoreceli(string a, string b, string c, string d, string e, string f)
{
    double webregular, webseldom, webfrequent;
    webregular = Say(a, b, c, "Regular", d, e, f, "");
    webseldom = Say(a, b, c, "Seldom", d, e, f, "");
    webfrequent = Say(a, b, c, "Frequent", d, e, f, "");

    double Pwebregular = (webregular) / (webregular + webseldom + webfrequent);
    double Pwebseldom = (webseldom) / (webregular + webseldom + webfrequent);
    double Pwebfrequent = (webfrequent) / (webregular + webseldom + webfrequent);

    double webregulargeccogunluk = Say(a, b, c, "Regular", d, e, f, "Gec_cogunluk");
    double webregularerkencogunluk = Say(a, b, c, "Regular", d, e, f, "Erken_cogunluk");
    double webregularinovator = Say(a, b, c, "Regular", d, e, f, "Inovator");
    double webregularerkenbenimseme = Say(a, b, c, "Regular", d, e, f, "Erken_benimseme");

    double webseldomgeccogunluk = Say(a, b, c, "Seldom", d, e, f, "Gec_cogunluk");
    double webseldomerkencogunluk = Say(a, b, c, "Seldom", d, e, f, "Erken_cogunluk");
    double webseldominovator = Say(a, b, c, "Seldom", d, e, f, "Inovator");
    double webseldomerkenbenimseme = Say(a, b, c, "Seldom", d, e, f, "Erken_benimseme");

    double webfrequentgeccogunluk = Say(a, b, c, "Frequent", d, e, f, "Gec_cogunluk");
    double webfrequenterkencogunluk = Say(a, b, c, "Frequent", d, e, f, "Erken_cogunluk");
    double webfrequentinovator = Say(a, b, c, "Frequent", d, e, f, "Inovator");
    double webfrequenterkenbenimseme = Say(a, b, c, "Frequent", d, e, f, "Erken_benimseme");
}

```

Figure 3.13: Counting of Website Activity Values For Relative Calculation

Then I have calculated total values of category based on result values which is required relative calculation to calculate value of each other attribute. For example, for the web seldom results four calculations, also four calculation for the web for regular and frequent web. Result attribute has four categorical value and that is why I have applied for value of each entry attributes –category values-. After getting total values from database through say() method, I have started to relative calculations. I have calculated each of categorical value of processed attribute based on result each categorical values. All categorical value probability which I have obtained from depended on four different result values is used for calculate entropy of categorical values of processed attribute. For example, such as hwebregular, hwebseldom, hwebfrequent. Ultimately, as we see in Figure 3.14 on the bottom line, I have got Website Activity relative entropy values with Websmartgoreceli variable. All probability, entropy calculation has resulted in this formula.

```

//Finding entropy
double pwebregulargeccogunluk =webregulargeccogunluk / (webregulargeccogunluk
+ webregularerkencogunluk + webregularinovator + webregularerkenbenimseme);
double pwebregularerkencogunluk = webregularerkencogunluk / (webregulargeccogunluk
+ webregularerkencogunluk + webregularinovator + webregularerkenbenimseme);
double pwebregularinovator = webregularinovator / (webregulargeccogunluk
+ webregularerkencogunluk + webregularinovator + webregularerkenbenimseme);
double pwebregularerkenbenimseme = webregularerkenbenimseme / (webregulargeccogunluk
+ webregularerkencogunluk + webregularinovator + webregularerkenbenimseme);

double pwebseldomgeccogunluk = webseldomgeccogunluk / (webseldomgeccogunluk
+ webseldomerkenbenimseme + webseldominovator + webseldomerkenbenimseme);
double pwebseldomerkenbenimseme = webseldomerkenbenimseme / (webseldomgeccogunluk
+ webseldomerkenbenimseme + webseldominovator + webseldomerkenbenimseme);
double pwebseldominovator = webseldominovator / (webseldomgeccogunluk
+ webseldomerkenbenimseme + webseldominovator + webseldomerkenbenimseme);
double pwebseldomerkenbenimseme = webseldomerkenbenimseme / (webseldomgeccogunluk
+ webseldomerkenbenimseme + webseldominovator + webseldomerkenbenimseme);

double pwebfrequentgeccogunluk = webfrequentgeccogunluk / (webfrequentgeccogunluk
+ webfrequenterkencogunluk + webfrequentinovator + webfrequenterkenbenimseme);
double pwebfrequenterkencogunluk = webfrequenterkencogunluk / (webfrequentgeccogunluk
+ webfrequenterkencogunluk + webfrequentinovator + webfrequenterkenbenimseme);
double pwebfrequentinovator = webfrequentinovator / (webfrequentgeccogunluk
+ webfrequenterkencogunluk + webfrequentinovator + webfrequenterkenbenimseme);
double pwebfrequenterkenbenimseme = webfrequenterkenbenimseme / (webfrequentgeccogunluk
+ webfrequenterkencogunluk + webfrequentinovator + webfrequenterkenbenimseme);

double hwebregular, hwebseldom, hwebfrequent;
hwebregular = -((pwebregulargeccogunluk * Math.Log(pwebregulargeccogunluk, 2))
+ (pwebregularerkencogunluk * Math.Log(pwebregularerkencogunluk, 2))
+ (pwebregularinovator * Math.Log(pwebregularinovator, 2))
+ (pwebregularerkenbenimseme * Math.Log(pwebregularerkenbenimseme, 2)));
hwebseldom = -((pwebseldomgeccogunluk * Math.Log(pwebseldomgeccogunluk, 2))
+ (pwebseldomerkenbenimseme * Math.Log(pwebseldomerkenbenimseme, 2))
+ (pwebseldominovator * Math.Log(pwebseldominovator, 2))
+ (pwebseldomerkenbenimseme * Math.Log(pwebseldomerkenbenimseme, 2)));
hwebfrequent = -((pwebfrequentgeccogunluk * Math.Log(pwebfrequentgeccogunluk, 2))
+ (pwebfrequenterkencogunluk * Math.Log(pwebfrequenterkencogunluk, 2))
+ (pwebfrequentinovator * Math.Log(pwebfrequentinovator, 2))
+ (pwebfrequenterkenbenimseme * Math.Log(pwebfrequenterkenbenimseme, 2)));

double websmartgoreceli = Pwebregular * hwebregular + Pwebseldom * hwebseldom + Pwebfrequent * hwebfrequent;

```

Figure 3.14: Relative Calculation of Website Activity

After fetch all count of these variables with help of parameters and stored procedure, as depended to results, required general entropy and relative entropy calculation for each dataset attribute is completed. Next aim is create node of decision tree with obtaining informatin gain of each attributes. I wanted to reflect decision tree on the screen in the form of TreeView with a button. That's why I have created objects via Smartphone class which has relative and general entropy calculations in the code block of button. Thus, I could reach all the previous methods. After checking exceptions, I could calculate gain values surely, as it appears in Figure 3.15. With “goreceli” object I have provided to access all attributes which need relative calculations. After a node is created I have used “branchresult” object to check categorical values that will come to nodes because sometimes tree has resulted directly with result attribute values. This required to check

process. When a node is occurred, "General Entropy" object is necessary for specifying a node to take place in the tree. With this object I have checked true node.

```
protected void Button1_Click(object sender, EventArgs e)
{
    SmartPhone goreceli = new SmartPhone();
    SmartPhone dalsonuc = new SmartPhone();
    SmartPhone genelEntropy = new SmartPhone();

```

Figure 3.15: Create Object From Smartphone Class

I have defined "kok" variable for root of tree as string. To determine to the root and the greatest gain information I have compared all calculation values of the attributes. This is true for each of the attributes that I have executed. In the result of these checks, If the greatest gain information value is belong executed attribute, this attribute will be dedicated as root in the decision tree. I have added this attribute to TreeView1 variable as root.

```
string kok;
if (genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.WebSiteEtkinlikGoreceli(" ", " ", " ", " ", " ", " ")
    > genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.AlinanDijitalTelefonGoreceli(" ", " ", " ", " ", " ", " ") &&
    genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.WebSiteEtkinlikGoreceli(" ", " ", " ", " ", " ", " ")
    > genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.VasGoreceli(" ", " ", " ", " ", " ", " ") &&
    genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.WebSiteEtkinlikGoreceli(" ", " ", " ", " ", " ", " ")
    > genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.OdemeSekliGoreceli(" ", " ", " ", " ", " ", " ") &&
    genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.WebSiteEtkinlikGoreceli(" ", " ", " ", " ", " ", " ")
    > genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.AydaAlinanElektronikGoreceli(" ", " ", " ", " ", " ", " ") &&
    genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.WebSiteEtkinlikGoreceli(" ", " ", " ", " ", " ", " ")
    > genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.MedeniHalGoreceli(" ", " ", " ", " ", " ", " ") &&
    genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.WebSiteEtkinlikGoreceli(" ", " ", " ", " ", " ", " ")
    > genelEntropy.GenelEntropy(" ", " ", " ", " ", " ", " ") - goreceli.CinsiyetGoreceli(" ", " ", " ", " ", " ", " "))
{
    kok = "WebsiteEtkinlik";
    TreeView1.Nodes.Add(new TreeNode(kok));

```

Figure 3.16: Controls of Determine of Root

After determining Website Activity as the root, I have made some checks with if structure, to branches of this determined attribute for categorical values which will take place in branches of tree. Also It is possible to be resulted directly from a branch. For example, Frequent branch is resulted directly in Inovator. Of course, this may not be applied to every branch. Because of this, I have checked all possibilities. See Figure 3.17 .

```

//For Frequent branch
TreeView1.Nodes[0].ChildNodes.Add(new TreeNode("Frequent"));
if(dalsonuc.Say("", "", "", "Frequent", "", "", "", "Gec_cogunluk")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Erken_cogunluk") &&
    dalsonuc.Say("", "", "", "Frequent", "", "", "", "Gec_cogunluk")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Inovator") &&
    dalsonuc.Say("", "", "", "Frequent", "", "", "", "Gec_cogunluk")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Erken_benimseme"))
    TreeView1.Nodes[0].ChildNodes[0].ChildNodes.Add(new TreeNode("Gec_cogunluk"));

else if (dalsonuc.Say("", "", "", "Frequent", "", "", "", "Erken_cogunluk")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Gec_cogunluk") &&
    dalsonuc.Say("", "", "", "Frequent", "", "", "", "Erken_cogunluk")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Inovator") &&
    dalsonuc.Say("", "", "", "Frequent", "", "", "", "Erken_cogunluk")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Erken_benimseme"))
    TreeView1.Nodes[0].ChildNodes[0].ChildNodes.Add(new TreeNode("Erken_cogunluk"));

else if (dalsonuc.Say("", "", "", "Frequent", "", "", "", "Inovator")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Gec_cogunluk") &&
    dalsonuc.Say("", "", "", "Frequent", "", "", "", "Inovator")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Erken_cogunluk") &&
    dalsonuc.Say("", "", "", "Frequent", "", "", "", "Inovator")
    > dalsonuc.Say("", "", "", "Frequent", "", "", "", "Erken_benimseme"))
    TreeView1.Nodes[0].ChildNodes[0].ChildNodes.Add(new TreeNode("Inovator"));

else
    TreeView1.Nodes[0].ChildNodes[0].ChildNodes.Add(new TreeNode("Erken_benimseme"));
//Finish of Frequent

```

Figure 3.17: Controls of Determine of Frequent Branch

I have also applied these calculation processes for Seldom and Regular, than, I have worked on whether a node will take place, If not, a categorical value will be resulted directly. Calculations which are made for determine the root are the same for find each nodes. Each of general entropy and relative entropy calculations are reused for next occurred tree place -node or branch- which is required for tree structure till all attributes are took place in decision tree.

4. RESULTS AND DISCUSSION

The results of the study can be evaluated from two different angles. The first perspective, In the data analyze study made with decision tree, using with columns applied C4.5 algorithm these have numerical values on wide range, for appropriate number of attributes works through such as multi line data can be done. application with a user interface which data is dynamically retrieved can be much more efficient. In respect to this there is an open side to develop. I have showed the results of the work with the TreeView on this project. I have directed user to the tree with the help of a button. In the directed screen as result of executed code implementation, conclusions of decision tree choosed level by level, gradually are accessible. In the result display, different decision tree works can be provided in terms of a variety of visual richness. Today, implementation of a decision tree to different data sets, gaining information have been realized with natural language processing tools such as Rapidminer, Weka but in my opinion decision tree implementations which is completed with this way is not enough. Because for decision trees it is not clear in some cases, for example, even if a decision tree is correctly formed calculated based on gain, a node of this tree can not be resulted with categorical value of result attribute. This resultless attribute required to take place by another node for next place. In tools It is not clear that this resultless node would be added from which branch and which checks would be made. In this study, I have made calculations as counted from database and exceptional tree controls using with control structure. While doing this, I have avoided to stop algorithm. The second perspective, is rules generated through calculations and code implementations in light of this gain information.

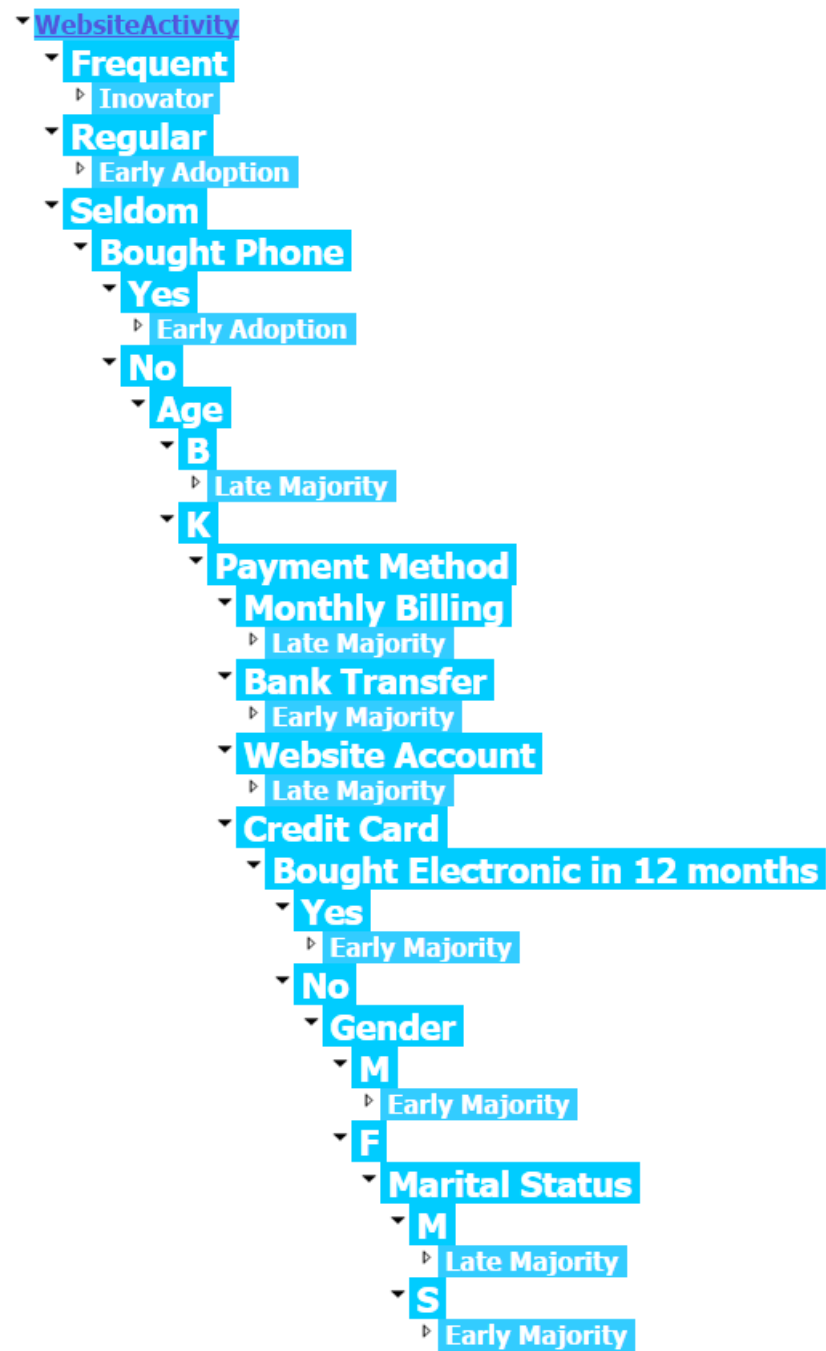


Figure 4.1: Decision Tree with Treeview

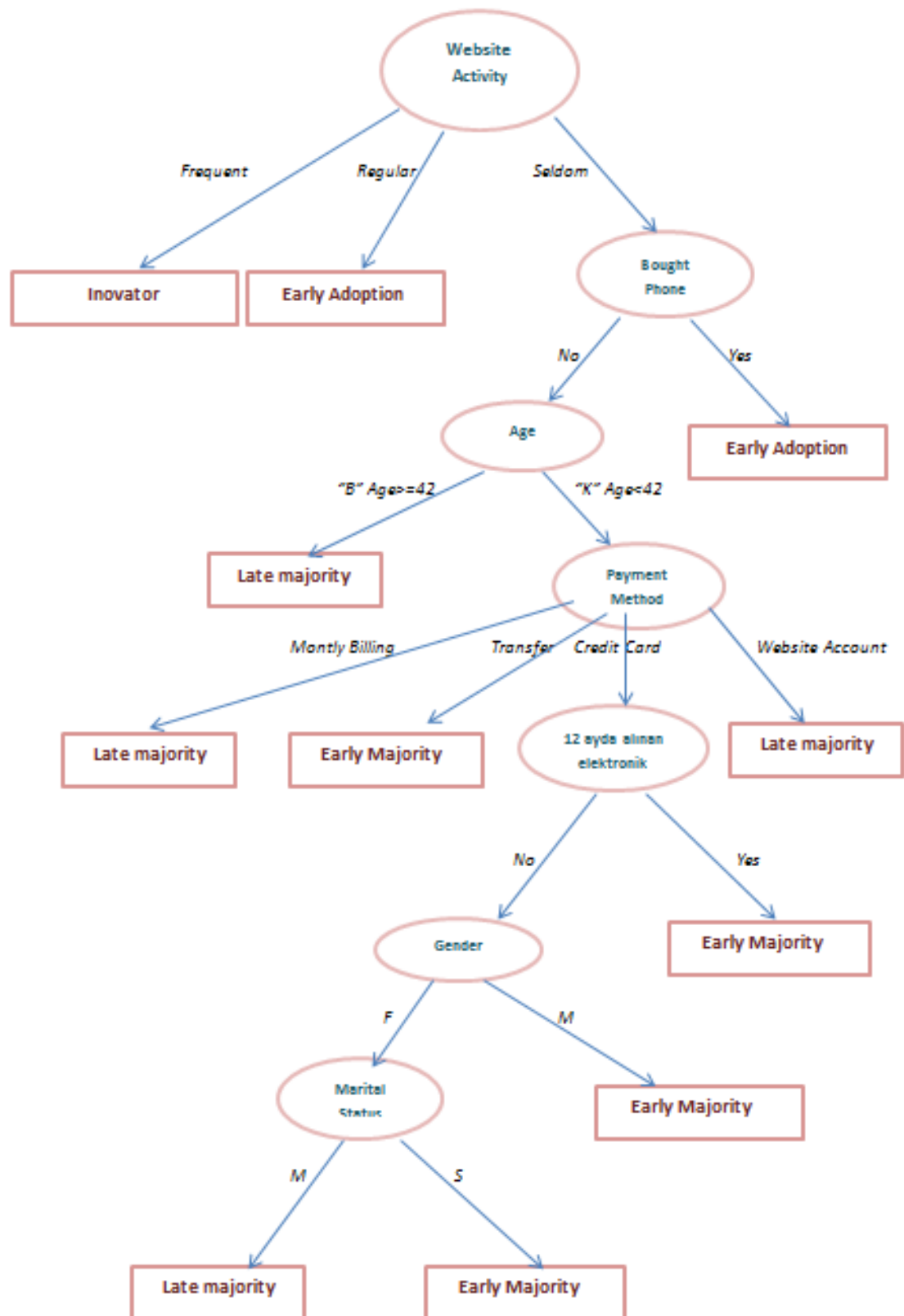


Figure 4.2: Other Representation of Decision Tree of Work

After developing our decision tree, I have made rules extraction on the last stage. make interpretation about results is possible whereby obtained rules. Considering that the most important part of decision tree is the root, The Website Activity is the most important thing that also seems from there. After a node does not go directly to result, another node is going to take place below of this node. It means that next state is dependent on this node. When commenting It should be considered as related together. Created rules are shown below which are numbered.

Table 4.1:Result Values of Smartphone Sale

Early Adoption:	Sales in 2. or 3. week
Inovator :	Sales in 1. week
Early Majority :	Sales after 3 weeks and before 2 month
Late majority :	Sales after first 2 months

1. *“If Website activity is Frequent, result is Inovator.”*

Users who frequently online on the internet, have bought in first week of sales. So a large majority of inovator is consisted by Frequent.

2. *“If Website activity is Regular, result is Early Adoption.”*

If you have loyal customers over the Internet, and they are in a regular follow-up, we see that they are in early adoption. We can say that customers who bought in the 2nd or 3rd week is inside of early adoption.

3. *“If Website activity is Seldom, Bought Phone is yes, result is Early Adoption.”*

People who has seldom website activity and never buy cellphone before are still pleased to shop. They still rely on keep company as dependable. Consequently these people are inside of early adoption.

4. *“If Website activity is Seldom, Bought Phone is no, Age is bigger than 42, result is Late majority.”*

We have people who has seldom website activity and did not buy phone is resulted with Age node. Age is C4.5 applied attribute because of its wide value range and its low gain information. For example we can imagine that age of many persons show variation person to person. Therefore we have calculated average of values of attribute, than we have divided people two categorical values whose age is bigger than forty two and smaller ones. - average value is forty two - . People who has forty two and more ages are resulted in late majority. People in late majority has been bought two months later of sale start. People who has less ages from forty two are resulted with another node that is payment method.

5. *“If Website activity is Seldom, Bought Phone is no, Age is lower than 42, Payment Method is Monthly Billing, result is Late majority. ”*

Values of “Age” attribute that lower than 42 is not resulted. When a branch is not resulted, It will be placed with another node. After, as lower than 42 branch we have Payment Method node. Payment Method attribute has three branches. A path in the decision tree must be finished with a result attribute. We have kept following Payment Method, one of branch of Payment method is Montly Billing. After Montly Billing branch is resulted with Late Majority.

6. *“If Website activity is Seldom, Bought Phone is no, Age is lower than 42, Payment Method is Bank Transfer, result is Early majority.”*

Bank Transfer which is another branch of Payment Method is resulted with Early Majority which is included people who buy 3 weeks later and 3 months before as of sale start. From this result also we can have inference as young people use bank transfer method.

7. *“If Website activity is Seldom, Bought Phone is no, Age is lower than 42, Payment Method is Website Account, result is Late majority.”*

Bank Transfer which is another branch of Payment Method is resulted with Early Majority which is included people who buy 3 weeks later and 3 months before as of sale start. From this result also we can have inference as young people use bank transfer method.

8. *“If Website activity is Seldom, Bought Phone is no, Age is lower than 42, Payment Method is Credit Card, Bought Electronic in 12 months is yes, result is Early majority.”*

Credit Card which the last branch of Payment Method is not directly resulted and another node is placed. This node "Bought Electronic in 12 months" attribute which has customers who is bought electronic product with credit card but never bought cellphone before, they are inside of Early Majority.

It means that these type customers are buying between after 3 weeks and 2 months ago as of buy start. Also these customers are under forty two ages.

9. *“If Website activity is Seldom, Bought Phone is no, Age is lower than 42, Payment Method is Credit Card, Bought Electronic in 12 months is no), Gender is M result is Early majority. ”*

The "No" branch which is branched from "Bought Electronic in 12 months" attribute is not resulted directly. Another node is placed here. This node is Gender, It seems that male customers who does not buy neither cellphones nor electronic products in twelve months probably buy from three weeks earliest t two months latest such customers are called Early Majority.

10. *“If Website activity is Seldom, Bought Phone is no, Age is lower than 42, Payment Method is Credit Card, Bought Electronic in 12 months is no, Gender is F, Marita Status is M result is Late majority. ”*

If the customers who does not buy neither cellphones nor electronic product in twelve months are female, Result depends Marital Status attribute. If women are married, result is Late Majority. Also these women have age under of forty two.

11. *“If Website activity is Seldom, Bought Phone is no, Age is lower than 42, Payment Method is Credit Card, Bought Electronic in 12 months is no, Gender is F, Marital Status is S, result is Early majority.”*

Again, If the customers who does not buy neither cell phones nor electronic product in twelve months are female, Result depends Marital Status attribute. If women are single, result is Late Majority. Also these women have age under of forty two.

BIBLIOGRAPHY

1. Walker, J. (2007, May 31). *importance of the internet*. Retrieved March 10, 2016, from [tibetangeeks: https://www.tibetangeeks.com/using_tech/importance/importance_of_internet.html](https://www.tibetangeeks.com/using_tech/importance/importance_of_internet.html)
2. KEMP, S. (2016, 01 27). *digital-in-2016*. Retrieved 01 28, 2016, from wearesocial: <http://wearesocial.com/uk/special-reports/digital-in-2016>
3. Eticaretport. (2016, 02 04). *eticaret-nedir-faydalari-nelerdir--s1865*. Retrieved 03 03, from <https://www.eticaretport.com/eticaret-nedir-faydalari-nelerdir--s1865.html>
4. Wikipedia. (2016, 02 17). *Decision_tree*. Retrieved 03 17, 2016, from wikipedia: https://en.wikipedia.org/wiki/Decision_tree
5. Wikipedia. (2016, 02 12). *Decision rules*. Retrieved 03 03, 2016, from wikipedia: https://en.wikipedia.org/wiki/Decision_tree#Decision_rules
6. Mitchell, T. M. (1997). Machine Learning . In T. M. Mitchell, *Machine Learning* (pp. 57-58). McGraw.
7. Roland, J. (1995). *Diffusion of innovations*. Retrieved from ehow: http://www.constitution.org/col/03317_diffusion.htm
8. VALENTE, Thomas W. Social network thresholds in the diffusion of innovations. *Social networks*, 1996, 18.1: 69-89.
9. RAM, S. ; JUNG, Hyung-Shik. Innovativeness in product usage: A comparison of early adopters and early majority. *Psychology & Marketing*, 1994, 11.1: 57-67.
10. MARKOV, Zdravko; LAROSE, Daniel T. Frontmatter. *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage*, i-xvi.
11. Microsoft .NET Framework SDK QuickStart Tutorials Version 2.0. (2005). *treeview*. Retrieved 02 2016, from [quickstarts.asp.net: http://quickstarts.asp.net/quickstartv20/aspnet/doc/ctrlref/navigation/treeview.aspx](http://quickstarts.asp.net/quickstartv20/aspnet/doc/ctrlref/navigation/treeview.aspx)
12. YALÇIN, E. (2014, 11 16). *BAGLAN CLASS (tek_veri_getir)*. Retrieved 11 18, 2015, from erdiyalcin: <http://www.erdiyalcin.com/2014/11/baglan-class-tekverigetir.html>

13. YALÇIN, E. (2014, 11 16). *BAGLAN CLASS (tek_veri_getir)*. Retrieved 11 18, 2015, from erdiyalcin: <http://www.erdiyalcin.com/2014/11/baglan-class-tekverigetir.html>
14. Microsoft. (n.d.). *TreeView Class*. Retrieved 11 2015, from msdn.microsoft: [https://msdn.microsoft.com/enus/library/system.web.ui.webcontrols.treeview\(v=vs.10\).aspx](https://msdn.microsoft.com/enus/library/system.web.ui.webcontrols.treeview(v=vs.10).aspx)
15. RapidMiner Studio . (n.d.). Data and Result Visualization. In R. Studio, *RapidMiner Studio Manual* (pp. 86-90). <http://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf>.
16. SAS. (n.d.). Machine Learning. Retrieved 04 02, 2016, from sas: http://www.sas.com/en_us/insights/analytics/machine-learning.html
17. Hawks, D. D. (n.d.). *What Is a Decision Tree? - Examples, Advantages & Role in Management*. Retrieved 04 15, 2016, from study: <http://study.com/academy/lesson/what-is-a-decision-tree-examples-advantages-role-in-management.html>
18. Lefebvre, R. C. (n.d.). *On Social Marketing and Social Change*. Retrieved 05 15, 2016, from socialmarketing.blogs: http://socialmarketing.blogs.com/r_craig_lefebvres_social/2009/10/social-models-for-marketing-diffusion.html
19. Microsoft. (n.d.). *Microsoft Visual Studio Professional 2012*. Retrieved 04 15, 2016, from microsoft: <https://www.microsoft.com/en-us/download/details.aspx?id=30682>
20. Microsoft. (n.d.). *Microsoft® SQL Server® 2012 Express*. Retrieved 04 15, 2016, from Microsoft: <https://www.microsoft.com/en-US/download/details.aspx?id=29062#>
21. Tutorialspoint. (n.d.). *ASP.NET - Introduction*. Retrieved 04 15, 2016, from Tutorialspoint: http://www.tutorialspoint.com/asp.net/asp.net_introduction.htm
22. Microsoft. (n.d.). *TreeView Control (Visual Studio)*. Retrieved 04 15, 2016, from Microsoft: [https://msdn.microsoft.com/en-us/library/d4fz6xk2\(v=vs.80\).aspx](https://msdn.microsoft.com/en-us/library/d4fz6xk2(v=vs.80).aspx)
23. Microsoft. (n.d.). *TreeView Web Server Control Overview*. Retrieved 04 15, 2016, from Microsoft: [https://msdn.microsoft.com/en-us/library/e8z5184w\(v=vs.80\).aspx](https://msdn.microsoft.com/en-us/library/e8z5184w(v=vs.80).aspx)
24. Rouse, M. (2016, 02). *machine learning*. Retrieved 04 15, 2016, from whatis.techtarget: <http://whatis.techtarget.com/definition/machine-learning>

25. BY INC. STAFF . (n.d.). *Conducting Online Market Research: Tips and Tools*.

Retrieved 04 04, 2016, from inc: http://www.inc.com/guides/biz_online/online-market-research.html