THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY

ANALYSIS OF IMPACT ON ETHICAL VALUES OF INTERNET USE: A DATA MINING MODEL APPROACH

Master's Thesis

MERVE ARITÜRK

İSTANBUL, 2015

THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

INFORMATION TECHNOLOGIES

ANALYSIS OF IMPACT ON ETHICAL VALUES OF INTERNET USE: A DATA MINING MODEL APPROACH

Master's Thesis

MERVE ARITÜRK

Advisor: Asst. Prof. Dr. Dilek KARAHOCA

İSTANBUL, 2015

T.C.
BAHÇEŞEHİR UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE
INFORMATION TECHNOLOGIES

Name of the thesis ANALYSIS OF IMPACT ON ETHICAL VALUES OF INTERNET USE: A DATA MINING MODEL APPROACH

Name of the Student: Merve Arıtürk

Date of thesis defence: May 22, 2015

It has been accepted by the Institute of Science that this thesis fulfils the needed requirements as a master thesis.

Asoc. Prof. Dr. Nafiz ARICA
Institute Principal
Signature

I certify that this thesis has fulfilled all the requirements as a master thesis.

Prof. Dr. Adem KARAHOCA
Program Coordinator
Signature

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

Examining Committee Members	Signature
Supervisor	
Asst. Prof. Dr. Dilek Karahoca	……..
Member	
Prof. Dr. Adem Karahoca	……
Member	
Asst. Prof. Dr. İbrahim Pınar	……

PREFACE

I would like to express my deepest gratitude to my honourable thesis supervisor Dilek KARAHOCA, who provided me with help and support throughout the study; helped me constituting the content of the thesis; and shed light for my work with her knowledge and experience.

I would like to thank Bahcesehir University for reaching the data about the students who took Ethical Internet Usage Survey.

To my family who provided full moral and material support during the preparation of my thesis.

İstanbul, 2015 Merve ARITÜRK

ABSTRACT

ANALYSIS OF IMPACT ON ETHICAL VALUES OF INTERNET USE: A DATA MINING MODEL APPROACH

Merve Arıtürk

Information Technologies

Supervisor: Asst. Assoc. Dr. Dilek KARAHOCA

May 2015, 45 Pages

The purpose of this study, analyse the usage of Internet between awareness of information and faith with investigating the experiences of people who live in Turkey. The problem is trying to demonstrate the format of ethical behaviour for different users with their perspective.

In this study, the universe of the sample was reached by the social media platforms. The average age of the participants was calculated as 21. Scanning Model is used with the survey that was named as "Ethical Internet Usage" was performed on the Internet.

While maintaining this work, the questions which were on different surveys with the related to our work were analysed. After analysing, the questions were selected with the existing correlation value that significantly depends on the information. Then, the arff file is created with the factors and the data. According to the arff file, some data mining techniques are used to make predictions. The best result is gained from the NaiveBayes classification method.

Based on the results, women have more ethical behaviour than men. There is a strict relationship between the ethical personal behaviour and the education level of parents. According to the classification methods, the scale will be created with the results which are significant. If there is a new value or data about the research, the scale shows the user behaviour that is assuming based on users' demographical information.

Keywords: Ethics, Internet Ethics, Values of Faith, Data Mining

ÖZET

İNTERNET KULLANIMININ ETİK DEĞERLER ÜZERİNDEKİ ETKİSİNİN ANALİZİ: VERİ MADENCİLİĞİ MODELİ YAKLAŞIMI

Merve Arıtürk

Bilgi Teknolojileri

Tez Danışmanı: Yrd. Doç. Dr. Dilek KARAHOCA

Mayıs 2015, 45 Sayfa

Bu çalışmanın amacı, Türkiye'deki bireylerin etik internet kullanımı üzerindeki deneyimlerini araştırarak internet kullanımındaki bilgi farkındalığı durumunun inanç bağlantı analizinin yapılmasıdır.

Bilgi iletişim teknolojilerindeki gelişmelerle birlikte elektronik iletişim kaynaklarına olan talep, sosyal yaşam değerleri üzerinde farklı etkiler meydana getirmiştir.Sosyal medya kullanıcıları günlük yaşam faaliyetlerinin yanı sıra inanç sistemlerindeki değerleri paylaşmak , eleştirmek , yeniden yorumlamak gibi düşünsel aktiviteleri etkin kullanmaktadır.Bu araştırmada bireylerin etik internet kullanımı alışkanlıklarının sosyal medya ortamında değerlerler üzerindeki etkileri incelenmiştir. Çalışmanın örneklemini sosyal medya kullanıcılarından seçilerek, "Etik İnternet Kullanımı" anketi uygulanmıştır. Bu anketle veriler toplanarak; veri madenciliği yaklaşımı ile değerlendirilmiştir.

Çalışmalarda mevcut internet kullanımına ilişkin farklı anketlerden belirlenen araştırma boyutlarıyla ilgili sorular teker teker analiz edilerek ilgili araştırmalardaki mevcut korelasyon değerleri ve bilginin anlamlı olup olmamasına bağlı olarak sorular seçilmiştir.

Uygulanan anketin sonuçları incelenerek bireylerin etik olması ile interneti etik kullanması arasındaki ilişki irdelenmiş, veri madenciliği teknikleri kullanılarak analiz yapılmıştır. Araştırma sonuçları doğrultusunda önerilerde bulunulmuştur.

Anahtar Kelimeler: Etik, İnternet Etiği, İnanç Değerleri, Veri Madenciliği

TABLE OF CONTENTS

TABLES

FIGURES

1. INTRODUCTION

Nowadays, as the science world makes big progress; technology also develops swiftly in parallel. While technology takes part as one of the indispensables of people, usage of technology becomes a part of their lives. Developing technology has started to make people's life easier. Technological developments make it easier to receive faster results and to reach correct values in scientific research. Since the beginning of the 20th century science and technology keep going as an inseparable whole even today because thanks to electronic computers used since 1950's, a bond of connection has started between science and technology. Day by day, computers that develop in terms of technological features and hardware help shed light to research that were difficult to conduct before; and reach faster and more correct results in existing research. In addition, developing technology place an important role in revealing secret information. The infrastructure of the scientific comprised together with the Renaissance and Reform movements. All the negativity that tried to hinder science which started to develop with the Renaissance and Reform movements disappeared in time and helped technology develop which the help of technology people explored the world they are living in and problems in this world. Thanks to these developments scientific developments have accelerated when they merged with technology (Neill, 1989).

According to the development of the technology, the interaction between human and computer is gained much more importance. According to Gupta, Human Computer Interaction (HCI) is a teach that goes for a secured understanding and outlining of diverse interfaces in the middle of people and PCs in a manner that it characterizes frameworks that are pleasant to utilize, are drawing in and are available (Gupta, 2012).

According to Carroll, on a very basic level, HCI is about understanding and making exercises for individuals, exercises in which innovative curios assume essential parts. HCI starts where the configuration objective moves from considering antiquities simply regarding innovation arranged predicates like efficient , systematic , and portable to considering them as far as errand situated predicates like leverages earlier information of word transforming or supports crafty investigation (Carroll, 1996). From an HCI perspective, one could say that systems and applications are the set of activities they afford and support (Carroll 1994).

HCI helps us to comprehend why some product items are great and other programming is terrible (Blackwell, 2011). But unfortunately it is not an ensured equation for making a fruitful item. In this sense it is similar to structural engineering or item plan. Modellers and item originators require an intensive specialized handle of the materials they work with, however the achievement of their work relies on upon the innovative utilization of this specialized information. This innovativeness is an art expertise that is regularly adapted by meeting expectations with an expert architect in a studio, or from contextual analyses of effective outlines (Blackwell, 2011).

According to study of Te'eni, Carey and Zhang (2007), Man-Machine Interaction or Interfacing which was a thought of Human-Computer Interaction/Interfacing (HCI) was commonly addressed with the creating of PC, or more all-around machine, itself. The reason, in reality, is clear: most developed machines are futile unless they can be used genuinely by men. This central discord basically presents the essential terms that should be considered in the arrangement of HCI: convenience and usability (Te'eni, Carey and Zhang, 2007).

Why a system is truly arranged can in the end be portrayed by what the structure can do i.e. how the components of a system can help towards the achievement of the purpose behind the structure. Value of a structure is portrayed by the arranged of exercises or organizations that it provides for its customers. In any case, the estimation of usefulness is obvious just when it gets to be conceivable to be productively used by the client (B. Shneiderman and C. Plaisant, 2004). Usability of a system with a certain helpfulness is the compass and degree by which the structure can be used successfully and enough to satisfy certain goals for particular customers. The genuine ampleness of a structure is accomplished when there is a fitting concordance between the value and convenience of a system (Nielsen, 1994).

Having these thoughts at the highest point of the need rundown and considering that the terms PC, machine and structure are frequently used correspondingly as a piece of this association, HCI is an arrangement that should convey a fit between the customer, the machine and the obliged organizations in order to attain to a certain execution both in quality and optimality of the associations (Te'eni, 2006). Figuring out what makes a certain HCI arrangement awesome is fundamentally subjective and association

subordinate. Case in point, a flying machine part delineating mechanical assembly should give high precisions in connection and design of the parts while an arrangement altering programming may not need such precision. The accessible innovation could likewise influence how different sorts of HCI are planned for the same reason. One case is using summons, menus, graphical user interfaces (GUI), or virtual reality to get to functionalities of any given PC. In the accompanying fragment, a more organized survey of existing techniques and devices used to interface with PCs and the late advances in the field is displayed (Karray, Alemzadeh, Saleh and Arab, 2008).

Outlining for HCI signifies "Guaranteeing framework usefulness and ease of use, giving successful client collaboration bolster, and improving a lovely client experience." Furthermore, "The larger objective is to attain to both hierarchical and individual client adequacy and effectiveness. To achieve these objectives, administrators and designers need to be learned about the interchange among clients, assignments, assignment settings, data innovation (IT), and the situations in which frameworks are utilized" (Carey, 2004).

According to the human computer interaction, information behaviour is very important. Information Behaviour is the totality of human lead in association with sources and channels of information, including both dynamic and unapproachable information searching for, and information usage. Accordingly, it incorporates face – to – face correspondence with others, and in addition the aloof gathering of data as in, for instance, sitting in front of the TV promotions, with no expectation to follow up on the data given.

Information Seeking Behaviour (ISB) is the purposive searching for information as an after effect of a need to satisfy some goal. Over the compass of searching for, the individual may participate with manual information systems, (for instance, a day by day paper or a library), or with PC – based structures (for instance, the World Wide Web).

Data Searching Behaviour is the 'smaller scale level' of conduct utilized by the searcher in cooperating with data frameworks of numerous types. It comprises of every last one of cooperation with the framework, whether at the level of human PC cooperation (for instance, utilization of the mouse and taps on connections) or at the scholarly level (for instance, embracing a Boolean seek methodology or deciding the criteria for choosing

which of two books chose from adjoining places on a library rack is most valuable), which will likewise include mental acts, for example, judging the pertinence of information or data recovered.

Data Use Behaviour comprises of the physical and mental acts included in consolidating the data found into the individual's current information base. It may include, hence, physical acts, for example, stamping areas in a content to note their significance or centrality, and mental acts that include, for instance, examination of new data with existing information.

1.1 INFORMATION SEARCH (SEEKING) PROCESS (ISP)

Kuhlthau (1991, 1993) added to a model of the data looking for procedure that considers the data seeker's passionate, intellectual and physical encounters at diverse phases of the methodology. The model was spurred by an enthusiasm for researching why understudies act the way they do (Hyldegård, 2006). Kuhlthau (1991) needed to address a perceived crevice between data frameworks and clients common procedure of data trying to enhance data framework outline and aide the experiencing methodology made by mediators. The model has been attempted in two longitudinal studies and has further been weighed in two incomprehensible scale studies (Kuhlthau, 1991, 1993; Hyldegård, 2006).

As indicated by these studies, the ISP can be isolated into six stages, which separate data hunt down, methods for seeking and significance evaluations, while moving the seeker from the starting state of apparent information need to the target state of determination. This advancement is made by a movement of choices made through a brain boggling trade between three spaces of activity: physical (the moves made), subjective (ponderings about the procedure and content) and passionate (feelings experienced) (Hyldegård, 2006). As indicated by Kuhlthau (1991) who expands upon MacMullin (1984) the decisions made by the data seeker are likewise affected by natural requirements, for example, related knowledge, information and interest, data accessible, prerequisites of the issue furthermore, time administered for determination. These common criteria, in any case, have not been considered in the ISP model (Hyldegård, 2006).

According to Kuhlthau (1991, 1993) and Hyldegård (2006) the six stages and their accompanying tasks 5 are

1. Initiation—recognize information need.

2. Selection—identify general topic.

3. Exploration—investigate information on general.

4. Formulation—formulate focus.

5. Collection—gather information pertaining to focus.

6. Presentation—complete.

Table 1.1: Kuhlthau's (1991) model of information search [seeking] process.

Stages	**Initiation**	**Selection**	**Exploration**	**Formulation**	**Collection**	**Presentation**
Feelings (affective)	Uncertainty	Optimism	Confusion/ frustration/ doubt	Clarity	Sense of direction/ confidence	Relief/ satisfaction or disappointment
Thoughts (cognitive)	General / Vague →			Narrowed/ clearer	Increased interest	Clearer or focused
Action (physical)	Seeking background information		Seeking relevant information		Seeking relevant or focused information	

Source: This figure has been prepared by Merve Arıtürk.

In Table 1.1, the connections between the six phases of the ISP display and the interrelationships between the full of feeling and psychological encounters and physical activities are elucidated.

The ISP model demonstrates the data seeker's productive undertakings and exercises of discovering importance from data to amplify his/her condition of information on a specific issue or theme. As per Kuhlthau (1991), the data looking for procedure is started by instability coming about because of an absence of comprehension, a crevice in significance or a restricted development to tackle a certain issue. This will change after some time simultaneously with the seeker getting data and building intending to take care of the issue. Amid the introductory phases of the data looking for procedure the data seeker is normally feeling confounded, baffled and in uncertainty; in the last

stages he or she is ordinarily fulfilled, certain and mitigated (Kuhlthau, 1991). "Plan" speaks to the defining moment of the ISP where sentiments of vulnerability decrease and certainty increments. The errand here is to frame a centre from the data experienced. As of right now and all through whatever remains of the methodology, data looking for ordinarily begins to perish though composing begins to build, additionally connoting that the data seeker has begun entering the "presentation" stage (Kuhlthau, 1991). At the presentation organize the assignment is to finish the inquiry and plan to present or generally utilize the data from the "accumulation" stage. Emotions of alleviation are normal with a feeling of fulfilment if the pursuit has gone well or disillusionment in the event that it has not (Hyldegård, 2006). Further, as the learning state developments to clearer, more focused ponderings about the subject or issue, a relating development is routinely noted in feelings of extended conviction and sureness (Hyldegård, 2006).

Discoveries from Kuhlthau's studies, additionally upheld by Kracker (2002) and Kracker and Wang (2002) appear to affirm a connection between the emotional encounters and the psychological exercises of the model. What the physical exercises of the model is concerned, no solid connection has been found between the data looking for exercises and the intellectual and full of feeling encounters. A large portion of the members in Kuhlthau's studies did not achieve a centred point of view of their subject whenever amid the methodology. In spite of the fact that a critical change in contemplations was discovered (Kuhlthau, 1991), just 50 percentage of the members put forth centred expressions of their theme at midpoint, and at conclusion, just 25 percentage reported they were looking for data on a centred point of view.

As per Kuhlthau (1991), this may prompt the supposition that numerous individuals may enter the presentation or composing stage without an obviously, centred theme.

What's more, it may likewise show that different variables, other than the learning and data looking for exercises, influence the data seekers passionate and subjective encounters. The exploration enthusiasm for Kuhlthau's studies has centred on the people data looking for procedures of developing implying that is the learning and inquiry undertakings of the person. Notwithstanding, different errands, for example, the work undertaking, may influence the procedure too. In an audit of studies on errand

based data seeking, Vakkari (2003) and Hyldegård (2006) recognize assignments as a procedure comprising of a few stages and errand as an apriority condition however left with no further portrayal. In these cases, the purpose of flight is frequently seeking, not the work assignment that has delivered it. In the ISP model the assignment is conceptualized as far as the learning and seeking exercises and procedures included when setting up an examination proposition, which implies that the centre of the model is on these subtasks, instead of the work errand itself and the related exercises and courses of action. At the point when contemplating the data conduct of people going about as gathering individuals, which is centre of the study reported here, the social measurement of the work errand might likewise influence the individual gathering part and his or hers conduct as needs be.

Today, most of the people spend their times on the Internet. So, it is very important to use the Internet ethically. The most important thing is to know what the ethics is. Morals comprehensively elude to the ethical standards administering conduct the guidelines that direct what are correct or off-base. Yet as the beforehand noted samples represent, what constitutes right or wrong is subjective, characterized by gatherings with specific points. Such points underlie the crucial clash between researchers who seek after learning that they trust may advantage society and the privileges of examination members (Neuman, 2011; McBurney & White, 2012; Hegtvedt, 2014). Morals are characterized as "an arrangement of acknowledged convictions which control conduct, particularly such a framework in light of ethics". Ethics is exceedingly subjective, and can shift between societies, organizations, or even individual childhood. A large portion of the frameworks of morals that are set up are of religious or social beginning, and may display totally distinctive ideas of what is correct and what isn't right.

1.2 RESEARCH PURPOSE

The purpose of the research is to analysis of impact on ethical values of Internet usage with a data mining model approach. For the research, data from the students who had taken the survey on Bahcesehir University analysed. In this data gender, family history, school, departments, Internet usage and Internet connection will be examined. Data mining will be utilized for the data. Using the data mining methods, the relationship between ethical Internet usage and faith is analysed.

2. LITERATURE REVIEW

Ethics is defined as “a system of accepted beliefs which control behaviour, especially such a system based on morals.” Ethics is highly subjective, and can vary between cultures, businesses, or even individual upbringing. Many of the systems of ethics that are in place are of religious or cultural origin, and may present completely different concepts of what is right and what is wrong.

Ethics broadly refer to the moral principles governing behaviour—the rules that dictate what are right or wrong. Yet as the previously noted examples illustrate, what constitutes right or wrong is subjective, defined by groups with particular aims. Such aims underlie the fundamental conflict between scientists who pursue knowledge that they hope may benefit society and the rights of research participants (McBurney & White, 2012; Neuman, 2011; Hegtvedt, 2014). In the absence of moral absolutes, professional associations and others craft rules for what is proper and improper regarding scientific inquiry to ameliorate this conflict. The resulting ethics codes reflect philosophical ideas and attempt to bridge to regulatory requirements. The ethical conduct of research pertains to more than data collection involving human participants and encompasses more than simply complying with specific federal regulations protecting such participants.

Exploratory unfortunate behaviour discourses (e.g., Altman & Hernon, 1997; Neuman, 2011) attention on dishonest conduct regularly originating from the weights researchers feel to make their contentions and manufacture their vocations. Inability to recognize the deficiencies of one's exploration or to smother discoveries of "no distinction" may be gently deceptive practices. Taking alternate ways that include distorting or distorting information or exploration strategies, or really concealing negative discoveries, then again, is more unfortunate infringement. Such extortion delays exploratory advances and undermines people in general's trust in investigative attempts. Copyright infringement, another type of research wrongdoing, happens when a scientist guarantees as his or her own work done or composed by others (e.g., associates and understudies) without satisfactory reference. Albeit not actually unlawful if the "stolen" materials are not copyrighted (e.g., exhibiting Ibn Khaldun's words as one's own), unoriginality bargains research honesty, which charges researchers frankly, reasonable, and

deferential of others and to act in ways that don't risk their own or others' expert welfare.

Characterization of these practices as manifestations of logical unfortunate behaviour gets to some degree from philosophical standards like those hidden the sympathy toward the insurance of the welfare of human exploration members. Israel and Hay (2006) break down philosophical ways to deal with how individuals may choose what is ethically right—what ought to be done—in specific circumstances. One approach, focusing on the consequences of behaviour, comes from the writings of utilitarian philosopher John Stuart Mill and invokes a cost–benefit analysis. Essentially, if the benefits that arise from behaviour outweigh the risks or harm associated with that behaviour, then it is morally acceptable. This methodology, on the other hand, makes you wonder of what constitutes an advantage or damage. Interestingly, non-consequential methodologies, starting in progress of Immanuel Kant, recommend that what is correct is steady with human respect and worth. This point of view additionally underlines obligations, regardless of the outcomes in essence.

2.1 GENERAL ETHICAL CONCERNS

Approach parts on examination morals incorporate arrangements of concerns (e.g., Babbie, 2013; McBurney & White 2012; Neuman, 2011), fluctuating long and utilizing distinctive marks. Four interrelated points, then again, for the most part rise: externalization of examination members; potential damages; coercive, exploitative, or nosy practices; and upkeep of security or privacy. Every worry certainly or expressly raises issues of force and trust between the scientist and study member. Researchers procure the power to lead social research by maintaining their moral obligation to secure the hobbies of their members (Neuman, 2011).

2.2 INTERNET ETHICS AND ACADEMIC DISHONESTY

The issues inside Web morals are most usually talked about inside more extensive point of PC morals. Field (1985) characterized PC morals as a field concerned with "approach vacuums" and "applied tangles" in regards to the social and moral utilization of data innovation.

PC morals is an issue with incomprehensible writing; be that as it may, a significant part of the vital work in PC morals is given to proposing calculated structures for comprehension moral issues including PC innovation as expressed (Moor 1985).

According to Floridi and Sanders (2002), PC morals originated from viable concerns emerging regarding the effect of Data and Correspondence Innovations (ICT) on contemporary society. Given the definition, numerous issues, incorporating those including practices in scholastics that can be connected with the Web, ought to likewise be incorporated inside this measurement of Web and PC morals.

Akbulut et al. (2008) have distinguished and measured a few sorts of scholastic untrustworthiness practices that can be connected to the utilization of the Web. Among them are falseness, misrepresentation, unoriginality, wrongdoing and unapproved help. Aside from unapproved help, a large portion of the components measured brought about a sensibly high unwavering quality tests (an over 0.7) (Hair, Rolph, Ronald, & William, 1998). In investigating the elements that add to Web untrustworthiness or e-unscrupulousness measured by the creators, they distinguished and measured a few variables, for example, singular qualities, institutional arrangements, and companion weights as contributing components. Notwithstanding, the exploration did not show the outcomes in light of any relationship investigation keeping in mind the end goal to backing the presence of the connections.

In other research Szabo and Underwood (2004) found that a substantial number of students are involved in academically dishonest activities using the Internet. The alarming figures are found more prevalence for male students and for those in their first year of studies. The study also highlighted the influence of personal and situational factors in contributing towards engaging and avoiding plagiarism.

The ethical use of the Internet for conducting research is also found heavily discussed in the literature. The issues fall under concerns over protecting privacy, protecting harm, and validating data collected from subjects using the Internet (DeLorme, Zinkha & French, 2001; Walther, 2002). In such manner, scholarly untrustworthiness may happen among exploration understudies and also academicians either by getting invalid information or by submitting destructive conduct on the respondents while directing their examination utilizing the Web.

Freestone and Mitchell (2004) distinguished 24 distorted Web practices which were ordered into five components as unlawful exercises, flawed exercises, hacking exercises, human Web exchange and downloading. Despite the fact that the study was directed on era Y, a portion of the discoveries are truly significant to the connection of

the college understudies. This is on the grounds that the vast majority of the college understudies under study are inside the order of era Y.

The audit demonstrates the accessibility of examination, however not adequate, in territories identified with scholarly untruthfulness and the utilization of the Web. Some of these studies likewise give clarifications in the matter of why understudies are tending to take part in such exploitative conduct over the Web environment (Freestone & Mitchell, 2004; Szabo & Underwood, 2004; Walther, 2002).

Unethical Computer Use Behaviour Scale (UECUBS) is developed by Namlu (2004) because PC innovation understudies don't take a necessary course identified with morals amid their four year mull over in Turkey. At the end of this study, it is needed to bring issues to light in understudies and to give an instrument to future PC experts to depend on and accept that it must be acknowledged with their own particular codes of profound quality. While finishing that exploration, two assets are utilized as a guide: Ten Charges proposed in 1992 by The PC Morals Establishment (Rosenberg, 1997) and "Four Moral Issues of the Data Age" proposed (Bricklayer, 1986). When these two resources are combined, six headings are found out: privacy, accuracy, property, accessibility, social impact and net ethics. Dishonest PC conduct can be summed under five titles as per the consequences of the variable examination. These titles are as per the following: protected innovation, social effect, security and quality, net respectability and data uprightness. While developing this scale, four steps are followed. The principal step incorporates a survey of writing, the second includes constituting a thing pool, the third, assessment of master perspectives for substance legitimacy and the fourth and last step covers the factual examination of the things. This scale is applied on university students who are studying at Anadolu University. The study intended to build up a scale (UECUBS) to focus the unscrupulous PC use conduct. For this research, the most important factor is named as intellectual property. It is the most complex issues faced not only in the computer world but in the whole society with computers making it easier than ever before. The second most essential component is named as social effect following the related issues are either socially riotous or happened in social environment. The explanation behind this may be that PC understudies are somewhat delicate about social issues. Five variables together uncover that a fitting behaviour of correspondence and communication is the focus in any moral conduct. Untrustworthy

PC use conduct incorporates a dialog instead of a monolog and any sort of communication in the PC world ought to be defended under the heading of; protected innovation, social effect, wellbeing and quality, net respectability and data uprightness (Namlu, 2004).

In other examination Akbulut, Uysal and Odabaşı (2007) found that an exceptionally noteworthy communication between the project of study and sexual orientation was observed, which demonstrated that the distinction in the middle of guys and females did not take after a comparable example crosswise over diverse projects of study (Akbulut, Uysal and Odabaşı, 2007; Acılar, 2014).

The Internet Ethics Attitude Scale that is developed by Torun (2007) is applied for the students studying in secondary school. The validity and reliability studies of this scale are examined and the significant questions of the survey are selected and adapted to "Ethical Internet Usage Survey".

In this research, the perspectives of the Internet ethics, computer ethics and information ethics are analysed from different articles that contain the concepts of ethics. "The Scale of Ethical Internet" which is created by Torun (2007) is the starting point of this research. While creating this scale, 7 factors, which are shown in Table 2.1, are determined to evaluate.

Table 2.1: Factors of the scale

Sexual Content
Hacking and Disturbing
Plagiarism at Homework
Copyright
Computer Awareness
Violent Games
Virtual Honesty

Source: This figure has been prepared by Merve Arıtürk.

Six factors (sexual content, hacking and disturbing, plagiarism at homework, copyright, violent games, and virtual honesty) and the questions are shown in Table 2.2.

Table 2.2: The questions with the related factors

Name of the factor	*Questions*
Sexual Content	When pornographic web page appears, i close the page immediately
	I have a program which prevents access to pornographic sites.
Hacking and Disturbing	I want to send trojan (viruses) on the Internet to someone for learning private information.
	I have tried to hack someone's password of MSN, Facebook, twitter etc.
	Hacking someone's account of Facebook, twitter makes me happy.
	I use chat rooms to disturb people.
	I would love to pick on people in the chat rooms.
Plagiarism at Homework	The assignment can be submitted as the same with downloading from the Internet
	I would be disturbed when I download someone's homework/assignment from the Internet.
	It is really incorrect that download an assignment whose fee is paid.
	I can give my homework, which is downloaded from the Internet, to the teacher.
	It is wrong to download an assignment from the Internet to get higher points.
	I do copy and paste without specifying the source, while downloading a homework from the Internet
	Submitting a homework which is downloaded from the Internet is cheating.
Copyright	I can download everything from the Internet.
	It is really incorrect that downloading music (mp3) whose fee is not paid from the Internet.
	The music which is downloaded from the Internet should not be shared by the program
	I download music from the Internet without paying the fee.
	It is wrong to download a program which is cracked from the Internet.
	It can be possible to download an antivirus program, which is cracked from the Internet, instead of without paying fee.

Name of the factor	*Questions*
Violent Games	Online games which contain violence shall be forbidden.
	I play online games which contain violence.
	Students should not play online games that contain violence.
Virtual Honesty	Person should behave as you on the Internet.
	Person should not tell a lie in the chat rooms.
	It is not true that a person who is surfing on the Internet exaggerate his/her personal characteristics.
	I can tell small lies on the Internet because people cannot see/know me.
	It can be possible to chat someone with using someone's nickname.
	Reading emails of my girlfriend/boyfriend in a secret makes me happy.
	I'd like to deride another people in the chat rooms.
	It is suitable that students can share pornographic pictures on the Internet.
	The programs which crack programs that need a fee can be shared on the Internet.
	I'd like to introduce myself as a different person on the Internet.
	Hacking someone's computer feels me successful.

Source: This figure has been prepared by Merve Arıtürk.

According to these questions and factors, computer awareness and ethical values & computer usage factors are added to survey. The questions which are related with the computer awareness are shown in the Table 2.3.

Table 2.3: The Questions of Computer Awareness Factor

Name of the factor	*Questions*
Computer Awareness	While surfing on the Internet, I would be careful not to share my personal/private information.
	While surfing on the Internet, I can control my personal information that has not been copied.
	I have some specific information on safe roaming on the Internet
	I can share my email and telephone on the websites.
	Every day, I change my password which is used on the Internet.
	I reflect my troubles by using social media to reach my personal aim.

Source: This figure has been prepared by Merve Arıtürk.

According to Torun's result, women have more ethical behaviour than men. It is emphasized that the women are more sensitive than men about the copyright. However, if the age ratio increases, copyright status has declined. There are some varieties on the ethical behaviour according to the connection of the Internet. Increasing the duration of the Internet usage negatively effects on attitude of the Internet ethics. The relationship between the Internet attitude and income level of the students is found with the survey.
In addition to Torun's research, Kayak investigated the perception of the Internet ethics for university students. As the same with Torun's results, it is found that women are more ethical than men in Kayak's research. There is no relationship between the education level of parents and attitude of the Internet ethics of the student.

After analysing all researches about the attitude of the Internet ethics, the questions of computer usage and ethical values are designed and added to survey.

Table 2.4: The Questions of Ethical Values Factor

Ethical Values and Computer Usage	Do you consult your religious beliefs when you decide to solve social problems on the Internet?
	How often do you talk with your friends about religious issues on the Internet?
	When a disaster occurs in your life or you are upset, do you use Internet to pray?
	I can surf on the Internet to get to more information about my own religion.
	I think the Internet helps me for my religious obligations.
	When I want to learn something about other religions, I use the Internet.
	I use the Internet to remember some prayers.
	To be able to pray, watching videos from the Internet is useful.
	My family /I use the Internet to realize obligatory alms.
	When I enter into nude content websites, I feel like I have sinned.
	I have a look at some websites for consulting my sins.
	I consider myself as one depends on faith.

Source: This figure has been prepared by Merve Arıtürk.

3. DATA, METHODS & FINDINGS

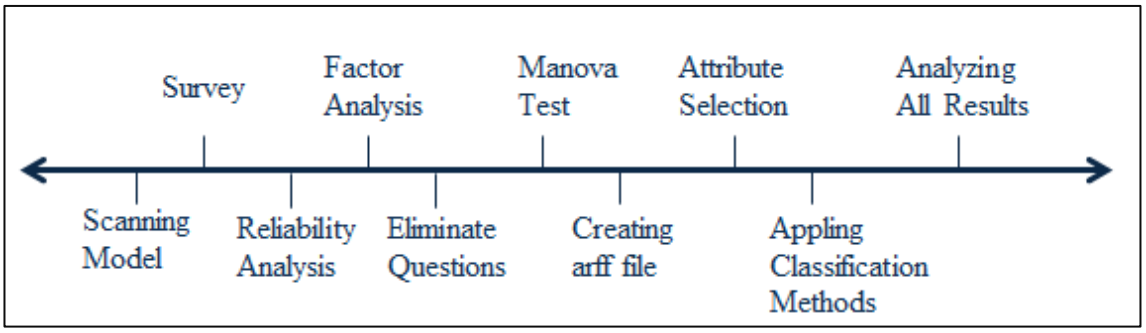
In the methodology section, problem solutions of the research have taken place, and the research model sampling method, data collection tools, data collection and data mining methods and applications in Weka application that are used in data analysis have been considered.

3.1 RESEARCH MODEL

While researching, the scanning model is used with the "Ethical Internet Usage" survey. Weka data mining software has been used for the purpose of applying the data mining models in the study. WEKA data mining software has been written in Java programming language, and can work on different systems such as Windows, Linux and Macintosh (Witten and Frank, 2000).

On the datasets of patients with respiratory problems on Weka data mining model more than one Weka model were run and utilized to choose the model with the highest accuracy. The research model which is used is shown in Figure 3.1.

Figure 3.1: Research Model



Source: This figure has been prepared by Merve Arıtürk.

3.2 DATA COLLECTION AND ANALYSIS

3.2.1 Sampling

The sample of the study consists of students who took the survey in 2012/2013 in Engineering Ethics lecture of Bahçeşehir University and friends who are reached from the personal social networks.

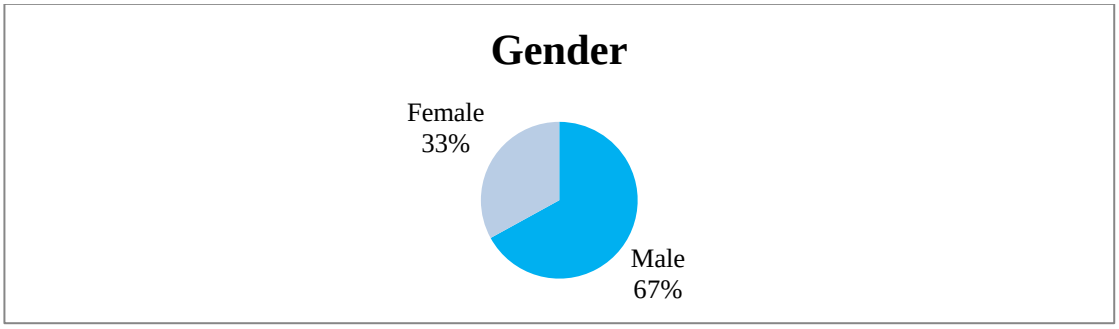
After the written consent was obtained from the students in order to get the test results of the students, gender, school, department, the Internet connection, average Internet usage, and education levels of parents' data values were noted.

In the study, participation was reached approximately 250 students who performed the ethical Internet usage survey.

3.2.2 Data Analysis and Results

Some people (248 people) attended in the "Ethical Internet Usage". 67,742 percent of respondents were male (168 people) and 32,258 (80 people) percent of respondents were female as shown in the Table 3.1.

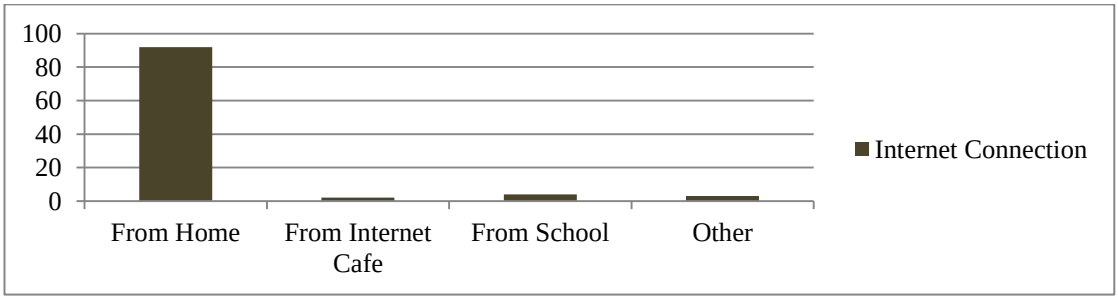
Table 3.1: Distribution of Gender



Source: This figure has been prepared by Merve Arıtürk.

People who attended to the survey connected to the Internet from home, Internet café, school and other areas. 92, 74 (230 people) percent of the respondents are connecting to the Internet from home and the distribution of connecting to the Internet is shown in the Table 3.2. Two people who attended to survey connected to the Internet from Internet café. Ten people connect to the Internet from the school. Six people connect to the Internet from other locations.

Table 3.2: Distribution of the Internet Connection

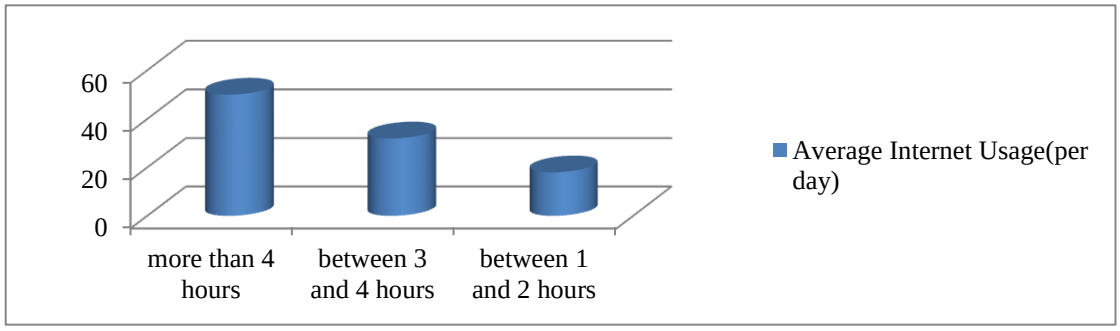


Source: This figure has been prepared by Merve Arıtürk.

When analysing the average Internet usage, 50 percent of the participants (124 people) spend their time on the Internet more than 4 hours per day. Average Internet usage is shown in Table 3.3.

According to average Internet usage field, 124 people who attended to survey used Internet more than 4 hours. Moreover, 81 people spent their time on the Internet between 3 and 4 hours per day. 43 people used Internet between 1 and 2 hours per day.

Table 3.3: Average Internet Usage (per day)



Source: This figure has been prepared by Merve Arıtürk.

Firstly, the data that were collected were analysed through SPSS program and it was determined which questions are more significant and related to this study. Factor analysis was done by the SPSS. The first step of the factor analysis was to calculate Kaisar – Meyer – Olkin (KMO) and Bartlett Test. According to test, the result had been found 0.814. The calculated KMO value was above 0.5. So, that means the volume of the sample was sufficient. The second step of the factor analysis was to investigate the results of KMO & Bartlett Test which was shown in the Figure 3.2. Detailed information about component analysis can be seen in Figure 3.2. The first column is question numbers which are on the survey. Extraction column is the relation number for the question.

Figure 3.2 : Component Loading (Communalities)

	Initial	Extraction
1	1	0,340
2	1	0,441
3	1	0,585
4	1	0,521
5	1	0,723
6	1	0,587
7	1	0,661
8	1	0,652
9	1	0,523
10	1	0,584
11	1	0,609
12	1	0,506
13	1	0,483
14	1	0,579
15	1	0,505
16	1	0,730
17	1	0,710
18	1	0,609
19	1	0,647
20	1	0,625
21	1	0,472
22	1	0,494
23	1	0,595
24	1	0,478
25	1	0,658
26	1	0,514
27	1	0,645
28	1	0,725
29	1	0,734
30	1	0,686
31	1	0,518
32	1	0,295
33	1	0,524
34	1	0,717
35	1	0,336
36	1	0,568
37	1	0,511
38	1	0,533
39	1	0,493
40	1	0,626
41	1	0,586
42	1	0,507
43	1	0,250
44	1	0,248
45	1	0,241
46	1	0,322
47	1	0,390
48	1	0,706
49	1	0,705
50	1	0,326
51	1	0,619
52	1	0,616
53	1	0,357
54	1	0,459
55	1	0,447
56	1	0,394

Source: This figure has been prepared by Merve Arıtürk.

According to communalities table, question 15 had been identified the 50 percentage of the common variance. When the other values in Figure 3.3 had examined, some questions had a common variance values which were under the 0.50. So, the questions which had a low common variance value removed from the survey. Then, factor

analysis was repeated (Removed questions are 1, 2, 13, 21, 22, 24, 32, 35, 39, 43, 44, 45, 46, 47, 50, 53, 54, 55, and 56).

Figure 3.3: Component Analysis after removing insignificant questions

Communalities

	Initial	Extraction
3	1	0,629
4	1	0,560
5	1	0,761
6	1	0,699
7	1	0,738
8	1	0,660
9	1	0,559
10	1	0,624
11	1	0,572
12	1	0,572
14	1	0,637
15	1	0,512
16	1	0,770
17	1	0,694
18	1	0,632
19	1	0,664
20	1	0,634
23	1	0,438
25	1	0,849
26	1	0,670
27	1	0,848
28	1	0,740
29	1	0,769
30	1	0,688
31	1	0,572
33	1	0,575
34	1	0,747
36	1	0,613
37	1	0,506
38	1	0,595
40	1	0,754
41	1	0,773
42	1	0,570
48	1	0,736
49	1	0,765
51	1	0,704
52	1	0,727

Source: This figure has been prepared by Merve Arıtürk.

After removing insignificant questions, the results are examined. The common variance of the question 23 is calculated as 0.438. So, the question 23 is removed and the factor analysis is repeated.

The question 55 has a common variance which explains 44 percentage of this research. To improve the results, some questions are added to survey (Added questions are 13, 21, 22, 24, 54 and 55). The same procedure has done for the factor analysis to calculate Cronbach's Alpha. The results of the reliability analysis are shown in the Figure 3.4.

Figure 3.4: The Results of Reliability Analysis

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,851	,856	44

Source: This figure has been prepared by Merve Arıtürk.

The questionnaire, which has 56 questions at the beginning, is reduced to 44 questions. As a result of factor analysis, it is clearly seen that the questions of survey are gathered in nine factors. After analysing the factors and the questions of this result, the questions that are related with the factor of "sexual content" are removed because of the insignificant values while doing the factor analysis. Before removing the questions, the survey has nine factors. When the questions of sexual content are removed, the number of factors is marked as eight. However, nine factors are found in the analysis. The factors and the questions are investigated then; the factor of "hacking and disturbing" is split into two factors that are "hacking" and "disturbing". The factors and the questions that are obtained from the factor analysis are shown in Figure 3.5.

Figure 3.5: Factors and Questions

	FACTORS								
	Hacking	Disturbing	Plagiarism at Homework	Copyright	Internet Addiction	Violent Games	Virtual Honesty	Computer Awareness	Ethical Values & Computer Usage
QUESTIONS	3	6	8	15	21	25	28	39	48
	4	7	9	16	22	26	29	40	49
	5	37	10	17	23	27	30	41	51
	33	42	11	18	24		31		52
	38		12	19			37		54
			13	20					55
			14						

Source: This figure has been prepared by Merve Arıtürk.

After eliminating the questions, the arff file is created for data mining. The file format is shown in Figure 3.6.

Figure 3.6: The Format of arff File

```
@RELATION merve

@ATTRIBUTE int_kullanimi  NUMERIC
@ATTRIBUTE cins   NUMERIC
@ATTRIBUTE int_bagl  NUMERIC
@ATTRIBUTE anne_egitim   NUMERIC
@ATTRIBUTE baba_egitim NUMERIC
@ATTRIBUTE cin_ice_geomean    NUMERIC
@ATTRIBUTE bilg_kors_rah_et_geomean NUMERIC
@ATTRIBUTE odev_asirma_geomean       NUMERIC
@ATTRIBUTE telif_hakki_geomean         NUMERIC
@ATTRIBUTE int_bagimliligi_geomean    NUMERIC
@ATTRIBUTE sid_iceren_oyun_geomean NUMERIC
@ATTRIBUTE sanal_durustluk_geomean  NUMERIC
@ATTRIBUTE bilg_farkindaligi_geomean NUMERIC
@ATTRIBUTE din_bilg_kullanimi_geomean        NUMERIC

@DATA
4.00,1.00,1.00,5.00,5.00,5.00,4.32,5.00,1.51,5.00,5.00,3.61,2.00,1.23
5.00,0.00,1.00,4.00,4.00,5.00,5.00,4.09,2.38,4.47,1.59,4.23,3.46,1.47
4.00,0.00,2.00,5.00,5.00,2.00,2.51,4.69,1.12,1.68,2.00,2.30,2.83,1.15
//There are many lines..
4.00,0.00,1.00,4.00,5.00,3.16,5.00,4.26,2.57,3.94,1.82,4.70,2.00,1.37
5.00,0.00,3.00,3.00,4.00,3.46,3.29,4.36,1.35,3.36,3.42,3.79,2.83,2.75
5.00,0.00,1.00,5.00,5.00,2.00,3.98,3.28,1.00,2.00,1.00,2.10,3.16,2.55
5.00,0.00,1.00,3.00,4.00,3.87,2.37,2.99,1.41,1.00,2.52,2.99,3.00,1.95
4.00,0.00,1.00,5.00,5.00,5.00,5.00,4.69,3.38,4.47,3.00,4.68,1.73,1.47
```

Source: This figure has been prepared by Merve Arıtürk.

3.3 FEATURE SELECTION AND IMPLEMENTING CLASSIFICATION METHODS

In the study, data mining classification methods were used for the detection and prediction of the ethical Internet usage of the people. Before applying the classification methods, firstly, all values are filtered with the NumericToNominal which is located on unsupervised attribute. After filtering, the main step of the classification which is selecting attributes is done. While selecting, attribute evaluator is selected InfoGainAttributeEval and search method is selected Ranker. The results of the selection are shown in Figure 3.7.

Figure 3.7: The Results of Attribute Selection

Ranked attributes:

0.59259	12	virtual_honesty_geomean
0.50146	9	copyright_geomean
0.42514	14	ethical_values_compt_usage_geomean
0.40617	8	plagirism_hw_geomean
0.38266	10	computer_addiction_geomean
0.30829	11	violent_games_geomean
0.20251	6	sexual_content_geomen
0.15843	7	hacking_disturbing_geomean
0.04587	13	computer_awareness_geomean
0.02701	1	computer_usage
0.02022	5	edu_level_father
0.01788	3	internet_connection
0.01507	2	gender
0.0097	4	edu_level_mother

Selected attributes: 12, 9, 14, 8, 10, 11, 6, 7, 13, 1, 5, 3, 2, 4 : 14

Source: This figure has been prepared by Merve Arıtürk.

According to the results, 6 attributes that are education level of mother and father, gender, connection and duration of the Internet and computer awareness are removed because of their ranks. 8 attributes that are virtual honesty, copyright, ethical values computer usage, plagiarism, computer addiction, violent games, sexual content and hacking and disturbing are selected After eliminating and selecting the attributes, some classification algorithms which are shown in Table 3.4 are applied on data.

Table 3.4: Applied classification methods

	Classification Methods
1	BayesNet
2	NaiveBayes
3	NaiveBayesUpdateable
4	RBF Network
5	SMO
6	KStar
7	OneR
8	Ridor
9	ID3

Source: This figure has been prepared by Merve Arıtürk.

3.3.1 BayesNet

In the data mining classification study on ethical Internet usage test, BayesNet classification model results reveal that 211 out of 248 ethical Internet usage data was classified accurately (85.08 percent), while 37 of them (14.92 percent) were classified inaccurately.

Figure 3.8: The results of BayesNet Algorithm

```
=== Summary ===

Correctly Classified Instances         211               85.0806 %
Incorrectly Classified Instances        37               14.9194 %
Kappa statistic                          0.6599
Mean absolute error                      0.1767
Root mean squared error                  0.3358
Relative absolute error                 39.6323 %
Root relative squared error             71.1456 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.747     0.097      0.795     0.747     0.77       0.903    0
                 0.903     0.253      0.876     0.903     0.89       0.903    1
Weighted Avg.    0.851     0.201      0.849     0.851     0.85       0.903

=== Confusion Matrix ===

   a   b   <-- classified as
  62  21 |   a = 0
  16 149 |   b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

Looking at the area of Detailed Accuracy by Class in Figure 3.8 in BayesNet model, values in Class 0 and 1 in ROC Area are observed to be close to 1. BayesNet model seems to have yielded successful results.

3.3.2 NaiveBayes

Another classification model, NaiveBayes model, was applied on the ethical Internet usage data, the results of NaiveBayes classification model classified 216 out of 248 (87.1 percent) of the data accurately, and 32 of the data (12.9 percent) inaccurately. When the Naïve Bayes algorithm is applied, the root mean squared error is found 0.3239. This algorithm is better than BayesNet algorithm.

Figure 3.9: The results of NaiveBayes algorithms

```
=== Summary ===

Correctly Classified Instances         216               87.0968 %
Incorrectly Classified Instances        32               12.9032 %
Kappa statistic                          0.7013
Mean absolute error                      0.1822
Root mean squared error                  0.3239
Relative absolute error                 40.8525 %
Root relative squared error             68.6359 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.747     0.067      0.849     0.747     0.795      0.912    0
                 0.933     0.253      0.88      0.933     0.906      0.912    1
Weighted Avg.    0.871     0.191      0.87      0.871     0.869      0.912

=== Confusion Matrix ===

   a   b   <-- classified as
  62  21 |   a = 0
  11 154 |   b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

In Figure 3.9, some detailed information about the Naive Bayes Network can be seen. In case of accurately testing the Naive Bayes model, cross- validation test was applied and on the data cluster of 216 participants using Internet ethically and different detailed information on Naive Bayes was obtained.

3.3.3 NaiveBayesUpdateable

When another classification model, NaiveBayesUpdateable model, was applied on the ethical Internet usage data, the results of this classification model classified 216 out of 248 (87.09 percent) of the ethical Internet usage data accurately, and 32 of the data (12.90 percent) inaccurately. These results are the same with the results of NaiveBayes modelling.

Figure 3.10: The Results of NaiveBayesUpdateable Algorithm

```
=== Summary ===

Correctly Classified Instances         216               87.0968 %
Incorrectly Classified Instances        32               12.9032 %
Kappa statistic                          0.7013
Mean absolute error                      0.1822
Root mean squared error                  0.3239
Relative absolute error                 40.8525 %
Root relative squared error             68.6359 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.747     0.067      0.849     0.747     0.795      0.912    0
                 0.933     0.253      0.88      0.933     0.906      0.912    1
Weighted Avg.    0.871     0.191      0.87      0.871     0.869      0.912

=== Confusion Matrix ===

   a   b   <-- classified as
  62  21 |   a = 0
  11 154 |   b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

3.3.4 Radial Basis Function (RBF) Network

In RBF Network, k-means clustering algorithm is used to provide the basis functions and regressions. If class attributes are nominal, it uses the given number of clusters each class. It standardizes all numeric attributes to zero mean and unit variance.

When another classification model, RBF Network model, was applied on the ethical Internet usage data, the results of this classification model classified 207 out of 248 (83.47 percent) of the ethical Internet usage data accurately, and 41 of the data (16.53 percent) inaccurately.

Figure 3.11: The results of RBF Network

```
=== Summary ===

Correctly Classified Instances         207               83.4677 %
Incorrectly Classified Instances        41               16.5323 %
Kappa statistic                          0.6321
Mean absolute error                      0.1715
Root mean squared error                  0.3861
Relative absolute error                 38.4591 %
Root relative squared error             81.8159 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.771     0.133      0.744     0.771     0.757      0.851    0
                 0.867     0.229      0.883     0.867     0.875      0.852    1
Weighted Avg.    0.835     0.197      0.836     0.835     0.835      0.851

=== Confusion Matrix ===

   a   b   <-- classified as
  64  19 |   a = 0
  22 143 |   b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

3.3.5 Sequential Minimal Optimization (SMO)

SMO actualizes John Platt's successive insignificant streamlining calculation for preparing a bolster vector classifier.

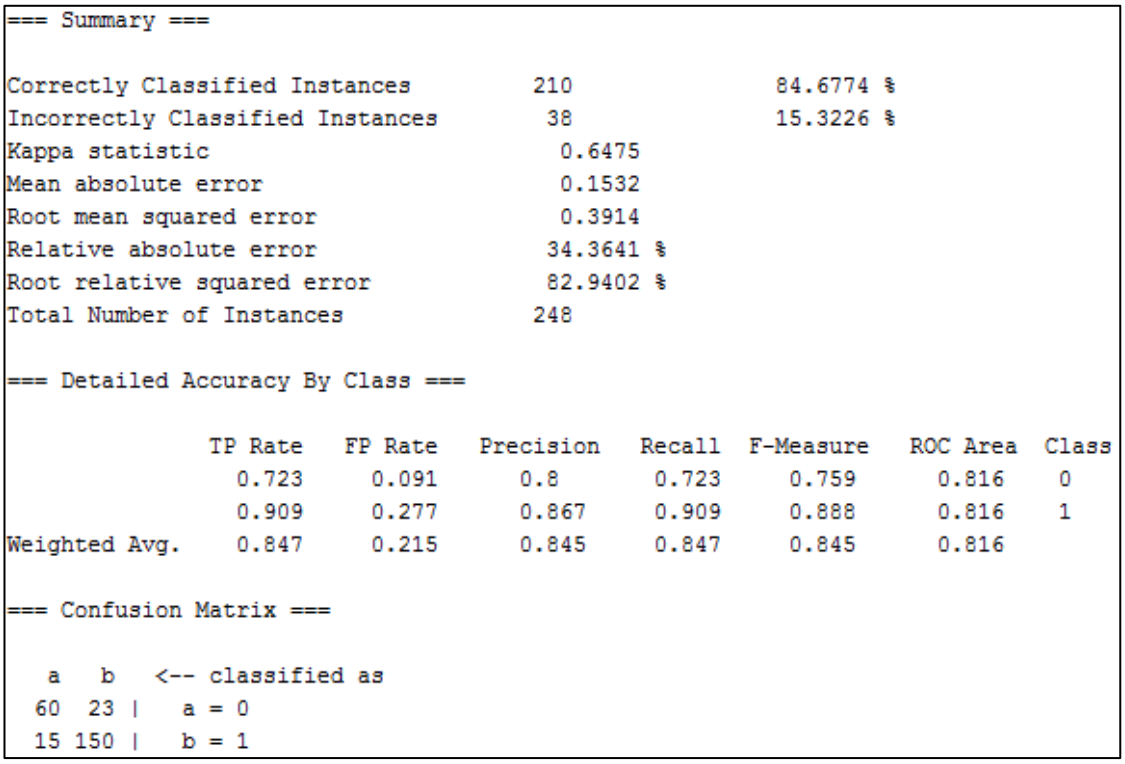
This execution internationally replaces every missing esteem and changes ostensible qualities into parallel ones. It additionally standardizes all traits of course (All things considered the coefficients in the yield are taking into account the standardized information, not the first information - this is critical for deciphering the classifier.).

Multi-class issues are tackled utilizing pairwise grouping (1-versus 1 and if logistic models are assembled pairwise coupling as indicated by Hastie and Tibshirani, 1998).

To get legitimate likelihood appraisals, utilize the choice that fits logistic relapse models to the yields of the bolster vector machine. In the multi-class case the anticipated probabilities are coupled utilizing Hastie and Tibshirani's pairwise coupling strategy.

At the point when another grouping model, SMO model, was connected on the moral web utilization information, the consequences of this arrangement model ordered 210 out of 248 (84.68 percent) of the moral web use information precisely, and 38 of the information (15.32 percent) incorrectly.

Figure 3.12: The results of SMO



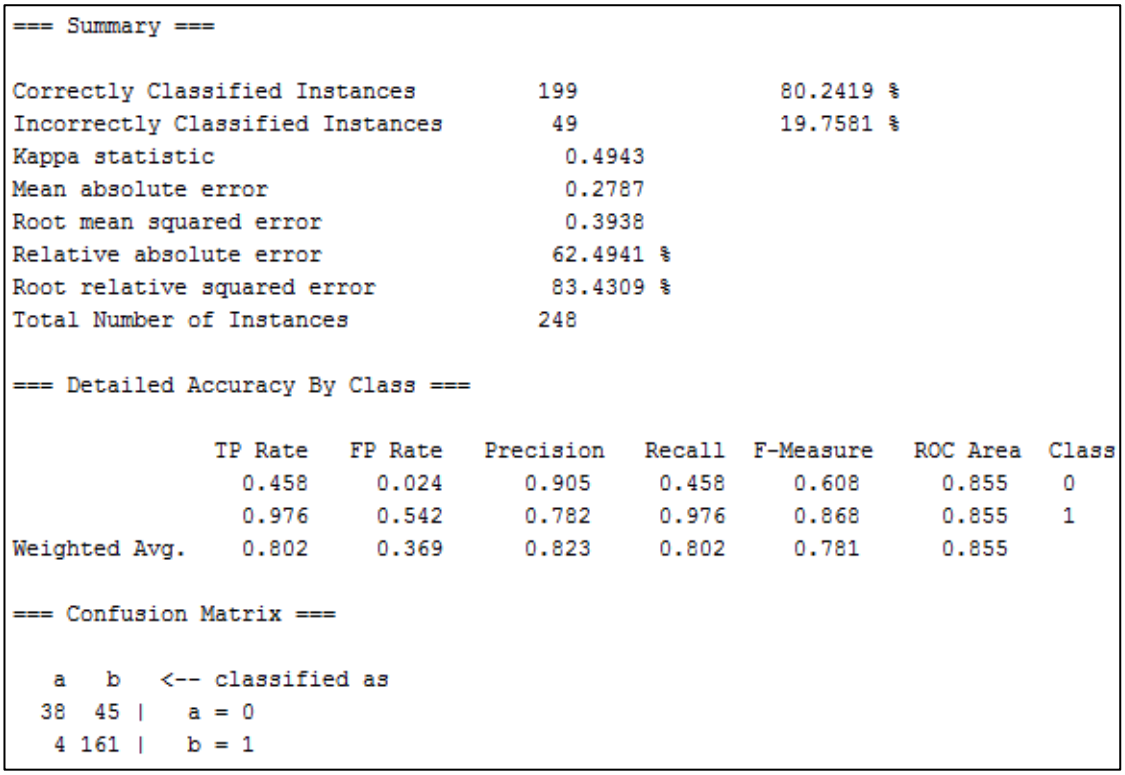
```
=== Summary ===

Correctly Classified Instances         210               84.6774 %
Incorrectly Classified Instances        38               15.3226 %
Kappa statistic                          0.6475
Mean absolute error                      0.1532
Root mean squared error                  0.3914
Relative absolute error                 34.3641 %
Root relative squared error             82.9402 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.723     0.091      0.8       0.723     0.759      0.816    0
                 0.909     0.277      0.867     0.909     0.888      0.816    1
Weighted Avg.    0.847     0.215      0.845     0.847     0.845      0.816

=== Confusion Matrix ===

   a   b   <-- classified as
  60  23 |   a = 0
  15 150 |   b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

3.3.6 KStar

When KStar model was applied on the ethical Internet usage data, the results of this classification model classified 199 out of 248 (80.24 percent) of the ethical Internet usage data accurately, and 49 of the data (19.76 percent) inaccurately.

Figure 3.13: The results of KStar



```
=== Summary ===

Correctly Classified Instances         199               80.2419 %
Incorrectly Classified Instances        49               19.7581 %
Kappa statistic                          0.4943
Mean absolute error                      0.2787
Root mean squared error                  0.3938
Relative absolute error                 62.4941 %
Root relative squared error             83.4309 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.458     0.024      0.905     0.458     0.608      0.855    0
                 0.976     0.542      0.782     0.976     0.868      0.855    1
Weighted Avg.    0.802     0.369      0.823     0.802     0.781      0.855

=== Confusion Matrix ===

   a   b   <-- classified as
  38  45 |   a = 0
   4 161 |   b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

3.3.7 OneR

When OneR model was applied on the ethical Internet usage data, the results of this classification model classified 177 out of 248 (71.37 percent) of the ethical Internet usage data accurately, and 71 of the data (28.63 percent) inaccurately. The root mean squared error is found 0.5351.

Figure 3.14: The results of OneR

```
=== Summary ===

Correctly Classified Instances         177               71.371  %
Incorrectly Classified Instances        71               28.629  %
Kappa statistic                          0.4019
Mean absolute error                      0.2863
Root mean squared error                  0.5351
Relative absolute error                 64.2066 %
Root relative squared error            113.3712 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.723     0.291       0.556     0.723     0.628      0.716    0
                 0.709     0.277       0.836     0.709     0.767      0.716    1
Weighted Avg.    0.714     0.282       0.742     0.714     0.721      0.716

=== Confusion Matrix ===

   a   b   <-- classified as
  60  23 |   a = 0
  48 117 |   b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

3.3.8. Ridor

When Ridor model was applied on the ethical Internet usage data, the results of this classification model classified 183 out of 248 (73.79 percent) of the ethical Internet usage data accurately, and 65 of the data (26.21 percent) inaccurately. The root mean squared error is found 0.512.

The result of Ridor modelling is shown in Figure 3.15.

Figure 3.15: The results of Ridor

```
=== Summary ===

Correctly Classified Instances         183               73.7903 %
Incorrectly Classified Instances        65               26.2097 %
Kappa statistic                          0.3198
Mean absolute error                      0.2621
Root mean squared error                  0.512
Relative absolute error                 58.7807 %
Root relative squared error            108.4751 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.337     0.061      0.737     0.337     0.463      0.638    0
                 0.939     0.663      0.738     0.939     0.827      0.638    1
Weighted Avg.    0.738     0.461      0.738     0.738     0.705      0.638

=== Confusion Matrix ===

   a   b   <-- classified as
  28  55 |   a = 0
  10 155 |   b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

3.3.9 ID3

When ID3 model was applied on the ethical Internet usage data, the results of this classification model classified 104 out of 248 (41, 94 percent) of the ethical Internet usage data accurately, and 31 of the data (12, 5 percent) inaccurately. The root mean squared error is found 0.4792.

The results of ID3 modelling are shown in the Figure 3.16.

Figure 3.16: The results of ID3

```
=== Summary ===

Correctly Classified Instances          87               35.0806 %
Incorrectly Classified Instances        30               12.0968 %
Kappa statistic                          0.2825
Mean absolute error                      0.2564
Root mean squared error                  0.5064
Relative absolute error                128.4779 %
Root relative squared error            164.867  %
UnClassified Instances                 131               52.8226 %
Total Number of Instances              248

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.387     0.128      0.522     0.387     0.444      0.539    0
                 0.872     0.613      0.798     0.872     0.833      0.613    1
Weighted Avg.    0.744     0.484      0.725     0.744     0.73       0.593

=== Confusion Matrix ===

  a  b   <-- classified as
 12 19 |  a = 0
 11 75 |  b = 1
```

Source: This figure has been prepared by Merve Arıtürk.

3. 4 PREDICTION PHASE

The data mining research conducted with Weka program was run for 248 participants. When the results obtained were considered, it is seen that the BayesNet classification model is the one with the highest accuracy. In terms of ROC area results, the best result is BayesNet model. All applied classification methods and their results are shown in Table 3.5.

Table 3.5: All methods and Results

	Classification Methods	CCI	ICI	KS	MAE	RMSE	RAE	RRSE	TNI
1	BayesNet	211	37	0.6599	0.1767	0.3358	39.63	71.14	248
2	NaiveBayes	216	32	0.7013	0.1822	0.3239	40.85	68.63	248
3	NaiveBayes Updateable	216	32	0.7013	0.1822	0.3239	40.85	68.63	248
4	RBF Network	207	41	0.6321	0.1715	0.3861	38.45	81.81	248
5	SMO	210	38	0.6475	0.1532	0.3914	34.36	82.94	248
6	KStar	199	49	0.4943	0.2787	0.3938	62.49	83.43	248
7	OneR	177	71	0.4019	0.2863	0.5351	64.20	113.37	248
8	Ridor	183	65	0.3198	0.2621	0.512	58.78	108.47	248
9	ID3	87	30	0.2825	0.2564	0.5064	128.47	164.86	248

Source: This figure has been prepared by Merve Arıtürk.

Some methods are selected randomly to calculate the sensitivity, specificity, precision and correctness. While calculating the sensitivity, firstly the confusion matrix is analysed according to the methods with the given formulas in Table 3.6.

Table 3.6: The formulas

	Formula
Sensitivity	TP / (TP + FN)
Specificity	TN / (FP + TN)
Precision	the value which is getting from the detailed accuracy by class table
Correctness	the percentage of the correctly classified instance

Source: This figure has been prepared by Merve Arıtürk.

Table 3.7: Randomly selected methods and their values

Classification Methods	Sensitivity	Specificity	Precision	Correctness
BayesNet	0.795	0.876	0.795	0.85
NaiveBayes	0.849	0.88	0.849	0.87
NaiveBayes Updateable	0.849	0.88	0.849	0.87
RBF Network	0.74	0.88	0.744	0.83
SMO	0.8	0.867	0.8	0.84
K*	0.9	0.821	0.905	0.80
One R	0.55	0.835	0.556	0.71
Ridor	0.736	0.738	0.737	0.73
ID3	0.36	0.797	0.522	0.35

Source: This figure has been prepared by Merve Arıtürk.

It is assumed that the closer the value in ROC area to 1, the better estimation that data mining model provides.

When the classification models in data mining were considered, the classification model with the worst result is REP Tree and Ibk model. Best results received in classification model are NaiveBayes, LBR and NBTree classification model, because they run the bayes algorithm. Logistic, Decision Stump, DTNB, LogitBoost and SMO classification models did not yielded good results as BayesNet and Naive Bayes classification models.

4. DISCUSSIONS & CONCLUSION

Before survey results are evaluated, the time which is spent on the Internet is expected to increase because of the development of the Internet and reducing the age that is accessed to the Internet. This situation causes to increase the number of the people that are malicious. Therefore, the number of cases that are clicking a malicious links while surfing on the Internet, chatting with the harmful people has increased. If the person spends so much time on the Internet, he or she has a big risk to face with the malicious people. When all these conditions are evaluated, the time that is spent on the Internet affect people adversely. After the research, the theory which is thought before the study is correct. The Internet usage is increased dramatically. The research shows that if a person uses the Internet too much, s/he will be behaving unethically.

It is expected that the ethical behaviour of the person will be increased with the educational level of the parents. According to previous research about the relationship between gender and the ethical behaviour will be the same. So, the ethical behaviour of women will be greater than the men. After results are examined, educational level of the parents and ethical behaviour has a direct proportion. If the parents have a high educational level, the person also has a high ethical behaviour. The ethical behaviour has differences by the gender. Women are more ethics than men.

The survey is mostly answered by students studying at the university. According with my personal experiences, it is expected that the factors of the survey that are plagiarism and copyright has a low mean.

When doing factor analysing with Statistical Package for the Social Sciences (SPSS) and selecting the attributes with WEKA, ethical values and computer usage factor has the lowest average, but this factor is used firstly. So, there is no theory about this topic. The aim of this research is to find the relation on the ethical values and computer usage. After this factor, the copyright has the lowest average. This result is expected.

To improve this study, firstly, the survey should be applied on whole country. Then the results will be examined. After that, it will be diversified with the religious. Then, the

results which are found with the country factor and the religious factor will be compared.

The practice of ethical behaviour in the digital media concepts is discussed in different dimensions, but the relationship between thought and behaviour is not analysed in the digital media platforms. The concepts of belief that are to give a value, an application, an interpretation and review to the actual dimension are discussed in this research.

When the research is analysed in terms of gender, it has found that women have a higher perception of ethical than men. 91.25 percentages of the women which are participated in the study are marked ethics, while 79.17 percentages of the men are marked ethics. These results are consistent with the results of Khazanchi (1996), Bisset and Simpson (1999), Mert (2003), Uysal (2006), Torun (2007) and Kayak (2011).

When the mother's education level and the ethical behaviour are examined, some differences are detected. In the research, 81 percentages of the participants whose mother is graduated from the university are signed as ethics. If the mother's education level is the high school, 88.4 percentages of the respondents are ethical. 88 percentages of the participants whose mother has a primary school graduates have an ethical behaviour. This finding is similar with the Torun's result (2007). According to his result, the education level of mother has a big role for the ethical behaviour of person. However, Kayak (2012)'s result says that there is no relationship between the education level of the mother and the ethical behaviour of person.

In this study, the education level of father is also analysed and the results is examined that there is no relationship between the level of the father education and the ethical behaviour. 80.35 percentages of the respondents whose father is graduated from the university have a high ethical behaviour. 85 percentages of the participants whose father is graduated from the high school are marked as an ethical. The same results have found by Torun (2007) and Kayak (2012).

When the Internet connection is examined, 50 percentages of the respondents use the Internet more than 4 hours per day. 32 percentages of the respondents spend their time between 3 and 4 hours per day. 17 percentages of the participants connect to the Internet between 1 and 2 hours per day. This result shows that the hour which is spent on the Internet is increasing day by day because of the development of the Internet in daily life. Spending time on the Internet has a radical change according to Kayak's (2012) result.

In order to develop this research, the scale will be created with the results which are significant. If there is a new value or data about the research, the scale shows the user behaviour that is assuming based on users' demographical information.

Based on these results, organizations about the computer ethics and the Internet ethics will be prepared to train people for increasing the sensitivity of the behaviour. The organization has 2 parts that are theoretical and practical experiences.

There will be lectures about the ethics and this lecture will start from the elementary school. So that, person will grow with the ethics. The topics of lecture will be ethics, computer ethics and the Internet ethics.

There will be some rules which are about ethical attributes. If person does not follow the rules, people will be punished like the traffic rules. This research will be done more comprehensive to fix unethical behaviour.

REFERENCES

Books

Andress, Jason, & Winterfeld Steve, 2014. *Cyber Warfare*, 2. Edition: Techniques, Tactics and Tools for Security Practitioners

Hair, J. F. J., Rolph, E. A., Ronald, L. T., & William, C. B. (1998). Multivariate data analysis (5th ed.). *Upper Saddle River, NJ: Prentice Hall*.

Witten, I., Frank, E., & Hall, M., 2011. *Data Mining*. 3. Edition. USA : Morgan Kaufmann Publishers.

Periodicals

Acılar, A., Aydemir, M. (2009). Freshman Students' Attitudes toward Issues of Computer Ethics, Education Proceedings, 2, **pp**. 213-222.

Akbulut, Y., Sendag, S., Birinci, G., Kilicer, K., Sahin, M. C., & Odabasi, H. F. (2008). Exploring the types and reasons of Internet-triggered academic dishonesty among Turkish undergraduate students: Development of Internet-Triggered Academic Dishonesty Scale (ITADS), *Computers & Education*, **51**(1), 463–473.

Altman, E. & Hernon, P. (Eds.). (1997). Research misconduct: Issues, implications, and strategies. *Greenwich, CT: Ablex*.

Babbie, E. (2013). The practice of social research. *Belmont, CA: Wadsworth*.

Bissett, A., & Shipton, G. (1999). An investigation into gender differences in the ethical attitudes of IT professionals. ETHICOMP)), Rome.

Blackwell, A. (2011) Human Computer Interaction – Lecture Notes. *Cambridge Computer Science Tripos, Part II*

Carey, J., D. Galletta, J. Kim, D. Te'eni, B.Wildemuth, and P. Zhang. (2004). "The Role ofHuman–Computer Interaction in Management Information Systems Curicula:A Call to Action." Communications of the Association for Information Systems, Vol. 13, pp. 357–379.

Caroll, J. M. (1994), Making Use a Design Representation. *Communications of the ACM*, **37**, 29 - 35.

Carroll, J. M. (1996), Becoming social: Expanding scenario-based approaches in HCI, Behavior & Information Technology, **15**(4), **pp**. 266-275

DeLorme, D. E., Zinkha, G. M., & French, W. (2001). Ethics in the Internet: Issues related with qualitative research. *Journal of Business Ethics*, **33**, 271–286.

Floridi, L., & Sanders, J. W. (2002). Mapping the foundationalist debate in computer ethics. *Ethics and Information Technology*, **4**, 1–9

Freestone, O., & Mitchell, V. W. (2004). Generation Y attitudes towards e-ethics and internet related misbehaviors. *Journal of Business Ethics*, **54**, 121–128

Gupta, R. 2012. Human Computer Interaction – A Modern Overview. *Int. J. Computer Technology*. **3** (5), 1736 - 1740

Hand, D.J., 1998. Data Mining. *The American Statistician.* **52**, pp. 112-118.

Hegtvedt, A. Karen, 2014. Ethics and Experiments, **25, pp** 23-51.

Hyldegård, J., 2006. Collaborative information behaviour—exploring Kuhlthau's Information Search Process model in a group-based educational setting. Information Processing & Management, **42**(1), **pp** 276–298.

Israel, M., & Hay, I. (2006). Research ethics for social scientists: Between ethical conduct and regulatory compliance. *Newbury Park, CA: Sage*.

Jacobs, P., 1999. Data Mining. *Harvard Management Update.* **4** (10), s. 8.

Karray, F., Alemzadeh, M., Saleh, J. A., Arab, M. N. (2008). 2008Human-Computer Interaction: Overview on State of the Art. *International Journal on Smart Sensing and Intelligent System.* 1 (1).

Khazanchi, D. (1996). *Unethical behavior in information systems: the gender factor.* Journal of Bussiness Ethics, 14(741-9).

Kracker, J. (2002). Research anxiety and students perceptions of research: an experiment. Part I. Effect of teaching Kuhlthau's ISP model. Journal of the American Society for Information Science and Technology, 53(4), 282–294.

Kracker, J., & Wang, P. (2002). Research anxiety and students perceptions of research: an experiment. Part II. Content analysis of their writings on two experiences. Journal of the American Society for Information Science and Technology, 53(4), 295–307.

Kuhlthau, C. C. (1991). Inside the search process: seeking meaning from the users perspective. Journal of the American Society for Information Science, 42(5), 361–371.

Kulthau, C. (1993). Seeking meaning. Norvood, NJ: Ablex.

Kitler, R. ve Wang, W., 1998. The Emerging Role of Data Mining. *Solid State Technology***. 42** (11), p. 45.

MacMullin, S. E., & Taylor, R. S. (1984). Problem dimensions and information traits. The Information Society, 3, 91–111.

Mason, R. (1986). Four ethical issues of the information age. *MIS Quartely*, **10**(1), 5 - 11

McBurney, D. H., & White, T. L. (2012). Research methods. *Belmont, CA: Wadsworth/Thomson.*

Moor, J. H. (1985). What is computer ethics? Metaphilosophy, **16**(4), 266–275.

Namlu, A. G., & Odabaşı, H. F. (2004). Unethical computer using behavior scale: A study of reliability and validity on Turkish university students. *ScienceDirect*

Neuman, W. L. (2011). Basics of social research: Qualitative and quantitative approaches. *Boston: Pearson.*

Nielsen J., (1994). Usability Engineering, *Morgan Kaufman*, San Francisco.

Nor, S.A.K., Nurul, H.A.Z.,Yakinah M.N.(2009).Exploring the relationship between Internet ethics in university students and the big five model of personality, *Elsevier*

Özekes, S., 2003. Data Mining Models and Application Areas. *İstanbul Commerce University Journal of Science.* **3** (3), pp. 65-82

Rosenberg, R. S. (1997). The social impact of computers. *San Diego: Academic Press*

Savaş, S., Topaloğlu, N. & Yılmaz, M., 2012. Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri. *İstanbul Commerce University Journal of Science.* **11** (21), pp. 1-23.

Shearer, C., 2000. The New Blueprint for Data Mining. *Journal of Data Warehousing.* **5** (4), pp. 13-23

Shneiderman B. and Plaisant C., (2004). Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th edition), *Pearson/Addison-Wesley*, Boston.

Szabo, A., & Underwood, J. (2004). Cybercheats: Is information and communication technology fuelling academic dishonesty? Active Learning in Higher Education, **5**(2),180–199.

Te'eni D., (2006) "Designs that fit: an overview of fit conceptualization in HCI", in P. Zhang and D. Galletta (eds), Human-Computer Interaction and Management *Information Systems: Foundations*, M.E. Sharpe, Armonk.

Te'eni D., Carey J. and Zhang P., (2007).Human Computer Interaction: Developing Effective Organizational Information Systems, *John Wiley & Sons, Hoboken*.

Vakkari, P. (2003). Task-based information searching. In B. Cronin (Ed.), Annual review of information science and technology 2003 (Vol. 37, pp. 413–464). Medford, NJ.

Other References

Akgöbek, Ö., & Çakır, F., *Veri Madenciliğinde Bir Uzman Sistem Tasarımı.* [online], http://ab.org.tr/ab09/kitap/akgobek_cakir_AB09.pdf [accessed 15 April 2014].

Albayrak, M., 2008. EEG Sinyallerindeki Epileptiform Aktivitenin Veri Madenciliği Süreci ile Tespiti. *Doktora Tezi*. Sakarya : Sakarya University, FBE.

Cambridge Advanced Learner's Dictionary. Definition of ethic noun from Cambridge dictionary online. Cambridge dictionary [online] http://dictionary.cambridge.org/ictionary/british/ethic?q¼ethic [accessed 03 July 2014].

Competence Software. Section 406 – Code of ethics for senior financial officers. Sarbanes Oxley simplified, http://www.sarbanesoxleysimplified.com/sarbox/compact/htmlact/sec406.html; [accessed 03 July 2014].

Dener, M., Dörterler M., & Orman, A., 2009. *Açık kaynak kodlu veri madenciliği programları: weka'da örnek uygulama.* [online], http://ab.org.tr/ab09/bildiri/42.pdf [accessed 10 May 2014].

Howell, David C. 2009. SPSS Manual [online] http://www.uvm.edu/~dhowell/fundamentals7/SPSSManual/SPSSLongerManual/SPSSChapter1.pdf [accessed 10 October 2014]

İnan, O., (2003). Veri Madenciliği. *Yüksek Lisans Tezi*. Konya : Selçuk University FBE.

Kalikov, A., (2006). Veri madenciliği ve bir e-ticaret uygulaması. *Master Thesis*. Ankara : Gazi University, FBE.

Neill, M., 1985. *Teknolojinin gelişimi.* [online], http://www.diyadinnet.com/YararliBilgiler-214&Bilgi=bilim-ve-teknoloji [accessed 5 June 2014].

Patterson, D., Turner, D., Concepcion, A., & Lynch, R. 2008. Performance comparison of the data reduction system. *Proceedings of the SPIE symposium on defense and security.* USA : Orlando

Payalan, F., 2013. *Weka ile veri madenciliği.* [online],

http://yfpayalan.blogspot.com.tr/2013/03/weka-ile-veri-madenciligi.html [accessed 10 May 2014].

Piramuthu, S., 1998. *Evaluating feature selection methods for learning in data mining applications.* [online], http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=648324 [accessed 5 May 2014].