T.R.
VAN YÜZÜNCÜ YIL UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

EXAMINE THE INTERNET ADDICTION LEVELS OF STUDENTS IN TÜRKİYE AND IRAQ COMPARATIVELY WITH THE MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) METHOD

Ph.D. THESIS

Hewa Ghafor HASSAN
Supervisor: Prof. Dr. Murat KAYRİ
Second Supervisor: Ass. Prof. Dr. Hikmet ŞEVGIN

VAN – 2023

T.R.
VAN YÜZÜNCÜ YIL UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

EXAMINE THE INTERNET ADDICTION LEVELS OF STUDENTS IN TÜRKİYE AND IRAQ COMPARATIVELY WITH THE MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) METHOD

Ph.D. THESIS

Hewa Ghafor HASSAN

VAN – 2023

ACCEPTANCE AND APPROVAL PAGE

This thesis entitled Examine the Internet Addiction Levels of Students in Türkiye and Iraq Comparatively with the Multivariate Adaptive Regression Splines (MARS) Method” presented by Hewa Ghafor Hassan under supervision of Prof. Dr. Murat Kayri in the department of Computer and Instructional Technology Education and second supervisor Ass. Prof. Dr. Hikmet Şevgin in the department of Educational Sciences has been accepted as a Ph.D. thesis according to Legislations of Graduate Higher Education on 15/08/2023 with unanimity / majority of votes members of jury.

Chair: Prof. Dr. Fevzi ERDOĞAN	Signature:
Member: Prof. Dr. Murat KAYRİ	Signature:
Member: Prof. Dr. Sıddık KESKİN	Signature:
Member: Prof. Dr. SİNAN ÇALIK	Signature:
Member: Doktor Öğretim Üyesi Erdal EKER	Signature:

This thesis has been approved by the committee of The Institute of Natural and Applied Sciences on ..……../.........../…...….. with decision number of ....

Signature

………………….

Director of Institute

ETHICAL DECLARATION

I declare that all the information in this thesis has been obtained and presented within the framework of ethical behavior and academic rules, and that in this thesis, which has been prepared in accordance with the thesis writing rules, all kinds of statements and information that do not belong to me have been fully cited.

Signature

Hewa Ghafor HASSAN

ABSTRACT

EXAMINE THE INTERNET ADDICTION LEVELS OF STUDENTS IN TÜRKİYE AND IRAQ COMPARATIVELY WITH THE MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) METHOD

HASSAN, Hewa Ghafor
Ph.D. Thesis, Department of Statistics
Supervisor: Prof. Dr. Murat KAYRİ
Second Supervisor: Ass. Prof. Dr. Hikmet ŞEVGIN
August 2023, 132 Pages

In recent years, the internet has changed modern life in many ways. Internet technologies have transformed the whole field of education, health, defense industry, and health into a new format. However, the internet has changed social life and human-human, human-machine interaction, and these interactions have created addictions at various levels. Internet addiction is accepted as a current and serious disease that can affect mental health. The aim of this study is to model the factors affecting the addiction levels of students from two countries (Türkiye, Iraq) using the Multivariate Adaptive Regression Extensions method. The primary purpose of this study is to monitor the performance of MARS, which is one of the multivariate statistical methods, on a data set and to test its predictive ability on the data types used in the research. On other hands, the main purpose of the study is to test the applicability of MARS on the variable types and data characteristics used in the research and to reveal the effectiveness of this method to the researchers who will work with this type of data. The secondary aim of the study is to model the factors that trigger internet addiction, which is seen as an important disease and danger of this age, through some measurement tools. Here, it was also desired to examine whether the factors affecting internet addiction differ between cultures. Therefore, the sample of the study consisted of university students from Türkiye and Iraq.

The MARS method used in the thesis is a non-parametric data mining technique. MARS puts a knot at each point where linearity will end in the relationship between the variables and creates appropriate functions for each linear extension it obtains. In this respect, it can model the cause-effect relationship in a more rational space.

The data set for the thesis study consisted of the internet addiction Scale and the data obtained from the questionnaires containing demographic/personal information. The sample of the study consists of 2235 students, 1220 (427 boys and 793 girls) from Türkiye and 1015 (465 boys and 550 girls) from Iraq, using the random sampling method. The data within the scope of the study were benefited from the SPM 8.2 program. As a result of MARS analysis; while 9 (nine) basic functions were obtained for students in Türkiye, the number of basic functions was observed as 20 (twenty) for the data set consisting of Iraq. In the model created by MARS; It has been determined that the most important factor affecting the internet addiction level of both Turkish and Iraqi students is "daily internet usage time". The details of the study are in the Findings and Conclusion sections of the thesis. We hope that the findings on the factors affecting the level of internet addiction with the MARS method will contribute to the literature, and it is recommended that such data mining methods be applied to life-oriented data together with the theory.

Keywords: Data Mining, Internet Addiction, MARS, Statistics

ÖZET

TÜRKİYE VE IRAK'TAKİ ÖĞRENCİLERİN İNTERNET BAĞIMLILIK DÜZEYLERİNİN ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON UZANIMLARI (MARS) YÖNTEMİ İLE KARŞILAŞTIRMALI OLARAK İNCELENMESİ

HASAN, Hewa Ghafor
Doktora Tezi, İstatistik Anabilim Dalı
Danışman: Prof. Dr. Murat KAYRİ
İkinci Danışman: Doktor Öğretim Üyesi Hikmet ŞEVGİN
Ağustos 2023, 132 Sayfalar

Son yıllarda internet, modern hayatı birçok açıdan değiştirmiştir. Internet teknolojileri; eğitim, sağlık, savunma sanayi, endüstri ve sağlık alanının tümünü yeni bir formata dönüştürmüştür. Bununla birlikte internet, sosyal hayatı ve insan-insan, insan-makine etkileşimini değiştirmiş ve bu etkileşimler çeşitli düzeylerde bağımlılıklar oluşturmuştur. İnternet bağımlılığı, ruh sağlığını etkileyebilecek güncel ve ciddi bir durum olarak kabul edilmektedir. Bu çalışmanın amacı, Çok Değişkenli Uyarlanabilir Regresyon Uzanımları yöntemi ile iki ülke (Türkiye, Irak) öğrencilerinin bağımlılık düzeylerini etkileyen faktörleri modellemektir. Bu çalışmanın birincil amacı; çok değişkenli istatistik yöntemlerinden olan MARS'ın uygulamalı olarak bir veri seti üzerinden performansını izlemek, araştırmada kullanılan ve ölçeklerden elde edilen veri tipleri üzerindeki tahminleme yeteneğini test etmektir. Araştırmada kullanılan değişken tipleri ve veri karakteristiği üzerinde MARS'ın uygulanabilirliğini test etmek ve bu tipteki verilerle çalışacak araştırmacılara bu yöntemin etkililiğini ortaya koymak çalışmanın esas amacı olarak görülmüştür. Çalışmanın ikincil amacı ise; bu çağın önemli bir hastalığı ve tehlikesi olarak görülen internet bağımlılığını tetikleyen faktörleri bazı ölçme araçları üzerinden modellemektir. Burada, internet bağımlılığını etkileyen faktörlerin kültürler arasında farklılık gösterip göstermediği de incelenmek istenmiştir. Bu nedenle, çalışmanın örneklemini Türkiye ve Irak'tan üniversite öğrencileri oluşturmuştur.

Tez çalışması kapsamında kullanılan MARS yöntemi, parametrik olmayan bir veri madenciliği yöntemidir. MARS, değişkenler arasındaki ilişkide doğrusallığın biteceği her bir noktaya bir düğüm (knote) koyup, elde ettiği her bir doğrusal uzanım için uygun fonksiyonlar oluşturmaktadır. Bu yönüyle neden-sonuç ilişkisini daha rasyonel bir uzayda modelleyebilmektedir.

Tez çalışmasına ait veri setini İnternet Bağımlılık Ölçeği ve demografik/kişisel bilgileri içeren anket formlarından elde edilen veriler oluşturmuştur. Araştırmanın örneklemi, 1220'si (427 erkek ve 793 kız) Türkiye ve 1015'i (465 erkek ve 550 kız) Irak'tan olmak üzere seçkisiz örnekleme yöntemiyle toplamda 2235 öğrenciden oluşmaktadır. Çalışma kapsamındaki verilerin analizinde, SPM 8.2 programından istifade edilmiştir. MARS analizi sonucunda; Türkiye'deki öğrenciler için 9 (dokuz) temel fonksiyon elde edilirken, Irak'tan oluşan veri seti için ise temel fonksiyon sayısı 20 (yirmi) olarak gözlenmiştir. MARS tarafından oluşturulan modelde; hem Türkiyeli hem de Iraklı öğrencilerin internet bağımlılık düzeyini etkileyen en önemli faktörün "günlük internet kullanım süresi" olduğu tespit edilmiştir. Çalışmaya ait detaylar ise tezin Bulgular ve Sonuç bölümlerinde yer almaktadır. MARS yöntemi ile internet bağımlılık düzeyini etkileyen faktörlere ilişkin bulguların literatüre katkı sağlayacağını ümit ediyor ve bu tür veri madenciliği yöntemlerinin teorisi ile birlikte yaşama dönük verilere uygulamalı olarak yansıması tavsiye edilmektedir.

Anahtar kelimeler: İnternet Bağımlılığı, İstatistik, MARS, Veri Madenciliği

ACKNOWLEDGMENT

I want to offer my profound gratitude to the following individuals, without the help whom this research would not have been possible:

First, I like to start by thanking my Ph.D. advisors, Prof. Dr. Murat KAYRİ and Ass. Prof. Dr. Hikmet ŞEVGİN, for their constant support, direction, and inspiration. Their knowledge, suggestions, and insights have been really helpful in guiding my research and helping me improve my research techniques.

Additionally, I would like to express my appreciation to my committee members, Prof. Dr. Sıddık KESKİN, Prof. Dr. Fevzi ERDOĞAN, and Prof. Dr. Murat KAYRİ, for their insightful comments, idea, and suggestions that helped me construct and polish this thesis. My research and argument development greatly benefited from their knowledge and helpful feedback.

I would like to express my gratefulness to Assoc. Prof. Yener Altun, in the Department of Business Administration at Van Yuzuncu Yil University, who has provided me with a supportive and collaborative research environment. Their feedback and camaraderie have been invaluable throughout this process.

Finally, I want to thank my family and friends for their love, encouragement, and support throughout my Ph.D. journey. Their steadfast confidence in me has aided me as a constant source of motivation and inspiration.

I want to thank everyone for their contributions to this study and their encouragement throughout my Ph.D. program.

2023

Hewa Ghafor HASSAN

CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

SYMBOLS AND ABBREVIATIONS

Some symbols and abbreviations used in this thesis are presented below, along with their descriptions.

Symbols	**Description**
$\boldsymbol{\alpha}$	Constant Value, Intercept
n	Number of Observation or Sample size
r	Simple Correlation
SD	Standard Deviation
$\boldsymbol{\bar{x}}$	Mean
$\boldsymbol{X_i}$	Represents Independent Variables
$\boldsymbol{Y_i}$	Represents Dependent Variables

Abbreviations	**Description**
ADHD	Attention Deficit Hyperactivity Disorder
ANOVA	Analysis of Variance
AUC	Area Under Curve
BS	Basis Function
CART	Classification And Regression Trees
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CM	Confusion Matrix
CVS	Computer Vision Syndrome
DQ	Diagnostic Questionnaire
EFA	Exploratory Factor Analysis
FN	False Negative
FP	False Positive
FPR	False Positive Rates
GCV	Generalized Cross Validation

Abbreviations	Description
GFI	Goodness of Fit Index
GLM	Generalized Linear Model
IA	Internet Addiction
IAT	Internet Addiction Test
IRABI	Internet Related Addictive Behavior Inventory
LR	Linear Regression
LSE	Least Squares Estimation
MARS	Multivariate Adaptive Regression Splines
MLE	Maximum Likelihood Estimation
NFI	Normed Fit Index
NNFI	Non-Normed Fit Index
PIU	Pathological Internet Use
PIUS	Pathological Internet Use Scale
RMSEA	Root Mean Square Error of Approximation
ROC	Receiver Operator Characteristic Curve
RSI	Repetitive Strain injury
RSS	Residual Sum of Squares
SPSS	Statistical Package for Social Sciences
SRMR	Standardized Root Mean Square Residual
TN	True Negative
TP	True Positive
TPR	True Positive Rates

LIST OF APPENDIX

Page

1. INTRODUCTION

Firstly, one should define what is meant by the Internet. It is generally considered to be an international network of computers initially instigated, according to Schneider et al. (2006), by the US Department of Defence in the 1960s. The main reason was the military. As is widely acknowledged, now it is mainly the general public as well as others who benefit and use it daily. It has given rise to mass communication, social networking, data streaming, and suchlike, as well as news and personal broadcasting applications such as Telegram.

According to recent statistics by Rosliza et al. (2018), around 40% of the world's inhabitants – or over 3 billion people – had access to the Internet in 2016, while Karma et al. (2018) cite a figure of 3.5 billion Internet users in 2018, an increase of 500 million over two years. The same research noted that the majority were young people. Clement (2020) provides figures as of July 2020, namely 4.57 billion Internet users or 59% of the worldwide population. The rate of Internet access is thus increasing dramatically. Although there are many possible factors relating to why Internet use is increasing, for example, ease of access to cheaper and more user-friendly technology, the recent COVID-19 pandemic could be a significant reason for the steep increase as citizens were often locked down and unable to contribute in the daily face-to-face life.

On the one hand, the internet is very useful for people due to it can help us to gain and transfer information and knowledge, communicate more widely, and suchlike, however on the other hand, according to Buchholz (2009), concern has been growing over some of the risks including addiction, especially among students in tertiary stage education per (Chou and Hsiao, 2000).

To this end, we can realize that there is the concept of Internet Addiction (IA) how it could be applied to its use. There are several factors around this, including if IA is a realistic type of addiction compared to others, for example, drugs; there are disagreements about the relevant diagnostic criteria and also how exactly IA is defined per se. Initially, IA started out being based on established principles and frameworks, including by Greenfield (1999) – pathological gambling – technology such as TV addiction – Griffiths (1995) and substance abuse and dependence – (Anderson, 2001). The Diagnostic and Statistical Manual of Mental Disorders (4th Edition), condensed and frequently referred to as DSM-IV from the American Psychiatric Association (1994), has been used in some studies as the baseline measurement and reference point whereby IA has been classified. Lam et al. (2009) found that it is students who are more likely to be the most vulnerable type of people to suffer from IA, nevertheless there are few other studies relating to treatment or providing any solution.

As mentioned per the DSM-IV, IA can be classified and identified as akin to pathological gambling as well as substance abuse in terms of the following:

- preoccupation with the Internet to the extent that it causes major academic or professional, relationship, or other social issues;
- maladaptive preoccupation with it;
- duration of use is far longer than originally intended or wished for at the outset.

Nonetheless, research still needs to be done to recognize whether IA exists simply as symptomatic of some other mental health condition; whether Internet use causes psychological detriment; or what it actually is, according to Zboralski et al. (2009). Having noted that, it is generally agreed per Griffiths (1995; 2000), that using the Internet can generate an addiction to it. Kaltiala-Heino et al. (2004) found that in their empirical research with students that some had become addicted, and in other epidemiological studies looking at school-aged children, there was a wide range in IA. Milani et al. (2009) found a prevalence of 36.7% in Italy, while by contrast, Tsitsika et al. (2009) found only 1%. Naturally, there are many questions as to why the results differ therefore widely given similar types of society, and that may be due to the actual classification criteria, sampling strategies, and how the data was collected. It does, however, pose interesting questions.

Various researchers have cited and been interested in IA, as well as teachers, health professionals, and suchlike. This has given rise to correspondingly various studies, many of which will be examined in the Literature Review below. They range from Lin and Tsai (1999), who looked at interpersonal and relationship difficulties, to general professional and academic performance difficulties per Kim et al. (2010). Furthermore, two strands of thought seem to be prevalent, namely, as mentioned previously that IA could be a symptom of an existing mental disorder Yen et al. (2007) while per Lin and Tsai (1999) in their research on Taiwanese high-school students, it was not observed as a heavy influence on relationships with their peers in reality and at the same time opening up the possibility to meet new friends online and thus being a positive thing.

Researchers such as Gunuc and Kayri (2010) are among those who have studied the apparent phenomenon. They devised an internet addiction Level as a diagnostic tool with which to describe levels of potential internet addiction. Their tool comprised 35 questions with a minimum score overall of 35 and a maximum of 175, since it is designed to be a scale. Their original study based on results from 754 respondents, and whereas it is scalar, those with a higher number are correspondingly seen to have greater issues with dependency. Their results showed that children who came from families of higher socioeconomic status fared worse i.e., they were more likely to suffer IA. The results were as follows: 26.7% were clearly addicted; similar percentages (27.8%) were not addicted, while almost the majority, at 45.5%, were observed to be at danger of addiction. Furthermore, a study by Kayri (2010) have been indicated that the daily Internet use time on average and also the intended use of the Internet is to be considered as the main influences for internet addiction among the Turkish secondary school students, which lead to found that 26.7% of student determined to have internet addiction and 45.5% of students were recognized as being in a risk group.

Internet addiction can be attributed to many factors, as mentioned in several previous studies over time. However, one major issue over the last two years or more cannot be ignored, namely the COVID-19 pandemic. It goes without saying that internet usage rose specifically during the early phases of the pandemic, with the concomitant increase in remote working and lack of contact with family members separated from each other through lockdowns and suchlike. (Tayyar Şaşmaz et al., 2014; Anusha et al., 2016; Sowndarya and Pattar, 2018; Garmah and Rida, 2020) have examined the rate and incidence of IA among university-age students in various countries; Nagaur (2020) and Dong et al. (2020) add to these studies by pointing out that the pandemic has predisposed students to a higher risk of IA due to taking their studies online away from campuses, for example.

The aim of the study is to develop a scale for internet addiction and examine the factors that can affect the addiction status of individuals, moreover, how addiction to the internet affects students in different in both sample (Türkiye and Iraq), and lastly how much of an overall issue internet addition is among these students in general.

Added to these stated objectives above, the study has various aims. One complicating issue is that there is no generally accepted definition as to what constitutes internet addiction, hereafter stated by the acronym "IA". It is unclear what the main universally accepted criteria have been adopted for measuring the various rates of addiction. Of course, it is universally known by society that internet usage is prevalent and undeniably a necessity of modern life and that the young and students in particular, given the type of hierarchical and traditional society that exists in both Türkiye and Iraq, are much heavier users of the internet in particular social-media and the various messaging and video type applications for example. Nevertheless, given that, one of the keys aims of the study has come from this, is that no actual empirical data exist as to whether internet use affects students positively or negatively or indeed as to the actual rates of IA themselves. Such data could assist in an evidence-based approach to positive interventions mitigating any potential effects, be they psychological, social, or otherwise, caused by IA.

To begin with, the literature review will be used to determine some kind of initial definition for IA. Having established this definition will be the basis for what survey takes place on the two student cohorts at Van Yuzuncu Yil University and Soran University. Lastly, something called "Multivariate Adaptive Regression Splines" (hereinafter denoted "MARS") is going to be the main mathematical modelling tool for analysis of the obtained data.

Therefore, to re-state the aims of the current study in a clearer way, it will firstly draw on the literature review and MARS to obtain a concrete definition of IA; secondly, examine across different faculties the rate of IA in both cohorts – Van and Soran; thirdly attempt to identify how students are influenced by various factors in the use of the internet.

2. LITERATURE REVIEW

2.1. Usage of the Internet

2.1.1. History of the Internet

As mentioned in the Introduction, it is commonly accepted that the Internet as a concept date from the 1960s and is now firmly established in modern society. It is a generally accepted fact that Internet usage and access have increased exponentially over time Schneider et al. (2006). Even by 2010, according to the Internet World Stats (2011) 28.7% of the global population used the Internet.

Schneider et al., (2006) can be cited as defining the Internet as a large-scale network of computers used to link global populations in order to access and disseminate information by various means for financial transactions and educational purposes.

2.1.2. Nature of the Internet

According to Greenfield (1999), a standard protocol is used to create the Internet Connection Network – or Internet. Given the simulating content, speed of connection, and suchlike, it is these things which could give rise to IA. As a result of the Internet facilitates such ease of communication, almost limitless sources and types of information and content, and is almost universally available, together with more user-friendly applications and interfaces means that access is really very simple and continues to get simpler. This was per Chou (2001), and twenty or more years later, the situation has clearly developed from that – a universally known fact. Since many Internet applications are interactive, Young (1998) stated that it is these features that impact IA and not the Internet itself.

2.1.3. The Nature of those who Use the Internet

In terms of trying to explain why certain types of people become addicted to the Internet, various studies have been conducted. These include Suler (2000), who posited that personal needs, mitigation of loneliness, and the need for a sense of belonging might be to blame, while others such as Chou et al. (1999) determined that it is the sense of escapism through what can be almost impersonal online communications that provide personal satisfaction. Morahan-Martin and Schumacher (2000) detailed their findings that Internet users find it a comfortable and convenient means for social exchange, which means that those with IA feel more satisfied with this way of communication.

2.1.4. Internet Use Worldwide

The price of smartphones is steadily falling, and digital infrastructure is getting better, which is causing an increase in connectedness among international internet audiences. Many formerly underserved areas are now reaping the benefits of mobile internet access. As of 2022, East Asia contributed to around 1.2 billion of the world's internet users, next to Southern Asia with just over one billion. Five billion people were online as of April 2022, according to statistics as it has been shown in be (Figure 2.1).

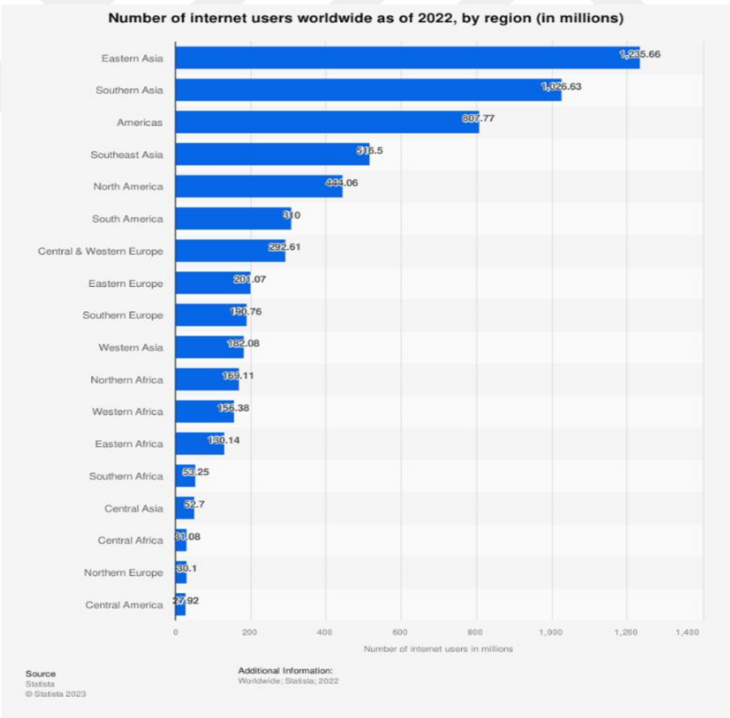


Figure 2.1 World internet users by regions in 2022

According to a Kepios investigation Simon (2021a) the number of internet users has increased by more than a factor of two over the previous ten years, rising from 2.18 billion at the beginning of 2012 to 4.95 billion at the beginning of 2022.

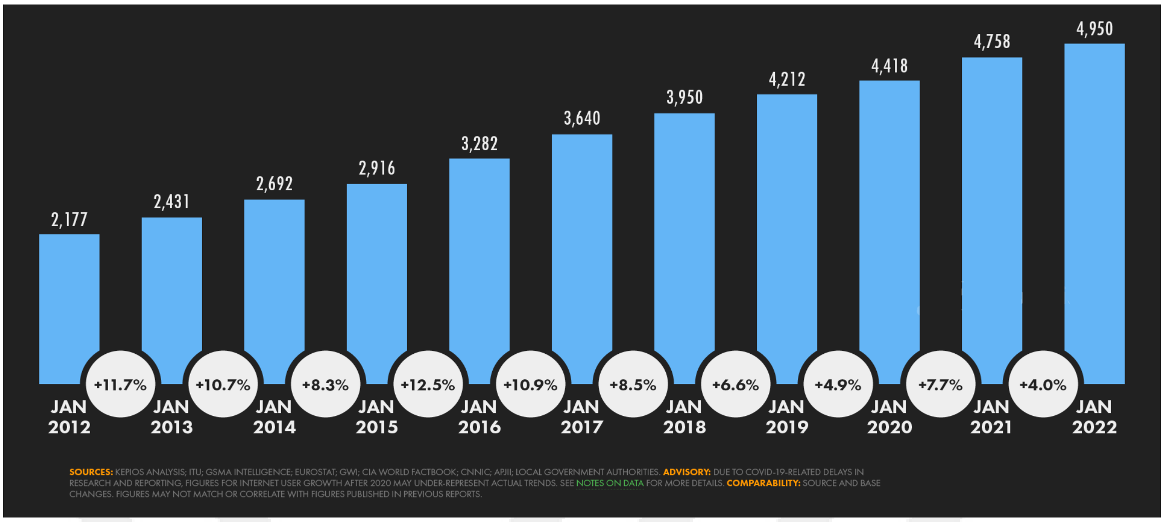


Figure 2.2 Number of internet users worldwide from 2012 to 2022 (Simon, 2021a)

According to the most recent data, there have been an additional 192 million internet users during the previous 12 months, translating to a mere 4.0 percent annual growth in 2021. Although there is a strong suspicion that these numbers do not accurately reflect the rise in internet users over the past year, this lower growth rate is possibly the result of difficulties with data collection and reporting throughout the ongoing COVID-19 pandemic.

2.2. Prevalence of Adolescent Internet Addiction

According to Hull and Proulx (2022) internet addiction is considered to affect 6% of people worldwide, even though only 39% of people have access to the internet. It appears that the prevalence of internet addiction varies significantly between countries. In Southeast Asia, between 0 and 47.4% of students reported serious problematic Internet use (PIU) or an addiction to the Internet, as opposed to 7.4% to 46.4% who noted they may have a possible internet addiction (Balhara et al., 2018).

There have been various studies into IA within the adolescent age groups. Pallanti et al. (2006) discovered that age is of no consequence when it comes to IA, and neither is it the case with social class. Nevertheless, research often tends to focus on adolescents, given that they are at a critical period of development in life. To this end Van Rooij et al.

(2010) showed that in the West, the Internet was the most typical way to spend leisure time for these adolescents, and they demonstrated that it is even more important than watching television for those over the age of 14, while for those aged 11 to 15, in general, it is the most common leisure time pursuit. Furthermore, Lin et al. (2009) showed that it is older adolescents (as exemplified above with the Dutch study) who are more using the Internet than younger ones.

A study by Sayed et al. (2022) identified that 38.5% of Egyptian university students had internet addiction symptoms, in order to transition from being adolescents to adults, university students must go through a transitional stage. A lot of tension, depression, and worry may be brought on by this shift for some of them. Researchers (Amr et al., 2018; Kavitha, 2021) anticipated that students at Al-Mansura University from Egypt would experience depression and anxiety at prevalence rates of 28.3% and 21.2%, respectively, in their studies. Furthermore, according to statistics published by Ibrahim et al. (2012), 37.6% of Asyut University students experienced moderate to severe depression symptoms. Additionally, a study conducted by AlQahtani et al. (2015) in Saudi Arabia estimated that among university students there, the prevalence rates of depression, anxiety, and stress were 48.1%, 58.9%, and 40.4%, respectively.

According to Ko et al. (2009), they claim that 19.8% of global adolescents are addicted to the Internet, with the first usage beginning at pre-teen and teenage levels. Lin and Yu (2008) cited the fact that about half of all young people were online, and further, Forrester Research's 2005 study into adolescent Internet usage revealed that of those aged between 12 and 17 in the USA, most were online day-to-day and for around eleven hours a week. Additionally, a Taiwanese study from 2011by the Taiwan Network Information Center TNIC (2011) demonstrated that of a total number of Internet users of 15 million, those under 20 comprised about 1/6th (2.86 million) with the 12 – 15 age range having Internet participation of 98% and the 16 – 20 range at 95.6%.

Research undertaken by Pallanti et al. (2006) showed that in their sample, 5.4% of the suffered from IA. Their study looked at 275 students whose age was an average of 16.67 +/- 1.85 years, with an approximately equal distribution of the sexes – 52.4% male and 47.6% female. While Internet access in Italy was shown to be less than elsewhere, it was observed that in China, IA was becoming a thoughtful problem between their adolescents. China Internet Network Information Centre CNNIC (2006) had a study that presented 123 million individuals online, and of this number, 14.9% were those under 18 years old. From this, they drew the conclusion that IA is becoming more serious. IA among Taiwanese students per Chou and Hsiao was at a rate of 5.9% while almost double the number of their Chinese counterparts per Wu and Zhu – 10.6% were suffering IA, (Cao et al., 2007). Other interesting statistics related to problematic usage of the internet come from Chebbi et al. (2005), who discovered that in Taiwan, 73.7% of online crime is theft and 20.2% is fraud. This is related to online gaming, and of course, the age of offenders is adolescent or teenager (Wan and Chiou, 2007).

A South Korean study undertaken by Park (2008) demonstrated that internet usage is most common among adolescents compared to other age groups. They found that 97.3% of those between 6 and 19 in that country were online in 2005. They also found that in a sample of 903 youths in South Korea, 10.7% of the adolescents showed high levels on the IA scale and thus were at prime risk for IA. It underlined the general observation that the Internet almost forms one of the corner stones of South Korean society.

2.3. Theoretical Framework for Internet

A 2008 study by Douglas et al. (2008) proposed a key conceptual framework for IA. They observed that certain push and pull factors are at play. Push factors are generally considered as inner needs and thus primary internal motivation. On the other hand, pull factors can include how attractive the Internet is perceived to be. There is thus a relationship between push and pull factors and thus the strictness of IA, including Internet overdoing and overdependence on it.

As previously mentioned, while negative effects of IA can include deteriorating relationship and professional activity, financial and academic issues, and may possibly result in deviant behaviour types, it is also quite possible for someone with IA to recognize their behaviour patterns and take steps to mitigate or improve it. These can be termed control strategies, which can reduce addictive behaviours and even the temptation to commit an online crime, which would be one kind of deviant behaviour. Of course, some people have a propensity towards such deviant behaviour compared with others. Thus, it is necessary to examine previous behaviour patterns on the psychological level.

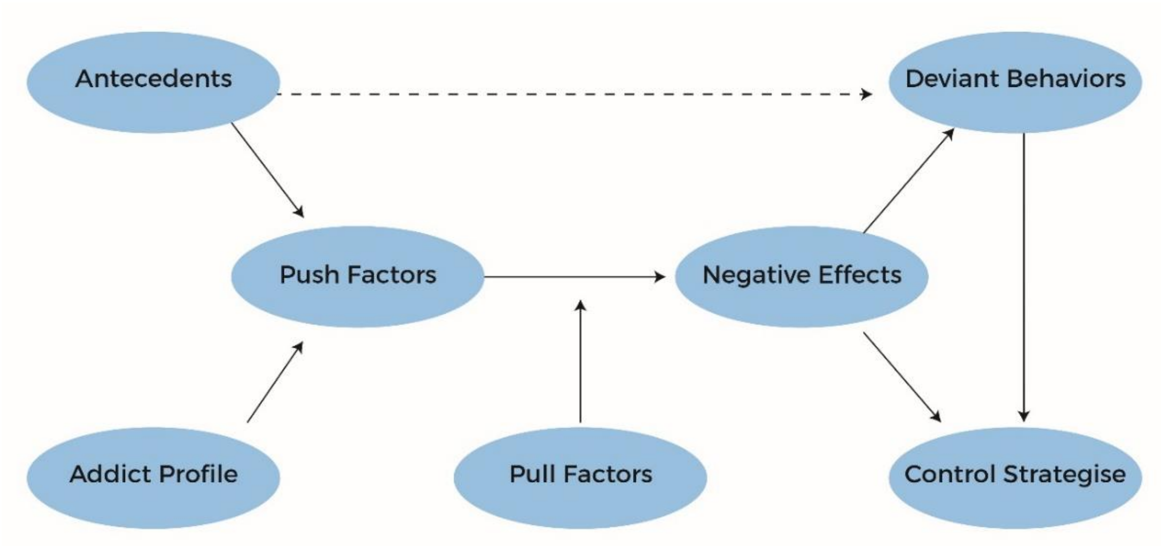


Figure 2.3 Conceptual Model of internet addiction (Douglas et al., 2008)

One key study by Preece (2000), corroborating findings obtained by Wellman and Gulia (1999), showed that one major factor behind such addiction is socialization. Further, Grohol (2005), as cited by Douglas et al. (2008), mentioned about the sociability of the Internet leading to people spending excessive time on gaming, chat, forums, and suchlike.

Pathological Internet Use (PIU) has been proposed by Davis (2001). This is a phenomenon whereby a maladaptive response is either maintained or intensified by certain types of behaviour, which in this case would be the use of the Internet. Cognitive symptoms could come before and cause the resulting behavioural/ affective issues due to effective or behavioural symptoms had previously been considered as the manifestation of PIU. The study by Pratarelli et al. (2006) outlined to this extent a four factors framework for PIU and its psychopathology. These are as follows.

a) Problematic behavioural styles leading to dysfunctional behaviour that correlates with Internet overuse;
b) How the Internet is used functionally – is it productive or meaningful;
c) How users gain sexual gratification, social advancement, or living out fantasies, particularly among those who are introverted;
d) People with little interest in or engagement with the Internet is thus not dependent on it.

Moving back a step-in order to obtain further relevant background, Niemz (2005) explained that cognitive theory does not necessarily play a large part in addiction studies and treatment in three kinds of models. These are related to behavioural type, pathology, and adaptive factors. Where research looks at IA from a cognitive perspective related to human motivation, it is aimed at examining the ways in which motivations differ between addicts and those who are not addicted.

According to Lepper (1973), who researched from the cognitive approach, motivation is described psychologically as either intrinsic or extrinsic. The first means the kinds of motivations from within ourselves, for example, we learn a language since we want to or like it. Extrinsic motivators are those external to us and can include praise from others as well as employment or financial incentives to perform or act cited by (Wan and Chiou, 2007).

The Cognitive-Behavioural Model, per Davis et al. (2002), examines depression and anxiety, which are known to be psychological issues. According to him, such atypical behaviour then leads to or results in pathological behaviour. The study does however distinguish between specific and generalized types of addiction. This latter form manifests itself in more socially interactive situations and can lead to higher levels of PIU (AIM Conference Center, 2008).

2.4. Internet Addiction

According to Ali et al. (2017) the most popular form of media in the world now is the internet, which differs from other forms of media. It has entered every aspect of life. With the advent of the internet, a number of facets of everyday life have changed, including how people occupy themselves and communicate with one another, Al Mukhaini (2021). There are now more than 4 billion active internet users worldwide, a rise in Internet usage that has been exponential (Ramón-Arbués et al., 2021).

4.65 billion people (58.7% of the world's population) will be using social media, 5.32 billion (67%) will be using mobile devices, and 5.00 billion people will be using the internet globally in April 2022. A total of 196 million more people is becoming internet users per year, or 4.1%. Online linked activities occupy 6 hours and 57 minutes of time every day on average. Internet usage is higher among younger individuals than among older ones, with young women using it the most frequently (Simon, 2021b).

Young (1996) is generally considered as the first study to explore IA. Results obtained showed that of a study cohort of 496 general Internet users, 79.88% of them could be classified as dependent on it through analysis of a Diagnostic Questionnaire (DQ) sent via e-mail and followed up with phone interviews. Goldberg (1996) first coined the term internet addiction, and it was from that point that it became to be recognized more widely. There are, as is well known, various other types of addiction, namely to technology, substances, pathological gambling, and suchlike (Horvath, 2004).

Young (1997) recognize IA as the inability to use the Internet without it causing a problem to the individual, however as has been previously mentioned, the term IA itself is inconsistently applied within the literature. On the one hand, Hur (2006) term the issue 'Internet Addiction Disorder'. On the other hand, (Lin and Tsai, 1999; Chou, 2001; Nalwa and Anand, 2003; Cao and Su, 2006; Kima et al., 2006; Young, 2006; Ko et al., 2007; Lam et al., 2009; Yen et al., 2009; Thomas and Martin, 2010) use the term 'Internet Addiction'.

Given that the majority refer to it as 'Internet Addiction', it has become the most generally accepted term for those behaviour types associated with problematic Internet use. PIU, as previously mentioned above, is another term – 'Problematic Internet Use' as defined by (Morahan-Martin and Schumacher, 2000; Davis, 2001; Niemz et al., 2007; Milani et al., 2009). Other terms include 'Internet Dependency' per Sharma and Sharma (2018), Internet Addictive Behaviour per Li and Chung, (2006), and Compulsive Internet use per (Meerkerk et al., 2009).

It can therefore be stated that given the multitude of terms expressed above and, in the literature, no standard diagnostic criteria exist for IA, having supposed that, most agree on its existence. "Excessive use of the Internet may not be problematic in most cases, nevertheless the limited case study evidence suggests that for some individuals, excessive Internet use is a real addiction and of genuine concern." Quoting from (Griffiths, 1998).

As previously established and restated, the precise definition of IA has not been agreed upon. It has been compared and related to other types of addictive behaviour. Two examples may be given of differences: Young (1996) mentioned about the inability to control Internet use and how it affects our daily life, relationships and suchlike, and could be used to aid a general outlook on life. The second may be cited from Goldberg (1996) who based it on a parallel to substance abuse. We can therefore conclude that the varying definitions and lack of a standard model obtain from the fact that different researchers are from different academic disciplines, still have tried to draw parallels with Internet use through their own work. The end result is naturally overlap, minor disagreement, and lack of overall clarity.

It is not to state however that researchers have been content to let this lack of overall clarity rest. Young (1999) developed an eight points Diagnostic Questionnaire (DQ) which drew on research from pathological gambling and thus comprised a series of binary questions probing into time spent using the Internet, how it impacted their lives, preoccupation with it, and suchlike. It was developed further into the internet addiction Test, or IAT. Young (2004) was a starting point, and further diagnostic tests have come into being such as the Internet Related Addictive Behaviour Inventory, or IRABI – (Brenner, 1997). Others include the PIUS (Pathological Internet Use Scale) per Morahan-Martin and Schumacher (2000) the internet addiction level for Taiwanese High-School Students, or IAST per Lin and Tsai (1999) and CIAS (the Chinese Internet Addiction Scale) of 1999.

All of these diagnostic tests have used either scalar method, different types of questions and possible types of answers, various upper limits on which to identify problematic behaviour, varying deviations from the normal and suchlike – all the various kinds of methods in normal data gathering and analysis of results. This therefore shows there is still no consistent approach to internet addiction classification.

2.5. Symptoms of Addiction to the Internet

Attention Deficit Hyperactivity Disorder (ADHD), depression and other mental problems can be connected with high and problematise levels of Internet use. Ju et al. (2008) ascertained that adolescents addicted to the Internet may experience poor future mental health outcomes.

To evaluate the key symptoms of internet addiction, such as tolerance, impairment of control, obsession, and excessive online time, Sharma and Sharma (2018) created several diagnostic criteria for internet addiction. College students are more likely to experience the prevalence, which can vary according on age, sex, and ethnicity, (Pujazon-Zazik and Park, 2010). Internet addicts are found to have a significant prevalence of personality disorders (Dalbudak et al., 2014).

2.6. Causes of IA

According to Ahmed Z. (2023) the development of internet addiction is influenced by a number of factors, including genetics, structural brain alterations, environmental influences, and existing mental health disorders. A growing body of research shows that computer addiction is genetically and biologically predisposed.

While it may be true that a person's psychological makeup has a definite bearing on prediction to IA, socioeconomic and demographic issues as with any mental health condition may be equally to blame. Zboralski, (2009) wrote about family issues, social breakdown and lack of stable or complete family being key red flags in terms of potential IA. We can state this being the case as mentioned before with high levels of Chinese IA given their until recently "One Child Policy". According to Zboralski (2009) study, IA can also affect parent-child relationships, marriages and suchlike. Young (1999) had already mentioned that those with IA are prone to anger management issues. Kayri (2010) showed that the family with high socioeconomic levels which tend to have less children and high monthly income, since mostly both parents are often working, in such a socioeconomic level parent more likely to satisfy their kid’s technological needs.

2.7. Impact of Internet Addiction

According to Bowditch et al. (2018), the Internet used in moderation can assist with emotional issues and help raise levels of self-esteem as per (Brailovskaia et al., 2019; Dong et al., 2020) the other side of the coin is that over- and uncontrolled use can precipitate addiction to it. This addiction can manifest itself in a number of ways, as has been demonstrated by various studies. These include psychological (Cheng et al., 2018; El Asam et al., 2019); academic, as demonstrated per Islam et al., (2020), or physical types, per Tepecik Böyükbaş et al. (2019). It is therefore the case that more research needs to be done into how one can minimise the extent to which student and young people could become psychologically damaged by excessive Internet use, developing research-based strategies to prevent it. The COVID-19 pandemic, somewhat perversely, provides an excellent 'laboratory' in which IA can be examined as relating to Internet access and use among university age students.

While there can be many positive effects of using the Internet e.g., better connections, education and suchlike, negatively the results can be poorer relationships (or more superficial ones), social issues and impacts of work or academic performance. Positive and negative effects are detailed below.

2.7.1. Positive Effects of Internet Use

2.7.1.1. Job Hunting

This is one of the most valuable outcomes for people in that they can look for work from home, or in their own time, and are not limited to the local area. Metzger (2007) mentions the different types of job sites on the Internet.

2.7.1.2. Entertainment and Communication

Ellison et al. (2007) have mentioned about the way in the modern connected world, relationships with people in different countries can be formed as well as existing long-distance ones maintained; there is a range of applications for suchlike; and that various entertainment options about for example video streaming or gaming with users in any part of the world.

2.7.1.3. Collection and Dissemination of Information

The world of news is now very much "24/7" and in a hyperconnected world, all manner of information is available at our fingertips, as mentioned by (Rice, 2006). All manner of messaging possibilities abounds, often for free, and information is no longer the preserve of the elite.

2.7.2. Negative Effects of Internet Use

2.7.2.1. Relationship and Social Issues

There are numerous elements that can contribute to internet addiction development. The prevailing consensus is that internet addiction is similar to other addictive behaviours in that both internal and environmental variables play a role in how they develop. A shortage of social connection is one external element that can contribute to internet addiction. Social support is defined as the material and mental care or assistance from others through difficult times or emergencies (Kim and Hwang, 2022).

The amount of social assistance college students receive and their level of mental health are closely correlated. According to studies, perceived support from family and teachers greatly predicts college students' life happiness (Yalçin, 2011). College students, on the other hand, typically feel that they get less help than they require, and this discrepancy is linked to greater levels of depression (Rankin et al., 2018).

Furthermore, cyber-bullying, as the name suggests, is bullying, although in the virtual world, and this can be a significant issue in the modern world as a result of it is not "seen" instead "out of sight".

2.7.2.2. Issues with Academic Performance

Sayed et al. (2022) have demonstrated that as long as students frequently spend a lot of time online, internet addiction may have an impact on how well they learn. while a study by Nalwa and Anand (2003) showed that among those who are online for excessive periods, behavioural, social and other issues result. These relate to the above relationship and social issues.

2.7.2.3. Effect on Health

Computer Vision Syndrome, or CVS, Repetitive Strain Injury, or RSI and eating disorders are just some of the negative physical effects of IA and Internet overuse. CVS is typically characterized, according to (Young, 1998; Chou and Hsiao, 2000) as resulting in delayed visual response times, itchy and sore eyes. Another big issue is that related to musculoskeletal problems that result from poor posture and ergonomics. It is now the case that in many countries, Health and Safety legislation requires employers to assist computer users to mitigate against such issues. The World Health Organization (WHO) claims that psychiatric diseases are one of the leading causes of disability worldwide, (Noorbala et al., 2017). Stress, anxiety, and sadness are all instances of mental health issues. Each of them is viewed as posing a risk to public health, particularly for the young population.

2.8. The COVID-19 Pandemic and its Relationship with IA

According to studies by (Dong et al., 2020; Duan et al., 2020; Gomez-Galan et al., 2020), one major impact of the lockdowns and social boundaries caused by COVID-19 has been the concomitant and exponential growth in the use of digital entertainment, use of the Internet – mainly through increased use of and dependence on social media as a form of news information as well as maintaining contact with family and friends and attempting to mitigate against the effects of loneliness, for example. This has been corroborated by information gathered by Nielsen Global Media (2020). As has been previously underlined, excessive Internet usage leads to addiction and Griffiths (2000) has described this as people being unable to take negative consequences into mind when increasing their use beyond acceptable norms, for example in studying and suchlike, where it is of great use, per (Dong et al., 2020).

Research by Cerniglia et al. (2017) exposed that for those in younger generations, addiction to the internet could manifest itself in various psychosocial issues; another study by Moreno et al. (2013) showed that internet overuse manifests in depression, stress and anxiety, which has been corroborated by both (Sharma and Sharma, 2018; Błachnio et al., 2019) in respect of its effects on university age students. Same study demonstrated that lower levels of overall life satisfaction obtained for participants through a cross sectional analysis of cohorts in both Italy and the USA as a result of their Internet overuse and possible addiction.

3. MATERIALS AND METHODS

In this study, one of the two main types of statistical data analysis is employed. By cause of the we must always arrive at a conclusion that is valid according to the data, certain assumptions must hold true. In data analysis, we can either look at what occurred in the past and attempt to predict the future – predictive analysis – or else examine the data and situation as it is – inferentially – attempting to ascertain or derive the rules and relationships that appear to derive from a particular phenomenon as revealed by the data. We would then do various modelling and analysis based on this situation rather than looking to the past.

Nevertheless, in data analysis, whether it predictive or inferential, the same basic tenets apply. There is a set of input variables which are altered in some way, to provide the output variables, and the means of doing is through the Data Generating Mechanism as illustrated by the following diagram below.

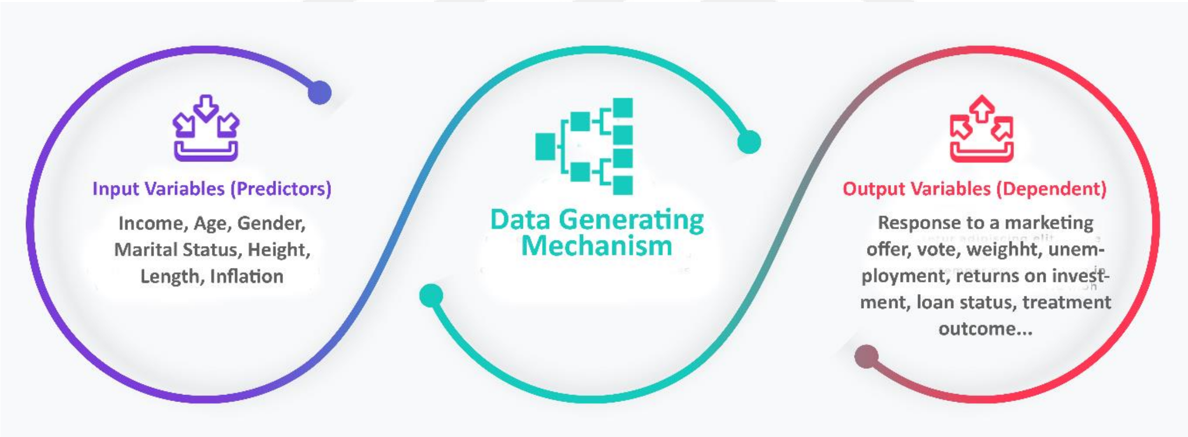


Figure 3.1 Data Generating Mechanism (Kibet, 2012)

Data scientists and statisticians generally agree that the Data Generating Mechanism can be of two different kinds. It can either be algorithmic, where one attempts to predict the future based on past behaviour, for example the input. This is most typically used in internet advertising, for example – showing browsers advertisements based on their browsing history and assumed interests. Or, it can be stochastic, whereby the output is related to the input through various types of statistical regression, for example.

The basic principles of statistical modelling and analysis are fairly standard. Data is examined whether through its presentation in the form of graphs, descriptive statistics, diagrams and suchlike. This first step is a sort of test which establishes the reliability of the data and identifies outliers, strange variances and the like – basically, if it 'looks right'. All data must be verified and authenticated. The second step is to try and form the basis of variable relationships and thus a potential model (Lee et al., 2006).

This study adopts Multivariate Adaptive Regression Spline (MARS), which is a sort of regression-analysis that uses recursive partitioning. Below will be explained both the advantages and disadvantages of recursive partitioning in order to create reliable data sets. Firstly, let us look at what recursive partitioning is. The idea is to divide data into a set of sub-regions, and analyse apiece sub-region in turn. The way this happens is that partitions are set up between each of the sub-regions, and is done recursively. That is to say that beginning from the domain as a whole, it is divided into two, with a kind of partition between each one. These two are then divided again, and the process continues recursively until an appropriate number of sub-divisions are in hand.

Following that, the reverse occurs, whereby according to Breiman et al. (2014) they generate the ideal or optimal set with the exclusion of lack of fit or excess number of sub-regions. The sub-regions are thus recombined. As a corollary, it is preferred that the variables with less impact on the overall result are excluded or set aside, and are placed into a local variable sub-set selection. However, with the process of basic recursive partitioning on linear functions, there tend to be limitations on the degree to which data can be interpreted effectively. The model is not necessarily constant. It is nevertheless possible to use classification trees as a means of impacting the rate of retention. Note that one principle of recursive partitioning is that there is no real continuity at the boundaries of sub-regions, however we can use approximations that are piece-wise smooth and constant which aids in the interpretation, thus representation as a binary tree is possible (Zhang and Goh, 2016).

As to the disadvantages of this method, since there is no continuity between sub-region boundaries, accuracy of approximation is obviously affected. This can especially be the case where the underlying function itself is continuous. This limits the effectiveness, as does the case where several non-zero coefficients exist, and furthermore, if only a proportionally small number of the total number of variables form the basis of the number of dominant interactions.

As has been mentioned above, as the data sub-regions are partitioned, it is therefore not possible to ascertain with concrete certainty from the model is either linear or additive (which would be a simple model) or whether the variables are related through complex interactions. To this end, MARS is employed as a way of overcoming some of the limitations described above. Let us turn therefore to a description of some principles of MARS (Taylan et al., 2018).

In describing the concepts of MARS, firstly let us consider the Linear Regression Model or LRM. This is where a regression model contains fitted parameters in a linear relationship.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \tag{3.1}$$

Equation 3.1 give LRM basic equation, where y is the dependent (response) variable, and X_j forms the set of regressor variables. These are the independent (predictor) variables, and form a series where j=1, 2, up to k. A random error component, ε, is introduced; while errors in themselves are assumed as being of normal distribution where the unknown constant variance is σ^2 and the mean is 0. All errors are assumed as being independent of each other and uncorrelated.

β_0 is the intercept while the various regression coefficients form the other parameters $\beta_j (j = 1; 2; \dots; k)$.

β_1 is used to represent how we expect y (the dependent variable) to change per unit chance of x_1 when the set of regressor variables that remain $x_j (j = 1; 2; \dots; j \neq 1)$ stay constant.

The purpose of β_1 and σ^2, neither of which can be said to be fully known, is to try and use the model to provide solutions to problems in the real world. In doing so, the Maximum Likelihood Estimation (MLE) and Least Squares Estimation (LSE) are employed when dealing with regression parameters that are unknown.

3.1. Generalized Linear Model

Where prevailing assumptions about constant variance and normality are unsatisfied, a combination of the binary forms of regression model (linear and non-linear) are used, and this framework is called a Generalized Linear Model (GLM). The basic advantage of GLM is that response distributions which are non-normal can be incorporated. To elaborate on this, the mean of y (the dependent variable) can depend linearly on a predictor (independent) through a non-linear link function. It further means that any exponential family of distributions can be part y, the probability distribution.

The family of GLMs include statistical models such as the following – log-linear models of multinomial data; classical linear models with normal errors; probit and logistic models for binary data. For the reason that GLMs incorporate dependent variables, a relevant link function, and a response probability distribution, other classical models such as Gamma, normal, binomial and Poisson can be thus termed.

Let us look at the fundamental algebraic structure of a GLM. It

$$h(\mu_i) = X_i^T \beta \tag{3.2}$$

where h is a smooth monotonic "link function", x, β is a vector with unknown parameters and $\mu_i = E(Y_i)$, is the input dependent variable of predictors.

There are some distributional assumptions imposed within a GLM, namely that any distribution taken from an exponential family of the form

$$Y_i \sim f_{Yi}(y_i, \theta_i, \emptyset) exp\left\{\left[\frac{\theta_i y_i - b_i(\theta_i)}{a_i(\emptyset)}\right] + c_i(y_i; \emptyset)\right\} \ (i = 1,2, \dots, N) \tag{3.3}$$

can represent response variables Y in Equation 3.3. Here, a_i, b_i, and C_i are arbitrary functions, $\emptyset$ is an arbitrary “scale” parameter and θ_i is a natural parameter. Furthermore, by incorporating the log-likelihood double-differential equation

$$\theta_i, \mu_i = E(Y_i) = b_i{}'(\theta_{i)} \text{ and } Var(Y_i) = b_i{}''(\theta_i).a_i(\emptyset) \tag{3.4}$$

Equation 3.4 represent the log-likelihood double-differential equation, furthermore we can also obtain a general expression for the mean and variance of dependent variable *Y*, (Wood, 2006).

3.2. Accuracy Measures and Regression Models that are Non-Parametric

The algebraic relationship

$$y = f(x) + \varepsilon \tag{3.5}$$

Describe in Equation 3.5 is a general non-parametric model of regression, and here $x = (x_1, x_2, ..., X_k)^T$.

By contrast to traditional regression analysis, where one aims to estimate the parameters of the model, non-parametric regression has as its aim the direct estimation of *f*, the regression function. It is assumed that with error term ε, constant variance σ^2 and a mean of 0, one assumes by default that it is continuous and smooth. Nevertheless, in some instances not relevant here, it does not necessarily have to be smooth.

The Equation 3.6 describes the additive regression model

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_k(x_k) + \varepsilon \tag{3.6}$$

Equation 3.6 in which β_0 constitutes the intercept (or unknown bias) and one assumes the inherent smoothness of the partial regressive functions f_j (j =1, 2,...,k). The aim is to estimate, from the data, the functions f_j (j =1, 2,...,k) and β_0.

Of course, there is a kind of variation model of the additive regression type, which is called the Semiparametric Regression Model. Here, unknown functions of dependent variables such as

$$y = \beta_0 + \beta_1 x_1 + f_1(x_1) + f_2(x_2) + \cdots + f_k(x_k) + \varepsilon \tag{3.7}$$

Equation 3.7 gives the unknown functions of dependent variables, which are separated additively. Furthermore, functions that are unknown, appearing as terms of higher dimension, for example

$$y = \beta_0 + f_{12}(x_1, x_2) + f_1(x_1) + f_2(x_2) + \cdots + f_k(x_k) + \varepsilon \tag{3.8}$$

may interact as predictor variables, and in the case of non-parametric regression of a general type, may also apply.

3.3. Multivariate Adaptive Regression Splines (MARS)

Friedman (1991) developed a technique of data mining known as MARS. Its main function is in the solution to problems of the regression type. MARS is similar to regression in that the aim is use of the least squares method to optimize the fit of a dependent variable, by contrast, it is possible with MARS to specify other and more complex functions in addition to the traditional additive and linear ones. This is done through a combination of both tree-techniques and regression.

Likewise, to the GLM, a non-parametric procedure, where no functional relationship exists between the two variables (dependent and independent), a relationship is defined using a piece-wise portioning approach like trees. On the other hand, the advantage of MARS is that nodes can be split at each step as opposed to only the terminal steps; additive and linear relationships can be captured; and two types of outcomes (continuous or categorical) can obtain, due to the range of predictors is greater. In doing so, data is separated or divided into various “splines”, which are separated on the basis of equivalent intervals. Then, for each spline, further sub-grouping of data occurs. The process is thus similar, however more detailed than in previous approaches. Sub-groups may be separated by “knots”, and occur in various different places, while the Basis Function (BF) incorporates the data relevant to each sub-group.

Thus, as mentioned, the process is more detailed, however in the same genre. In this case, the data itself determine the location of knots, the number of BF themselves, and suchlike, through the expansion of the spine product BFs. By their very nature, which is where there is a difference between MARS and the linear models, the relationship between splines is a smooth and continuous curve. It can be seen therefore that the step function utilized in trees is substituted by a series of piece-wise polynomials to the n^{th} degree. The knots are located at the abscissa points of the joins between the derivatives of each polynomial. A continuous model is thus obtained through use of a truncated power spline.

While a recursive partitioning approach to regression has its uses, continuous models and their derivatives obtain from the use of MARS, as opposed to discrete ones. With multi-variable interactions and relationships between variables that are either or nearly additive, better and more efficient modelling occurs; and further, these variables and their relationships may be identified. Hastie and Tibshirani (1990) and their work on Classification and Regression Trees (CART), which is a technique of recursive partitioning, is a direct predecessor to MARS in both spirit and genre. High level data may incorporate patterns or interactions that are otherwise difficult to discern, therefore as a result of it being a continuous model, it has many advantages over other such methods.

The fundamental idea behind MARS is that we can approximate a non-linear model by using separate slopes of regression, separated by distinct intervals, in the independent variable space. As mentioned above, these spaces are called “knots”, and it can be seen that the regression line and its slope alter when such knots are crossed. MARS is both simultaneous and discrete in that it examines single variables and their inter-relatedness, while taking stock of the overall bigger picture, as it were.

3.3.1. Basis Functions

The foundation of such data analysis is to determine the relationships between dependent and independent variables. Therefore, with the idea of knots, these can be discovered by applying the use of multivariate, adaptive regression splines. Basis Functions in MARS are usually found in distinct pairs, and the Equation 3.9 for Basis Function 1 (BF1) in terms of a variable elevation can be described as follows

$$\mathrm{BF}_1 = \max(0, \mathrm{elevation} - 219) \tag{3.9}$$

If one were to consider two sets of data for elevation variables, as described above, considering the BF tend to occur in pairs, we could see that a first set constituted elevations below a threshold margin (in this case for example 219 metres), and the second set constituted those which are higher than this point. Elevation and slope degree are unrelated in this example. Following this concept, due to MARS allows us to discard various areas by assigning them a zero value, it means that a particular set of BF may be employed as predictors in the spline regression model for the primary data (Friedman, 1991).

Relationships between predictor and response variables within the MARS system comprise a double sided truncated function as in this graph below.

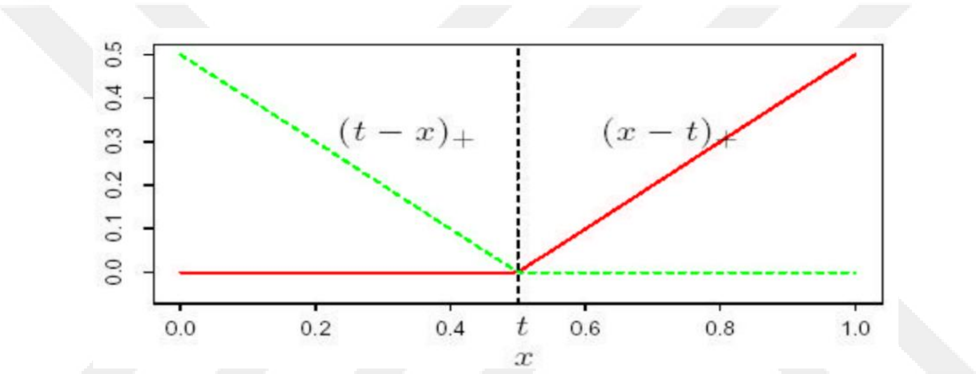


Figure 3.2 The BFs used by MARS $(x\text{-}t)_+$ and $(t\text{-}x)_+$

Per Hastie et al. (2001) this graph describes two basis functions $(t\text{-}x)_+$ and $(x\text{-}t)_+$ In this example, parameter t is the knot of the BF defining sections, obtained from the input data, of the piece-wise linear regression. Only positive results of each equation are permitted; evaluation to 0 is the other possibility if negative.

The BF collection is

$$\begin{gathered} C = \left\{ \left(X_j - t\right)_+, \left(t - X_j\right)_+ \right\} \\ t \in \left\{x_{1j}, x_{2j}, \dots, x_{Nj}\right\} \\ j = 1,2, \dots, p \end{gathered} \tag{3.10}$$

It has noted that if the input values are distinct, then there are $2Np$ BF.

Here is another example, which comes out through multiplying 2 MARS BF that are piece-wise and linear. It can be seen from the diagram that if both component functions are non-zero, then the result is only non-zero over a small range.

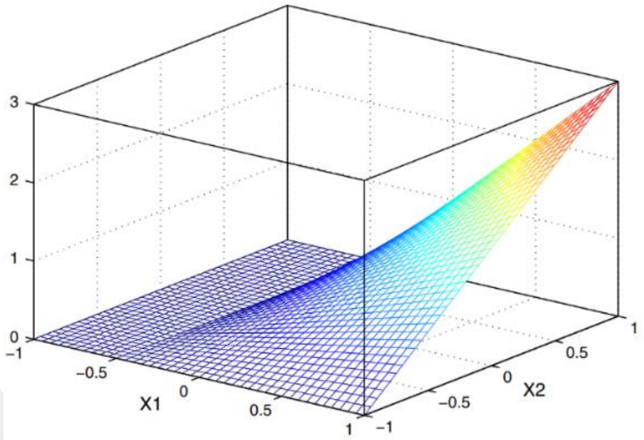


Figure 3.3 Two-way interaction basis function

3.3.2. Friedman's MARS Model

Friedman (1991) described a flexible regression model that was non-parametric, and used for high level data. Hastie et al. (2001) described a general equation for MARS being as follows

$$y = f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X) + \varepsilon \tag{3.11}$$

In the Equation 3.11, y is defined as relating to X in that it is a predictor variable (which may or may not relate to its interactions with other predictor variables). β_0 is the intercept parameter and β_m is the weighted sum of *hm(X)* which are one or more BF. The number of variables in the model is *M*.

It drew on Friedman (1991) which defined the main system as

$$y = f\left(x_1, \dots, x_p\right) + \varepsilon \tag{3.12}$$

Through the use of the LSE, predictions based on the various inputs are generated and combined, by using the parameters of the model. Let us further elaborate on the Basis Function (BF), denoted $\beta_i(x)$.

The BF together with the model parameters which are estimated through least squares estimation are combined to produce the predictions given the inputs.

Each basis function $\beta_i(x)$ may constitute one of these three states. It may be a "hinge" function, which as the name suggests is of type either/or: *max(0, x-const)* or *max(0, const-x)*. According to MARS, the knots in these "hinge" functions are determined automatically from the MARS process. Secondly, $\beta_i(x)$ could the product of two or more of the "hinge" functions which would be the case when there are two or a greater number of variables. And finally, the intercept alone would be where $\beta_i(x)$ is 1. Note that the BF may be of different forms including either a simple constant or at the other end of the spectrum, hinge functions which constitute multiple sub-functions and their product.

Let us look at what happens to the locations of knots (which is where input and predictor values occur), and the interactions between the variables themselves. Starting with one BF taken from the range of BF allowed, additional BF is added to the search input into the MARS algorithm. This means that for all the BF spanning every value of each predictor, a weighted average is chosen by MARS in order to obtain a goodness of fit criteria that is maximized through use of the least squares method.

According to the process of MARS, the algorithm is used under the assumption that function *f* is locally smooth. Where the standard concept of smoothness is not extant, Friedman (1991) elaborated on the initial concept of MARS through the function estimate

$$f(x) = \sum_{m=1}^{M} a_m \, I(x \in A_m), M \leq K \tag{3.13}$$

Where there is a variable, *x*, in a single category, given $x \in \{c_1, \dots, c_k\}$. Here, ($A_1$,..., A_m) constitute sub-sets of {c_1, ..., c_k } and *I* is the indicator function. If *M* is a smaller number, then the estimate is seen as smoother. Continuous and categorical explanatory variables can also be accommodated through this expanded model.

3.3.3. Hinge Functions

In discussing hinge functions, which typically occur in pairs, we must also consider "knots" and "kinks" or sharp turns within one dimension. As has been explained previously, where two splines intersect, we find a knot. It is at this point where two different regression models meet. The hinge functions are of the general form max(0, x-c) or max(0, c-x) in which *c* is the constant giving the value for the kink; 0 is the function's minimum value; and *x* is the variable itself. Here is a typical pair of hinge functions. Functions which are not linear may also obtain through their multiplication. We can see the location of the knot as being x = 3.1 in this particular graph.

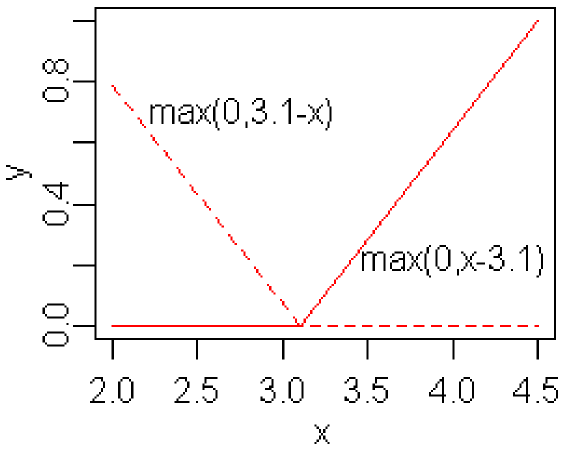


Figure 3.4 A mirrored-pair of hinge-functions with a knot at x=3.1

3.3.4. The MARS Approach to a Process of Model Building

When modelling statistically using MARS, a two-stage approach is utilized. It can itself be compared as akin to that of the process using recursively partitioned trees. These two stages are called the forward and backwards passes. As has been described above, MARS incorporates piece-wise linear BF, on which are constructed double sided functions truncated by the predictors, *x*, in the format $(x-t)_+$ and $(t-x)_+$ and the value of the knot is at *t*. Where $x \in \mathbb{R}$, a truncation occurs in these two functions below

$$(x-t)_+ = \begin{cases} x-t & if\ x > t \\ 0 & otherwise \end{cases} \tag{3.14}$$

and

$$(t - x)_+ = \begin{cases} t - x & if\ x < t \\ 0 & otherwise \end{cases} \tag{3.15}$$

At each t-value from Equation 3.14 and 3.15, it can be observed from first figure above, that each function is piece-wise and linear. As indicated above, only positive values are allowed otherwise the value is 0 by default. This is termed a reflected pair. By induction, and logically from the equations, for each reflected pair there is an input x_j with a knot. It follows that a collection of base functions exists

$$c = \left\{ \left(x_j - t\right)_+, \left(t - x_j\right)_+ \middle| t \in \left\{x_{1j}, x_{2j}, \dots, x_{Nj}\right\}, \quad j \in \{1,2, \dots, p\} \right\} \tag{3.16}$$

2Np base functions may exist in the case where all input values are different even though all the base functions relate to a single x_j and may still be seen as being a function in the entire space of $\mathbb{R}^P$. When it becomes necessary to consider the spline fitting in higher dimensions, by using basis functions we can generalize it into the various tensor products relating to each univariate spline function. So, for basis functions that are multivariate, the algebraic representation is

$$B_m(x) = \prod_{k=1}^{km} \left[s_{km} \left(x_{v(km)} - t_{km} \right) \right]_+ \tag{3.17}$$

Here in Equation 3.17, K_m is the total number of truncated linear functions in the m^{th} basis function, $x_Y(km)$ is the input variable which corresponds to the truncated linear function at the k^{th} position of in the m^{th} basis function while t_{km} is the knot value which corresponds to it and finally $s_{km} \in \{\mp\}$, Yerlikaya (2008).

With linear regression of the step-wise form, functions and corresponding products from the set C as opposed to the initial value are used. Building a MARS model is similar, and is of this particular form

$$f(x) = \beta_0 - \sum_{m=1}^{M} \beta_m \beta_m(x) + \varepsilon \tag{3.18}$$

In the Equation 3.18, from the set *C*, $B_m(x)$ is either a function the product of two or more functions. With a choice of B_m it is possible to estimate β_m which is the coefficient, by using the method of minimizing the least squares regression and its sum. It is important nonetheless to consider $\beta_m(x)$ and how it can be structured, since within *C*, one only finds candidate functions and the function which is constant: $\beta_0(x) = 1$. β_m and its product must be considered for every step of the process within the model collection (*M*) and the related reflective pair. If what obtains from this process is input into our model (*M*), these next results

$$\hat{\beta}_{M+1} B(x)\left(x_j - t\right)_+ + \hat{\beta}_{M+2} B_1(x)\left(t - x_j\right)_+, B_1 \in M \tag{3.19}$$

The great advantage of this procedure is that training errors can be minimized. In this equation, $\hat{\beta}_{M+1}$ and $\hat{\beta}_{M+2}$ are considered coefficients, along with other *M+1* coefficients, which we can determine using the method of least squares with the refined outcome input into MARS. It is like an iterative process in that the process begins again until finally the highest allowable amount of items in *M* results. The optimum model, $\hat{f}_\alpha$, with the perfect value α, is the desired outcome for a backwards elimination process that removes the preceding terms causing minimal increases in the error of residual squares. *Generalized cross-validation (GCV)* is a parameter employed to reduce computational cost, or more specifically, its optimal value, α. In a similar process, an adjusted residual sum of squares imposes a penalty on those models of excess complexity. Above, where we wanted to avoid an overfit to the data of the total model, MARS is useful in that it provides an optimal compromise between underfit and overfit. When a particular variable is removed, the amount of degradation can be ascertained using GCV. Utilizing MARS is advantageous, as it helps to ensure a trade-off between variance and bias in terms of the criterion of best fit: with a model that is global and over-arching, it may have low variance and variability, by contrast a local and more specialized model has high variability. The real function which underlies the basis functions *f(x)* draws on the above to provide approximations to either linear expansion or its non-linear counterpart.

An encapsulation of dependent variable y with M number of terms incorporated within the MARS concept is as follows (whereby M is the number of terms over which summation occurs; the model's parameters are β_m and β_o and the number of knots related to each BF is t.

$$y = f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m + H_{km}\big(x_{v(k,m)}\big) \tag{3.20}$$

Function H from Equation 3.20 may be defined according to

$$H_{km}\big(x_{v(k,m)}\big) = \prod_{k=1}^{k} h_{km} \tag{3.21}$$

In the Equation 3.21, $x_{v(k,m)}$ is the predictor to the k^{th} iteration of the m^{th} product. The difference between additive and pair-wise interactive models lies in the value of K. For the latter, $K=1$ and the former, $K=2$. To commence the process, it is necessary to use candidate functions (all those in the set C) and $h_o(X) = 1$, the constant function.

Let us examine the two types of process within MARS: forward pass and backwards pass.

3.3.5. The Forward Pass

By beginning with an initial basis function, "layers" are added according to each pair of BF. The aim of this is to reduce as much as possible the residual error in terms of the sum of squares. To do so, it has to find the appropriate pair. These are always identical except for the case of a mirrored hinge or rather another side of this hinge is incorporated within the function. For each successive BF, where a hinge is defined in terms a knot and a variable, the constant may be multiplied by the successive new hinge. It is therefore necessary for the below process in the search procedure to occur, in three parts. Firstly, terms which already exist (parent terms), all variables (so that a new variable is chosen for the succeeding BF), and finally every value of every variable (so as to constitute a new hinge function's knot). When the residual error resulting from each step is, therefore minimized such that it is not possible to continue, or the maximal number of terms has been reached, is when the process ends. Using a heuristic enables reduction of the number of parent terms it is necessary to consider, according to the inherent nature of the hinge functions (Taylan et al., 2018).

3.3.6. The Backwards Pass

The other kind of way that MARS works is with the backwards pass. By contrast to the forwards pass, this technique is advantageous in that it is able to look at past data and events, and is thus able to pare away those inputs that are not as useful in generalizing to new data. By cause of that, while it generates a good and descriptive model, this overfitting is not as just explained seen as ideal. It bases future models on past data and does not have good predictability.

As with CART, terms are deleted individually, with that term which is worst in efficiency removed. This enables the sub-sets in the model to be cross validated using the GCV criteria, and also avoid the issue of over-fitting though giving "penalties" to unsuitable new entrants in the process and reliable testing of data fidelity. Thus, in being able to delete any prior term, it is superior to the forwards pass that can only see as far ahead as the next pair. Furthermore, the backwards pass is able to remove one side of a hinge pair.

3.3.7. Generalized Cross Validation (GCV)

As mentioned previously, an ideal model is one which balances complexity against goodness of fit, and since the backwards pass utilizes GCV with lower values of GCV being better, it is possible to have the backwards pass and GCV estimate future performance based on new data. The key word here is 'estimate', whereas new data is not there when the model is built, thus it is more desirable to understand how a particular performance comes from that new data.

Using Raw Residual Sum of Squares (RSS) on historical i.e., training data is disadvantageous in that increasing terms in MARS causes the RSS to increase. Thus, due to that with a very large model, optimal performance is reduced, and the backwards pass always tends towards the biggest model, this causes problems and becomes unwieldy.

Presenting the GCV formula below, N is the number of rows in a particular matrix, x, and RSS is the Residual Sum of Squares.

$$GCV = RSS/(N * (1 - Effective\ Number\ of\ Parameters/N)^\wedge 2) \tag{3.22}$$

Equation 3.22 gives the GCV formula and for the particular situation of MARS, we can define the most efficient parameter number as being the (Number of MARS Terms + Penalty) * (Number of MARS Terms – 1) / 2, where the penalty is around the 2 or 3 or mark.

The addition of knots is penalized in this formula, since the (Number of Mars Terms – 1) / 2 is the number of hinge function knots. Thus, with training data, we can keep in account the flexibility of the model, through the ability of GCV to increase RSS training data. On the other hand, an overly flexible model can result in extraneous information i.e., noise, diverting from what we want to drill down into.

To illustrate GCV error, we can incorporate both model complexity and residual error and find its specific measure. It is described in Equation 3.23 as

$$GCV = \frac{\sum_{i=1}^{N}\left(y_i - f(x_i)\right)^2}{\left(1 - \frac{C}{N}\right)^2} \tag{3.23}$$

with

$$C = 1 + cd \tag{3.24}$$

Defining the terms from Equation 3.24, we have the “penalty” for adding a BF as c, N being the number of cases in a data-set, and d being the number of effective degrees of freedom. According to Hastie et al (2001) and as mentioned above, the most effective C-values are around a d of 2 or 3.

Using GCV techniques within MARS, an RSS process ajusted by a penalty that relates to model complexity , Friedman (1991), gives the possibility of evaluating model appropriateness. Model complexity is reflected in the value of the denominator, while the numerator comprises the average RSS error. An estimate within the regression model for the error variance incorporating a penalty factor is the basis of the process carried out within this type of regression. The GCV statistics simply substitute for the GCV R-squared value.

Craven and Wahba (1978) first described the process of GCV for error approximation using a formula that determines the ability to validate results through repeatedly leaving a previous term out. Friedman (1991) drew on this to describe the concept of MARS.

3.3.8. MARS: its Relative Advantages and Disadvantages

As is the case with, quite many models in data science and statistics, therefore it is with stating with absolute certaintly that one kind of regression modelling technique is vastly superior to any other. In terms of the positives of MARS, it uses piece-wise linear functions that generate continuous models. This leads to it being able, like CART, to be a fully efficient and automated process, having categorised both dependent and independent variables within a reasonably short time-frame without drawing too heavily on computational resources. It is most efficient at the providing answers to complext problems with multitudinous input variables, and is able to process non-linear relationships thus giving clear interpretations. It is especially good at analysing situations where the number of input variables is high and complex. Graphs can be produced of the results, and the response variable can be determined through both interactive and additive variables.

On the other side, the key disadvantages of MARS are that due to the over-fitting that obtains from analysis of the training data, in which the various interactions and non-linearities are identified, it may not be able to predict forwards as best hoped. It is nonetheless possible, having said that, to mitigate against the level of over-fit through using a smaller number of basis functions and giving higher penalties for each knot (Yerlikaya, 2008). Another disadvantage is that a data-set of sufficient size must be available; which can lead to complications in the balance between variance and relative bias. This means that although MARS generally uses less computational resources, boosting or bagging can be more successful at generating accurate predictions.

4. DATA COLLECTION

The data collection and analysis section is a crucial component of any research study or project, providing a systematic approach to gathering and examining information to draw meaningful insights and conclusions. This section encompasses the methodologies, tools, and techniques employed to collect relevant data, ensuring its reliability and validity.

4.1. Research Questions

- Classification performances of MARS data mining method in predicting internet addiction levels of university students according to data obtained from both countries (Türkiye -Iraq);
    - Does it differ according to the correct classification rate?
    - Does it differ according to the specificity rate?
    - Does it differ according to the sensitivity rate?
    - Does it differ according to the accuracy rate?
    - Does it differ according to the F1-Statistics?
    - Does it differ in terms of the area under the ROC curve, that is, the rate of misclassification?
- What are the most important predictors of internet addiction of university students in Türkiye based on MARS data mining method?
- What are the most important predictors of internet addiction of university students in Iraq based on MARS data mining method?

4.2. Research Sample

The aim of the study is to examine a scale for internet addiction which is developed by Gunuc and Kayri (2010), to examine the factors that can affect the addiction status of individuals. The sample of the study is at the international level, and cosmopolitan universities from each country (Türkiye and Iraq) and a university from mentioned countries were selected, which is Van Yuzuncu Yil University from Türkiye and Soran University from Iraq. The target audience was determined as university students. The fact that the sample was composed of adolescents is due to the fact that internet addiction is mostly seen in adolescents and individuals in this period are open to all kinds of influences. The number of individuals in the sample is 2235; 1220 students from Van Yuzuncu Yil University, 1015 from Soran University were assigned by random sampling method. The age range of the sample ranged from 18-67 and the mean age was found to be 21.66.

4.3. Research Context

According to a study by Derbyshire et al. (2013), students of college age are becoming increasingly reliant on technology both to communicate with their peers, also to navigate everyday life. As we know, with the recent COVID-19 pandemic, technology use has only become more embedded in daily life for all, in a very short space of time. This includes the introduction and often replacement of college classes with online interaction.

4.4. Research Design

This research was designed in relational screening model, which is one of the general screening models. Relational screening models are research designs that try to define the relationships between the variables and identify the presence and/or quantity of change between two or more variables. Furthermore, relational screening model was not a widely known statistical method. However, potential attributes might include identifying variable changes, determining relationships between variables, using correlation analysis, and aiding in prediction. For the latest information, refer to recent research sources.

4.4.1. Measuring Tools

A questionnaire (Appendix 1) was partially designed by Gunuc and Kayri (2010) and the researcher with a total of 54 items to ask. It was semi- rather than fully-structured in form. The idea was to look at the extent to which internet addiction was a factor for the students at this school. It was divided into three separate sections, and responses required grading according to a five-point Likert scale. James and Lee (2011) this is a well-established and commonly used scale that obviates simple yes or no responses, which do not provide sufficiently detailed responses.

The three sections were as follows. For the first, it looked at the classic social and demographic backgrounds of the participants including the usual questions relating to socioeconomic status, age and gender. This comprised 5 closed ended questions. The second addressed the effects of the COVID-19 pandemic; more specifically, how it had changed their use of the Internet by asking about length of time spent on it, how has use changed, whether they had it at home, on what sort of device they accessed it, and suchlike. This comprised 10 questions both closed and open in style. The third part incorporated a scale for internet addiction, where four diagnostic areas were explored. These were, in no particular order: social isolation, difficulty controlling use, the effects of withholding use, and its impact on daily life. The details of participants were only seen by the researcher for reasons of confidentiality.

4.4.2. How the Research was Carried-out

Preceding the survey, informed, formal consent was obtained from all participants, whether they undertook the questionnaire online or through one of the paper forms distributed in person. The questionnaires were returned, data extracted and then analysed. Permission had also to be sought from the administrations of both Soran and Van Yuzuncu Yil University.

Since the language of the internet addiction scale is Turkish, it has been converted into English by taking into account the scale adaptation steps, afterward that it can be applied to students who study 100% English at Soran University. Thus, it was necessary for thorough review and comparison between the translations to ensure that the questions were of the same line of enquiry. The questionnaire was then piloted among a group of 250 students at Soran University, before full distribution.

Employing a technique of random sampling, the link for the questionnaire was distributed to students at 5 of the faculties at Soran – i.e., 50 students from each, totalling

250 as mentioned. A time limit was imposed of a week, with a reminder message two day later.

4.5. Analysis of the Data from Pilot Study

Analysing the data statistically using quantitative methods since the research is one-by-one quantitative study; where responses were open ended, general trends were identified and thus quantified for statistical analysis.

4.5.1. What the Research's Pilot Study Revealed

As a result of the Confirmatory Factor Analysis applied on the data as a result of the application to 250 students, it was observed that the structure was under the factor specified in each item as stated in the original scale (RMSEA: 0.060, CFI: 0.95, NFI: 0.90, NNFI: 0.95 and GFI: 0.81). Again, the reliability values (McDonald's ω) obtained for the four dimensions were obtained as 0.829, 0.839, 0.833 and 0.795, respectively. It was concluded that the internet addiction scale, which was adapted into datasets with these values, was valid and reliable.

4.5.2. Testing the Data Set's Assumptions

Below will be set out a commentary on how certain operations on the data set may affect results in different ways. Before any kind of scientific analysis is carried out, one has to assume various criteria that hold, therefore both the reliability and validity of measurement may be in question.

Before one of the two main methods of statistical analysis (parametric or non-parametric) can be applied to a data set, various assumptions have to be made as just stated. According to at least two papers by Mishra et al. (2019) it is necessary to have four specific requirements be met. These are generally seen by them as:

- Lack of multi-co-linearity: there are no errors in independent variables;
- Normality: the data show a normal distribution;

- Linearity: with independent variables constant, the relationship between dependent and independent variables is linear; and
- Homogeneity: any combination of the independent variable may be allowed causing subsequent variance in the dependent variable.

Sayed et al. (2022) it is also possible to employ a non-parametric approach. For the reason that a parametric approach only holds for when the data are under a normal distribution; in the opposite situation, per Hazra (2017) different confidence intervals and values can result – with the predictable outcome that results may differ from reality, which is clearly undesirable.

Researchers generally use two main methods when deciding whether data exhibit a normal distribution or not. These are the Kolmogorov-Smirnov test where $n \geq 50$ and for less than that, the Shapiro-Wilk test. It is necessary for data to be from a normally distributed population and the null hypothesis to hold, with $p<0.05$.

With the results of the present study, both of the below tests exhibit $p<0.05$ for the Kolmogorov-Smirnov test as shown below in Tables 4.1 and 4.2 below, meaning that a normal distribution does not apply.

Table 4.1 Normality test for the dependent variable for the Türkiye’s sample

	Kolmogrov-Smirnov		
Internet Addiction	Statistic	*df*	*p*
	0.063	1220	0.001

Table 4.2 Normality test for the dependent variable for the Iraq’s sample

	Kolmogrov-Smirnov		
Internet Addiction	Statistic	*df*	*p*
	0.056	1015	0.001

As can be seen above, a parametric approach to data analysis is inappropriate given the significance values in both data sets: the Normality assumption does not hold, meaning that at least one of the four required assumptions have not been met. Thus, a non-parametric approach needs to be taken. It should be noted however that there is no requirement to test data set assumptions when employing a MARS approach.

4.6. Measurements of Reliability and Validity

While there is a certain temptation to do so, we should not see reliability and validity as independent of one another. It is logical to assume that something ought to be valid and reliable, according to Biolcati-Rinaldi et al. (2018) if we are to depend on is as representing an attribute or concept with any degree of certainty. This ensures that a well-designed study produces consistent and accurate results of data analysis.

According to Chetwynd E. (2022) researchers should be aware of the need for reliability and validity in evaluating the literature, therefore they can choose the most appropriate basis for their studies in order to be seen as credible.

4.6.1. The Validity of the Data Sets

Buntins et al. (2017) validity can be seen as the degree to which a certain instrument measures what it claims to, with a concurrent degree of inherent reliability and clarity. A well-constructed research instrument is needed, according to Geldhof et al. (2014) in order to further add credibility to the study, and to ascertain whether the scientific method has been employed correctly. According to Lucas Gren (2018) every scientific study must demonstrate validity; clarity; be dependable and trustworthy; and follow the accepted method of generating research findings.

According to the exploratory factor analysis obtained by Gunuc and Kayri (2010) which has been apply to the internet addiction scale in regarding the construct validity of the scale, after the exploratory factor analysis the scale consists of four sub-dimensions (Deprivation, Control Difficulty, Impairment in Functioning and finally Social Isolation). Subsequently of conducting exploratory factor analysis, in order to show that the results and structure of this study are valid, confirmatory factor analysis was applied with the obtained data.

4.6.1.1. Confirmatory Factor Analysis

Lee et al. (2006) described Exploratory Factor Analysis, or EFA, and it was developed into Confirmatory Factor Analysis, or CFA, as described by (Douglas, 2017). The aims of CFA are to test a hypothesis by means of initial examination, testing and drawing the relevant conclusions. In the present study, it was necessary to determine whether data were of a multivariate normal distribution.

To carry out the required analysis and build an appropriate model, the package LISREL8.80 was used. Factor scaling with a variance factor of 1 was employed. The following factors were also examined within the data analysis: Normed Fit Index (NFI), Non-Normed Fit Index (NNFI), Root Mean Square Error of Approximation (RMSEA), Standardised Root Mean Square Residual (SRMR), Goodness of Fit Index (GFI), Chi squared (x^2) and Comparative Fit Index (CFI).

It is worth mentioning though that since sample size is a major factor affecting x^2, it was discarded. This follows the approach taken by Tibshirani (2001) where all of the other above factors were however used. Values as a reference point of NFI, GFI, NNFI and CFI of ≥0.90 and SRMR and RMSEA of ≤0.08 were incorporated into the calculations. Table 4.3 below depicts the results of our Confirmatory Factor Analysis, while Appendix 2 is where the path diagrams may be located.

Table 4.3 Results of confirmatory factor analysis

Sample	RMSEA	SRMR	CFI	GFI	NFI	NNFI
Türkiye Sample	0.074	0.062	0.97	0.81	0.97	0.97
Iraq Sample	0.069	0.058	0.95	0.83	0.94	0.95

According to Schumacker and Lomax (2004), values of SRMR and RMSEA are ≤0.08, there is an acceptable fit. As shown above, for both sets of data (Türkiye and Iraq), we can conclude these SRMR and RMSEA values as being acceptable and within the boundary.

Additionally, as pointed out above, if all the other CFA results apart from CFI, GFI, NFI, NNFI exhibit a figure of ≥0.90 then it is clear that there is an acceptable fit to the data. We can observe from the table above that with the exception of the GFI, which is nonetheless close to 0.90, that statement holds true. It can thus be concluded that having applied a range of indices to the data, the data can be considered valid; the model is suitable; and thus, the data is validated.

4.6.2. Measurement and its Reliability

Below follows a discussion of two types of reliability coefficient: the McDonald ω and the Cronbach α. They both relate to scales of the Likert type.

Let us first consider the α coefficient. This is employed where test items include either isomorphic or parallel and τ-equivalent measures, and is used as an unbiased method of estimating the degree of data reliability. By contrast, McDonald's ω coefficient is employed if we are to measure items of differing sensitivity and size and the structure itself i.e., measurements which are congeneric.

According to Geldhof et al. (2014), the former α coefficient generates biased results and thus the ω coefficient is preferred. Yurdugul (2006) observes that the relative difference carries through to the results of factor analysis. In order to calculate both the Cronbach and McDonald values, Jamovi 0.9.0.3 was used. Drilling down further into the data, Table 4.4 below presents the various sub-categories of Scale Reliability with respect to the Turkish data.

Table 4.4 Reliability statistics for the Türkiye's sample

	α (Cronbach)	ω (McDonald)
Internet Addiction Scale	0.957	0.959

As can be seen from the Table 4.5 above with an α value of 0.957 and a ω value of 0.959, it is clear that the data can be considered valid and thus reliable based on the discussion above.

Table 4.5 Scale reliability statistics for the various sub-categories of Türkiye dataset

	α (Cronbach)	ω (McDonald)
Deprivation	0.887	0.888
Control Difficulty	0.898	0.904
Impairment in Functioning	0.908	0.91
Social Isolation	0.878	0.883

Examining Table 4.5 above, we can state that both α and ω values are almost the same. This indicates reliability and consistency and thus data validity.

Let us turn to the corresponding data for Iraq, and the results obtained as to sample reliability. From Table 4.6 below, the following points may be noted. For both coefficients, they are marginally lower than for the Turkish data, nevertheless are much of a muchness. Again, given the discussion above, the data may be considered reliable, consistent and valid.

Table 4.6 Reliability statistics for the Iraqi sample

	α (Cronbach)	ω (McDonald)
Internet Addiction Scale	0.936	0.937

In Table 4.7 below, similar values for each coefficient have obtained, and the same conclusions as for the Turkish data hold, i.e., the data is reliable, consistent and valid. Much the same statements can be made for both sets of data.

Table 4.7 Scale reliability statistics for the various sub-categories of Iraqi dataset

	α (Cronbach)	ω (McDonald)
Deprivation	0.847	0.848
Control Difficulty	0.859	0.863
Impairment in Functioning	0.853	0.856

Table 4.7 Scale reliability statistics for the various sub-categories of Iraqi dataset (continued)

	α (Cronbach)	ω (McDonald)
Social Isolation	0.813	0.814

5. ANALYSIS OF DATA

This section serves as a vital component of the research study, where raw data is transformed into meaningful insights and valuable information. This section involves the application of statistical and computational techniques to organize, interpret, and derive conclusions from the collected data. By employing appropriate analytical methods, researchers can identify patterns, trends, relationships, and correlations within the data, providing a deeper understanding of the research objectives.

5.1. Demographic Descriptive Statistics

In the section, it was aimed to examine the relationship between the addiction variable and some demographic variables. The percentage of variables obtained from the sample, such as country, gender, stage, age, parental education, parental occupation, number of siblings, smoking status, family income, internet ownership at home, the most frequent use of internet, annual and daily use of internet, also question relate to the COVID-19 pandemic, and frequency distributions are given in this section.

Table 5.1 Frequency distribution of individuals according to countries

	Iraq	Türkiye
Number of Participants	1015	1220

As seen in Table 5.1, data were collected from two universities of two different countries which are Türkiye and Iraq. The number of participants from each sample varies of each which is 1015 for Iraq and 1220 for Türkiye.

Table 5.2 Frequency distribution of individuals by gender

	Iraq	Türkiye
	n (%)	n (%)
Male	465 (45.81%)	427 (35%)
Female	550 (54.19%)	793 (65%)

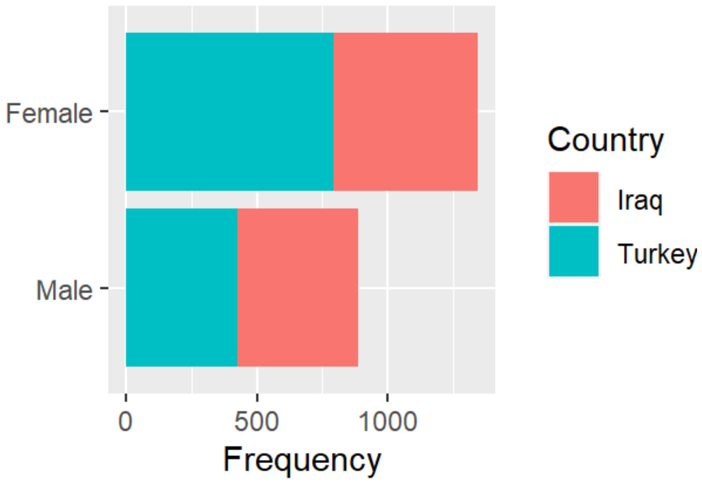


Figure 5.1 Frequency distribution of individuals by gender

As can be seen from Table 5.2, 892 (39.9%) male students and 1343 (60.1%) female students participated in the study which has been shown by each country. The fact that the number of female students was close to the number of male students in the random sampling minimized the possibility of a single factor in terms of gender being effective in internet addiction.

Table 5.3 Frequency distributions of individuals according to their fathers' educational status

Study Level	Iraq	Türkiye
	n (%)	n (%)
Not-Literate	144 (14.19%)	56 (4.59%)
Literate	0	51 (4.18%)
Primary School	247 (24.33%)	356 (29.18%)
Middle School	218 (21.48%)	235 (19.26%)
High School	196 (19.31%)	306 (25.08%)
Diploma or Bachelor	187 (18.42%)	189 (15.49%)
Master	13 (1.28%)	19 (1.56%)
Ph.D.	10 (0.99%)	8 (0.66)
Total	1015	1220

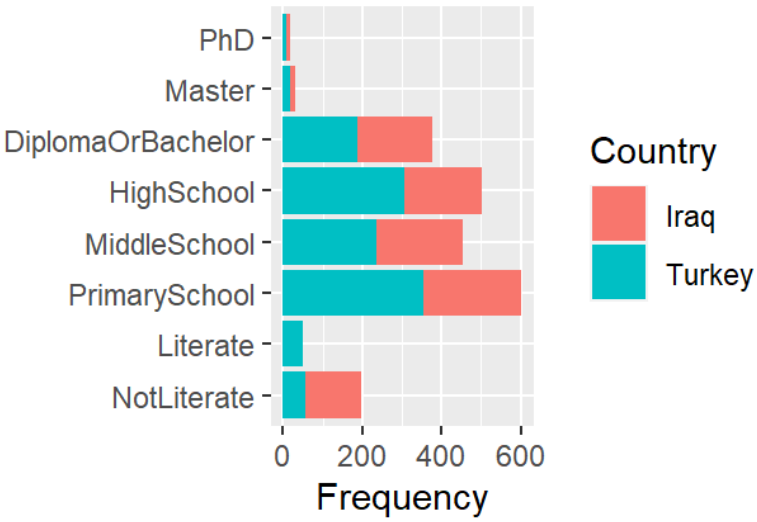


Figure 5.2 Frequency distributions of individuals according to their fathers' educational status

From Table 5.3 the frequency distributions of individuals' fathers' educational status are seen. It is observed that the educational status of the fathers of the individuals is generally low and they are predominantly composed of middle school, high school and Diploma or Bachelor graduates.

Table 5.4 Frequency distributions of individuals according to their mothers' educational status

Study Level	Iraq	Türkiye
	n (%)	n (%)
Not-Literate	434 (42.76%)	257 (21.06%)
Literate	0 (0%)	84 (6.88%)
Primary School	283 (27.88%)	434 (35.57%)
Middle School	141 (13.89%)	183 (15%)
High School	73 (7.19%)	180 (14.75%)
Diploma or Bachelor	81 (7.98%)	67 (5.49%)
Master	2 (0.1%)	10 (0.82%)
Ph.D.	1 (0.09%)	5 (0.41%)
Total	1015	1220

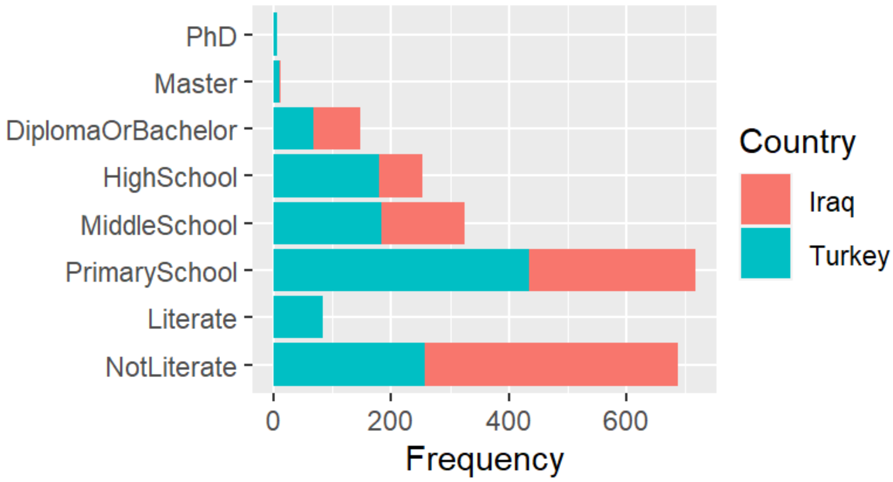


Figure 5.3 Frequency distributions of individuals according to their mothers' educational status

From Table 5.4 the frequency distributions of individuals' mothers' educational status are seen. It is observed that the educational status of the fathers of the individuals is generally low and they are predominantly composed of primary school, middle school and high school graduates, and primary school holds the greatest ranking amongst the two countries, with 283, 434 mothers' of students from both Türkiye and Iraq having completed primary school.

Table 5.5 Frequency distributions of individuals according to their fathers' occupations

Working Status	Iraq	Türkiye
	n (%)	n (%)
Not-Live	18 (1.77%)	23 (1.89%)
Public Personnel	549 (54.09%)	146 (11.97%)
Private Sector Employee	31 (3.05%)	45 (3.69%)
Self-employed-artisan-businessowner	349 (34.38%)	771 (63.2%)
Retired Working	6 (0.59%)	1 (0.08%)
Retired Not-working	62 (6.11%)	169 (13.85%)
Not-working, Unemployed	0 (0%)	65 (5.33%)
Total	1015	1220

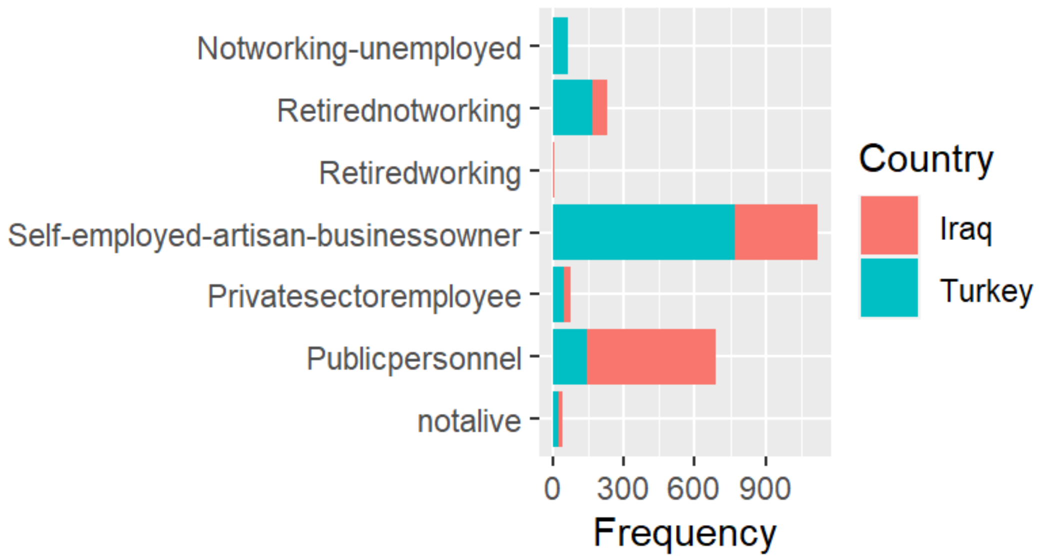


Figure 5.4 Frequency distributions of individuals according to their fathers' occupations

Table 5.5 demonstrates that the majority of fathers' occupations in the two countries are distinct from one another, with 549 (54.09%) fathers of Iraqi students working as public employees and 771 (63.2%) fathers of Turkish students working as self-employed artisan business owner. Father unemployment rates in Iraq and Türkiye are 0 and 65, respectively.

Table 5.6 Frequency distributions of individuals according to their mothers’ occupations

Working Status	Iraq	Türkiye
	n (%)	n (%)
Not-Live	5 (0.49%)	5 (0.41%)
Public Personnel	115 (11.33%)	31 (2.54%)
Private Sector Employee	14 (1.38%)	13 (1.07%)
Self-employed-artisan-businessowner	11 (1.08%)	145 (11.89%)
Retired Working	3 (0.3%)	10 (0.82%)
Retired Not-working	865 (85.22%)	959 (78.61%)
Housewife	2 (0.2%)	57 (4.67%)
Not-working, Unemployed	0 (0%)	0 (0%)
Total	1015	1220

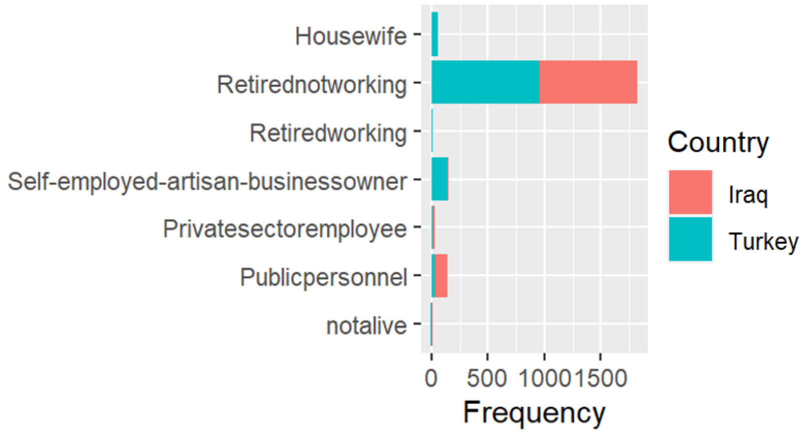


Figure 5.5 Frequency distributions of individuals according to their mothers' occupations

Table 5.6 shows that the majority of mothers' professions is retired nonworking in the two countries, with 865 (85.2%) and 959 (78.6%) mothers of both countries are retired nonworking, and these number may lead to misleading the information, like the majority of students misunderstand between retired nonworking and housewife which is true in particular in the case of Iraq.

Table 5.7 Frequency distributions of individuals according to smoking

	Iraq	Türkiye
	n (%)	n (%)
Yes	131 (12.91%)	252 (20.66%)
No	816 (80.39%)	834 (68.36%)
Sometimes	68 (6.7%)	134 (10.89%)
Total	1015	1220

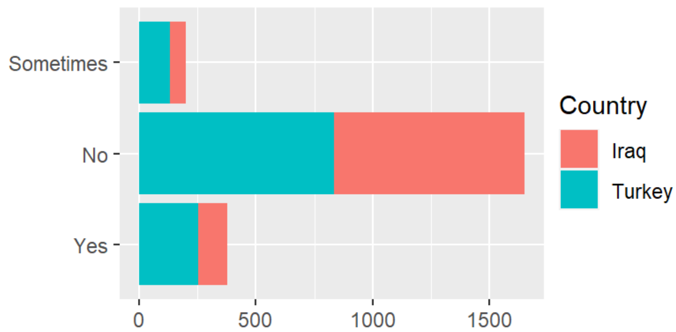


Figure 5.6 Frequency distributions of individuals according to smoking

The state of their smoking was inquired of the students. The number of non-smokers was discovered to be 816 (80.4%) and 834 (68.4%) students from each Iraq and Türkiye, with a stunning majority, as shown in Table 5.7. The reason for this rate's high prevalence is that people tend to conceal their smoking habits.

Table 5.8 Frequency distributions according to the income distribution of families

	Iraq	Türkiye
	n (%)	n (%)
Less than 200$	114 (11.23%)	310 (25.41%)
200-400	199 (19.61%)	483 (39.59%)
400-600	239 (23.55%)	226 (18.52%)
600-1000	244 (24.04%)	142 (11.64%)
1000-1500	141 (13.89%)	36 (2.95%)
More than 1500$	78 (7.68%)	23 (1.89%)
Total	1015	1220

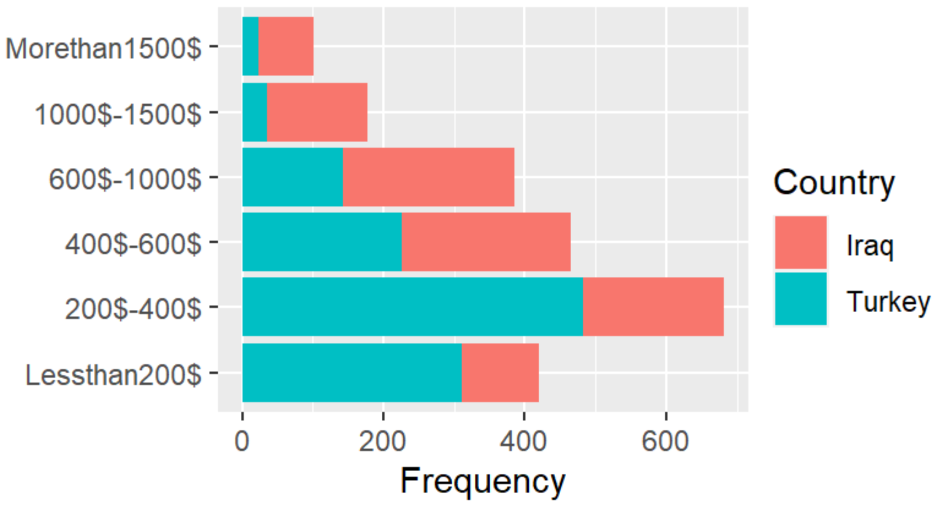


Figure 5.7 Frequency distributions according to the income distribution of families

Since internet access involves a certain cost and therefore the relationship between family income and internet addiction is wanted to be examined, the distribution of individuals' family incomes has been obtained. As can be seen in Table 5.8, it is observed that the income distribution of families is gathered at low values. Accordingly, the income level in the sample has gained weight between 400$ and 600$ for Iraq sample and 200$-

400$ for Türkiye sample, and there are 239 (23.5%) and 483 (39.6%) families in both countries in these income range. The determination of family incomes having such low values is important for this study in examining the sociological and psychological conditions of the students.

Table 5.9 Frequency distributions according of having internet at home

	Iraq	Türkiye
	n (%)	n (%)
Yes	939 (92.51%)	919 (75.33%)
No	76 (7.49%)	301 (24.67%)
Total	1015	1220

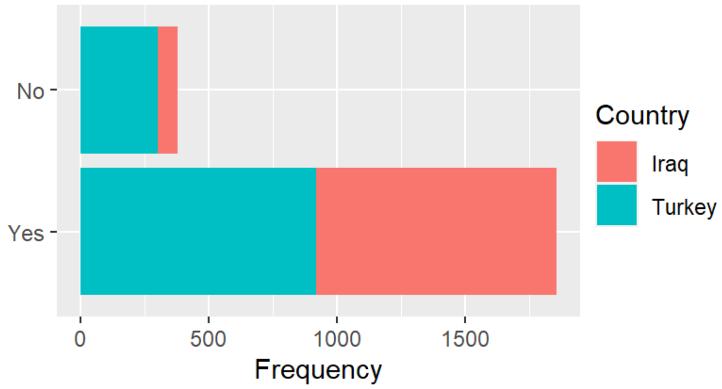


Figure 5.8 Frequency distributions according of having internet at home

Table 5.9 demonstrate that 939 (92.1%) 919 (75.3%) of students from each Iraq and Türkiye have an internet in their home.

Table 5.10 Frequency distributions according of individuals according to the most frequent use of the internet

	Iraq	Türkiye
	n (%)	n (%)
Research	318 (31.33%)	249 (20.41%)
On-line Education	67 (6.6%)	220 (18.03%)
Chat	225 (22.17%)	297 (24.34%)

Table 5.11 Frequency distributions according of individuals according to the most frequent use of the internet (continued)

	Iraq	Türkiye
	n (%)	n (%)
News	240 (23.65%)	61 (5%)
Music-Movie	103 (10.15%)	293 (24.02%)
Game	40 (3.94%)	76 (6.23%)
Pornography	3 (0.3%)	4 (0.33%)
Shopping	11 (1.08%)	16 (1.31%)
Gambling	8 (0.79%)	4 (0.33%)
Total	1015	1220

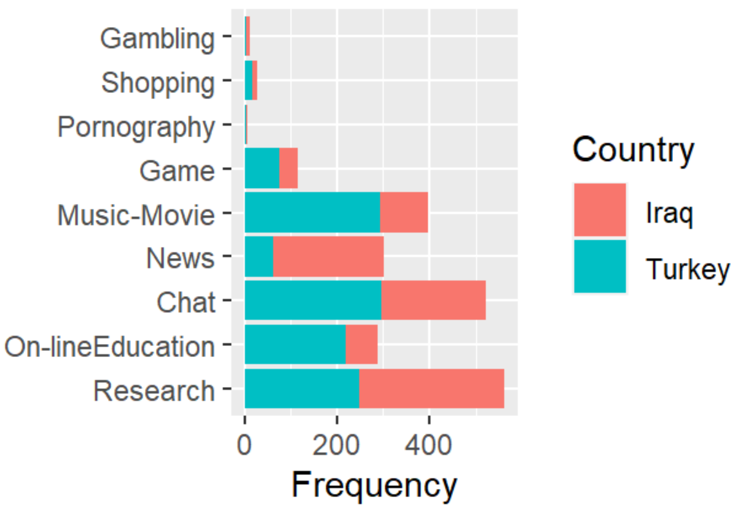


Figure 5.9 Frequency distributions according of individuals according to the most frequent use of the internet

Individuals were asked to indicate their purpose of using the Internet. However, it is thought that most of the internet usage purposes in Table 5.10 are used by the individual. Therefore, in order to obtain more precise results, individuals were asked to indicate the purpose of using the Internet most frequently. According to this; when Table 5.10 is examined, it is seen that the majority of individuals use the Internet for "Research" and "Chat" for Iraqi sample, and “Chat” and “Music-Movie” for Türkiye sample. In addition, the fact that a higher rate is expected in the use of the internet for pornographic purposes, nevertheless at a low rate of 0.3% for both Iraq and Türkiye, is interpreted as the fact that individuals cannot easily state this purpose in the scale forum due to embarrassment and disgust. The distribution in the most common usage purpose is shown in (Figure 5.9).

Table 5.12 Frequency distributions according to pandemic forced individual to use more internet

	Iraq	Türkiye
	n (%)	n (%)
Yes	651 (64.14%)	1055 (86.48%)
No	364 (35.86%)	165 (13.52%)
Total	1015	1220

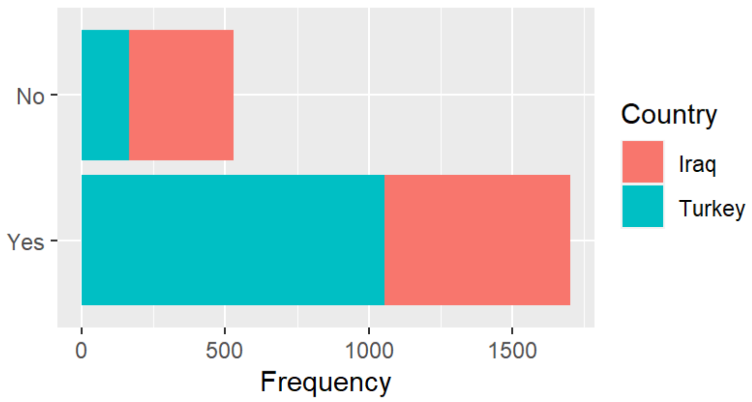


Figure 5.10 Frequency distributions according to pandemic forced individual to use more Internet.

During the pandemic COVID-19 which most people stay all time at home and change the lifestyle of people and education, student asked to explain the relation between pandemic and using more internet in that era. As it can be seen from table 5.11, the most student 651 (64.14%) and 1055(86.5%) of Iraq and Türkiye respectively, clarify that the pandemic COVID-19 forced them to use more time on internet.

Table 5.13 Frequency distributions according to infecting with COVID-19

	Iraq	Türkiye
	n (%)	n (%)
Yes	377 (37.14%)	429 (35.16%)
No	638 (62.86%)	791 (64.84%)

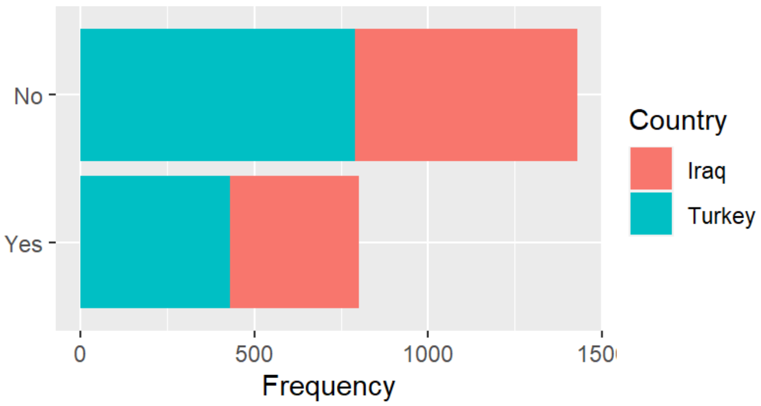


Figure 5.11 Frequency distributions according to infecting with COVID-19

From the Table 5.12 students from both countries were asked to clarify that are they get infected with COVID-19, as it can be demonstrate from above table, most student 638 (62.9%) and 791 (64.8%) from Iraq and Türkiye correspondingly answered with no, which lead the most students do not get infected during pandemic COVID-19.

Table 5.14 Frequency distributions according to the type of device for study

	Iraq	Türkiye
	n (%)	n (%)
Laptop	229 (22.56%)	437 (35.82%)
PC	30 (2.96%)	81 (6.64%)
Smart Phone	686 (67.59%)	678 (55.57%)
Tablet	70 (6.9%)	24 (1.97%)
Total	1015	1220

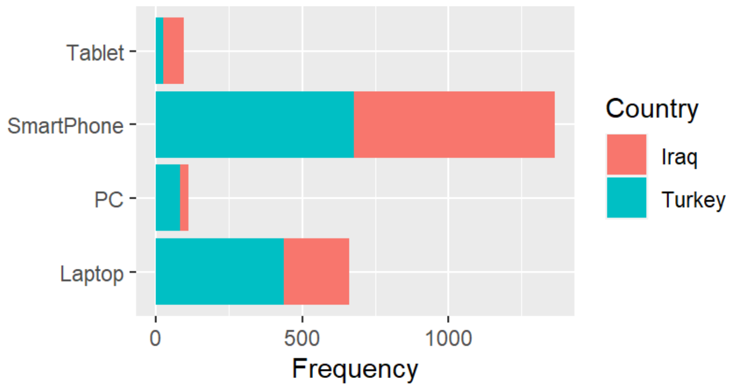


Figure 5.12 Frequency distributions according to the type of device for study

From Table 5.13 it can be seen that most student from both countries use smart phone for study online or using Internet.

Table 5.15 Frequency distributions according using internet during COVID-19 lockdown

	Iraq	Türkiye
	n (%)	n (%)
Less than 1 hour in a day	88 (8.67%)	30 (2.46%)
1-2 Hours a day	116 (11.43%)	70 (5.74%)
2-4 Hours a day	168 (16.55%)	185 (15.16%)
4-6 Hours a day	202 (19.9%)	327 (26.8%)
6-8 Hours a day	218 (21.48%)	329 (26.97%)
More than 8 hours in a day	223 (21.97%)	279 (22.87%)
Total	1015	1220

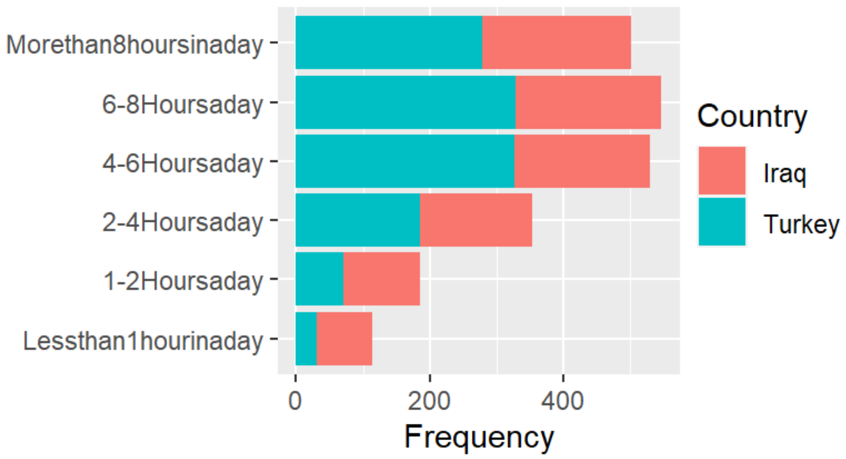


Figure 5.13 Frequency distributions according using internet during COVID-19 lockdown

From Table 5.14 it can be seen that most student from Iraq use Internet more than eight hours a day, and in the other hands, student form Türkiye use Internet between six to eight hours per day, which we can conclude that the students from Iraq, use Internet more than Türkiye and they will face internet addiction more against Turkish Students.

Table 5.16 Frequency distributions according using internet before COVID-19 lockdown

	Iraq	Türkiye
	n (%)	n (%)
Less than 1 hour in a day	142 (13.99%)	69 (5.66%)
1-2 Hours a day	228 (22.46%)	243 (19.92%)
2-4 Hours a day	299 (29.46%)	404 (33.11%)
4-6 Hours a day	159 (15.67%)	326 (26.72%)
6-8 Hours a day	109 (10.74%)	100 (8.2%)
More than 8 hours in a day	78 (7.68%)	78 (6.39%)
Total	1015	1220

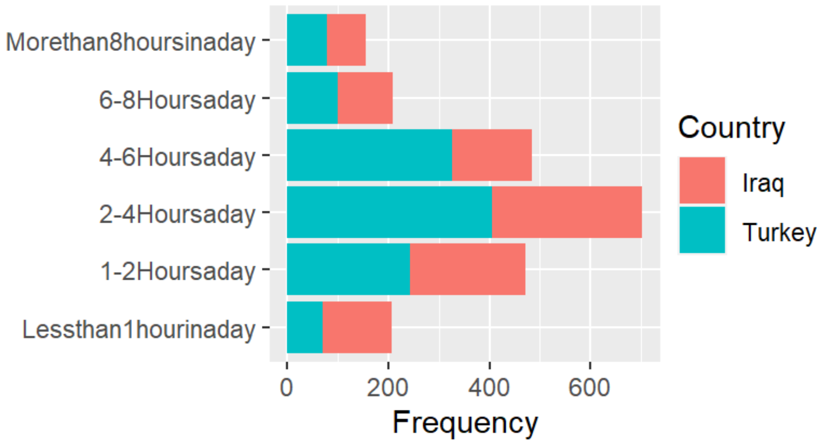
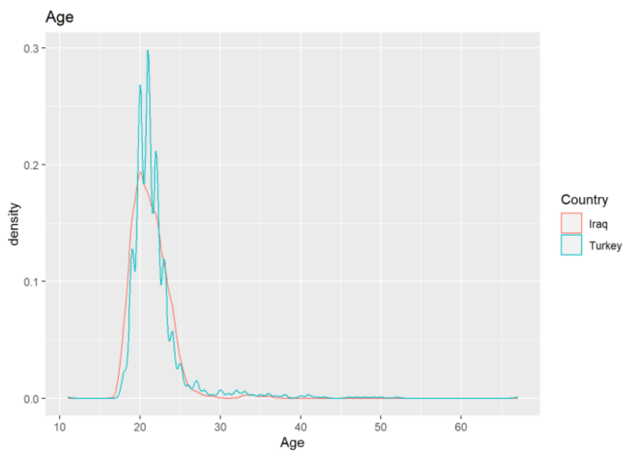


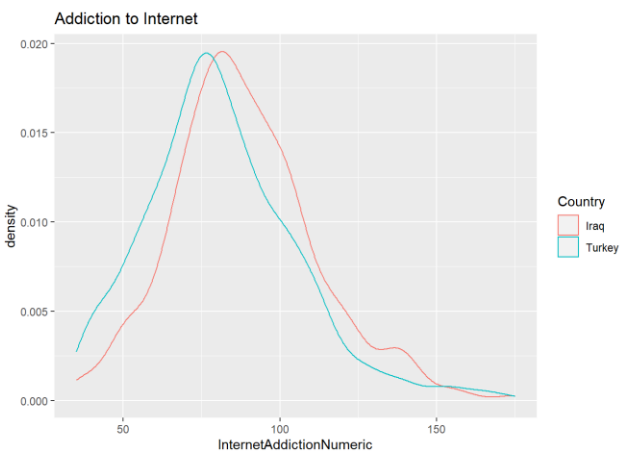
Figure 5.14 Frequency distributions according using internet before COVID-19 lockdown

From the Table 5.15 it was clarified that the pandemic has big impact to use more internet among student in both countries, and from Table 5.14 and 5.15 it in obvious that there is huge different in using internet among students before and during COVID-19 pandemic, which from Iraqi sample before pandemic the high rank of hourly usage is between 2-4 hours a day, that have been changed to more than 8 hours in a day. In other hand, from Türkiye sample before pandemic the high rank of hourly usage is between 2-4 hours a day, that have been changed to 6-8 hours a day.

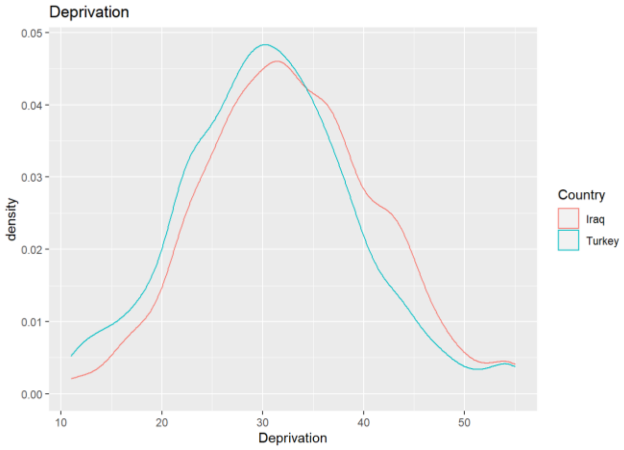
From above state it claims that the pandemic has strong effect of using Internet, which lead significantly to internet addiction among students of both countries.



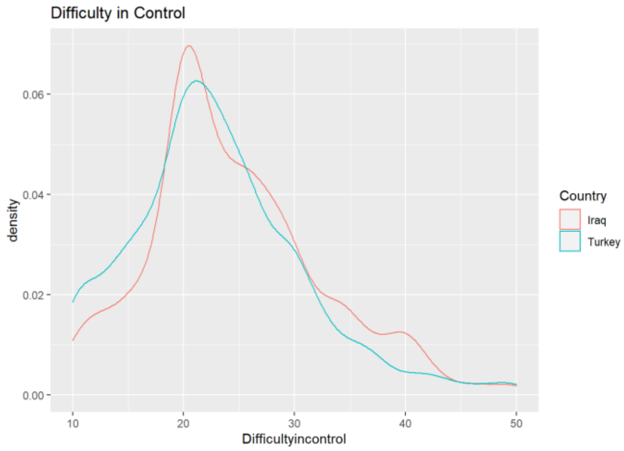
a. density plot of Age for both countries



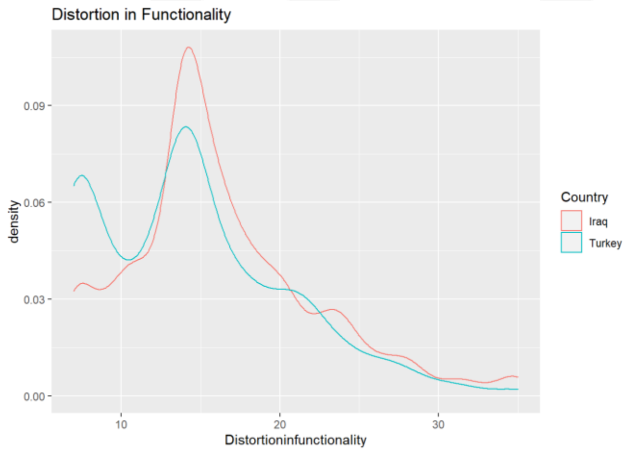
b. density plot of Addiction to internet for both countries



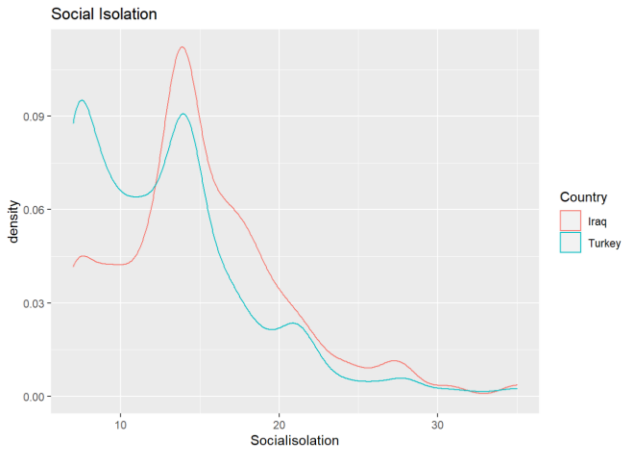
c. density plot of Deprivation for both countries



d. density plot of Difficulty in Control for both countries



e. density plot of Distortion in Functionality for both countries



f. density plot of Social Isolation for both countries

Figure 5.15 Density plot for both data sets

From the above density plots which help to visualize the overall shape of a distribution, and as a result the we have got from the shape of density plot of (age, Internet addiction, Deprivation, Difficulty in control, Distortion in functionality, Social isolation) we can conclude that the both country almost have the same shape and they are almost equal.

Regarding the Pairs plots (Appendix 3), it demonstrates how the upper panel will display the correlation between both the continuous variables, the bottom panel will present the continuous variables' scatter plots, and the diagonal panel will display the continuous variables' density plots. The data frame's columns can be chosen using the columns argument before being plotted. Pairs plot is a useful visualization tool for researcher, it shows the relationship between every variable and every other one.

Before datamining we need to get the correlation matrix, we apply the correlation matrix between four dimension of internet addiction (Deprivation, Difficulty in control, Distortion in functionality, Social isolation) and independent variables

which it is demonstrate from the above plots, and we can state the following outcome;

- There is a relationship (linear relationship) between country and all dependent variables (Internet Addiction and its components; Deprivation, Difficulty in control, Distortion in functionality, Social isolation), and here *** shows that the p-values are less than 0.01.
- Stage, Siblings Number, Smoke, Being infected with COVID-19, Quality of online education during COVID-19 do not have significant correlation with dependent variables.
- Gender has significant correlation with the variables Internet Addiction, Deprivation, Distortion in functionality, Social isolation.
- Age has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality.
- Mother and Father Education do not have significant correlation with dependent variables.
- Father Job has significant correlation with the variables Internet Addiction, Deprivation, Distortion in functionality, Social isolation.
- Mother Job and Family Income has significant correlation with Deprivation.
- Having Internet at Home has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Social isolation.
- Purpose of using Internet has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality.
- Hours of using Internet has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality, Social isolation.

- Force of Using Internet During COVID-19 has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality.
- Electronic Device has significant correlation with the variables Internet Addiction, Difficulty in control, Distortion in functionality, Social isolation.
- Hours of Using Internet During COVID-19 has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality, Social isolation.
- Hours of Using Internet before COVID-19 has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality, Social isolation.

5.2. Correlation between Independent Variables

From our data set we have 19 independent variables, and Spearman method have been used for finding the correlation coefficients and p-values of categorical variables. After finding the result of the correlation between variables which have been shown in the appendix 4, we can conclude that the most of features (predictors) are significantly correlated to each other.

5.3. Two Step Cluster Analysis

The scale total scores that we consistently obtained for each individual in the internet addiction scale were categorised by a two-step clustering analysis (two-step cluster). Two-step clustering analysis allows the data set to be divided into homogeneous subgroups. In other words, it aims to divide the heterogeneous data set into homogeneous subclasses or clusters. It is reported in the literature that statistical studies obtained from the clusters thus formed have healthier results (Kayri, 2007). The general purpose of this technique is to classify ungrouped data according to their similarities and assist the researcher in obtaining relevant, useful and summative information (Harrigan, 1985). The internet addiction scale is a 5-point Likert scale, with a minimum score of 35 and a maximum score of 175 for each participant. After the two-stage cluster analysis, 2 clusters with a cut-off score of 93 points (including 93) for the Türkiye sample and 88 points (including 88) for the Iraq sample were obtained. These clusters are named as "non-dependent" and "dependent". Information about the clusters and their frequencies obtained for Türkiye and Iraq are presented in Table 5.16.

Table 5.17 Two-step clustering analysis

		Türkiye		Iraq	
		N	%	*n*	%
Cluster	Non-dependent	893	73.2	542	53.4
	Dependent	327	26.8	473	46.6

In Table 5.16, the data set consisting of 1220 students forming the sample of Türkiye is gathered under 2 clusters, 893 students (73.20%) in the first cluster and 327 students (26.80%) in the second cluster show common characteristics. On the other hand, the data set consisting of 1015 students forming the Iraq sample is gathered under 2 clusters, 542 students (53.40%) in the first cluster and 473 students (46.60%) in the second cluster show common characteristics. on the contrary, for both data sets, we can say that the individuals in one cluster differed from the individuals in the other cluster. According to these similarities, internet addiction was made categorical and the MARS model was established and analyses were carried out.

5.4. Confusion Matrix

A Confusion Matrix (CM) is a device which allows patterns to be recognised and allocated to their relevant class, and to see at a glance whether a system has confused them. According to binary logic, results can either be positive or negative and true or false. Thus, we have True Positive and True Negative (TP and TN), with the opposite being either False Positive or False Negative (FP or FN).

As underlined by Concha et al (2014), the ideal is clearly that the sum of TP and TN be close to that shown by the pattern; and that the sum of FP and FN be as close to zero as possible. The device is most usefully used for binary classifications, for example success/failure, yes/no and suchlike. As shown below in (Figure 5.16), a Confusion Matrix has been presented. It shows outcomes according to actual and estimated results. A CM can also be used in determining the AUC-ROC curve, precision, recall and accuracy of data. If the number of classification problems is greater than two, one simply sums the outcomes accordingly.

	Predicted			
Actual		**Negative (N)**	**Positive (N)**	**Total**
	Negative (N)	TN	FP	TN+FP
	Positive (N)	FN	TP	FN+TP
	Total	TN+FN	FP+TP	TP+FP+TN+FN

Figure 5.16 Confusion matrix

5.4.1. Accuracy

Accuracy can quite simply be defined as the ration between how many input samples there were versus the number of predictions which were correct according to the model. It therefore follows that when analysing the AUC-ROC curve, the True and False Positive Rates (TPR and FPR) are determined.

The following formulae apply when determining TPR and FPR:

$$TPR = \frac{TP}{TP + FN} \quad \text{and} \quad FPR = \frac{FP}{FP + TN} \tag{5.1}$$

It thus follows that accuracy is:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{5.2}$$

5.4.2. Precision

As the name suggests, the degree of precision shows how well predictions from the data turn out to be correct in reality, and is key when the FN rate is of lesser concern than that of the FP. The formula for precision is shown below:

$$Precision = \frac{TP}{TP + FP} \tag{5.3}$$

5.4.3. Sensitivity

The TPR or sensitivity shows the degree to which a model is able to respond and predict appropriately, and the ideal would be a situation where there are a low number of FN such that the sensitivity is as close to 1 as possible.

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.4}$$

5.4.4. F1 Score

When analysing data and indeed when using scientific instruments in general, we can either increase sensitivity or precision, not both. There always has to be a balance between the two. We therefore use something called the F1 score to illustrate the situation. As it uses a harmonic mean, the greatest value is where the sensitivity value is the same as that for precision.

$$F = \frac{2 * Sensitivity * Precision}{Sensitivity + Precision} \tag{5.5}$$

5.4.5. Area Under the Receiver Operator Characteristic Curve (ROC)

Kim and Hwang (2020), the means by which the optimum classifier can be determined is through use of the Receiver Operator Characteristic (ROC). Concha et al. (2014) explains that the balance between TPR and FPR can be illustrated by the ROC curve. It helps us to select, organise and visualise the various classifiers according to their degree of robustness, and this measure can be shown by the Area Under Curve (AUC).

When examining the ROC curve, on a graph of distribution against probability, we can ascertain the level of performance relating to a certain classifier. As the ideal is a high TPR and correspondingly low FPR, there will be different distributions for each. A classification threshold is required as well as the ratio of false negatives to false positives. A typical situation is shown below in (Figure 5.17):

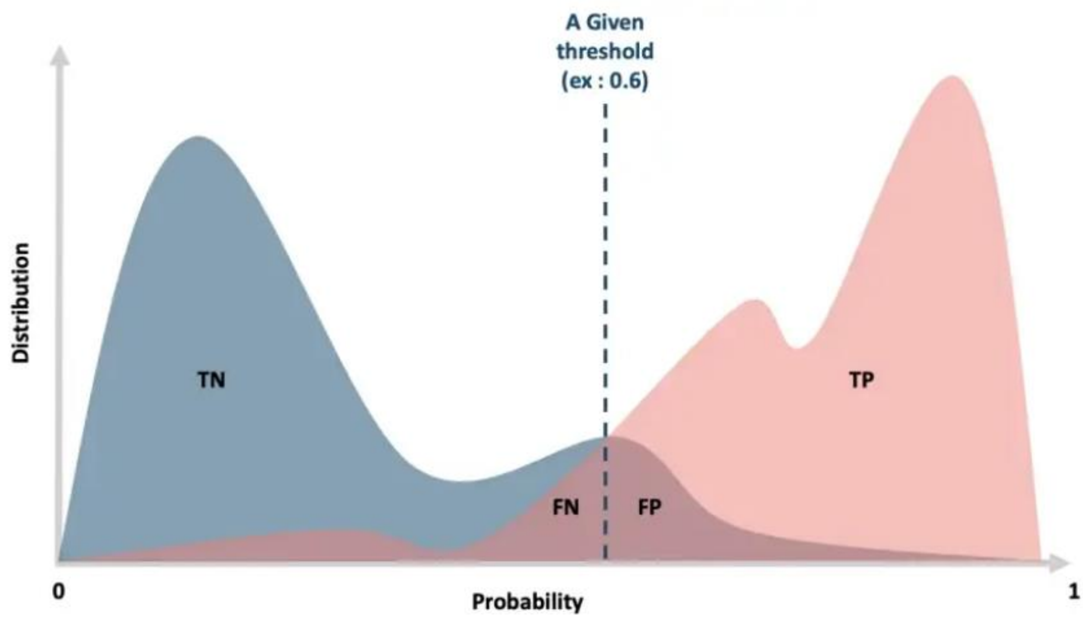


Figure 5.17 Distribution against probability

5.4.6. Area Under the Curve (AUC)

The AUC is a measure which is superior to the ROC curve in that it is more explicit in allowing the comparison and contrast between two models, not simply just that of TRP and FPR against each other. It details the level of performance across all thresholds of classification. With a value between 0 and 1, it shows how effective the model is in a clear and transparent way. It shows how well initial predictions relate to actual outcomes. Taken from Kim and Hwang (2020) is an example of a typical ROC graph as revealed below in (Figure 5.18). It is clear that of three different classifiers A, B and C, the one with the best TPR against FPR and thus performance is A, then B and finally C.

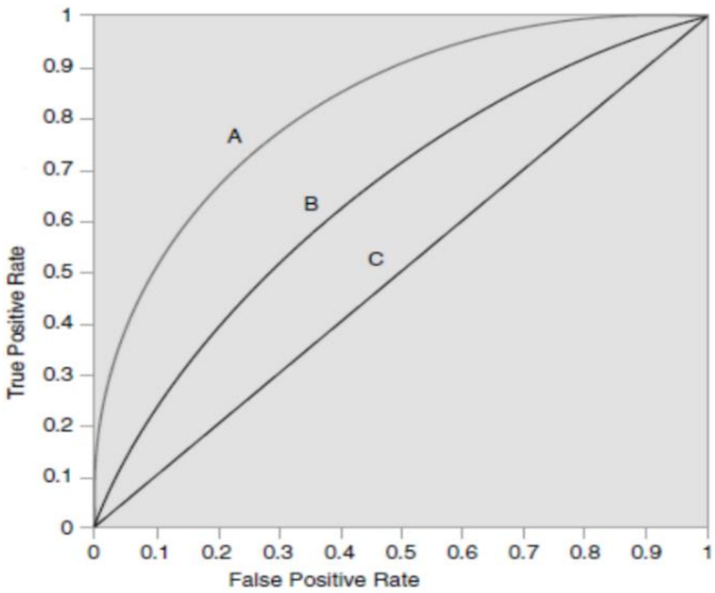


Figure 5.18 A typical ROC plot

6. FINDINGS AND COMMENTS

In this section, the results of the MARS data mining method performed on the data of Türkiye and Iraq are given separately, and comparatively based on each sub-problem.

6.1. Classification Performance of MARS Data Mining Method in Predicting Internet Addiction Levels of University Students

In this section, the findings of the MARS data mining method regarding the Türkiye and Iraq data set are given in order, firstly Türkiye and secondly Iraq, in terms of classification performance. The SPM 8.2 program's default values were used as a default value for the MARS data mining study, and operations were carried out taking the complete data set into account, regardless of learning data and test data. As a result of the analysis made with the MARS data mining method for the Türkiye dataset, when students are classified as "addicted/not addicted" in the name of internet addiction, Table 6.1 lists the number of students who fit into each of these groupings.

Table 6.1 Confusion matrix obtained from MARS analysis (Türkiye sample)

	Independent	Dependent	Total Number of Students
Independent	614	279	893
Dependent	118	209	327
Total Number of Students	732	488	1220

As seen in Table 6.1, from Confusion Matrix obtained from MARS analysis of the Türkiye sample, MARS analysis method classified 209 students out of 327 students from Türkiye sample in the Dependent category on the basis of internet addiction scale, and classified 118 students in the independent category. Moreover, in the Türkiye sample, it classified 614 students out of 893 students in the independent category on the basis of internet addiction scale, and classified 279 students in the dependent category.

As a result of the analysis made with MARS data mining method for the Iraq data set, when students are classified as "addicted/not addicted" in the name of internet addiction, Table 6.2 lists the number of students who fit into each of these groupings.

Table 6.2 Confusion matrix obtained from MARS analysis (Iraq sample)

	Independent	Dependent	Total Number of Students
Independent	405	137	543
Dependent	156	317	473
Total Number of Students	561	454	1015

As seen in Table 6.2, from Confusion Matrix obtained from MARS analysis of the Iraq sample, MARS analysis method classified 317 students out of 473 students from Iraq sample in the Dependent category on the basis of internet addiction scale, and classified 156 students in the independent category. Moreover, in the Iraq Sample, it classified 405 students out of 543 students in the independent category on the basis of internet addiction scale, and classified 137 students in the dependent category.

The classification performance of the MARS data mining method as a result of comparing both data sets with each other in the form of actual class range and estimated class range is given in Table 6.3.

Table 6.3 MARS analysis result classification performance rates in Türkiye and Iraq samples

Criteria	Türkiye	Iraq
Accurate Classification Rate	67.46%	71.13%
Specificity Ratio/ Specificity	63.91%	67.02%
Sensitivity Rate	68.76%	74.72%
Precision Rate	83.88%	72.19%
F1-Score	75.57%	73.44%
Area Under the ROC Curve (AUC)	71.28%	80.23%

In Table 6.3, model classification rates of accuracy, specificity, sensitivity, precision, F1-statistic values and AUC values of the ROC curve are given, as a result of MARS analysis method. As can be seen in Table 6.3, it was observed that beside the slight differences there was similarity in the percentage levels considered.

In this direction, the dataset values, which are the default values of the program, were taken into account for the above-mentioned criteria and are interpreted comparatively below.

6.1.1. In Terms of Correct Classification Rate

In terms of correct classification rate, MARS analysis method differs among both countries, while the correct classification rate obtained with the MARS analysis method for Türkiye and Iraq is 67.46% and 71.13% respectively, it is seen that the sample from Iraq has a higher correct classification rate than the Türkiye sample. In other words, the MARS analysis method for Iraq sample, classified a dependent variable in the dependent category and independent variable in the independent variable category with higher accuracy.

6.1.2. In Terms of Specificity Rate

In terms of Specificity Rate, MARS analysis method differs among both countries, while the specificity rate obtained with the MARS analysis method for Türkiye and Iraq is 63.91% and 67.02% respectively, it is seen that the sample from Iraq has a higher specificity rate than the Türkiye sample. In other words, 67.02% of the students from Iraq sample predicted by the MARS analysis method in the independent category are actually in the independent category. Therefore, the MARS analysis method has a higher rate of estimating the true negative from Iraq sample.

6.1.3. In Terms of Sensitivity Rate

In terms of sensitivity rate, MARS analysis method differs among both countries, while the sensitivity rate obtained with the MARS analysis method for Türkiye and Iraq is 68.76% and 74.72% respectively, it is seen that the sample from Iraq sample has a higher sensitivity rate than the Türkiye sample. In other words, 74.72% of the students predicted by the MARS analysis method in the dependent category are actually in the dependent category. Therefore, the MARS analysis method has a higher rate of estimating the correct positive from Iraq sample.

6.1.4. In Terms of Precision Rate

In terms of precision, MARS analysis method differs among both countries, while the precision rate obtained with the MARS analysis method for Türkiye and Iraq is 83.88% and 72.19% respectively, it is seen that the sample from Türkiye sample has a higher precision rate than the Iraq sample. In other words, 83.88% of the students of Türkiye's sample, who were in the dependent category with the MARS analysis method were classified in the Dependent category.

Therefore, this value obtained by the MARS analysis method, which is higher than the MARS analysis method for Iraq Sample, gives the ratio of the students who were predicted correctly, among all those who were predicted successfully (Şevgin and Önen, 2022).

6.1.5. In Terms of F1-Statistic

In terms of F1–Statistic classification rate for both countries, the F1-statistic is a statistical measure obtained as a result of the harmonic average of precision and sensitivity measures. MARS analysis method differs among both countries, while the F1-statistic obtained with the MARS analysis method for Türkiye and Iraq is 75.57% and 73.44% respectively, it is seen that the sample from Türkiye has a higher F1–Statistic rate than the Iraq sample.

In other words, the F-1 statistical result for Türkiye sample, which is the harmonic mean of MARS analysis method precision and sensitivity, showed a higher classification success. Furthermore, the MARS analysis method for Türkiye sample was higher than the MARS analysis method for Iraq sample in identifying dependent and distinguishing independent category.

6.1.6. In Terms of the Area Under the ROC Curve (AUC)

With regard to the area under the ROC curve (AUC), the larger the area under the ROC curve, the higher the classification success rate of the model. As the area under the ROC curve, MARS analysis method differs among both countries, while the area under the ROC curve obtained with the MARS analysis method for Türkiye and Iraq is 71.28% and 80.23% respectively, it is seen that the sample from Iraq sample has a higher specificity rate than the Türkiye sample. As this curve approaches the upper left corner, the better the performance of the classifier. ROC curve graphs of MARS for both Türkiye and Iraq are given in (Figures 6.1 and 6.2).

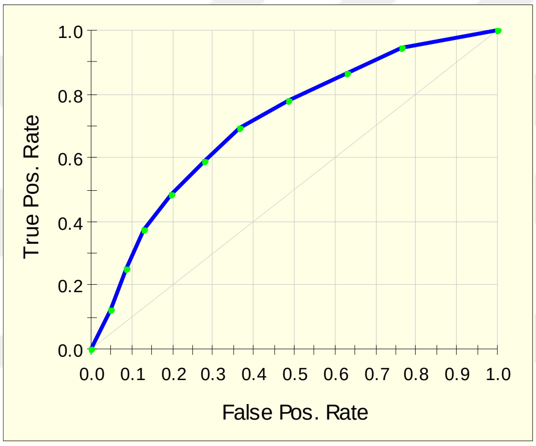


Figure 6.1 ROC curve graphs of MARS analyze for Türkiye sample for 9 basis function

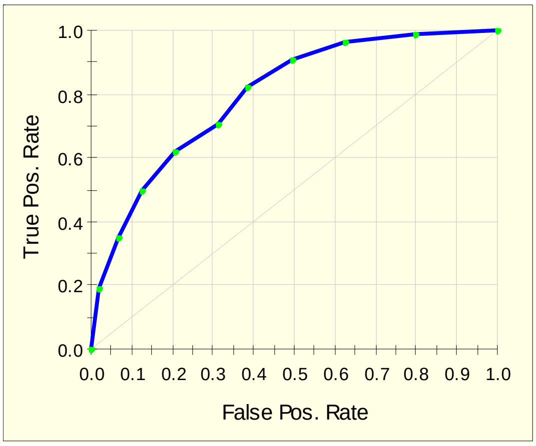


Figure 6.2 ROC curve graphs of MARS analyze for Iraq sample for 20 basis functions

As observed in (Figures 6.1 and 6.2), the MARS analysis method for Iraq sample has a higher area under the ROC curve compared to the MARS analysis method for Türkiye sample.

In other words, according to Şevgin and Önen (2022) the MARS analysis method Iraq's sample minimized the false positive-rate compared to the MARS analysis method for Türkiye sample, whereas rising the true positive-rate to a higher level. Moreover, the MARS analysis technique for Iraq sample classified dependent and independent type with less error than the MARS analysis method for Türkiye sample.

All the finding above showed that the MARS analysis method for Iraq sample offers higher rates and achieves better than the MARS analysis method for Türkiye sample in terms of classification performance.

6.2. The Key Predictors for the Türkiye Sample Based on MARS Analysis.

In the MARS analysis method, the model setup phase was created based on the established (default) values. This research was carried out with the participation of 1220 students from Türkiye and 1015 students from Iraq. Moreover, before starting the MARS analysis method, determining the maximum number of fundamental functions is a point to be considered. The determination of the maximum number of fundamental functions is realized by determining the smallest GCV value that will be given in the analysis by the number of basic functions entered in various trials, at least more than twice the number of independent variables (Şevgin and Önen, 2022). Below are the GCV values for determining the maximum number of fundamental functions in Table 6.4.

Table 6.4 GCV values for determining the maximum number of basic functions

Maximum Number of Basic Functions	GCV Value	Maximum Number of Basic Functions	GCV Value
35	0.18015	42	0.18015
36	0.18002	43	0.18022
37	0.18012	44	0.18027
38	0.17999	45	0.18033
39	0.18009	46	0.18025

Table 6.4 GCV values for determining the maximum number of basic functions (continued)

Maximum Number of Basic Functions	GCV Value	Maximum Number of Basic Functions	GCV Value
40	0.18017	47	0.1803
41	0.18006	48	0.18032

As can be seen in Table 6.4, the maximum number of basic functions with the lowest value was determined as 38.

Since the whole data are used as learning data, the lowest GCV value represents the number of basic functions for which the most appropriate model will be created. This is valid for the default case where the entire data set is used completely.

Since the data set is considered as learning data, the lowest GCV value obtained is shown in (Figure 6.3).

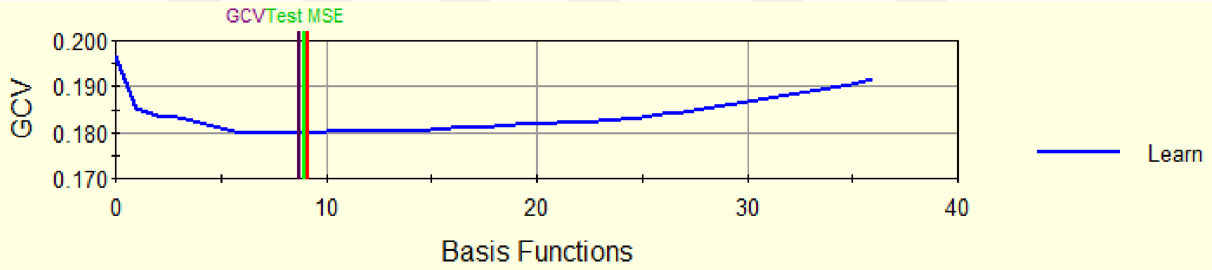


Figure 6.3 The number of basic functions indicated by the GCV value obtained as a result of the MARS analysis

The number of basic functions corresponding to the lowest GCV value was obtained as 9. Information showing the lowest GCV and MSE result values obtained for the established MARS model is given in Table 6.5.

Table 6.5 Result values for the established MARS model

Basis Functions	N Predictors	N Inputs	Effective Parameters	GCV	GCV R- Sq	Learn MSE	Learn R- Sq	Learn ROC
36	14	13	97	0.119155	0.02524	0.1623	0.17272	0.76008
35	14	13	94.33	0.19069	0.02963	0.16234	0.17253	0.75985
34	14	13	91.67	0.18984	0.03397	0.16238	0.17233	0.76019
33	14	13	89	0.18901	0.03818	0.16244	0.17204	0.75939
32	14	13	86.33	0.1882	0.04232	0.1625	0.17171	0.75912
31	14	13	83.67	0.18741	0.04631	0.16259	0.17127	0.75809
30	14	13	81	0.18669	0.04999	0.16272	0.17059	0.75588
29	14	13	78.33	0.18588	0.0541	0.16278	0.17031	0.75677
28	14	13	75.67	0.18518	0.0577	0.16292	0.1696	0.75594
27	14	13	73	0.18457	0.06075	0.16315	0.16843	0.75538
26	14	13	70.33	0.18391	0.06412	0.16332	0.16755	0.75412
25	14	13	67.67	0.18348	0.06634	0.16369	0.16567	0.75407
24	14	13	65	0.18294	0.06907	0.16397	0.16425	0.75353
23	14	13	62.33	0.18256	0.07101	0.16438	0.16214	0.75192
22	14	13	59.67	0.18231	0.0723	0.16491	0.15945	0.7506
21	14	13	57	0.18213	0.07318	0.16551	0.15638	0.7454
20	14	13	54.33	0.18193	0.07423	0.16608	0.15347	0.74156
19	14	13	51.67	0.18164	0.07568	0.16658	0.15092	0.73636
18	14	13	49	0.18135	0.07718	0.16707	0.14842	0.73651
17	13	12	46.33	0.18117	0.07806	0.16767	0.14536	0.73321
16	12	11	43.67	0.18094	0.07924	0.16822	0.14257	0.73225
15	12	11	41	0.18047	0.08164	0.16854	0.14092	0.73262
14	10	10	38.33	0.1803	0.08251	0.16915	0.13785	0.73333
13	10	10	35.67	0.18024	0.08281	0.16986	0.13423	0.72987
12	10	10	33	0.18016	0.08319	0.17055	0.13069	0.72771
11	9	9	30.33	0.18024	0.08281	0.17139	0.12642	0.72234
10	9	9	27.67	0.18033	0.08233	0.17225	0.12204	0.7191
9	8	8	25	0.17999	0.08408	0.17269	0.11979	0.71278
8	8	8	22.33	0.18001	0.084	0.17348	0.11578	0.71306
7	8	8	19.67	0.18006	0.08373	0.1743	0.11158	0.71042
6	7	7	17	0.18006	0.08372	0.17508	0.10762	0.70418
5	6	6	14.33	0.18112	0.07833	0.17689	0.09839	0.69551
4	5	5	11.67	0.18222	0.07275	0.17875	0.0889	0.68883
3	4	4	9	0.18329	0.06727	0.1806	0.07947	0.67736
2	4	4	6.33	0.18373	0.06503	0.18183	0.07319	0.67129
1	2	2	3.67	0.18524	0.05735	0.18413	0.06147	0.65942
0	0	0	1	0.19651		0.19619		

The number of basic functions in the first column of Table 6.5, the number of independent variables in determining each regression equation in the second column, the number of arguments used in the third column, the effective parameters value in the fourth column, and the GCV values in the fifth column, the GCV-R^2 values in the sixth column, the MSE values of the learning data in the seventh column, and the eighth column is the value for R^2 and at the end in the ninth column is the value under the curve ROC.

The MARS analysis method, which starts model building with 0 basic functions, reaches its most complex level with 36 basic functions. Then, the basic functions that had no impact on the model were removed from the model with the backward pruning process, and the model using 8 variables with 9 basic functions and GCV value of 0.17999 indicated by the green bar formed the most appropriate model.

As a result of the variance analysis performed to define the relative contributions of each independent variable and the interactions between the variables, it was observed that the most appropriate model consisted of 8 functions, as seen in (Figure 6.4).

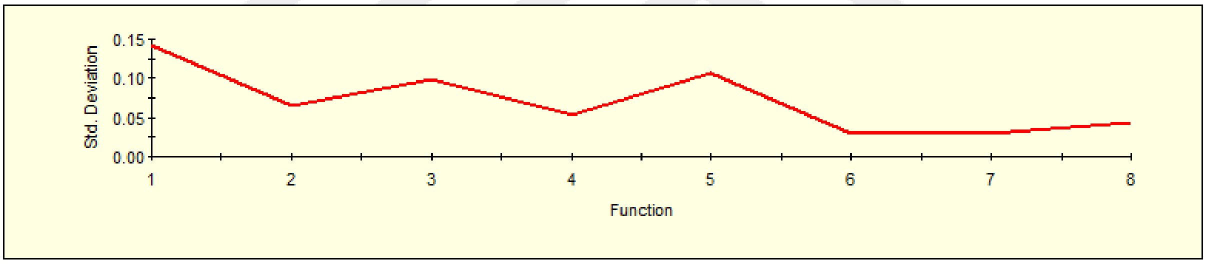


Figure 6.4 Analysis of variance plot for the best model

The contribution of the independent variables in the model of the variance analysis graph created for the most appropriate model was calculated by estimating the adjusted R^2 of the model as a result of excluding the ANOVA function at each step. The variance analysis information regarding these graphically illustrated functions is given in Table 6.6.

Table 6.6 Variance analysis for the optimal model

Function	Standard Deviation	Cost of omission	No. of Basis Functions	No. of Effective Parameters	Variables
1	0.14136	0.18308	2	5.333	NUMBER OF BROTHERS, DAILY INTERNET USE
2	0.06553	0.1821	1	2.667	MOTHER EDUCATION, DAILY INTERNET USE
3	0.09935	0.18177	1	2.667	AGE, DAILY INTERNET USE
4	0.05405	0.1821	1	2.667	MOTHER EDUCATION, SMOKING
5	0.10707	0.18168	1	2.667	DAILY INTERNET USE, DAILY INTERNET USE IN COVID19 QUARANTINE
6	0.03038	0.18001	1	2.667	DAILY INTERNET USE, QUALITY OF ONLINE EDUCATION IN THE COVID19 EPIDEMIC
7	0.03011	0.18011	1	2.667	DAILY INTERNET USE, DEVICE PREFERENCE FOR ONLINE IN QUARANTINE
8	0.04381	0.18113	1	2.667	SMOKING, THE QUALITY OF ONLINE EDUCATION IN THE COVID19 EPIDEMIC

In Table 6.6, the predictor variables in the functions and the number of basic functions and the values indicating the loss that will occur if the predictor variables are removed from the model are given. For example, if the number of siblings and daily internet use variables are removed from the model, the missing value in the estimations is 0.18308 and a fundamental function will decrease. Moreover, as can be seen from the table, it is seen that some predictor variables are included in the model with more than one basic function alone and combined with more than one variable and included in the model with a single basic function. The most suitable model obtained as a result of the forward step and backward step applications of the MARS analysis method is called the final model. Table 6.7 shows the table information about the final model.

Table 6.7 Final model created for the most appropriate model

Basis Function	Coefficients	Variable	Sign in	Parent Sign	parent	knots
0	0.1103					
4	-0.0350	NUMBER OF SIBILINGS	+	-	DAILY INTERNET USAGE	11.0000
5	-0.0062	NUMBER OF SIBILINGS	+	-	DAILY INTERNET USAGE	11.0000
6	0.0116	MOTHER EDUCATION	-	-	DAILY INTERNET USAGE	3.0000
9	0.0042	AGE	+	+	DAILY INTERNET USAGE	30.0000
11	0.0767	MOTHER EDUCATION	-	+	SMOKING	5.0000
12	0.0085	DAILY INTERNET USAGE DURING COVID19 QUARANTINE	+	+	DAILY INTERNET USAGE	1.0000
13	0.0062	THE QUALITY OF ONLINEEDUCATION IN THE COVID19 OUTPUT	-	+	DAILY INTERNET USAGE	6.0000
15	-0.0453	IN QUARANTINEONLINETO THE DEVICE CHOICE	+	+	DAILY INTERNET USAGE	3.0000
18	0.0565	THE QUALITY OF ONLINEEDUCATION IN THE COVID19 OUTPUT	-	+	SMOKING	4.0000

In Table 6.7, the nodal points where the slope of the variables and pairs of variable change, the basic functions of the variables, the coefficients of the basis functions and the signs indicating the direction of the variables. In the final model; The knot values formed by each predictor variable or variables that are independent of all variables, the directions of their slopes, the model coefficients that contribute to the regression equation as a result of their multiplication with the basic function, and information showing which variables the basic functions are distributed. The basic function equations consisting of predictor variables and node values are presented in Table 6.8.

Table 6.8 Fundamental equations of functions for the optimal model

BASIC FUNCTION
BF1 = max (0, DAILY INTERNET USAGE - 1);
BF2 = max (0, SMOKING - 2);
BF3 = max (0, 2 - SMOKING);
BF4 = max (0, NUMBER - 11) * BF1;
BF5 = max (0, 11 - NUMBER OF SISTER) * BF1;
BF6 = max (0, MOTHEREDUCATION - 3) * BF1;
BF9 = max (0, 30 - AGE) * BF1;
BF11 = max (0.5 - MOTHEREDUCATION) * BF2;
BF12 = max (0, DAILY INTERNET USAGE DURING COVID19 QUARANTINE - 1) * BF1;
BF13 = max (0, QUALITY OF ONLINEEDUCATION IN THE COVID19 OUTPUT - 6) * BF1;
BF15 = max (0, DEPARTMENT CHOICE FOR ONLINE IN QUARANTINE - 3) * BF1;
BF18 = max (0.4 - QUALITY OF ONLINE EDUCATION IN THE COVID19 OUTPUT) * BF3;

As can be seen in Table 6.8, the maximum number of basic functions determined based on the GCV value in the MARS analysis method was determined as 18. Then, 9 of the 18 basic functions were used to establish the most appropriate model. There are 12 basic functions in Table 6.8 alone. The basic functions BF1, BF2 and BF3 were used in the creation of other basic functions and were not included in the final model. The closed representation of the most appropriate model established with basic functions for internet addiction category is shown in Table 6.9.

Table 6.9 Internet addiction model table for the most appropriate model

MODEL INTERNETADDICTIONCATEGORY = BF4 BF5 BF6 BF9 BF11 BF12 BF13 BF15 BF18;

The MARS analysis method, multiplied by the model coefficients of the 9 basic functions that it uses while creating the most appropriate model, gives the regression equation in Table 6.10. Multiplication of each fundamental function by its coefficient gives its contribution to the model.

Table 6.10 Regression equation for optimal model

Y = 0.11026 – 0.0350244 * 0.00620594 * BF5 + 0.0116034 * BF6 + 0.00422424 * BF9 + 0.0766942 * BF11 + 0.00848087 * BF12 + 0.0061938 * BF13 – 0.0453293 * BF15 + 0.0564568 *

In this equation, the F and p-values calculated for the 3 degrees of freedom were calculated as F=18.29641 and the p-value was calculated as $p = 0.001$, as can be seen in Appendix 5. This shows that the model is meaningful. The data obtained with the internet addiction Scale for Türkiye sample were analyzed with the MARS analysis method and the most important predictors of internet addiction category were tried to be determined. The predictor variables included in the analysis and the significance levels of these variables on the variable predicted within themselves in the established model are given in Table 6.11.

Table 6.11 MARS analysis method table of significance levels of variables for Turkish sample

Variable	Score
DAILY INTERNET USAGE	100
MOTHEREDUCATION	50.82
NUMBER OF SIBLINS	47.58
SMOKING	45.52
AGE	36.07
DAILY INTERNET USAGE DURING COVID19 QUARANTINE	35.18
THE QUALITY OF ONLINEEDUCATION IN THE COVID19 OUTPUT	25.74
IN QUARANTINEONLINETO THE DEVICE CHOICE	9.47
DAILY USE OF THE INTERNET BEFORE COVID19QUARANT	0
FATHERHOOD	0
MOTHER'S PROFESSION	0
GENDER	0
FATHEREDUCATION	0
FAMILY INCOME	0
COVID19PANDEMIDEINTERNETUSE	0
Have you had COVID19	0
HOME INTERNET	0
INTERNETUSEPURPOSE	0
MOTHER'S PROFESSION_mis	0
Class	0

The most important predictors obtained on Türkiye sample as a result of the analysis made with the MARS analysis method are given in Table 6.11. Accordingly, the most important predictors of internet addiction Scale for Türkiye sample are respectively; Daily Internet Usage, Mothers' Education, Number of Siblings, Smoking, Age, Daily Internet Usage During COVID-19 Quarantine, The Quality of Online education in the COVID-19 Pandemic, the device used, The MARS analysis method's variables connected to the dependent variable were ranked according to their importance, starting from 100 points. The variable with the highest relationship with internet addiction scale was the self-efficacy perception variable, while the variable with the lowest relationship was the type of device to use the internet. Variables that have little or no relationship with internet addiction scale and are not included in the analysis are respectively; Daily use of the Internet before COVID-19 pandemic, Fatherhood, Mother's profession, Gender, Fathers' education, Family income, COVID-19 pandemic and internet use, have you effected by COVID-19, do you have internet at home, Internet use purpose, Mothers' Profession, and Stage of study were obtained as insignificant.

6.3. The Key Predictors for the Iraq Sample Based on MARS Analysis.

Below are the GCV values for determining the maximum number of fundamental functions for Iraq sample in Table 6.12.

Table 6.12 GCV values for determining the maximum number of basic functions for Iraq sample

Maximum Number of Basic Functions	GCV Value	Maximum Number of Basic Functions	GCV Value
35	0.20480	42	0.20527
36	0.20510	43	0.20518
37	0.20467	44	0.20550
38	0.20496	45	0.20480
39	0.20499	46	0.20501
40	0.20536	47	0.20512
41	0.20559	48	0.20516

As can be seen in Table 6.12, the maximum number of basic functions for Iraq sample with the lowest value was determined as 37.

In cases where the data set is separated as learning data and test data, the lowest point of the GCV value represents the number of basic functions for which the most appropriate model will be created (Salfort Systeam, 2018). In our cases, the Iraq sample is used as learning data, the lowest GCV value represents the number of basic functions for which the most appropriate model will be created. This is valid for the default case where the entire data set is used completely.

Since the data set is considered as learning data, the lowest GCV value obtained is shown in (Figure 6.5).

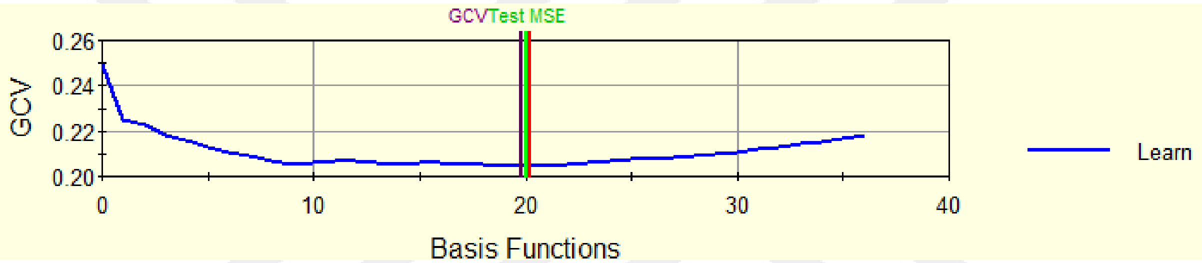


Figure 6.5 The number of basic functions indicated by the GCV value obtained as a result of the MARS analysis of Iraq sample

The number of basic functions corresponding to the lowest GCV value for Iraq sample was obtained as 20. Information showing the lowest GCV and MSE result values obtained for the established MARS model for Iraq sample is given in Table 6.13.

Table 6.13 Result values for the established MARS model for Iraq sample

Basis Functions	N Predictors	N Inputs	Effective Parameters	GCV	GCV R- Sq	Learn MSE	Learn R- Sq	Learn ROC
36	14	14	100	0.218	0.1256	0.17718	0.288	0.8144
35	14	14	97.25	0.2167	0.1308	0.17719	0.288	0.8144
34	14	14	94.5	0.2155	0.1357	0.17724	0.2878	0.814
33	14	14	91.75	0.2143	0.1405	0.17732	0.2874	0.8143
32	14	14	89	0.2131	0.1452	0.17739	0.2871	0.8141
31	14	14	86.25	0.212	0.1496	0.17753	0.2866	0,81350
30	14	14	83.5	0.211	0.1539	0.17768	0.286	0.8138
29	14	14	80.75	0.21	0.1579	0.17789	0.2851	0.8118
28	14	14	78	0.2091	0.1615	0.17817	0.284	0.8108
27	14	14	75.25	0.2085	0.164	0.17869	0.2819	0.8094

Table 6.13 Result values for the established MARS model for Iraq sample (continued)

Basis Functions	N Predictors	N Inputs	Effective Parameters	GCV	GCV R- Sq	Learn MSE	Learn R- Sq	Learn ROC
26	14	14	72.5	0.208	0.1659	0.17932	0.2794	0.8074
25	14	14	69.75	0.2075	0.1678	0.17997	0.2768	0.8054
24	14	14	67	0.2067	0.1709	0.18033	0.2753	0.8034
23	14	14	64.25	0.2059	0.1741	0.18067	0.274	0.802
22	14	14	61.5	0.2054	0.176	0.1813	0.2714	0.803
21	14	14	58.75	0.2049	0.1781	0.18189	0.2691	0.8042
20	14	14	56,00	0.2047	0.1791	0.18271	0.2658	0.8023
19	14	14	53.25	0.2048	0.1785	0.1839	0.261	0.7992
18	13	13	50,50	0.205	0.1778	0.18512	0.2561	0.7948
17	13	13	47.75	0.2054	0.1763	0.18652	0.2505	0.794
16	13	13	45	0.2059	0.1743	0.18802	0.2444	0.7897
15	13	13	42.25	0.206	0.174	0.18916	0.2399	0.7841
14	12	12	39.5	0.2055	0.1757	0.18985	0.2371	0.7803
13	11	11	36.75	0.2062	0.1729	0.11916	0.2302	0.7742
12	10	10	34	0.2067	0.171	0.19309	0.224	0.7676
11	10	10	31.25	0.2071	0.1694	0.19455	0.2182	0.7639
10	9	9	28.5	0.2063	0.1725	0.19489	0.2168	0.7634
9	8	8	25.75	0.2053	0.1765	0.19504	0.2162	0.7628
8	7	7	23	0.207	0.1697	0.19775	0.2053	0.7613
7	7	7	20.25	0.2088	0.1624	0,20059	0.1939	0.7583
6	6	6	17.5	0.2109	0.1541	0.2037	0.1814	0.7516
5	6	6	14.75	0.2128	0.1467	0.20661	0.1697	0.7437
4	5	5	12	0.2156	0.1353	0.21052	0.154	0.7188
3	4	4	9.25	0.2182	0.125	0.21421	0.1392	0.7034
2	4	4	6.5	0.2227	0.1067	0.21989	0.1164	0.6796
1	2	2	3.75	0.2253	0.0965	0.22362	0.1014	0.6591
0	0	0	1	0.2493		0.24884		

The number of basic functions in the first column of Table 6.13, the number of independent (predictor) variables in determining each regression equation in the second column, the number of arguments used in the third column, the effective parameters value in the fourth column, and the GCV values in the fifth column, the GCV-R^2 values in the sixth column, the MSE values of the learning data in the seventh column, and the eighth column is the value for R^2 and at the end in the ninth column is the value under the curve ROC.

The MARS analysis method for Iraq sample, which starts model building with 0 basic functions, reaches its most complex level with 36 basic functions. Then, the basic functions that did not contribute to the model were removed from the model with the backward pruning process, and the model using 14 predictor variables with 20 basic functions and GCV value of 0.20467 indicated by the green bar formed the most appropriate model.

As a result of the variance analysis for Iraq sample performed to define the relative contributions of each independent variable and the interactions between the variables, it was observed that the most appropriate model consisted of 15 functions, as seen in (Figure 6.6).

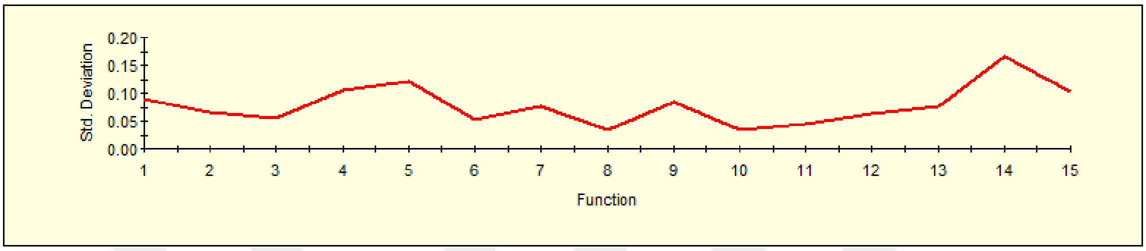


Figure 6.6 Analysis of variance plot for the best model for Iraq sample

The contribution of the independent variables in the model of the variance analysis graph created for the most appropriate model was calculated by estimating the adjusted R^2 of the model as a result of excluding the ANOVA function at each step. The variance analysis information regarding these graphically illustrated functions is given in Table 6.14.

Table 6.14 Variance analysis for the optimal model of the Iraq sample

Function	Standard Deviation	Cost of duty	No of basis Functions	No of Effective Parameters	Variables
1	0.08903	0.20538	1	2.750	DAILYINTERNETUSAGE
2	0.06707	0.20693	1	2.750	FAMILYINCOME, DAILYINTERNETUSAGE
3	0.05464	0.20584	1	2.750	INTERNETUSAGEPURPOSE, DAILYINTERNETUSAGE
4	0.1503	0.20693	2	5.500	DAILYINTERNETUSAGE, DAILYINTERNETUSAGEINCOVID19QUARANTINE
5	0.12145	0.20921	2	5.500	INTERNETUSAGECOVID19PANDEMIC, QUALITYONLINEEDUCATIONCOVID19PANDEMIC
6	0.05282	0.20631	1	2.750	AGE, DAILYINTERNETUSAGE
7	0.07625	0.20544	2	5.500	MOTHER'SEDUCATION, INTERNETUSAGECOVID19PANDEMIC
8	0.03528	0.20483	1	2.750	DAILYINTERNETUSAGE, DEVICEPREFERENCEFORONLINEINQUARANTINE
9	0.08515	0.20734	2	5.500	COVID19PREQUARANTINEDAILYINTERNETUSAGE, QUALITYONLINEEDUCATIONCOVID19PANDEMIC
10	0.03457	0.20483	1	2.750	NUMBEROFSIBLINGS, COVID19PREQUARANTINEDAILYINTERNETUSAGE

Table 6.14 Variance analysis for the optimal model of the Iraq sample (continued)

Function	Standard Deviation	Cost of duty	No of basis Functions	No of Effective Parameters	Variables
11	0.04430	0.20538	1	2.750	INTERNETHOME, INTERNETUSAGECOVID19PANDEMIC
12	0.06437	0.20503	1	2.750	INTERNETUSAGEPURPOSE, INTERNETUSAGECOVID19PANDEMIC
13	0.07758	0.20531	1	2.750	FATHERSJOB, DAILYINTERNETUSAGE
14	0.16633	0.20765	1	2.750	MOTHERSJOB, DAILYINTERNETUSAGE
15	0.1193	0.20541	2	5.500	NUMBEROFSIBLINGS, DAILYINTERNETUSAGE

In Table 6.14, the predictor variables in the functions and the number of basic functions and the values indicating the loss that will occur if the predictor variables are removed from the model are given. For example, if daily internet use variable is removed from the model, the missing value in the estimations is 0.20538 and a fundamental function will decrease. In addition, as can be seen from the table, it is seen that some predictor variables are included in the model with more than one basic function alone and combined with more than one variable and included in the model with a single basic function. The most suitable model obtained as a result of the forward step and backward step applications of the MARS analysis method is called the final model. Table 6.15 shows the table information about the final model of the Iraq sample.

Table 6.15 Final model created for the most appropriate model of the Iraq sample

Basis Function	coefficients	Variable	Sign in	Parent Sign	Parent	knots
0	0.2830					
1	0.0365	DAILYINTERNETUSAGE	+			3.0000
4	0.2705	FAMILYINCOME	-	-	DAILYINTERNETUSAGE	2.0000
7	-0.0167	INTERNETUSAGEPURPOSE	+	+	DAILYINTERNETUSAGE	3.0000
8	0.0176	DAILYINTERNETUSAGEINCOVID19QUARANTINE	-	+	DAILYINTERNETUSAGE	4.0000
9	0.0359	DAILYINTERNETUSAGEINCOVID19QUARANTINE	+	-	DAILYINTERNETUSAGE	4.0000
10	-0.0573	QUALITYONLINEEDUCATIONCOVID19PANDEMIC	+	+	INTERNETUSAGECOVID19PANDEMIC	2.0000
11	-0.3137	QUALITYONLINEEDUCATIONCOVID19PANDEMIC	-	+	INTERNETUSAGECOVID19PANDEMIC	2.0000
12	0.0157	AGE	+	+	DAILYINTERNETUSAGE	22.0000
14	0.8877	MOTHER'SEDUCATION	-	+	INTERNETUSAGECOVID19PANDEMIC	6.0000

Table 6.15 Final model created for the most appropriate model of the Iraq sample (continued)

Basis Function	coefficients	Variable	Signin	Parent Sign	Parent	knots
15	0.0362	MOTHER'SEDUCATION	+	+	INTERNETUSAGECOVID19PANDEMIC	6.0000
16	-0.0451	DEVICEPREFERENCEFORONLINEINQUARANTINE	-	+	DAILYINTERNETUSAGE	3.0000
19	0.0097	QUALITYONLINEEDUCATIONCOVID19PANDEMIC	+	+	COVID19PREQUARANTINEDAILYINTERNETUSAGE	2.0000
20	0.0689	QUALITYONLINEEDUCATIONCOVID19PANDEMIC	-	+	COVID19PREQUARANTINEDAILYINTERNETUSAGE	2.0000
23	0.0521	NUMBEROFSIBLINGS	+	+	COVID19PREQUARANTINEDAILYINTERNETUSAGE	10.0000
25	0.2395	INTERNETHOME	-	+	INTERNETUSAGECOVID19PANDEMIC	1.0000
27	-0.0360	INTERNETUSAGEPURPOSE	+	+	INTERNETUSAGECOVID19PANDEMIC	6.0000
28	-0.0669	FATHERSJOB	-	+	DAILYINTERNETUSAGE	1.0000
29	0.0575	MOTHERSJOB	+	-	DAILYINTERNETUSAGE	1.0000
30	0.2636	NUMBEROFSIBLINGS	+	+	DAILYINTERNETUSAGE	5.0000
32	-0.2517	NUMBEROFSIBLINGS	-	+	DAILYINTERNETUSAGE	4.0000

In Table 6.15, the nodal points where the slope of the variables and pairs of variable change, the basic functions of the variables, the coefficients of the basis functions and the signs indicating the direction of the variables. In the final model; The knot values formed by each predictor variable or variables that are independent of all variables, the directions of their slopes, the model coefficients that contribute to the regression equation as a result of their multiplication with the basic function, and information showing which variables the basic functions are distributed. The basic function equations consisting of predictor variables and node values are presented in Table 6.16.

Table 6.16 Fundamental equations of functions for the optimal model of Iraq sample

BASIC FUNCTION
BF1 = max (0, DAILYINTERNETUSAGE - 3);
BF2 = max (0, 3 - DAILYINTERNETUSAGE);
BF4 = max (0, 2 - FAMILY INCOME) * BF2;
BF5 = max (0, INTERNETUSAGECOVID19PANDEMIC - 1);
BF7 = max (0, 3 - INTERNETUSAGEPURPOSE) * BF1;

Table 6.17 Fundamental equations of functions for the optimal model of Iraq sample (continued)

BASIC FUNCTION
BF8 = max (0, DAILYINTERNETUSAGEINCOVID19QUARANTINE - 4) * BF1;
BF9 = max (0.4 - DAILYINTERNETUSAGEINCOVID19QUARANTINE) * BF1;
BF10 = max (0, QUALITYONLINEEDUCATIONCOVID19PANDEMIC - 2) * BF5;
BF11 = max (0, 2 - QUALITYONLINEEDUCATIONCOVID19PANDEMIC) * BF5;
BF12 = max (0, AGE - 22) * BF1;
BF14 = max (0, MOTHER'SEDUCATION - 6) * BF5;
BF15 = max (0.6 - MOTHER'SEDUCATION) * BF5;
BF16 = max (0, DEVICEPREFERENCEFORONLINEINQUARANTINE - 3) * BF1;
BF18 = max (0, COVID19PREQUARANTINEDAILYINTERNETUSAGE - 1);
BF19 = max (0, QUALITYONLINEEDUCATIONCOVID19PANDEMIC - 2) * BF18;
BF20 = max (0, 2 - QUALITYONLINEEDUCATIONCOVID19PANDEMIC) * BF18;
BF23 = max (0, NUMBEROFSIBLINGS - 10) * BF18;
BF25 = max (0, INTERNETHOME - 1) * BF5;
BF27 = max (0.6 - INTERNETUSAGEPURPOSE) * BF5;
BF28 = max (0, FATHERSJOB - 1) * BF2;
BF29 = max (0, MOTHERSJOB - 1) * BF2;
BF30 = max (0, NUMBEROFSIBLINGS - 5) * BF2;
BF32 = max (0, NUMBEROFSIBLINGS - 4) * BF2;

As can be seen in Table 6.16, the maximum number of basic functions of Iraq sample determined based on the GCV value in the MARS analysis method was determined as 32. Then, 20 of the 32 basic functions were used to establish the most appropriate model. There are 23 basic functions in Table 6.16 alone. The basic functions BF2, BF5 and BF18 were used in the creation of other basic functions and were not included in the final model. The closed representation of the most appropriate model established with basic functions for internet addiction category is shown in Table 6.17.

Table 6.18 Internet addiction model table for the most appropriate model of Iraq sample

MODEL INTERNETADDICTIONCATEGORIC = BF1 BF4 BF7 BF8 BF9 BF10 BF11 BF12 BF14 BF15 BF16 BF19 BF20 BF23 BF25 BF27 BF28 BF29 BF30 BF32;

The MARS analysis method, multiplied by the model coefficients of the 20 basic functions that it uses while creating the most appropriate model, gives the regression equation in Table 6.18. Multiplication of each fundamental function by its coefficient gives its contribution to the model.

Table 6.19 Regression equation for optimal model of Iraq sample

Y = 0.282973 + 0.0365357 * BF1 + 0.270469 * BF4
- 0.0166532 * BF7 + 0.0176 * BF8 + 0.0358795 * BF9
- 0.0572517 * BF10 - 0.313663 * BF11
+ 0.0157315 * BF12 + 0.887684 * BF14
+ 0.0361724 * BF15 - 0.0450909 * BF16
+ 0.097489 * BF19 + 0.0689124 * BF20
+ 0.0521286 * BF23 + 0.23951 * BF25
- 0.0359754 * BF27 - 0.0669053 * BF28
+ 0.0575376 * BF29 + 0.263582 * BF30
- 0.251731 * BF32;

In this equation, the F and p-values calculated for the 3 degrees of freedom were calculated as F=17.98916 and the p-value was calculated as p = 0.001, as can be seen in Appendix 6. This shows that the model is statistically significant. The data obtained with the internet addiction Scale for Iraq sample were analyzed with the MARS analysis method and the most important predictors of internet addiction category were tried to be determined. The predictor variables included in the analysis and the significance levels of these variables on the variable predicted within themselves in the established model are given in Table 6.19.

Table 6.20 MARS analysis method table of significance levels of variables for Iraq sample

Variable	Score
DAILYINTERNETUSAGE	100,00
INTERNETUSAGECOVID19PANDEMIC	66,55
QUALITYONLINEEDUCATIONCOVID19PANDEMIC	43,16
COVID19PREQUARANTINEDAILYINTERNETUSAGE	35,92
MOTHERSJOB	35,87
INTERNETUSAGEPURPOSE	34,79

Table 6.19 MARS analysis method table of significance levels of variables for Iraq sample (continued)

Variable	Score
FAMILYİNCOME	31,24
DAILYINTERNETUSAGEINCOVID19QUARANTINE	31,20
AGE	26,55
NUMBEROFSIBLINGS	19,59
MOTHER'SEDUCATION	18,21
İNTERNETHOME	17,44
FATHERSJOB	16,64
DEVICEPREFERENCEFORONLINEINQUARANTINE	8,32
GENDER	0,00
FATHER'SEDUCATION	0,00
HAVEYOUHADCOVID19	0,00
DOYOUSMOKE	0,00
STAGE	0,00

The most important predictors obtained on Iraq sample as a result of the analysis made with the MARS analysis method for Iraq sample are given in Table 6.19. Accordingly, the most important predictors of internet addiction Scale for Iraq sample are respectively; Daily Internet Usage, Internet Usage COVID-19 Pandemic, Quality Online Education COVID-19 Pandemic, COVID-19 Pre-Quarantine Daily Internet Usage, Mothers' Job, internet Usage Purpose, Family Income, Daily Internet Usage in COVID-19 quarantine, Age, Number of Siblings, Mothers' Education, Internet Home, Fathers' Job, Device Preference for Online in Quarantine.

The variables associated with the dependent variable in the MARS analysis method were ranked according to their importance, starting from 100 points. The variable with the highest relationship with internet addiction scale was the self-efficacy perception variable, while the variable with the lowest relationship was Device Preference for Online in Quarantine. Variables that have little or no relationship with internet addiction scale and are not included in the analysis are respectively; Gender, Fathers' Education, Have You Had COVID-19, Do You Smoke and Stage of study were obtained as insignificant.

7. DISCUSSION AND CONCLUSION

In this chapter, both implications for theory, future research and practice are discussed, prior to which was presented a discussion of what emerged during the study. The key takeaways are that internet addiction certainly impacts students at tertiary level in terms of their grades; cognitive and social impacts are also observed at the same time. Further research into the area is clearly needed, especially in today's fast paced world where technological innovations occur with increasing frequency. The present study therefore is of use for those interested in how the fields of student outcomes, technology use (with its associated distractions) and internet addiction interlink.

Within the present study, it was shown that internet addiction is a relatively recent yet growing phenomenon. It was suggested that ways should be found for teachers and others to help identify and manage those displaying such addiction, given that it is not a traditional area of research. If internet addiction comes to be accepted by those at higher levels within the educational community – as well as those students who might suffer from it – better support to students may be the result, which would benefit them on their educational journey.

Classification performances of MARS data mining method in predicting internet addiction levels of university students according to data obtained from both countries (Türkiye -Iraq), and it solved several question which arise by this study such as Does it differ according to the Correct classification, Specificity, Sensitivity, Accuracy rate?, does it differ according to the F1-Statistics?, and also, does it differ in terms of the area under the ROC curve, that is, the rate of misclassification?, and the last and important question, what are the most important predictors of internet addiction of university students in (Türkiye \ Iraq) based on MARS data mining method?

A cross-sectional and non-experimental survey was conducted to investigate the internet addiction scale, as a result of this study, a 54-item have been asked which is 35 items for internet addiction scale, 13 item for socio-demographic and 6 item for COVID-19 was developed.

Several limitations were identified within the present study. The first was that it is somewhat difficult to quantify just exactly what constitutes a measure of excessive Internet usage. It could not be shown to any degree of accuracy whether varying time of usage can affect levels of internet addiction. The second hindrance was that the literature in the area relating to Iraqi students in particular is scarce.

The aim of the study is to develop an existing scale for internet addiction and to examine the factors that can affect the addiction status of individuals. The sample of the study is at the international level, and cosmopolitan universities from each country (Türkiye and Iraq) and a university from mentioned countries were selected, which is Van Yuzuncu Yil University from Türkiye and Soran University from Iraq. The target audience was determined as university students. The fact that the sample was composed of adolescents is due to the fact that internet addiction is mostly seen in adolescents and individuals in this period are open to all kinds of influences. The number of individuals in the sample is 2235; 1220 students from Van Yuzuncu Yil University, 1015 from Soran University were assigned by random sampling method. The age range of the sample ranged from 18-67 and the mean age was found to be 21.66.

Test for normality have been apply to the dataset by using Kolmogrov -Smirnov text, the test for both sample exhibit $p<0.05$, meaning that a normal distribution does not apply, a parametric approach to data analysis is inappropriate given the significance values in both data sets. Thus, a non-parametric approach needs to be taken. It should be noted however that there is no requirement to test data set assumptions when employing a MARS approach.

Scale internal consistency measured using The Cronbach's alpha (α) was found to be 0.957 for Türkiye sample and 00.936 for Iraq sample. Exploratory factor analysis was applied to the scale regarding the construct validity of the scale, the scale consists of four sub-dimensions. In order to show that the results and structure of this study are valid, confirmatory factor analysis was applied with the obtained data. These four sub-factors are; It was named as “Deprivation”, “Control Difficulty”, “Distortion in Functioning” and “Social Isolation”.

The Cronbach's alpha (α) reliability coefficients for the four sub-factors were found to be 0.887 for the first sub-factor (Deprivation), 0.898 for the second sub-factor (Control Difficulty), 0.908 for the third sub-factor (Distortion in Functioning), and 0.878 for the fourth sub-factor (Social Isolation) for Turkish sample, where in other hand, 0.847 for the first sub-factor (Deprivation), 0.859 for the second sub-factor (Control Difficulty), 0.853 for the third sub-factor (Impairment in Functioning), and 0.813 for the fourth sub-factor (Social Isolation) for Turkish sample. These indicate reliability and consistency and those data validity.

The suitability of the dataset was tested with pilot study, as a result of the Confirmatory Factor Analysis performed on the data as a result of the application to 250 students, it was seen that the structure was under the factor specified in each item as stated in the original scale (RMSEA: 0.060, CFI: 0.95, NFI: 0.90, NNFI: 0.95 and GFI: 0.81). Again, the reliability values (McDonald's ω) obtained for the four dimensions were obtained as 0.829, 0.839, 0.833 and 0.795, respectively. It was concluded that the internet addiction scale, which was adapted into datasets with these values, was valid and reliable.

The scale is in five-point Likert type and likert-style expressions; “I totally agree”, “I agree”, “I’m undecided”, “I don’t agree”, “I strongly disagree”. The scale items are scored from 5 to 1, with 5 points corresponding to the "I totally agree" degree and 1 point to the "I strongly disagree" degree. All items in the scale are oriented towards addiction and there is no need for any score transpose.

A two-stage clustering analysis was performed on the total item scores of the individuals for both samples, and the addiction status was divided into two groups. After the two-stage cluster analysis, 2 clusters with a cut-off score of 93 points (including 93) for the Türkiye sample and 88 points (including 88) for the Iraq sample were obtained. These clusters are named as “non-dependent” and “dependent”. the data set consisting of 1220 students forming the sample of Türkiye is gathered under 2 clusters, 893 students (73.20%) in the first cluster and 327 students (26.80%) in the second cluster show common characteristics. On the other hand, the data set consisting of 1015 students forming the Iraq sample is gathered under 2 clusters, 542 students (53.40%) in the first cluster and 473 students (46.60%) in the second cluster show common characteristics. on the contrary, for both data sets, we can say that the individuals in one cluster differed from the individuals in the other cluster. According to these similarities, internet addiction was made categorical and the MARS model was established and analyzes were carried out.

In this study, besides the descriptive statistics on some demographic variables, the relationships between these variables and internet addiction status were also examined and the following results were obtained:

- There is a relationship (linear relationship) between country and all dependent variables (Internet Addiction and its components; Deprivation, Difficulty in control, Distortion in functionality, Social isolation).
- Stage, Siblings Number, Smoke, Being infected with COVID-19, Quality of online education during COVID-19 do not have significant correlation with dependent variables.
- Gender has significant correlation with the variables Internet Addiction, Deprivation, Distortion in functionality, Social isolation.
- Age has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality.
- Mother and Father Education do not have significant correlation with dependent variables.
- Father Job has significant correlation with the variables Internet Addiction, Deprivation, Distortion in functionality, Social isolation.
- Mother Job and Family Income has significant correlation with Deprivation.
- Having Internet at Home has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Social isolation.
- Purpose of using Internet has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality.
- Hours of using Internet has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality, Social isolation.
- Force of Using Internet During COVID-19 has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality.
- Electronic Device has significant correlation with the variables Internet Addiction, Difficulty in control, Distortion in functionality, Social isolation.
- Hours of Using Internet During COVID-19 has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality, Social isolation.

- Hours of Using Internet before COVID-19 has significant correlation with the variables Internet Addiction, Deprivation, Difficulty in control, Distortion in functionality, Social isolation.

As a result of the analysis made with MARS data mining method for the Iraq and Türkiye datasets which has been shown in the confusion matrix, when students are classified as “addicted/ not-addicted” in the name of the internet addiction, the number of students whom have been classified as dependent are 209 and 317 students for Türkiye and Iraq respectively, in addition, the number of students whom have been classified as Not-dependent are 614 and 405 students for Türkiye and Iraq respectively. As a consequence of these results, it can be state that the students from Iraq are more addicted to the internet.

As a result of using the MARS analysis approach, model classification rates for accuracy, specificity, sensitivity, and precision, as well as F1-statistic values and AUC values of the ROC curve, are provided. In terms of correct classification rate, MARS analysis method differs among both countries, while the correct classification rate obtained with the MARS analysis method for Türkiye and Iraq is 67.46% and 71.13% respectively, it is seen that the sample from Iraq has a higher correct classification rate than the Türkiye sample. Moreover, in terms of Specificity Rate, MARS analysis method differs among both countries, while the specificity rate obtained with the MARS analysis method for Türkiye and Iraq is 63.91% and 67.02% respectively, it is seen that the sample from Iraq has a higher specificity rate than the Türkiye sample. In addition, in terms of sensitivity rate, MARS analysis method differs among both countries, while the sensitivity rate obtained with the MARS analysis method for Türkiye and Iraq is 68.76% and 74.72% respectively, it is seen that the sample from Iraq sample has a higher sensitivity rate than the Türkiye sample. In terms of precision, MARS analysis method differs among both countries, while the precision rate obtained with the MARS analysis method for Türkiye and Iraq is 83.88% and 72.19% respectively, it is seen that the sample from Türkiye sample has a higher precision rate than the Iraq sample. In terms of F1–Statistic, the F1-statistic is a statistical measure obtained as a result of the harmonic average of precision and precision measures. MARS analysis method differs among both countries, while the F1-statistic obtained with the MARS analysis method for Türkiye and Iraq is 75.57% and 73.44% respectively, it is seen that the sample from Türkiye has a higher F1–Statistic rate than the Iraq sample. In other words, the F-1 statistical result for Türkiye sample, which is the harmonic mean of MARS analysis method sensitivity and precision, showed a higher classification success. Furthermore, the MARS analysis method for Türkiye sample was higher than the MARS analysis method for Iraq sample in identifying dependent and distinguishing independent category. Furthermore, As the area under the ROC curve, MARS analysis method differs among both countries, while the area under the ROC curve obtained with the MARS analysis method for Türkiye and Iraq is 71.28% and 80.23% respectively, it is seen that the sample from Iraq sample has a higher specificity rate than the Türkiye sample. In other words, the MARS analysis method for Iraq sample classified dependent and independent category with fewer errors than the MARS analysis method.

Before starting the MARS analysis method for determining the most important predictors, determining the maximum number of fundamental functions is a point to be considered. The determination of the maximum number of fundamental functions is realized by determining the smallest GCV value that will be given in the analysis by the number of basic functions entered, as a result the 0.179 and 0.205 are lowest GCV values for Türkiye and Iraq, which determined the 38 and 37 are maximum number of basic functions for both datasets, Türkiye and Iraq respectively. In addition, the number of basic functions corresponding to the lowest GCV in both data set are 9 for Türkiye sample and 20 for Iraq sample.

The MARS analysis method, which starts model building with 0 basic functions, reaches its most complex level with 36 basic functions. Then, the basic functions that did not contribute to the model were removed from the model with the backward pruning process, and the model using 8 predictors with 9 basic functions and GCV value of 0.17999 for Türkiye, and 14 predictors with 20 basic functions and GCV value of 0.20467 for Iraq.

As a result of the variance analysis performed to define the relative contributions of each independent variable and the interactions between the variables, it was observed that the most appropriate model consisted of 8 and 15 functions for Türkiye and Iraq correspondingly.

The contribution of the independent variables in the model of the variance analysis graph created for the most appropriate model was calculated by estimating the adjusted R^2 of the model as a result of excluding the ANOVA function at each step. Therefore, after applying variance analysis for the optimal model, the predictor variables in the functions and the number of basic functions and the values indicating the loss that will occur if the predictor variables are removed from the model are provided, it stated that, from Türkiye sample, if the number of siblings and daily internet use variables are removed from the model, the missing value in the estimations is 0.18308 and a fundamental function will decrease. In the other hand, from Iraq sample, if daily internet use variable is removed from the model, the missing value in the estimations is 0.20538 and a fundamental function will decrease.

The most suitable model obtained as a result of the forward step and backward step applications of the MARS analysis method is called the final model, which is created for both samples for the most appropriate model.

According to fundamental equations of functions for the optimal model, the maximum number of basic functions determined based on the GCV value in the MARS analysis method was determined as 18. Then, 9 of the 18 basic functions were used to establish the most appropriate model for Türkiye sample. Conversely, the maximum number of basic functions of Iraq sample determined based on the GCV value in the MARS analysis method was determined as 32. Then, 20 of the 32 basic functions were used to establish the most appropriate model. The basic functions BF1, BF2 and BF3 from Türkiye sample, and the basic functions BF2, BF5 and BF18 from Iraq sample, were used in the creation of other basic functions and were not included in the final model.

Regression equation for optimal model, the F and p-values calculated as $F=18.29641$ and $F=17.98916$ for Türkiye and Iraq, the p-value was calculated for both of them as $p = 0.001$, this indicate that the models are statistically significant.

The data obtained with the internet addiction scale for Türkiye and Iraq samples were analyzed with the MARS analysis method and the most important predictors of internet addiction category were determined. The predictor variables included in the analysis and the significance levels of these variables on the variable predicted within themselves in the established model

The most important predictors obtained on Türkiye sample as a result of the analysis made with the MARS analysis method are given in Table 6.11. Accordingly, the most important predictors of internet addiction Scale for Türkiye sample are respectively; Daily Internet Usage, Mothers' Education, Number of Siblings, Smoking, Age, Daily Internet Usage During COVID-19 Quarantine, The Quality of Online education in the COVID-19 Pandemic, the device used, the variables associated with the dependent variable in the MARS analysis method were ranked according to their importance, starting from 100 points. The variable with the highest relationship with internet addiction scale was the self-efficacy perception variable, while the variable with the lowest relationship was the type of device to use the internet. Variables that have little or no relationship with internet addiction scale and are not included in the analysis are respectively; Daily use of the Internet before COVID-19 pandemic, Fatherhood, Mother's profession, Gender, Fathers' education, Family income, COVID-19 pandemic and internet use, have you effected by COVID-19, do you have internet at home, Internet use purpose, Mothers' Profession, and Stage of study were obtained as insignificant.

The most important predictors obtained on Iraq sample as a result of the analysis made with the MARS analysis method for Iraq sample are given in Table 6.19. Accordingly, the most important predictors of internet addiction Scale for Iraq sample are respectively; Daily Internet Usage, Internet Usage COVID-19 Pandemic, Quality Online Education COVID-19 Pandemic, COVID-19 Pre-Quarantine Daily Internet Usage, Mothers' Job, internet Usage Purpose, Family Income, Daily Internet Usage in COVID-19 quarantine, Age, Number of Siblings, Mothers' Education, Internet Home, Fathers' Job, Device Preference for Online in Quarantine. The variables associated with the dependent variable in the MARS analysis method were ranked according to their importance, starting from 100 points. The variable with the highest relationship with internet addiction scale was the self-efficacy perception variable, while the variable with the lowest relationship was Device Preference for Online in Quarantine. Variables that have little or no relationship with internet addiction scale and are not included in the analysis are respectively; Gender, Fathers' Education, Have You Had COVID-19, Do You Smoke and Stage of study were obtained as insignificant.

The study's objectives are to examine an existing scale for measuring internet addiction as well as researching into the variables that may influence a student's level of addiction of Turkish and Iraqi university students, to evaluate the relationships between participant characteristics, socioeconomic level and internet addiction. The findings are different to other previous study, regarding the socioeconomic level which as this study found that the socioeconomic level family such as family income does not have an effect to raise the internet addiction among adolescence, additionally, other studies have out in Türkiye demonstrated its significance (Gunuc and Kayri, 2016).

Regarding the internet addiction, 26.8% and 46.6% of students from Türkiye and Iraq respectively has internet addiction. These results are comparable to those for young adults for whom earlier research had been done (Sayed et al., 2022).

Concerning the gender, as have been shown in this study the gender has been considered as effectless factor on internet addiction in both countries, while some studies found similarity result and indicate that there is no difference between the genders in internet addiction incidence (Ni, 2009), others indicate that internet addiction incidence increased more in male. Given that students frequently spend a lot of time online, several studies have suggested that internet addiction may have an impact on how well they learn, (Macur et al., 2016). However, daily internet usage before and after COVID-19 has been shown to increase as a result of the pandemic, and in our findings, we found a positive correlation between the force to use more internet variable and internet addiction. This could be explained by the possibility that a significant portion of the students' reported internet use was spent on their studies before and after the pandemic, which would mean that they used the internet more frequently to complete their assignments.

One of the goals of the current study was to examine the effects of the sub-categories of deprivation, control difficulty, impairment in functioning, and social isolation developed by Gunuc and Kayri (2010), it has been established that these sub-categories have a significant relationship to internet addiction among university students in both countries.

In addition, regarding the key factor which has affect to internet addiction by conducting MARS data mining technique for both countries is Daily Internet Usage, however, the students from Iraq spend more time using Internet compare to students form Türkiye, however still this variable has crucial effect on internet addiction for both countries, These results concur with those that have already been published, a study by Kayri (2010) has shown that the average daily time spent online as well as the purpose for which users access the Internet should be taken into account as the key influences for internet addiction among Turkish secondary school students.

With the present model of MARS drawing on nineteen predicted variables forming the scale of Internet Dependency, as well as the dependent variable, one aim of this study was to demonstrate how MARS can be employed for such studies as well as epidemiology as a whole. The predicted variable of $p<0.5$ is a key takeaway and shows how it is affected by the various predictors such as Daily Internet Usage, Mother education, Number of Siblings, Age, Daily internet Usage During COVID-19. It is shown, too, that magnitude of each variable may impact the model, however is nevertheless a prime advantage of MARS as against other methods.

Recommendations arising from the present study are that MARS is a tool of great significance for work in the field of epidemiology, as has been shown by the positive outcomes in terms of statistical analysis. Since it is able to generate predictions relatively free of bias due to the lack of need for a continuous/discrete scale, one can say that MARS does that by drawing on and analyzing data at a much higher level of granularity than other statistical methods.

It is crucial to spot internet addiction in adolescents since it is frequently linked to mental health issues. Various therapy approaches might be used to treat IA as well as prevalent psychological conditions such Deprivation, Control Difficulty, Impairment in Functioning and Social Isolation. On the basis of that, it is suggested that educators raise students' understanding of the health hazards linked with using the internet that could have an impact on their mental health.

Findings from this study have important clinical implications in general and university students specially. Ministry of higher education, mental health authorities, should be aware of these high rates of internet addiction among the university students in both countries, and should create awareness among students regarding internet addiction and its potential harms; this could be included in foundation course of curriculum implementation support program for students. We Initiative should be taken to provide wide opportunities for students to involve in extracurricular activities and interact with friends. There should be provision of counsellor for emotional and mental support of students as they are overburden with studies and long posting schedules. Measures should be undertaken to prevent further increase in rates and manage the possible cases.

REFERENCES

Ahmed, Z. (2023). Internet Addiction: Symptoms, Causes and Effects. Access date: August 8, 2023. Access address: https://diamondrehabthailand.com/what-is-internet-addiction/#:~:text=The%20causes%20of%20internet%20addiction%20are%20genetics%2C%20structural%20brain%20changes,to%20addictive%20behaviors%20around%20computers.

AIM Conference Center (2008). Internet Addiction research in Asia. **Inernational Workshop and Conference, Digital Dependency**, Shanghai, China.

Al Mukhaini, A. M., Al Houqani, F. A., & Al Kindi, R. M. (2021). Internet addiction and depression among postgraduate residents: a cross-sectional survey. *Sultan Qaboos University Medical Journal*, 21(3), 408.

Ali, R., Mohammed, N., & Aly, H. (2017). Internet addiction among medical students of Sohag University, Egypt. *Journal of Egyptian Public Health Association*, 92(2), 86-95.

AlQahtani, A. A. A., Nahar, S., AlAhmari, S. M., & AlQahtani, K. S. A. (2015). Association between obesity and mental disorders among male students of King Khalid University, Abha, Saudi Arabia. *Saudi Journal of Obesity*, 3(2), 48.

American Psychiatric Association. (1994). ***Diagnostic and Statistical Manual of Mental Disorders (DSM)*** (4th ed.). American Psychiatric Association: Washington, DC.

American Psychiatric Association. (2000a). ***Diagnostic and Statistical Manual of Mental Disorders*** (DSM-IV-TR). American Psychiatric Association: Washington, DC.

American Psychiatric Association. (2000b). ***Diagnostic and Statistical Manual of Mental Disorders*** (DSM-IV-TR): American Psychiatric Association: Washington, DC.

Amr, M., Hady El Gilany, A., & El-Hawary, A. (2018). Does gender predict medical students' stress in mansoura, egypt?. *Medical education online*, 13, 12. doi.org/10.3885/meo.2018.Res00273

Anderson, K. J. (2001). Internet Use among College Students: An exploratory study. *Journal of American College Health*, 50(1), 21-26.

Anusha Prabhakaran, M.C., Rajvee Patel, V., Jaishree Ganjiwale, D. & Somashekhar Nimbalkar, M. (2016). Factors associated with internet addiction among school-going adolescents in Vadodara. *J Family Med Prim Care*, 5(4), 765–769.

Balhara, Y. P. S., Mahapatra, A., Sharma, P., & Bhargava, R. (2018). Problematic internet use among students in South-East Asia: Current state of evidence. *Indian journal of public health*, 62(3), 197-210.

Biolcati-Rinaldi, F., Molteni, F., Salini, S. (2018). ***Assessing the Reliability and Validity of Google Scholar Indicators. The Case of Social Sciences in Italy. In: Bonaccorsi, A. (eds) The Evaluation of Research in Social Sciences and Humanities.*** Springer, Cham.

Błachnio, A., Przepiorka, A., Benvenuti, M., Mazzoni, E., & Seidman, G. (2019). Relations between Facebook intrusion, Internet addiction, life satisfaction, and self-esteem: A study in Italy and the USA. *International Journal of Mental Health and Addiction*, 17(4), 793–805. https://doi.org/10.1007/s11469-018-0038-y

Bowditch, L., Chapman, J., & Naweed, A. (2018). Do coping strategies moderate the relationship between escapism and negative gaming outcomes in World of Warcraft (MMORPG) players?. *Computers in Human Behavior*, 86, 69-76.

Brailovskaia, J., Rohmann, E., Bierhoff, H. W., Margraf, J., & Köllner, V. (2019). Relationships between addictive Facebook use, depressiveness, insomnia, and positive mental health in an inpatient sample: A German longitudinal study. *Journal of Behavioral Addictions*, 8(4), 703-713.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (2014) ***Classification and Regression Trees***. Chapman and Hall: Wadsworth, New York.

Brenner, V. (1997). Psychology of computer use. XLVII. Parameters of Internet use, abuse and addiction: the first 90 days of the Internet Usage Survey. *Psychological Reports*, 80(3 Pt 1), 679-882.

Buchholz, L. (2009). Teen Internet addicts more likely to self harm. Access date: July 6, 2022. Access address: http://abcnews.go.com/US/wireStory?id=9245921

Buntins, M., Buntins, K., & Eggert, F. (2017). Clarifying the concept of validity: From measurement to everyday language. *Theory & Psychology*, 27(5), 703–710. https://doi.org/10.1177/0959354317702256

Cao, F. L., & Su, L. Y. (2006). Internet addiction among Chinese adolescents: Prevalence and psychological features. *Child: Care, Health and Development*, 33(3), 275-281.

Cao, F.L. , Su, L. Y., Liu, T. Q., & Gao, X. P. (2007). The relationship between impulsivity and internet addiction in a sample of Chinese adolescents. *Journal of European Psychiatry*, 22, 466 471.

Cerniglia, L., Zoratto, F., Cimino, S., Laviola, G., Ammaniti, M., & Adriani, W. (2017). Internet Addiction in adolescence: Neurobiological, psychosocial and clinical issues. *Neuroscience & Biobehavioral Reviews*, 76, 174–184. https://doi.org/10.1016/j.neubiorev.2016.12.024

Chebbi, P., Koong, K. S., & Liu, L. (2005). Some Observation on Internet Addiction Disorder Research. *Journal of Information System Education*, 11, 3-4.

Cheng, Y. S., Tseng, P. T., Lin, P. Y., Chen, T. Y., Stubbs, B., Carvalho, A. F., Wu, C. K., Chen, Y. W., & Wu, M. K. (2018). Internet addiction and its relationship with suicidal behaviors: A meta-analysis of multinational observational studies. *Journal of Clinical Psychiatry*, 79(4), 17r11761.

Chetwynd E. (2022), Critical Analysis of Reliability and Validity in Literature Reviews. *Journal of Human Lactation*. 2022;38(3):392-396. doi:10.1177/08903344221100201

Chou, C. (2001). Internet heavy use and addiction among Taiwanese college students: An online interview study. *Cyberpsychology and Behavior*, 4(5), 573-585.

Chou, C., & Hsiao, M. C. (2000). Internet addiction, usage, gratification, and pleasure experience: the Taiwan college students‟ case. *Computers and Education*, 35, 65-80.

Chou, C., Chou, J., & Tyan, N. N. (1999). An exploratory study of Internet addiction, usage and communication pleasure: The Taiwan‟s case. *Journal of Educational Telecommunications*, 5(1), 47-64.

Clement, J. (2020). *Number of Internet Users in Nigeria from 2018-2022*. Nigeria. https://www.statista.com/statistics/183849/internet-users-nigeria/

CNNIC (). Statistical Report on Internet Development in China. China Internet Network Information Centre: China. https://www.cnnic.com.cn/IDR/ReportDownloads/201706/P020170608523740585924.pdf.

Concha, A., Mills, D.S., Feugier, A., Zulch, H., Guest, C., Harris, R. and Pike, T.W., (2014). Using sniffing behavior to differentiate true negative from false negative responses in trained scent-detection dogs. *Chemical senses*, 39(9), pp.749-754.

Craven, P., Wahba, G. (1978) Smoothing noisy data with spline functions. *Numer. Math*. 31, 377–403. https://doi.org/10.1007/BF01404567.

Dalbudak, E., Evren, C., Aldemir, S., & Evren, B. (2014). The severity of Internet addiction risk and its relationship with the severity of borderline personality features, childhood traumas, dissociative experiences, depression and anxiety symptoms among Turkish university students. *Psychiatry research*, 219(3), 577-582.

Davis, R. A. (2001). A cognitive-behavioral model of pathological internet use. Journal of Computer in Human Behavior. *ScienceDirect*.17, 187-195.

Davis, R. A., Flett, G. L., & Besser, A. (2002). Validation of a new scale for measuring problematic internet use: implications for pre-employment screening. *Cyberpsychology & behavior*, 5(4), 331–345.

Derbyshire, K.L., Lust, K.A.. Schreiber, L., Odlaug, B.L., Christenson, G., Golden, D.J., Grant, J.E. (2013). Problematic Internet use and associated risks in a college sample. *Comprehensive Psychiatry*, 54(5), 415-422

Dong, H., Yang, F., Lu, X., & Hao, W. (2020). Internet Addiction and Related Psychological Factors Among Children and Adolescents in China During the Coronavirus Disease 2019 (COVID-19) Epidemic. *Frontiers in psychiatry*, 11, 00751. https://doi.org/10.3389/fpsyt.2020.00751

Douglas Curran-Everett (2017) Explorations in statistics: the assumption of normality, *Adv Physiol Educ*, 41: 449–453, doi:10.1152/advan.00064.2017.

Douglas, A. C., Mills, J. E., Niang, M., Stepchenkova, S., Byun, S., Ruffini, C., Lee, S. K. et al. (2008). Internet addiction: Meta-synthesis of qualitative research for the decade 1996 -2006. *Journal of Computer in Human Behavior*, 24, 3027-3044.

Duan, L., Shao, X., Wang, Y., Huang, Y., Miao, J., Yang, X., & Zhu, G. (2020). An investigation of mental health status of children and adolescents in china during the outbreak of COVID-19. *Journal of Affective Disorders*, 275, 112–118. https://doi.org/10.1016/j.jad.2020.06.029

El Asam, A., Samara, M., & Terry, P. (2019). Problematic internet use and mental health among British children and adolescents. *Addictive Behaviors*, 90, 428-436.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook"friends": Social capital and college students‟ use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), 1143–1168.

Friedman. (1991) Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19 (1) 1 - 67.

Garmah Mohamed &Rida Bernouss (2020). A cross-sectional study on Internet addiction among Moroccan high school students, its prevalence and association with poor scholastic performance. *International Journal of Adolescence and Youth*, 25(1), 479-490.

Geldhof, G.J., Preacher, K.J. and Zyphur, M.J., (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological methods*, 19(1), p.72.

Goldberg, I. (1996). Internet addiction support group: is there truth in jest?. Access date: October 13, 2022. Access address: http://users.rider.edu/~suler/psycyber/supportgp.html

Gomez-Galan, J., Martínez-Lopez, J. Á., Lazaro-Perez, C., & Sarasola Sanchez-Serrano, J. L. (2020). Social networks consumption and addiction in college students during the COVID-19 pandemic: educational approach to responsible use. *Sustainability*, 12(18), 7737. https://doi.org/10.3390/su12187737

Greenfield, D. N. (1999). Psychological characteristics of compulsive internet use: A preliminary analysis. *Cyberpsychology and Behavior*, 8(2), 403-412.

Griffiths, M. (1995). Technological addictions. ***Clinical Psychology Forum***, 76, 14-19.

Griffiths, M. (1998). Internet addiction: Does it really exist? In J. Gackenbach (Ed.), ***Psychology and the Internet: Intrapersonal, interpersonal and transpersonal applications*** (pp. 61-75). Academic Press: New York.

Griffiths, M. (2000***)***. Does Internet and computer" addiction" exist? Some case study evidence. *CyberPsychology and Behavior*,3(2), 211–218. https://doi.org/10.1089/109493100316067

Grohol, J. M. (2005). Internet addiction guide. Hawker, S., Payne, S., Kerr, C., Hardey, M., & Powell, J. (2002). Appraising the evidence: Reviewing disparate data systematically. *Qualitative Health Research*, 12(9), 1284–1299.

Gunuc, S. & Kayri, M. (2010). Internet addiction profile and development of Internet Addiction Scale in Türkiye: Validity and reliability study. *Hacettepe University Journal of Education Faculty*, 39, 220–232.

Gunuc, S. & Kayri, M. (2016), Comparing Internet Addiction in Students with High and Low Socioeconomic Status Levels, Addicta: *The Turkish Journal on Addictions*, 3(2), 177-183.

Harrigan, K.R. (1985), An application of clustering for strategic group analysis. *Strat. Mgmt. J*., 6: 55-73. https://doi.org/10.1002/smj.4250060105

Hastie, T., Tibshirani, R., Friedman, J., (2001). ***The Elements of Statistical learning: Data Mining, Inference, and Prediction***. Springer, New York.

Hastie and Tibshirani (1990) Generalized Additive Models. Statistical Science, 1(3) 297-310. https://doi.org/10.1214/ss/1177013604.

Hazra A. (2017). Using the confidence interval confidently. *J Thorac Dis*. 2017 Oct;9(10):4125-4130. doi: 10.21037/jtd.2017.09.14.

Horvath, C. W. (2004). Measuring Television Addictions. *Journal of Broadcasting & Electronic Media*, 48(3), 378-398.

Hull, M., & Proulx, D. A. (2022). Internet addiction facts and statistics. The recovery village. Access date: June 22, 2022. Access address: https://www.therecoveryvillage.com/process-addiction/internet-addiction/internet-addiction-statistics/.

Hur, M. (2006). Demographic, habitual, and socioeconomic determinants of internet addiction disorder: An empirical study of Korean teenagers. *Cyberpsychology and Behavior*, 9(5), 514-525.

Ibrahim, A. K., Kelly, S. J., & Glazebrook, C. (2012). Analysis of an Egyptian study on the socioeconomic distribution of depressive symptoms among undergraduates. *Social psychiatry and psychiatric epidemiology*, 47, 927-937.

Internet World Stats News (2011). 2010 Internet Statistics Update. Access date: 15 December 2022. Access address: https://www.internetworldstats.com/pr/edi061.htm.

Islam, M. I., Biswas, R. K., & Khanam, R. (2020). Effect of internet use and electronic game-play on academic performance of Australian children. *Scientific Reports*, 10(1), https://doi.org/10.1038/s41598-020-78916-9

James and Lee (2011) Using Likert-Type Scales in the Social Sciences, *Journal of Adult Education*, V. 40, No. 1.
Ju, Y. Y., Chih, H. K., Cheng, F. Y., Sue, H. C., Wei, L. C., & Cheng, C. C. (2008). Psychiatric symptoms in adolescents with internet addiction: Comparison with substance abuse. *Psychiatry and clinical neurosciences*, 62, 9-16.
Kaltiala-Heino, R., Lintonen, T., & Rimpela, A. (2004). Internet addiction? Potentially problematic use of the Internet in a population of 12-18 year-old adolescents. *Addiction Research and Theory*, 12(1), 89-96.
Karma Tenzin, ThinleyDorji, Mongal S. Gurung, PelzomDorji, Sandip Tamang, Umesh Pradhan, GampoDorji (2018). Prevalence of Internet Addiction and Associated Psychological Co-morbidities among College Students in Bhutan. *Journal of Nepal Medical Association*, 56, 210.
Kavitha. (2021), Way to Develop the Faculty Competence. *Asian Journal of Nursing Education and Research* 1(2): April-June 2021; Page 45–47.
Kayri M. (2010). The analysis of internet addiction scale using multivariate adaptive regression splines. *Iranian journal of public health*, 39(4), 51–63.
Kayri M (2007). Two-step clustering analysis in researches: A case study. *Eurasian Journal of Educational Research*, 7(28): 89- 99.
Kibet E (2012). *A Multıvarıate Adaptıve Regressıon Splınes Approach to Predıct the Treatment Outcomes of Tuberculosıs Patıents in Kenya*, Master thesis. University of Nairobi, Kenya.
Kim, J. and Hwang, I.C., (2020). Drawing guidelines for receiver operating characteristic curve in preparation of manuscripts. *Journal of Korean medical science*, 35(24).
Kim, J. H. (2008). The effect of a R/T group counseling program on the internet addiction level and self-esteem of internet addiction university students. *International Journal of Reality Therapy*, 27(2), 4-12.
Kim, J. H., Hui, H. L. C., Lau, C. H., Kan, P., Cheuk, K. K., & Griffiths, M. (2010). Brief report: Predictors of heavy Internet use and associations with health promoting and health risk behaviors among Hong Kong university students. *Journal of Adolescence*, 33(1), 215-220.
Kima, K., Ryub, E., Chonb, M., Yeunb, E., Choic, S., Seod, J., & Namd, B. (2006). Internet addiction in Korean adolescents and its relation to depression and suicidal ideation: A questionnaire survey. *International Journal of Nursing Studies*, 43(2), 185-192.
Ko, C. H., Yen, J. Y., Liu, S. C., Huang, C. F., & Yen, C. F. (2009). The associations between aggressive behavior and internet addiction and online activities in adolescents. *Journal of Adolescents Health*, 44, 598-605.
Ko, C. H., Yen, J. Y., Yen, C. F., Lin, H. C., & Yang, M. J. (2007). Factors predictive for incidence and remission of Internet addiction in young adolescents: a prospective study. *Cyberpsychology and Behavior*, 10(4), 545-551.
Kraut, R., Patterson, M., Landmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological wellbeing? *American Psychologist*, 53(9), 1017-1031.
Lam, L. T., Peng, Z. W., Mai, J. C., & Ing , J. (2009). Factors Associated with Internet Addiction among Adolescents. *Cyberpsychology and Behavior*, 12(5), 551-555.
Lee T, Chiu C, Chou Y, Lu C (2006) Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis* 50, 1113–1130.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, 28(1), 129–137. https://doi.org/10.1037/h0035519.

Li, M., Chung, T. (2006) Internet Function and Internet Addictive Behavior. *Computers in Human Behavior*, 22(6),1067-1071.

Lin, C. H.,Yu, S. F. (2008). Adolescent internet usage in Taiwan: Exploring gender differences. *Journal of Adolescence,* 43(170), 317-331.

Lin, C. H., Lin, S. L. & Wu, C.P. (2009). The effects of parental monitoring and leisure boredom on adolescents' internet addiction. *Journal of Adolescence*, 44 (176), 993-1004.

Lin, S. J., & Tsai, C. C. (1999). Internet Addiction among High Schoolers in Taiwan. **107th American Psychology Association Annual Convention**, Boston, USA.

Lucas Gren. (2018). Standards of validity and the validity of standards in behavioral software engineering research: the perspective of psychological test theory. **In Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '18).** Association for Computing Machinery, New York, USA.

Macur, M., Király, O., Maraz, A., Nagygyörgy, K., & Demetrovics, Z. (2016). Prevalence of problematic internet use in Slovenia. *Slovenian Journal of Public Health*, 55(3), 202-211.

Meerkerk, G. J., Van Den Eijnden, R., Vermulst, A. A., & Garretsen, H. F. L. (2009). The Compulsive Internet Use Scale (CIUS): Some Psychometric Properties. *Cyberpsychology and Behavior,* 12(1), 1-6.

Metzger, M. J. (2007). Communication privacy management in electronic commerce. *Journal of Computer-Mediated Communication*, 12(2), 335-361.

Milani, L., Di Blasio, P., & Osualdella, D. (2009). Quality of interpersonal relationships and problematic Internet use in adolescence. *Cyberpsychology and Behavior*, 12(6), 681-684.

Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. (2019) Descriptive statistics and normality tests for statistical data. *Ann Card Anaesth*, 22(1):67-72. doi: 10.4103/aca.ACA_157_18.

Morahan-Martin, J., & Schumacher, P. (2000). Incidence and correlates of pathological Internet use among college students. Computers in Human Behavior, *Computers in Human Behavior*,16(1), 13

Moreno, M. A., Jelenchick, L. A., & Christakis, D. A. (2013). Problematic internet use among older adolescents: A conceptual framework. *Computers in Human Behavior*, 29(4), 1879–1887. https://doi.org/10.1016/j.chb.2013.01.053

Nagaur, A. (2020). Internet Addiction and Mental Health among University students during COVID-19 lockdown. *MuktShabd Journal*, 2347-3150.

Nalwa, K., & Anand, A. P. (2003). Internet Addiction in Students: A Cause of Concern. *Cyberpsychology and Behavior*, 6(6), 653-656.

Ni, X., Yan, H., Chen, S., & Liu, Z. (2009). Factors influencing internet addiction in a sample of freshmen university students in China. *Cyberpsychology & behavior*, 12(3), 327-330.

Nielsen Global Media. (2020). The impact of COVID-19 on media consumption across North Asia. Access date: December 25, 2022. Access address:

https://www.nielsen.com/wp-content/uploads/sites/3/2020/03/The-Impact-of-COVID-19-on-Media-Consumption-Across-North-Asia.pdf

Niemz, K., Griffiths, M., & Banyard, P. (2005). Prevalence of pathological internet use among university students and correlations with self-esteem, the General Health Questionnaire (GHQ), and disinhibition. *Cyberpsychology and Behavior*, 8(6), 562-570.

Noorbala, A. A., Yazdi, S. B., Yasamy, M. T., & Mohammad, K. (2017). Mental health survey of the adult population in Iran. *The British Journal of Psychiatry*, 184(1), 70-73.

Pallanti, S., Bernardi, S. & Quercioli, L. (2006). The shorter PROMIS questionnaire and the internet addiction scale in the assessment of multiple addictions in a high-school population: Prevalence and related disability. *Journal of MBL Communication Inc*.

Park, S. K., Kim, J. Y. & Cho, C. B. (2008). Prevalence of internet addiction and correlations with family factors among South Korean adolescents. *Journal of Adolescents*, 43 (172), 895-909.

Preece, J. (2000). Online communities: Designing usability, supporting sociability. Chichester: Wiley. Quayle, E., & Taylor, M. (2003). Model of problematic Internet use in people with a sexual interest in children. *CyberPsychology and Behavior*, 6(1), 93–106.

Pujazon-Zazik, M., & Park, M. J. (2010). To tweet, or not to tweet: gender differences and potential positive and negative health outcomes of adolescents' social internet use. *American journal of men's health*, 4(1), 77-85.

Ramón-Arbués, E., Granada-López, J. M., Martínez-Abadía, B., Echániz-Serrano, E., Antón-Solanas, I., & Nash, M. (2021). Prevalence and factors associated with problematic internet use in a population of Spanish university students. *International Journal of Environmental Research and Public Health*, 18(14), 7620.

Rankin, J. A., Paisley, C. A., Mulla, M. M., and Tomeny, T. S. (2018). Unmet social support needs among college students: relations between social support discrepancy and depressive and anxiety symptoms. *J. Couns. Psychol*. 65, 474–489. doi: 10.1037/cou0000269.

Rice, R. E. (2006). Influences, usage, and outcomes of Internet health information searching: Multivariate results from the Pew surveys. *International Journal of Medical Informatics*, 75(1), 8-28.

Rosliza, A.M., Ragubathi, M.N., Mohamad Yusoff, M.K.A, Shaharuddin, M.S. (2018). Internet Addiction among Undergraduate Students: Evidence from a Malaysian Public University. *International Medical Journal Malaysia*,17(2).

Salford Systeam, (2018). Salford predictive modeller: introducing MARS, USA. Access date: April 02, 2023. Access address: http://media.salford-systems.com/pdf/spm8/IntroMARS_v_8_2.pdf

Sayed, M., Naiim, C. M., Aboelsaad, M., & Ibrahim, M. K. (2022). Internet addiction and relationships with depression, anxiety, stress and academic performance among Egypt pharmacy students: a cross-sectional designed study. *BMC public health*, 22(1), 1826. https://doi.org/10.1186/s12889-022-14140-6

Schneider, G. P., Evans, J., & Pinard, K. T. (2006). ***The Internet Fourth Edition-Illustrated Introductory***. Thomson Course Technology: United States of America.

Schumacker, R.E. and Lomax, R.G. (2004) ***A beginner's guide to structural equation modeling*** (2nd Edition). Lawrence Erlbaum Associates: Mahwah.

Sharma, A., & Sharma, R. (2018). Internet addiction and psychological well-being among college students: A cross-sectional study from Central India. *Journal of Family Medicine and Primary Care*,7(1), 147–151. https://doi.org/10.4103/2Fjfmpc.jfmpc_189_17
Sheskin DJ (2011) ***Handbook of parametric and nonparametric statistical procedure***, 5th edn. CRC Press: the USA.
Simon K. (2021a). A Decade in Digital. Acess date: 7 August 2023. Access address: https://datareportal.com/reports/a-decade-in-digital.
Simon K. (2021b). Digital 2021 April Global Statshot Report. Acess date: 7 August 2023. Access address: https://datareportal.com/reports/digital-2021-april-global-statshot.
Sowndarya, T. A. &Pattar, M. (2018). Pattern of internet addiction among urban and rural school students, Mangaluru, India: a comparative cross-sectional study. *IJCP*, 5(5). DOI: http://dx.doi.org/10.18203/2349-3291 .
Suler, J. R. (2000). Psychotherapy in Cyberspace: A 5-dimensional model of online and computer-mediated psychotherapy. *CyberPsychology & Behavior*, 3(2), 151–159. https://doi.org/10.1089/109493100315996.
Şevgin, H., & Önen, E. (2022). Comparison of Classification Performances of MARS and BRT Data Mining Methods: ABİDE- 2016 Case. *Egitim ve Bilim*, 47(211), 195–222. https://doi.org/10.15390/EB.2022.10575.
TNIC (2011). *A Survey on Broadband Usage in Taiwan*. Tiwan Network Information Center: Tiwan. https://www.twnic.tw/download/200307/1101f.pdf
Taylan, P., Weber, G. W., & Yerlikaya, F. (2008). Continuous optimization applied in MARS for modern applications in finance, science and technology. **20th International Conference/Euro Mini Conference on Continuous Optimization and Knowledge-Based Technologies (EurOPT 2008)**, Neringa, LITHUANIA. https://hdl.handle.net/11511/55725.
Tayyar Şaşmaz, Seva Öner, A. Öner Kurt, Gülçin Yapıcı, Aylin Ertekin Yazıcı, Resul Buğdaycı& Mustafa Şiş. (2014). Prevalence and risk factors of Internet addiction in high school students. *European Journal of Public Health*, 24(1), 15–20.
Tepecik Böyükbaş, İ., Çitak Kurt, A. N., Tural Hesapçioğlu, S., & Uğurlu, M. (2019). Relationship between headache and internet addiction in children. *Turkish Journal of Medical Sciences*, 49(5), 1292-1297.
Thomas, N.J. and Martin, F.H. (2010), Video-arcade game, computer game and Internet activities of Australian students: Participation habits and prevalence of addiction. *Australian Journal of Psychology*, 62: 59-66. https://doi.org/10.1080/00049530902748283.
Tibshirani, R. Friedman, J., Hastie, T. (2001) ***An Introduction to Statistical Learning***. Springer Text in Statistics, Springer.
Tsitsika, A., Critselis, E., Kormas, G., Filippopoulou, A., Tounissidou, D., Freskou, A., Kafetzis, D. (2009). Internet use and misuse: a multivariate regression analysis of the predictive factors of internet use among Greek adolescents. *European Journal of Pediatrics*, 168(6), 655-665.
Van Rooij, Spijkerman, R., Van den Eijinden, R. J. J. M., T. J. , Vermulst, A. A. & Engels, R.C.M. E. (2010). Compulsive internet use among adolescents: Bidirectional parents child relationships. *Journals of Abnormal ChildPsychology*, 38, 77-89.
Wan, C. S. & Chiou, W. B. (2007). The motivations of adolescents who are addicted to online games: A cognitive perspective. *Journal of Adolescence*, 42 (165), 179-197.

Wellman, B., & Gulia, M. (1999). ***Net surfers don't ride alone: Virtual communities as communities. In P. Kollock & M. Smith*** (Eds.), Communities and Cyberspace. Routledge: New York.

Wood, S.N. (2006) ***Generalized Additive Models, An Introduction with R***, Chapman and Hall: New-York.

Kim, Y. and Xie, C.-S. (2022). Post-traumatic growth during COVID-19: the role of perceived social support, personality, and coping strategies. *Health* 10:224. doi: 10.3390/healthcare10020224.

Yalçın, İ. (2011). Social support and optimism as predictors of life satisfaction of college students. *Int. J. Adv. Couns*. 33, 79–87. doi: 10.1007/s10447-011-9113-9.

Yen, C. F., Ko, C. H., Chang, Y. P., Cheng, C. P., & Yen, J. Y. (2009). Multidimensional discriminative factors for Internet addiction among adolescents regarding gender and age. *Psychiatry and Clinical Neurosciences* 63(3), 357- 364.

Yen, C. F., Yen, J., Ko, C., Wu, H., & Yang, M. (2007). The comorbid psychiatric symptoms of Internet addiction: Attention Deficit and Hyperactivity Disorder (ADHD), depression, social phobia, and hostility. *Journal of Adolescent Health*, 41(1), 93-98.

Yerlikaya, F. (2008) *A New Contribution to Nonlinear Robust Regression and Classification with MARS and Its Applications to Data Mining for Quality Control in Manufacturing*, Master Thesis, METU, Ankara.

Young, K. (1996). Internet addiction: The emergence of a new clinical disorder. *Cyberpsychology and Behavior*, 1(3), 237-244.

Young, K. (1997). What makes the Internet addictive: Potential explanations for pathological Internet use. **the 105th annual conference of the American Psychological Association**, Chicago, IL, the USA.

Young, K. (1998). ***Caught in the net: How to recognize the signs of internet addiction and a winning strategy for recovery***. John Wiley & Sons: New York, the USA.

Young, K. (1999). Internet addiction: Symptoms, evaluation, and treatment. In L. VandeCreek & T. Jackson (Eds.), ***Innovations in Clinical Practice: A Source Book*** (Vol. 17, pp. 19-31). Professional Resource Press: Sarasota, FL.

Young, K. (2004). Internet Addiction: A new clinical phenomenon and its consequences. *American Behavioral Scientist*, 48(4), 402-415.

Young, K. (2006). Internet Addiction Test. Access date: March 03, 2022. Access address:http://www.netaddiction.com/index.php?option=com_bfquiz&view=onepage&catid=46&Itemid=106

Young, K., & Rogers, R. (1998). The relationship between depression and Internet addiction. *Cyberpsychology and Behavior*, 1(1), 25-28.

Yurdugül, H. (2006). The comparison of reliability coefficients in parallel, tauequivalent, and congeneric measurements. Ankara University, *Journal of Faculty of Educational Sciences*, 39(1), 15-37.

Zboralski, K., Orzechowska, A., Talarowska, M., Darmosz, A., Janiak, A., Janiak, M., Gałecki, P. (2009). The prevalence of computer and Internet addiction among pupils. *Postepy Hig Med Dosw*, 2(63), 8-12.

Zhang and Goh, (2016) Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci. Front.*, 7, pp. 45-52.

APPENDIX

A 1. Questionnaire Form

Dear Students

The main aim of this survey is to examine what is the Scale of Internet Addiction among students, and how much of an overall issue Internet addiction is among these students in different stages of study. This questionnaire is developed by researchers from Van Yüzüncü Yıl University. Completing the Questionnaire takes about ten minutes.

There are no right or wrong answers, this form consists of two parts, the first section consists of some socio-demographic and COVID-19 related questions. The second section consists of Internet Addiction Scale (35 item).

Please answer the questions sincerely, and don't leave the questions blank.

Data will be treated confidentially.

Hewa Ghafor Hassan   Prof. Dr. Murat Kayri

Socio-demographic Questions

- **What stage are you in?**
- **Your Gender:** O Male O Female
- **Your Age:**
- **Your Father's Education:** O Literate, O Not Literate, O Primary School, O Middle School, O High School, O University (Diploma/ Bachelor), O University (Master), O University (PhD)
- **Your Mother's Education:** O Literate, O Not Literate, O Primary School, O Middle School, O High School, O University (Diploma/ Bachelor), O University (Master), O University (PhD)
- **Your Father's Profession (Job):**
- **Your Mother's Profession (Job):**
- **Number of Siblings:**
- **Do you smoke:** O Yes, O No, O Sometimes
- **Family Income:** O less than 200$, O 200$-400$, O 400$-600$, O 600$-1000$, O 1000$-1500$, O More than 1500$
- **Do you have internet in your home?** O Yes, O No
- **For what purpose do you use the Internet the most?** (Select a single option)
    - Research (course, info)
    - On-line Education

- Chat
- News
- Music-Movie
- Game
- Pornography
- Shopping
- Gambling

- **How many hours do you use the Internet on average in one day?**
  - 1
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7
  - 8
  - 9
  - 10 and above

COVID-19 Questions

- **Are pandemic COVID-19 forced you to use more time on internet?** ☐ Yes ☐ No
- **Have you been infected with COVID-19?** ☐ Yes ☐ No
- **Which electronic device do you use to study online during lockdown?**
  - ☐ Laptop
  - ☐ PC
  - ☐ Smart phone
  - ☐ Tablet
- **How many hours do you spend on the Internet during COVID-19 lockdown?**
  - ☐ Less than 1 hour in a day
  - ☐ 1-2hours/ day
  - ☐ 2-4 hours/ day
  - ☐ 4-6 hours/ day
  - ☐ 6-8 hours/ day
  - ☐ More than 8 hours in a day
- **How many hours do you spend on the Internet before COVID-19 lockdown?**
  - ☐ Less than 1 hour in a day
  - ☐ 1-2hours/ day
  - ☐ 2-4 hours/ day
  - ☐ 4-6 hours/ day
  - ☐ 6-8 hours/ day
  - ☐ More than 8 hours in a day

- **How do you rate the quality/benefit of on-line education during COVID-19 pandemic? (1 is the lowest evaluation)**

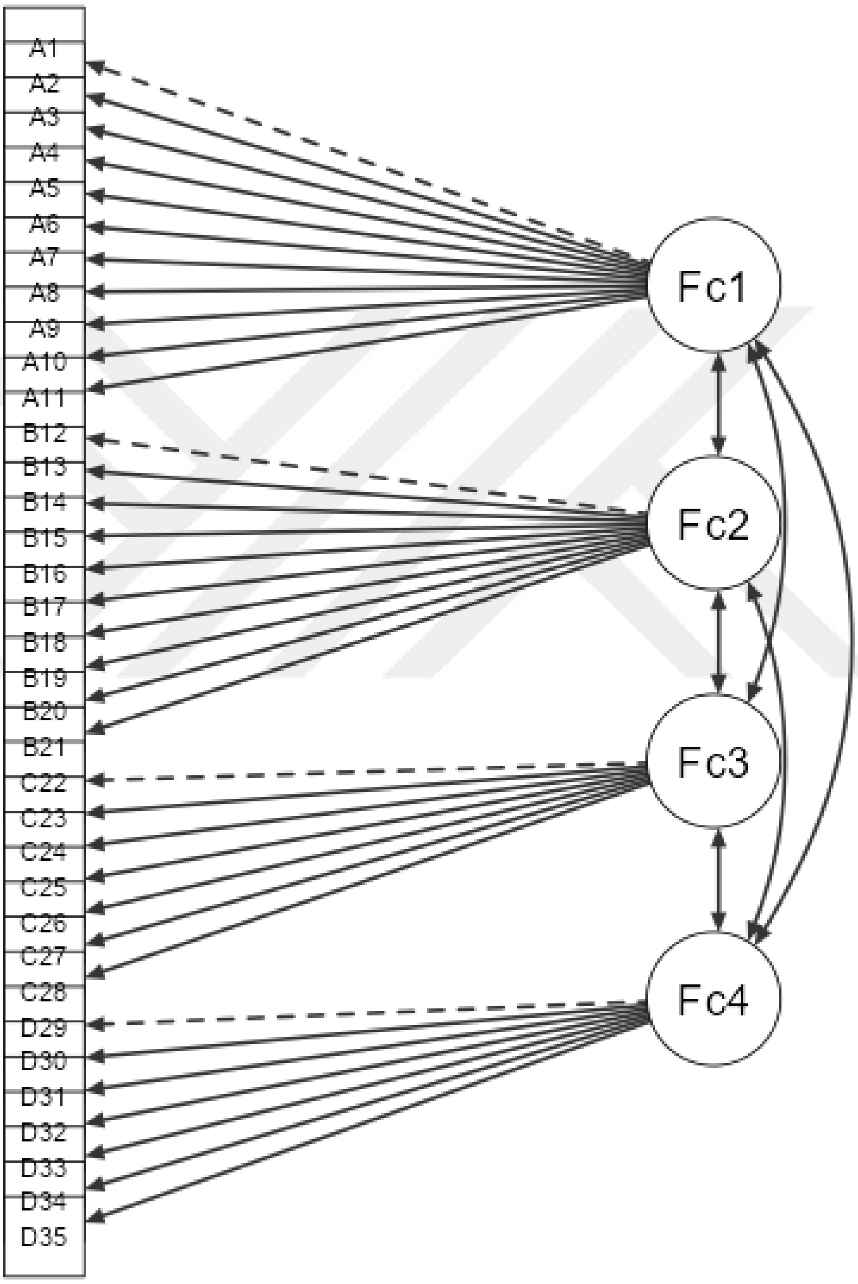
  ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

INTERNET ADDICTION SCALE By reading each of the following statements, you are kindly requested to mark "**I totally agree**" if this article is always correct for you, "**I agree**" if it is usually true, "**I am undecided**" if you are not sure, "**I do not agree**" if it is not usually true, and "**I strongly disagree**" if it is never correct	**I totally agree**	**I agree**	**I'm undecided**	**I don't agree**	**I strongly disagree**
Deprivation					
1. I feel tense/uneasy when I cannot use the Internet.					
2. I get nervous/outraged when I want to use the Internet but can't.					
3. I get nervous/outraged if the internet connection drops or slows down.					
4. When I use the Internet, I am happier/calmer than ever before.					
5. When I feel anxious or discussed, using the internet relaxes me.					
6. I get angry if someone removes me from the Internet.					
7. I tend to use the Internet to avoid my problems.					
8. I get angry if I can't access the Internet at the time I planned.					
9. When I have someone around me, I'd like to be alone and access the Internet.					
10. When I am not using the Internet, I look forward to accessing the Internet.					
11. I look for an Internet connection wherever I go.					
Difficulty in Control					
12. I have difficulty limiting or controlling my Internet use.					
13. When I wake up in the morning, my first thought is to access the Internet.					
14. Each time I would like to stay on the Internet longer than the last.					
15. I stay on the Internet longer than I plan.					
16. I think about the Internet even when I don't use the Internet.					
17. I do not feel or realize that I am hungry or thirsty while on the Internet.					
18. I cancel other plans to spend more time on the Internet.					
19. I can't get away from the Internet whenever I want.					
20. I can't get away from the Internet even if my parents call me.					
21. I sacrifice my sleep to use the Internet.					
Distortion in Functionality					
22. I have problems with my family due to my Internet use.					
23. I can't get away from the Internet even if my friends call me.					

24. I am less Interested in other activities (sports, cinema, reading books, etc.) due to my use of the Internet.					
25. I cannot fulfill or neglect my home/work/school responsibilities due to my Internet use.					
26. People around me complain about the time I spend on the Internet.					
27. I spend less time with my family because of my Internet use.					
28. I spend less time with my friends due to my Internet use.					
Social Isolation					
29. I have problems with my friends due to my Internet use.					
30. I prefer the friends I have made on the Internet to my friends in real life.					
31. I prefer to meet my friends in real life on the Internet rather than outside.					
32. I make my friends online.					
33. The Internet is my best friend.					
34. A life without the Internet seems meaningless/in vain to me.					
35. I have difficulty communicating face-to-face due to my Internet use.					

A 2. Path Diagram and It’s Value

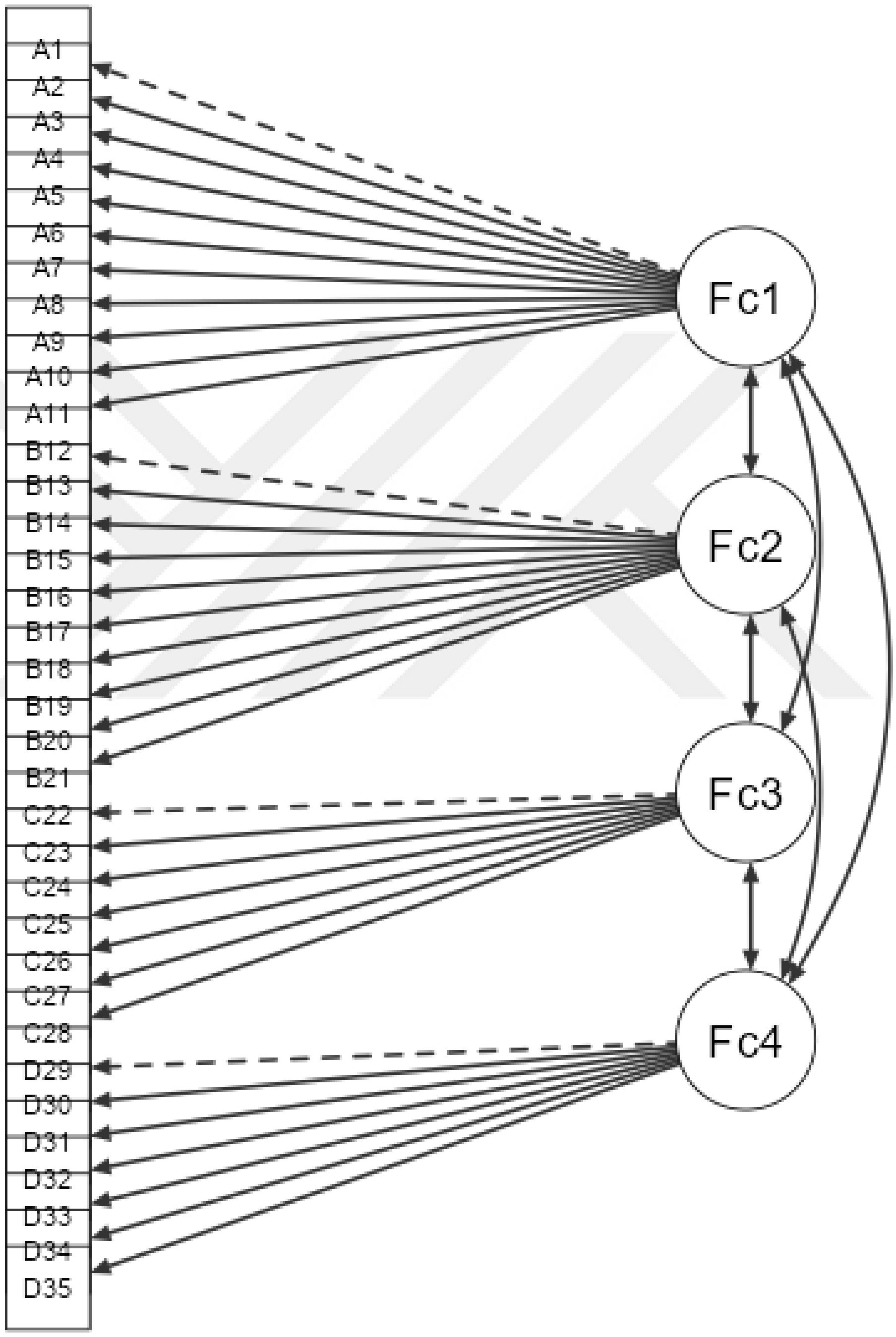
Path Diagram of the Turkish Sample



Values of Turkish Path Diagram

Substances	Estimate	Standard Estimate	p
A1	0.817	0.641	< .001
A2	0.837	0.56	< .001
A3	0.941	0.681	< .001
A4	0.674	0.65	< .001
A5	0.796	0.61	< .001
A6	0.497	0.419	< .001
A7	0.738	0.508	< .001
A8	0.782	0.571	< .001
A9	0.815	0.626	< .001
A10	0.563	0.434	< .001
A11	0.985	0.68	< .001
B12	0.735	0.562	< .001
B13	1.121	0.734	< .001
B14	0.404	0.436	< .001
B15	0.933	0.708	< .001
B16	0.441	0.434	< .001
B17	0.546	0.527	< .001
B18	0.346	0.396	< .001
B19	0.433	0.392	< .001
B20	0.316	0.383	< .001
B21	0.759	0.551	< .001
C22	0.5	0.416	< .001
C23	0.304	0.396	< .001
C24	0.691	0.504	< .001
C25	0.546	0.449	< .001
C26	0.472	0.382	< .001
C27	0.458	0.359	< .001
C28	0.37	0.359	< .001
D29	0.335	0.442	< .001
D30	0.483	0.473	< .001
D31	0.4	0.403	< .001
D32	0.479	0.528	< .001
D33	0.336	0.361	< .001
D34	0.876	0.655	< .001
D35	0.503	0.479	< .001

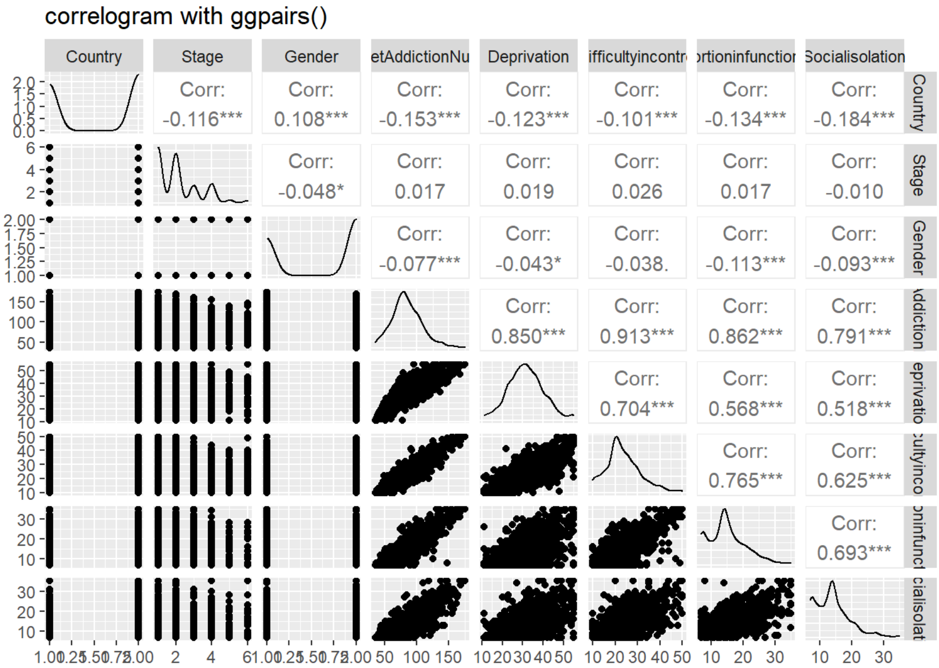
Path Diagram of the Iraqi Sample

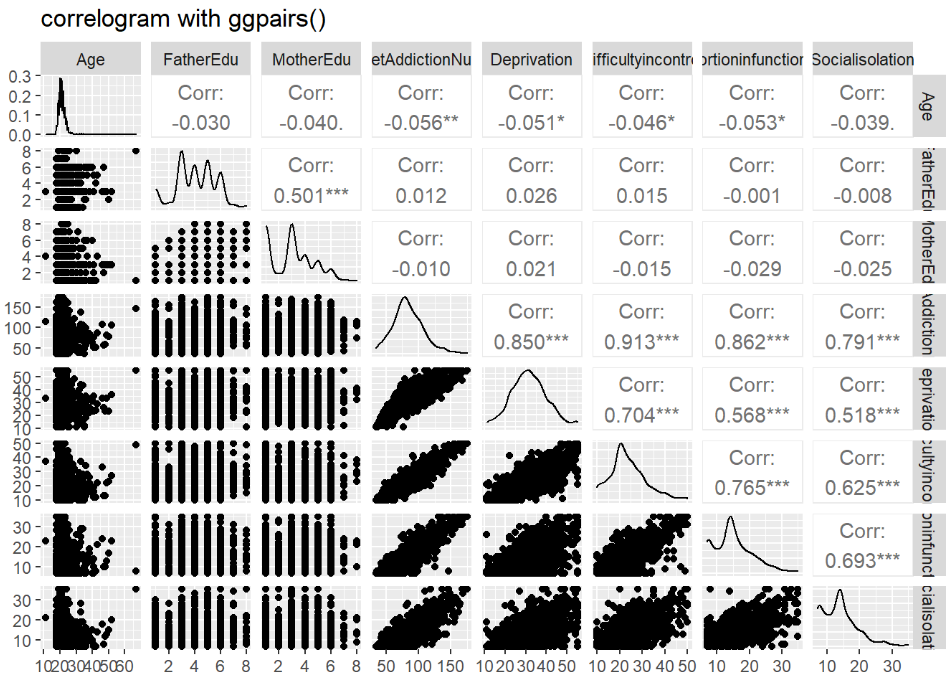


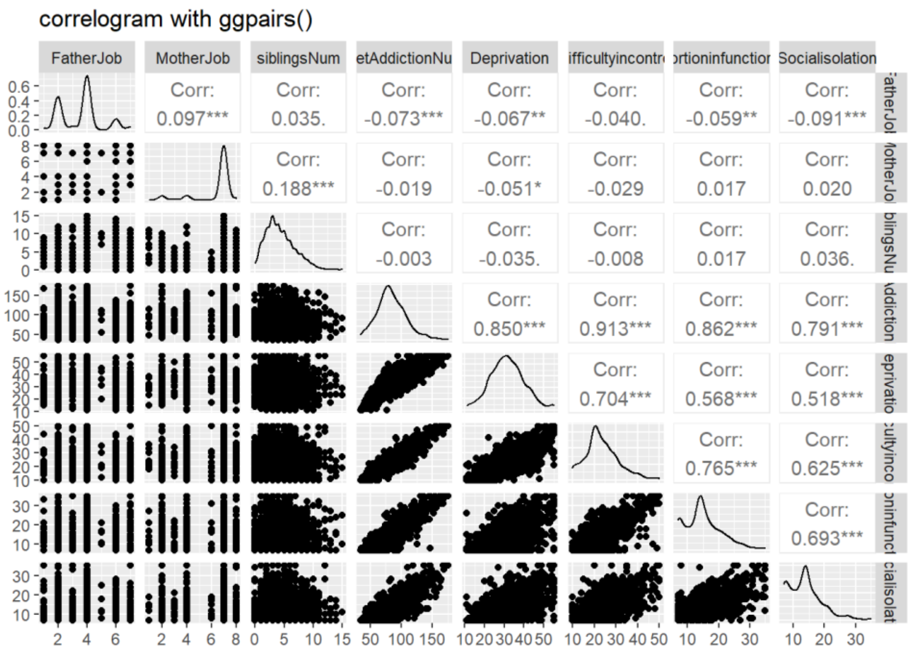
Values of Iraqi Path Diagram

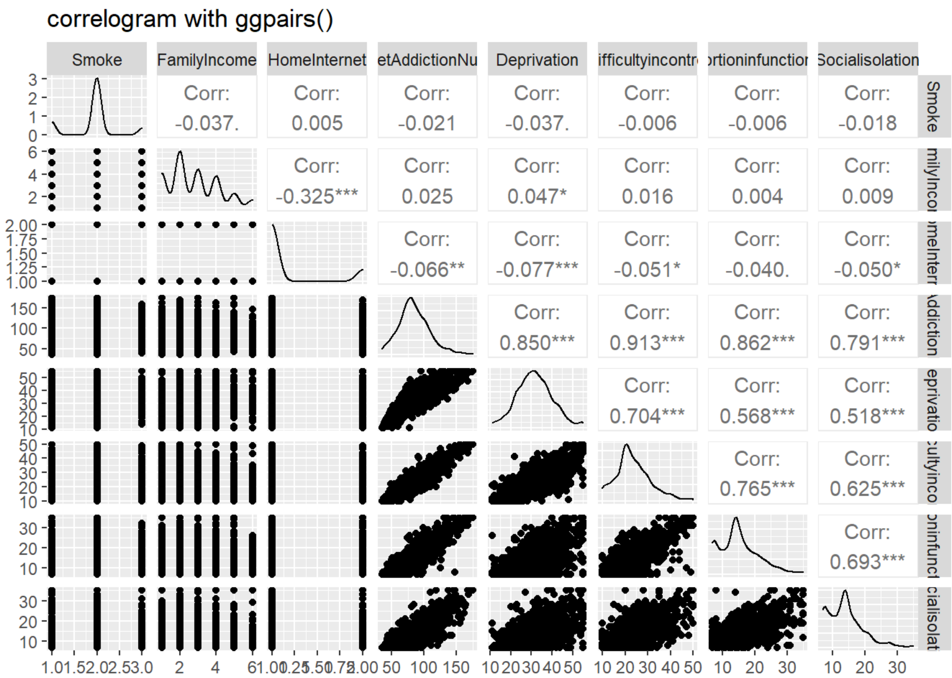
Substances	Estimate	Standard Estimate	p
A1	1.11	0.731	< .001
A2	0.895	0.577	< .001
A3	0.977	0.614	< .001
A4	0.927	0.683	< .001
A5	0.992	0.688	< .001
A6	0.966	0.6	< .001
A7	1,094	0.721	< .001
A8	1,070	0.702	< .001
A9	1.167	0.711	< .001
A10	0.872	0.616	< .001
A11	1.005	0.637	< .001
B12	1.032	0.664	< .001
B13	1,188	0.656	< .001
B14	0.675	0.514	< .001
B15	0.956	0.616	< .001
B16	0.71	0.528	< .001
B17	1.067	0.737	< .001
B18	0.656	0.573	< .001
B19	0.649	0.45	< .001
B20	0.643	0.632	< .001
B21	0.855	0.732	< .001
C22	0.738	0.631	< .001
C23	0.55	0.492	< .001
C24	1.108	0.658	< .001
C25	0.62	0.496	< .001
C26	0.7	0.506	< .001
C27	0.709	0.471	< .001
C28	0.698	0.516	< .001
D29	0.596	0.579	< .001
D30	0.709	0.555	< .001
D31	0.769	0.569	< .001
D32	0.752	0.657	< .001
D33	0.797	0.627	< .001
D34	0.929	0.65	< .001
D35	0.856	0.66	< .001

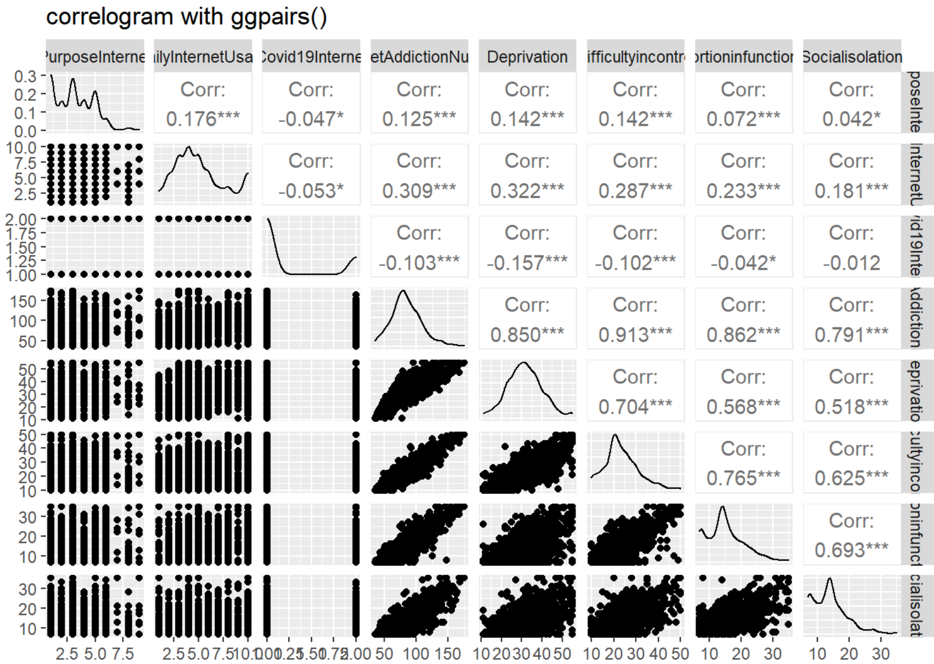
A 3. Pair Plots

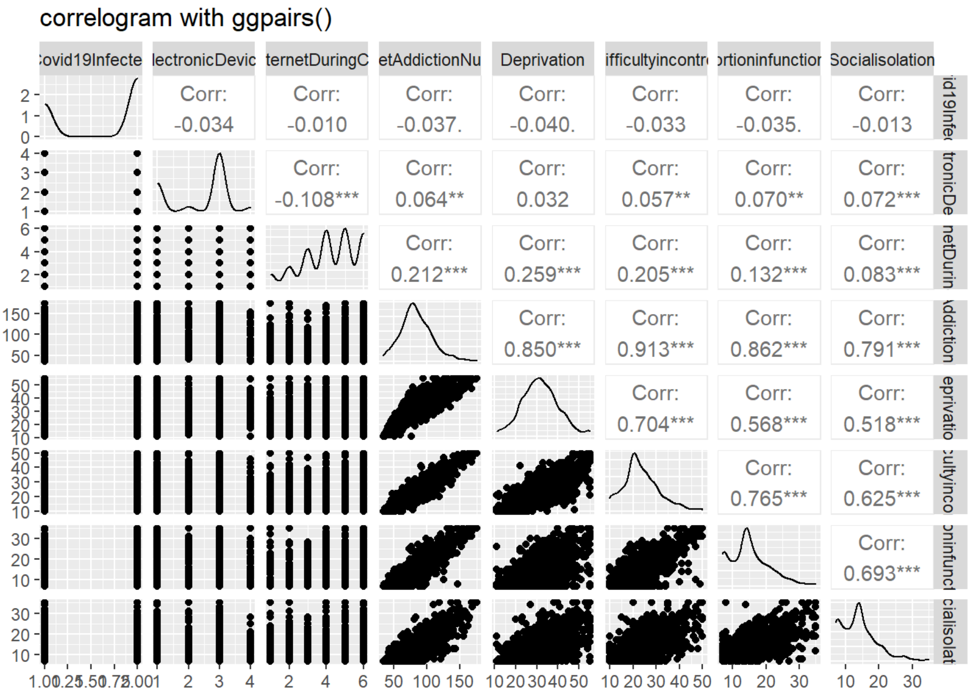


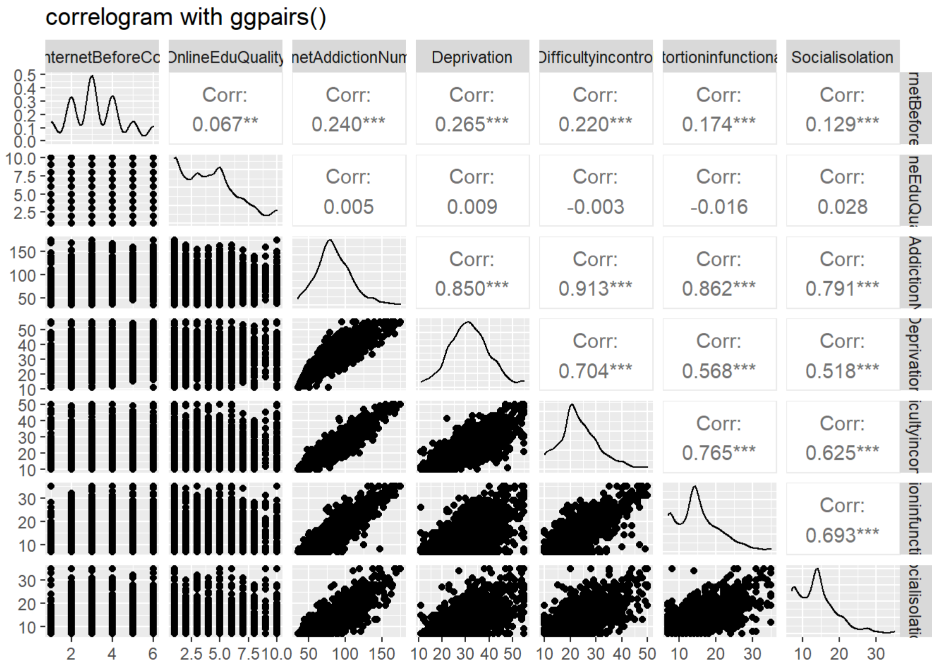



correlogram with ggpairs()
FatherJob
MotherJob
siblingsNum
Deprivation
Socialisolation


correlogram with ggpairs()
Smoke
FamilyIncome
HomeInternet
Deprivation
Socialisolation


correlogram with ggpairs()
PurposeInterne
ilyInternetUsa
Covid19Interne
etAddictionNu
Deprivation
ifficultyincontr
rtioninfunctior
Socialisolation
Corr: 0.176***
Corr: -0.047*
Corr: 0.125***
Corr: 0.142***
Corr: 0.142***
Corr: 0.072***
Corr: 0.042*
Corr: -0.053*
Corr: 0.309***
Corr: 0.322***
Corr: 0.287***
Corr: 0.233***
Corr: 0.181***
Corr: -0.103***
Corr: -0.157***
Corr: -0.102***
Corr: -0.042*
Corr: -0.012
Corr: 0.850***
Corr: 0.913***
Corr: 0.862***
Corr: 0.791***
Corr: 0.704***
Corr: 0.568***
Corr: 0.518***
Corr: 0.765***
Corr: 0.625***
Corr: 0.693***


correlogram with ggpairs()
Covid19Infecte
lectronicDevic
ternetDuringC
etAddictionNu
Deprivation
ifficultyincontr
rtioninfunctior
Socialisolation
Corr: -0.034
Corr: -0.010
Corr: -0.037.
Corr: -0.040.
Corr: -0.033
Corr: -0.035.
Corr: -0.013
Corr: -0.108***
Corr: 0.064**
Corr: 0.032
Corr: 0.057**
Corr: 0.070**
Corr: 0.072***
Corr: 0.212***
Corr: 0.259***
Corr: 0.205***
Corr: 0.132***
Corr: 0.083***
Corr: 0.850***
Corr: 0.913***
Corr: 0.862***
Corr: 0.791***
Corr: 0.704***
Corr: 0.568***
Corr: 0.518***
Corr: 0.765***
Corr: 0.625***
Corr: 0.693***


correlogram with ggpairs()
nternetBeforeCc
OnlineEduQuality
netAddictionNum
Deprivation
Difficultyincontro
tortioninfunctiona
Socialisolation
Corr: 0.067**
Corr: 0.240***
Corr: 0.265***
Corr: 0.220***
Corr: 0.174***
Corr: 0.129***
Corr: 0.005
Corr: 0.009
Corr: -0.003
Corr: -0.016
Corr: 0.028
Corr: 0.850***
Corr: 0.913***
Corr: 0.862***
Corr: 0.791***
Corr: 0.704***
Corr: 0.568***
Corr: 0.518***
Corr: 0.765***
Corr: 0.625***
Corr: 0.693***

A 4. Correlation between Independent Variables

```r
library(rstatix)
```

```
## Warning: package 'rstatix' was built under R version 4.1.3
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.1
```

```r
X = VNumData[,c(2:20)]
names(X)
```

```r
X = VNumData[,c(2:20)]
names(X)
```

```
##  [1] "Stage"                    "Gender"
##  [3] "Age"                      "FatherEdu"
##  [5] "MotherEdu"                "FatherJob"
##  [7] "MotherJob"                "siblingsNum"
##  [9] "Smoke"                    "FamilyIncome"
## [11] "HomeInternet"             "PurposeInternet"
## [13] "DailyInternetUsage"       "Covid19Internet"
## [15] "Covid19Infected"          "ElectronicDevice"
## [17] "HourInternetDuringCovid19" "HourInternetBeforeCovid19"
## [19] "OnlineEduQuality"
```

```r
cor.mat <- cor_mat(X, method="spearman")
print(cor.mat)
```

```
## # A tibble: 19 x 20
##    rowname          Stage  Gender     Age Fathe~1 Mothe~2 Fathe~3 Mothe~4 sibli~5
##  * <chr>            <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 Stage           1      -4.1e-2  0.57     0.046  8.5e-2  0.0044 -0.06   -0.063
##  2 Gender         -0.041   1   e+0 -0.12   -0.066 -4.5e-4  0.064   0.041   0.068
##  3 Age             0.57   -1.2e-1  1       -0.054 -2.6e-2  0.084   0.0056  0.071
##  4 FatherEdu       0.046  -6.6e-2 -0.054    1      4.9e-1 -0.24   -0.2    -0.33
##  5 MotherEdu       0.085  -4.5e-4 -0.026    0.49   1   e+0 -0.04   -0.36   -0.47
##  6 FatherJob       0.0044  6.4e-2  0.084   -0.24  -4   e-2  1       0.083  -0.017
##  7 MotherJob      -0.06    4.1e-2  0.0056  -0.2   -3.6e-1  0.083   1       0.23
##  8 siblingsNum    -0.063   6.8e-2  0.071   -0.33  -4.7e-1 -0.017   0.23    1
##  9 Smoke          -0.014   1.1e-1 -0.068   -0.037 -3.3e-2 -0.024   0.062   0.05
## 10 FamilyIncome    0.11   -1.7e-1 -0.032    0.31   2.1e-1 -0.36   -0.25   -0.11
## 11 HomeInternet   -0.034   5.6e-2  0.084   -0.22  -2.3e-1  0.16    0.12    0.24
## 12 PurposeInter~  0.0056 -4   e-2 -0.041    0.16   1.3e-1 -0.028  -0.069  -0.18
## 13 DailyInterne~  0.017  -3.6e-2 -0.082    0.13   1.5e-1 -0.044  -0.12   -0.19
## 14 Covid19Inter~ -0.078   1.2e-2 -0.023   -0.019 -2.7e-2 -0.13   -0.0046  0.1
## 15 Covid19Infec~ -0.073  -2.5e-2 -0.061   -0.11  -4   e-2  0.034   0.036  -0.0028
## 16 ElectronicDe~ -0.049   4.3e-2 -0.023   -0.14  -2.2e-1 -0.026   0.11    0.28
## 17 HourInternet~  0.089  -7.4e-2 -0.027    0.14   1.4e-1  0.0041 -0.13   -0.21
## 18 HourInternet~  0.06   -1   e-1 -0.028    0.15   1.6e-1 -0.027  -0.11   -0.21
## 19 OnlineEduQua~  0.12   -6.2e-2  0.059    0.11   1.4e-1 -0.071  -0.048  -0.14
## # ... with 11 more variables: Smoke <dbl>, FamilyIncome <dbl>,
## #   HomeInternet <dbl>, PurposeInternet <dbl>, DailyInternetUsage <dbl>,
## #   Covid19Internet <dbl>, Covid19Infected <dbl>, ElectronicDevice <dbl>,
## #   HourInternetDuringCovid19 <dbl>, HourInternetBeforeCovid19 <dbl>,
## #   OnlineEduQuality <dbl>, and abbreviated variable names 1: FatherEdu,
## #   2: MotherEdu, 3: FatherJob, 4: MotherJob, 5: siblingsNum
```

```r
cor.mat %>% cor_get_pval()
```

```
## # A tibble: 19 x 20
##    rowname         Stage   Gender       Age FatherEdu MotherEdu Father~1 Mother~2
##    <chr>           <dbl>    <dbl>     <dbl>     <dbl>     <dbl>    <dbl>    <dbl>
##  1 Stage      0          5.36e- 2 2.31e-192 2.9 e-  2 5.49e-  5 8.35e- 1 4.3 e- 3
##  2 Gender     5.36e-  2 0         2.02e-  8 1.78e-  3 9.83e-  1 2.66e- 3 5.03e- 2
##  3 Age        2.31e-192 2.02e- 8 0          1.15e-  2 2.21e-  1 7.64e- 5 7.9 e- 1
##  4 FatherEdu  2.9 e-  2 1.78e- 3 1.15e-  2 0         2.42e-135 5.28e-31 1.81e-21
##  5 MotherEdu  5.49e-  5 9.83e- 1 2.21e-  1 2.42e-135 0         5.58e- 2 6.23e-70
##  6 FatherJob  8.35e-  1 2.66e- 3 7.64e-  5 5.28e- 31 5.58e-  2 0        8.14e- 5
##  7 MotherJob  4.3 e-  3 5.03e- 2 7.9 e-  1 1.81e- 21 6.23e- 70 8.14e- 5 0
##  8 siblingsN~ 2.7 e-  3 1.33e- 3 8.01e-  4 1.46e- 59 1.33e-124 4.2 e- 1 1.27e-28
##  9 Smoke      5.16e-  1 2.66e- 7 1.29e-  3 8.12e-  2 1.17e-  1 2.66e- 1 3.16e- 3
## 10 FamilyInc~ 3.72e-  7 1.11e-16 1.35e-  1 7.36e- 52 4.09e- 23 7.94e-70 4.56e-32
## 11 HomeInter~ 1.11e-  1 7.77e- 3 6.69e-  5 1.32e- 26 3   e- 28 9.11e-15 9.83e- 9
## 12 PurposeIn~ 7.93e-  1 5.84e- 2 5.53e-  2 1.24e- 13 1.32e-  9 1.87e- 1 1.18e- 3
## 13 DailyInte~ 4.29e-  1 9.32e- 2 1.15e-  4 4.03e- 10 2.87e- 12 3.79e- 2 9.24e- 9
## 14 Covid19In~ 2.09e-  4 5.72e- 1 2.73e-  1 3.69e-  1 2.04e-  1 5.69e-10 8.27e- 1
## 15 Covid19In~ 5.39e-  4 2.34e- 1 3.92e-  3 5.87e-  7 6.19e-  2 1.06e- 1 8.94e- 2
## 16 Electroni~ 1.95e-  2 4.13e- 2 2.8 e-  1 1.33e- 11 3.59e- 25 2.27e- 1 1.79e- 7
## 17 HourInter~ 2.34e-  5 4.57e- 4 1.98e-  1 3.96e- 11 1.83e- 11 8.46e- 1 4.62e-10
## 18 HourInter~ 4.6 e-  3 1.49e- 6 1.87e-  1 3.84e- 13 6.2 e- 15 2.09e- 1 5.98e- 7
## 19 OnlineEdu~ 5.13e-  9 3.32e- 3 5.45e-  3 1.65e-  7 1.27e- 10 7.81e- 4 2.47e- 2
## # ... with 12 more variables: siblingsNum <dbl>, Smoke <dbl>,
## #   FamilyIncome <dbl>, HomeInternet <dbl>, PurposeInternet <dbl>,
## #   DailyInternetUsage <dbl>, Covid19Internet <dbl>, Covid19Infected <dbl>,
## #   ElectronicDevice <dbl>, HourInternetDuringCovid19 <dbl>,
## #   HourInternetBeforeCovid19 <dbl>, OnlineEduQuality <dbl>, and abbreviated
## #   variable names 1: FatherJob, 2: MotherJob
```

A 5. F-Statistics and p values of Basic Functions for Turkish Sample

MARS Regression: Training Data	
W: 1220.00	R-SQUARED: 0.1979
HAVE MEAN DEP: 0.62303	ADJ R-SQUARED: 0.11324
UNCENTERED R-SQUARED = R-0 SQUARED: 0.35571	

```
   parameter                    Estimate S.E. _ T - Ratio P-Value
------------------------------------------------------------------------------
 Constant | 0. 11026 0 .02203 5.00544 0.00000
 Basis Function 4 | -0. 03502 0 .01271 -2.75532 0.00595
 Basis Function 5 | -0. 00621 0 .00113 -5.49314 0.00000
 Basis Function 6 | 0.01160 0.00262 4.43474 0.00001
 Basis Function 9 | 0.00422 0 .00101 4.17059 0.00003
 Basis Function 11 | 0.07669 0.0 .01730 4.43253 0.00001
 Basis Function 12 | 0.00848 0.00207 4.09992 0.00004
 Basis Function 13 | 0.00619 0 .00264 2.34661 0.01911
 Basis Function 15 | -0. 04533 0 .01816 -2.49637 0.01268
 Basis Function 18 | 0.05646 0.0 .01561 3.61718 0.00031
------------------------------------------------------------------------------
F-STATISTIC = 18.29641 SE OF REGRESSION = 0.41727
P-VALUE = 0.00000 RESIDUAL SUM OF SQUARES = 210.68180
[ MDF,NDF ] = [ 9, 1210 ] REGRESSION SUM OF SQUARES = 28.67148
------------------------------------------------------------------------------
```

A 6. F-Statistics and p values of Basic Functions for Iraqi Sample

MARS Regression: Training Data	
W: 1015.00	R-SQUARED: 0.26576
HAVE MEAN DEP: 0.62303	ADJ R-SQUARED: 0.25099
UNCENTERED R-SQUARED = R-0 SQUARED: 0.60792	

```
   parameter                    Estimate S.E. _ T - Ratio P-Value
------------------------------------------------------------------------------
 Constant | 0.28297 0 .02635 10.73986 0.00000

 Basis Function 1 | 0.03654 0 .01207 3.02803 0.00253

 Basis Function 4 | 0. 27047 0 .06608 4.09280 0.00005

 Basis Function 7 | -0. 01665 0 0.00493 -3.38066 0.00075

 Basis Function 8 | 0.01760 0 .00607 2.90071 0.00381

 Basis Function 9 | 0.03588 0.00766 4.68581 0.00000

 Basis Function 10 | -0. 05725 0 .01163 -4.92279 0.00000

 Basis Function 11 | -0. 31366 0 .06494 -4.82980 0.00000

 Basis Function 12 | 0.01573 0.00425 3.69922 0.00023

 Basis Function 14 | 0.88768 0 .31140 2.85065 0.00445

 Basis Function 15 | 0.03617 0 .01210 2.99064 0.00285

 Basis Function 16 | -0. 04509 0 .01770 -2.54791 0.01099

 Basis Function 19 | 0.00975 0.00263 3.70192 0.00023

 Basis Function 20 | 0.06891 0.0 .01549 4.44743 0.00001

 Basis Function 23 | 0.05213 0 .02052 2.54064 0.01122

 Basis Function 25 | 0.23951 0 .07919 3.02466 0.00255

 Basis Function 27 | -0. 03598 0 .01318 -2.72957 0.00645

 Basis Function 28 | -0. 06691 0 .02250 -2.97384 0.00301

 Basis Function 29 | 0.05754 0 .01278 4.50186 0.00001

 Basis Function 30 | 0. 26359 0 .07708 3.41983 0.00065

 Basis Function 32 | -0. 25173 0 .06777 -3.71443 0.00021
------------------------------------------------------------------------------
F-STATISTIC = 17.98916 SE OF REGRESSION = 0.43194
P-VALUE = 0.00000 RESIDUAL SUM OF SQUARES = 185.45206

[ MDF,NDF ] = [ 20, 994 ] REGRESSION SUM OF SQUARES = 67.12528

-----------------------------------------------
```

EXTENDED TURKISH SUMMARY

(GENİŞLETİLMİŞ TÜRKÇE ÖZET)

TÜRKİYE VE IRAK'TAKİ ÖĞRENCİLERİN İNTERNET BAĞIMLILIK DÜZEYLERİNİN ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON UZANIMLARI (MARS) YÖNTEMİ İLE KARŞILAŞTIRMALI OLARAK İNCELENMESİ

HASAN, Hewa Ghafor
Doktora Tezi, İstatistik Anabilim Dalı
Danışman: Prof. Dr. Murat KAYRİ
İkinci Danışman: Doktor Öğretim Üyesi Hikmet ŞEVGİN
Ağustos 2023, 132 sayfa

Son yıllarda internet, modern hayatı birçok açıdan değiştirmiştir. Internet teknolojileri; eğitim, sağlık, savunma sanayi, endüstri ve sağlık alanının tümünü yeni bir formata dönüştürmüştür. Bununla birlikte internet, sosyal hayatı ve insan-insan, insan-makine etkileşimini değiştirmiş ve bu etkileşimler çeşitli düzeylerde bağımlılıklar oluşturmuştur. İnternet bağımlılığı, ruh sağlığını etkileyebilecek güncel ve ciddi bir durum olarak kabul edilmektedir.

Ramón-Arbués vd. (2021), dünya çapında 4 milyardan fazla aktif internet kullanıcısının olduğunu belirtmişlerdir. Bununla birlikte, Nisan 2022 tarihi itibariyle dünya genelinde 4,65 milyar kişi (dünya nüfusunun %58,7'si) sosyal medyayı, 5,32 milyarı (%67) mobil cihazları ve beş milyar kişi de genel amaçlı olarak interneti kullandıkları belirtilmiştir (Ramón-Arbués vd., 2021). Dolayısıyla internet kullanım oranı dramatik bir şekilde artmaktadır. İnternet kullanımının neden arttığına ilişkin bazı faktörler, daha ucuz ve daha kullanıcı dostu teknolojiye erişim kolaylığı olsa da Covid-19 salgını ile vatandaşların genellikle evlerde kalması ve yüz yüze hayata katılamaması, bu hızlı artışın önemli bir nedeni olabilir.

2022 yılı itibariyle, Doğu Asya dünyadaki internet kullanıcılarının yaklaşık 1,2 milyarını oluştururken, Güney Asya bir milyarın biraz üzerinde internet kullanıcısına sahiptir. Şekil 2.1'de gösterildiği gibi istatistiklere göre Nisan 2022 itibariyle beş milyar insan çevrimiçi olmuştur. Simon (2021a) tarafından yapılan bir araştırmaya göre (bkz şekil 2.2), internet kullanıcılarının sayısı son on yılda iki kattan fazla artarak 2012 başında 2,18 milyardan 2022 başında 4,95 milyara yükselmiştir. İnternet teknolojilerinin birçok yönden insan hayatına faydaları ve katkıları bulunmaktadır. Bu hususlar, iş bulmak, eğlence ve iletişim ile bilgilerin toplanması ve yayılması olarak sıralanabilir.

İnternet teknolojileri her ne kadar bilgi ve birikimi artırma yönünde yararlı görülse de kişilerin bağımlılık gibi riskleri de beraberinde getirdikleri alanyazında belirtilmiştir (Chou ve Hsiao, 2000). Alanyazında, İnternet bağımlılığının gerçekteki madde bağımlılığına eş bir durum olup olmadığı araştırılmış ve internet bağımlılığı; patolojik bağımlılık (uyuşturucu, yemek ve benzeri) ile aynı kabul edilmiştir (Griffifths, 1995; Greenfield, 1999; Anderson, 2001). Başka bir deyişle, internet bağımlılığı maddeyi kötüye kullanma ile benzer değerlendirilmiştir. İnternet Bağımlılığı (IB) kavramının var olduğunu ancak bunun kullanımının farklılaştığını görebiliriz. Bu konuda, IB'nin uyuşturucu gibi diğer bağımlılık türlerine kıyasla gerçekçi bir bağımlılık türü olup olmadığı; ilgili tanı kriterleri ve IB'nin kendi başına tam olarak nasıl tanımlandığı konusunda anlaşmazlıklar gibi çeşitli durumlar bulunmaktadır. Başlangıçta, IB, Greenfield (1999) tarafından patolojik kumar ya da TV bağımlılığı gibi teknoloji ile, Griffiths (1995) tarafından madde kullanımı ve bağımlılığı, Anderson (2001) tarafındansa yerleşik ilkelere ve çerçevelere dayandırılmaya çalışılmıştır. DSM-IV (Amerikan Psikiyatri Birliği, 1994) olarak adlandırılan Ruhsal Bozuklukların Tanısal ve İstatistiksel El Kitabı (4. Baskı), bazı çalışmalarda IB'nin sınıflandırıldığı temel ölçüm ve referans noktası olarak kullanılmıştır. Lam vd. (2009) İnternet Bağımlılığından mustarip olma olasılığı en yüksek olan kişilerin öğrenciler olduğunu bulmuşlardır, ancak tedavi veya herhangi bir çözüm sağlama ile ilgili çok az çalışma ortaya konmuştur.

Davis (2001) ve Young ve Rogers (1998), IA'yı interneti bireye sorun yaratmadan kullanamama olarak kabul eder, ancak daha önce bahsedildiği gibi, IA teriminin kendisi literatürde tutarsız bir şekilde uygulanmaktadır. Bir yandan Goldberg (1996) ve Hur (2006) konuyu 'İnternet Bağımlılığı Bozukluğu' olarak adlandırmaktadır. Öte yandan (Lin ve Tsai, 1999; Chou, 2001; Nalwa ve Anand, 2003; Cao ve Su, 2006; Kima vd., 2006; Young, 2006; Ko vd., 2007; Lam vd., 2009; Yen vd., 2009; Thomas ve Martin, 2010) "İnternet Bağımlılığı" terimini kullanır. İnternet bağımlılığının hoşgörü, kontrol bozukluğu, saplantı ve aşırı çevrimiçi zaman gibi temel semptomlarını değerlendirmek için Yong vd. (2014), internet bağımlılığı için çeşitli tanı kriterleri oluşturmuştur. İnternet bağımlılarında kişilik bozukluklarının önemli bir yaygınlığına sahip olduğu bulunmuştur (Dalbudak vd., 2014). Ahmed'e (2023) göre internet bağımlılığının gelişimi, genetik, yapısal beyin değişiklikleri, çevresel etkiler ve mevcut zihinsel sağlık bozuklukları gibi bir dizi faktörden etkilenir. Artan sayıda araştırma, bilgisayar bağımlılığının genetik ve biyolojik olarak yatkın olduğunu göstermektedir.

Ergen yaş gruplarında IB ile ilgili çeşitli çalışmalar yapılmıştır. Pallanti vd. (2006), IB söz konusu olduğunda yaşın hiçbir önemi olmadığını ve sosyal sınıf için de geçerli olmadığını vurgulamışlardır. Sayed vd. (2022), Mısırlı üniversite öğrencilerinin %38,5'inde İnternet Bağımlılığı belirtileri olduğunu tespit etti. Üniversite öğrencilerinin ergenlikten yetişkinliğe geçiş yapabilmeleri için bir geçiş aşamasından geçmeleri gerektiğini vurgulamıştır. Çin İnternet Ağı Bilgi Merkezi (2006), çevrimiçi olarak 123 milyon kişiyi sunan bir araştırma yaptı ve bu sayının %14,9'u 18 yaşın altındaki kişilerdi. Bundan, IB'nin daha ciddi hale geldiği sonucuna vardılar. Park vd. (2008) tarafından Güney Kore'de yapılan bir araştırma, internet kullanımının diğer yaş gruplarına kıyasla en çok ergenler arasında olduğunu göstermiştir. 2005'te o ülkedeki 6 ila 19 yaş arasındakilerin %97,3'ünün çevrimiçi olduğunu buldular.

Dünya Sağlık Örgütü (WHO), psikiyatrik hastalıkların dünya çapında engelliliğin önde gelen nedenlerinden biri olduğunu iddia etmektedir (Noorbala vd., 2017). Stres, kaygı ve üzüntü, zihinsel sağlık sorunlarının örnekleridir. Her biri, özellikle genç nüfus için halk sağlığı için bir risk oluşturuyor olarak görülüyor. Gomez-Galan vd. (2020), Duan vd. (2020) ve Dong vd. (2020), Covid-19'un neden olduğu kapanmaların ve sosyal sınırlamaların önemli bir etkisi olarak dijital eğlence kullanımında, İnternet kullanımında büyük artış olduğunu belirtmişlerdir. Bu, Nielsen Global Media (2020) tarafından toplanan bilgilerle doğrulanmıştır. Aşırı internet kullanımı bağımlılığa yol açmaktadır ve Griffiths (2000) bunu, insanların, örneğin ders çalışma vb. gibi durumlarda, kullanımlarını kabul edilebilir normların üzerine çıkardığında olumsuz sonuçları akıllarına getirememeleri olarak tanımlamıştır.

Çalışmanın amacı, internet bağımlılığına yönelik bir ölçek geliştirmek ve bireylerin bağımlılık durumunu etkileyebilecek faktörleri, ayrıca internet bağımlılığının öğrencileri her iki örneklemde (Türkiye ve Irak) farklı şekillerde nasıl etkilediğini ve son olarak ne kadar etkilediğini incelemektir. Genel bir sorunun internet eklenmesi genel olarak bu öğrenciler arasındadır.

Yukarıda belirtilen bu hedeflere ek olarak, çalışmanın çeşitli amaçları vardır. Karmaşık bir konu, bundan sonra “IB” kısaltması ile ifade edilecek olan internet bağımlılığını neyin oluşturduğuna dair genel kabul görmüş bir tanımın olmamasıdır. Çeşitli bağımlılık oranlarını ölçmek için evrensel olarak kabul edilen ana kriterlerin neler olduğu açık değildir. Elbette, internet kullanımının yaygın olduğu ve modern yaşamın yadsınamaz bir gerekliliği olduğu ve hem Türkiye'de hem de Irak'ta var olan hiyerarşik ve geleneksel toplum yapısı göz önüne alındığında, özellikle gençlerin ve öğrencilerin çok daha yoğun olduğu toplum tarafından evrensel olarak bilinmektedir. Bununla birlikte, çalışmanın temel amaçlarından birinin buradan geldiği göz önüne alındığında, internet kullanımının öğrencileri olumlu mu yoksa olumsuz mu etkilediğine veya gerçekten IB'nin gerçek oranlarına ilişkin gerçek ampirik verilerin bulunmamasıdır. Bu tür veriler, IB'nin neden olduğu psikolojik, sosyal veya başka türlü olası etkileri hafifleten olumlu müdahalelere kanıta dayalı bir yaklaşımda yardımcı olabilir.

İnternet Bağımlılığı kavramının tanımın oluşturulması, Van Yüzüncü Yıl Üniversitesi ve Soran Üniversitesi'ndeki iki öğrenci grubu üzerinde yapılacak anketin temelini oluşturmaktadır. Son olarak, “Çok Değişkenli Uyarlanabilir Regresyon Eğrileri” (MARS) adı verilen bir analiz türü, elde edilen verilerin analizi için ana matematiksel modelleme aracı olacaktır.

Bu çalışmada istatistiksel veri analizinin iki ana türünden biri kullanılmıştır. Bunun nedeni, her zaman verilere göre geçerli olan bir sonuca varmamız gerekirken, bazı varsayımların da doğru olması gerekir. Bununla birlikte ister tahmine dayalı ister çıkarımsal olsun, veri analizinde aynı temel ilkeler geçerlidir. Çıkış değişkenlerini sağlamak için bir şekilde değiştirilen bir dizi girdi değişkeni vardır ve bunu yapmanın yolu, Şekil 3.1'deki diyagramda gösterildiği gibi Veri Oluşturma Mekanizması aracılığıyladır.

Bu çalışma, özyinelemeli bölümlemeyi kullanan bir tür regresyon analizi olan MARS ile yürütülmüştür. İlk olarak özyinelemeli bölümlemenin ne olduğuna bakalım. Fikir, verileri bir dizi alt bölgeye bölmek ve sırayla her bir alt bölgeyi analiz etmektir. Bunun gerçekleşme şekli, bölümlerin her bir alt bölge arasında ayarlanması ve yinelemeli olarak yapılmasıdır. Yani, bir bütün olarak alandan başlayarak, her biri arasında bir tür bölümleme ile ikiye bölünmüştür. Bu ikisi daha sonra tekrar bölünür ve süreç, uygun sayıda alt bölüm elde edilene kadar yinelemeli olarak devam eder.

Bunu takiben, Breiman vd. (2014) göre, uyum eksikliği veya aşırı sayıda alt bölge hariç tutularak ideal veya optimal kümeyi oluştururlar. Böylece alt bölgeler yeniden birleştirilir. Sonuç olarak, genel sonuç üzerinde daha az etkisi olan değişkenlerin hariç tutulması veya bir kenara bırakılması ve yerel bir değişken alt kümesi seçimine yerleştirilmesi tercih edilir. Bununla birlikte, doğrusal fonksiyonlar üzerinde temel yinelemeli bölümleme süreciyle, verilerin etkin bir şekilde yorumlanabilme derecesinde sınırlamalar olma eğilimindedir. Model mutlaka sabit değildir. Yine de sınıflandırma ağaçlarını akılda tutma oranını etkileme aracı olarak kullanmak mümkündür. Özyinelemeli bölümlemenin bir ilkesinin, alt bölgelerin sınırlarında gerçek bir süreklilik olmaması olduğuna dikkat edilmelidir, ancak yorumlamaya yardımcı olan parçalı düzgün ve sabit yaklaşımlar kullanabiliriz, bu nedenle ikili ağaç olarak temsil mümkündür (Zhang ve Goh, 2016).

Bu yöntemin dezavantajları ise, alt bölge sınırları arasında süreklilik olmadığı için yaklaşıklığın doğruluğu açıkça etkilenmektedir. Bu, özellikle altta yatan işlevin kendisinin sürekli olduğu durum olabilir.

Yukarıda bahsedildiği gibi, veri alt bölgeleri bölümlenmiş olduğundan, modelin doğrusal mı yoksa toplamsal mı (basit bir model olurdu) olduğunu veya değişkenlerin karmaşık etkileşimler yoluyla ilişkili olup olmadığını somut bir kesinlikle tespit etmek mümkün değildir. Bu amaçla, MARS, yukarıda açıklanan bazı sınırlamaların üstesinden gelmenin bir yolu olarak kullanılır (Taylan vd., 2018). Friedman (1991), MARS olarak bilinen bir veri madenciliği tekniği geliştirdi. Ana işlevi, regresyon tipi problemlerin çözümündedir. MARS, amacın bağımlı bir değişkenin uyumunu optimize etmek için en küçük kareler yöntemini kullanmak olması bakımından regresyona benzer, ancak bunun aksine, MARS ile geleneksel toplama ve doğrusal olanlara ek olarak diğer ve daha karmaşık fonksiyonları belirtmek mümkündür. Bu hem ağaç teknikleri hem de regresyonun bir kombinasyonu yoluyla yapılır.

Parametrik olmayan bir prosedür olan GDM'ye benzer şekilde, iki değişken (bağımlı ve bağımsız) arasında işlevsel bir ilişkinin bulunmadığı durumlarda, ağaçlar gibi parçalı bir porsiyonlama yaklaşımı kullanılarak bir ilişki tanımlanır. Öte yandan, MARS'ın avantajı, düğümlerin yalnızca terminal adımlarının aksine her adımda bölünebilmesidir; toplamsal ve doğrusal ilişkiler yakalanabilir ve iki tür sonuç (sürekli veya kategorik) elde edilebilir, çünkü öngörücülerin aralığı daha fazladır. Bunu yaparken, veriler ayrılır veya eşdeğer aralıklara göre ayrılan çeşitli “spline”lara bölünür. Ardından, her spline için, verilerin daha fazla alt gruplaması gerçekleşir. Süreç bu nedenle benzerdir, ancak önceki yaklaşımlardan daha ayrıntılıdır.

Regresyona özyinelemeli bir bölümleme yaklaşımının kullanımları olsa da, sürekli modeller ve bunların türevleri, ayrık modellerin aksine MARS kullanımından elde edilir. Çok değişkenli etkileşimler ve değişkenler arasındaki ya da toplamaya yakın olan ilişkilerle, daha iyi ve daha verimli modelleme gerçekleşir ve ayrıca, bu değişkenler ve ilişkileri tanımlanabilir. Yüksek seviyeli veriler, aksi takdirde ayırt edilmesi zor olan kalıpları veya etkileşimleri içerebilir, bu nedenle sürekli bir model olduğundan, bu analizin diğer yöntemlere göre birçok avantajı vardır.

MARS'ın arkasındaki temel fikir, bağımsız değişken uzayında farklı aralıklarla ayrılmış ayrı regresyon eğimleri kullanarak doğrusal olmayan bir modele yaklaşabileceğimizdir. Yukarıda bahsedildiği gibi, bu boşluklara "düğüm" adı verilir ve bu düğümler geçildiğinde regresyon çizgisinin ve eğiminin değiştiği görülür. MARS, bir bakıma daha büyük resmin değerlendirmesini yaparken, tek değişkenleri ve aralarındaki ilişkileri incelemesi bakımından hem eş zamanlı hem de ayrıktır.

Veri toplama ve analiz bölümü, herhangi bir araştırma çalışmasının veya projesinin çok önemli bir bileşenidir ve anlamlı içgörüler ve sonuçlar çıkarmak için bilgi toplama ve incelemeye yönelik sistematik bir yaklaşım sağlar. Bu kapsamda araştırma soruları şu şekildedir:

• Her iki ülkeden (Türkiye-Irak) elde edilen verilere göre MARS veri madenciliği yönteminin üniversite öğrencilerinin internet bağımlılık düzeylerini tahmin etme performanslarının sınıflandırılması;

o Doğru sınıflandırma oranına göre farklılık gösteriyor mu?

o Spesifiklik oranına göre farklılık gösteriyor mu?

o Duyarlılık oranına göre farklılık gösteriyor mu?

o Doğruluk oranına göre farklılık gösteriyor mu?

o F1-İstatistiklerine göre farklılık gösteriyor mu?

o ROC eğrisi altında kalan alan, yani yanlış sınıflandırma oranı açısından farklılık gösteriyor mu?

• MARS veri madenciliği yöntemine göre Türkiye'deki üniversite öğrencilerinin internet bağımlılığının en önemli yordayıcıları nelerdir?

• MARS veri madenciliği yöntemine dayalı olarak Irak'taki üniversite öğrencilerinin internet bağımlılığının en önemli yordayıcıları nelerdir?

Ayrıca, bu çalışmanın amacı, Günüç ve Kayri (2010) tarafından geliştirilen internet bağımlılığı ölçeğinin bireylerin bağımlılık durumunu etkileyebilecek faktörleri incelemektir. Çalışmanın örneklemi uluslararası düzeyde olup, her ülkeden (Türkiye ve Irak) kozmopolit üniversiteler ve söz konusu ülkelerden Türkiye'den Van Yüzüncü Üniversitesi ve Irak'tan Soran Üniversitesi seçilmiştir. Hedef kitle üniversite öğrencileri olarak belirlendi. Örneklemin ergenlerden oluşması internet bağımlılığının daha çok ergenlerde görülmesi ve bu dönemdeki bireylerin her türlü etkiye açık olmasından kaynaklanmaktadır. Örneklemdeki birey sayısı 2235; Van Yüzüncü Üniversitesi'nden 1220, Soran Üniversitesi'nden 1015 öğrenci seçkisiz örnekleme yöntemiyle atanmıştır. Örneklemin yaş aralığı 11-67 arasında değişmekte olup, yaş ortalaması 21.66 olarak bulunmuştur.

Derbyshire ve diğerleri (2013) tarafından yapılan bir araştırmaya göre, üniversite çağındaki öğrenciler hem akranlarıyla iletişim kurmak hem de günlük yaşamda gezinmek için teknolojiye giderek daha fazla bağımlı hale geliyor. Bildiğimiz gibi, son COVID-19 salgını ile teknoloji kullanımı, çok kısa bir süre içinde herkes için günlük yaşamda daha fazla yerleşik hale geldi. Bu, üniversite sınıflarının çevrimiçi etkileşimle tanıtılmasını ve sıklıkla değiştirilmesini içerir.

Bu araştırma, genel tarama modellerinden biri olan ilişkisel tarama modelinde tasarlanmıştır. İlişkisel tarama modelleri, değişkenler arasındaki ilişkileri tanımlamaya çalışan ve iki veya daha fazla değişken arasındaki değişimin varlığını ve/veya miktarını belirlemeye çalışan araştırma tasarımlarıdır. Ayrıca ilişkisel tarama modeli yaygın olarak bilinen bir istatistiksel yöntem değildi. Bununla birlikte, potansiyel nitelikler, değişken değişikliklerini tanımlamayı, değişkenler arasındaki ilişkileri belirlemeyi, korelasyon analizini kullanmayı ve tahmine yardımcı olmayı içerebilir. En son bilgiler için, son araştırma kaynaklarına bakın.

Çalışmada kullanılan İnternet bağımlılığı ölçeğinin dili Türkçe olduğu için Soran Üniversitesi'nde %100 İngilizce eğitim gören öğrencilere uygulanabilmesi için ölçek uyarlama adımları dikkate alınarak İngilizceye çevrilmiştir. Bu nedenle, soruların aynı sorgulama hattında olduğundan emin olmak için çeviriler arasında kapsamlı bir inceleme ve karşılaştırma yapılmıştır. Anket daha sonra, tam olarak dağıtılmadan önce Soran Üniversitesi'ndeki 250 öğrenciden oluşan bir grup arasında pilot olarak uygulandı.

Rastgele örnekleme tekniği kullanılarak, anketin bağlantısı Soran'daki fakültelerin 5'indeki öğrencilere dağıtıldı – yani her birinden 50 öğrenci olmak üzere toplam 250 öğrenci. İki gün sonra bir hatırlatma mesajı ile bir hafta süre sınırı getirildi.

250 öğrenciye yapılan pilot uygulama sonucunda veriler üzerinde uygulanan Doğrulayıcı Faktör Analizi sonucunda yapının her bir maddede orijinal ölçekte belirtildiği gibi (RMSEA: 0.060, CFI: 0.95) belirtilen faktörün altında kaldığı görülmüştür. NFI: 0,90, NNFI: 0,95 ve GFI: 0,81). Yine dört boyut için elde edilen güvenilirlik değerleri (McDonald's ω) sırasıyla 0.829, 0.839, 0.833 ve 0.795 olarak elde edilmiştir. Bu değerler ile veri setlerine uyarlanan internet bağımlılığı ölçeğinin geçerli ve güvenilir olduğu sonucuna varılmıştır.

Bir veri setine iki ana istatistiksel analiz yönteminden (parametrik veya parametrik olmayan) biri uygulanmadan önce, az önce belirtildiği gibi çeşitli varsayımların yapılması gerekir. Mishra vd. (2019) ve Douglas'ın (2017) en az iki makalesine göre, dört özel şartın karşılanması gerekir. Bunlar genellikle onlar tarafından şu şekilde görülür:

• Çoklu doğrusallık eksikliği: bağımsız değişkenlerde hata yoktur;

• Normallik: veriler normal dağılım gösterir;

• Doğrusallık: bağımsız değişkenler sabitken, bağımlı ve bağımsız değişkenler arasındaki ilişki doğrusaldır;

• Homojenlik: bağımsız değişkenin herhangi bir kombinasyonunun, bağımlı değişkende müteakip varyansa neden olmasına izin verilebilir.

Sheskin (2011) parametrik olmayan bir yaklaşım kullanmak da mümkündür. Bunun nedeni, parametrik bir yaklaşımın yalnızca verilerin normal dağılım altında olduğu durumlar için geçerli olmasıdır; Hazra (2017) göre tersi durumda, farklı güven aralıkları ve değerler ortaya çıkabilir- sonuçların gerçeklikten farklı olabileceği öngörülebilir sonuçla, ki bu açıkça istenmeyen bir durumdur.

Araştırmacılar verilerin normal dağılım gösterip göstermediğine karar verirken genellikle iki ana yöntem kullanırlar. Bunlar $n \geq 50$ olduğunda Kolmogorov-Smirnov testi ve bunun altında ise Shapiro-Wilk testidir. Verilerin normal dağılan bir popülasyondan olması ve sıfır hipotezinin $P>0.05$ ile tutulması gerekir.

Mevcut çalışmanın sonuçlarıyla birlikte, aşağıdaki Tablo 4.1 ve 4.2'de gösterildiği gibi Kolmogorov-Smirnov testi için aşağıdaki testlerin her ikisi de P>0.05 göstermektedir, yani normal bir dağılım geçerli değildir.

Ölçeğin yapı geçerliğine ilişkin Günüç ve Kayri (2010) tarafından yapılan açımlayıcı faktör analizine göre ölçek dört alt boyuttan (Çekilme Kontrol Güçlüğü-İşlevsellikte Bozulma-Sosyal İzolasyon) oluşmaktadır. Bu çalışmanın sonuçlarının ve yapısının geçerli olduğunu göstermek için elde edilen verilerle doğrulayıcı faktör analizi uygulanmıştır.

Gerekli analizi yapmak ve uygun bir model oluşturmak için LISREL8.80 paketi kullanıldı. Varyans faktörü 1 olan faktör ölçeklemesi kullanıldı. Veri analizinde şu faktörler de incelenmiştir: (NFI), (NNFI), (RMSEA), (SRMR), (GFI), Ki kare (x^2) ve (CFI).

Örneklem büyüklüğü x^2'yi etkileyen önemli bir faktör olduğu için atıldığını belirtmekte fayda var. Bu, Tibshirani (2001) tarafından alınan ve yukarıdaki diğer tüm faktörlerin kullanıldığı yaklaşımı izler. NFI, GFI, NNFI ve CFI ≥0,90 ve SRMR ve RMSEA ≤0,08 referans noktası olarak değerler hesaplamalara dahil edilmiştir. Aşağıdaki Tablo 4.3, Doğrulayıcı Faktör Analizimizin sonuçlarını gösterirken, Ek 2'de yol diyagramları bulunabilir.

Geldhof vd. (2014) eski α katsayısı taraflı sonuçlar üretir ve bu nedenle ω katsayısı tercih edilir. Yurdugül (2006), göreli farkın faktör analizi sonuçlarına da yansıdığını gözlemlemektedir. Hem Cronbach hem de McDonald değerlerini hesaplamak için Jamovi 0.9.0.3 kullanıldı. Aşağıdaki Tablo 4.4'te verilerin daha derinine inildiğinde, Türkiye verilerine göre Ölçek Güvenilirliğinin çeşitli alt kategorileri sunulmaktadır. Tablo 4.5'ten, 0,957'lik bir α değeri ve 0,959'luk bir ω değeri ile, verilerin yukarıdaki tartışmaya dayanarak geçerli ve dolayısıyla güvenilir olarak kabul edilebileceği açıktır.

Bu çalışmaya ait veri setini İnternet Bağımlılık Ölçeği ve demografik/kişisel bilgileri içeren anket formlarından elde edilen veriler oluşturmuştur. Araştırmanın örneklemi, 1220’si (427 erkek ve 793 kız) Türkiye ve 1015’i (465 erkek ve 550 kız) Irak’tan olmak üzere seçkisiz örnekleme yöntemiyle toplamda 2235 öğrenciden oluşmaktadır. Çalışma kapsamındaki veriler, SPM 8.2 programından istifade edilmiştir.

MARS veri madenciliği çalışması için SPM 8.2 programının varsayılan değerleri varsayılan değerler olarak kullanılmış, öğrenme verileri ve test verileri dikkate alınmaksızın tüm veri seti dikkate alınarak işlemler gerçekleştirilmiştir. Türkiye veri seti için MARS veri madenciliği yöntemi ile yapılan analiz sonucunda öğrenciler internet bağımlılığı adına “bağımlı/bağımlı değil” olarak sınıflandırıldığında bu gruplamalara uyan öğrenci sayıları Tablo 6.1'de verilmektedir. Irak veri seti için MARS veri madenciliği yöntemi ile yapılan analiz sonucunda öğrenciler internet bağımlılığı adına “bağımlı/bağımlı değil” olarak sınıflandırıldığında bu gruplamalara uyan öğrenci sayıları Tablo 6.2'de verilmektedir.

MARS analizi sonucunda; Türkiye’deki öğrenciler için 9 (dokuz) temel fonksiyon elde edilmişken Irak’tan oluşan veri seti için ise temel fonksiyon sayısı 20 (yirmi) olarak gözlenmiştir. Bu çalışmanın amaçlarından biri de Günuç ve Kayri (2010) tarafından geliştirilen yoksunluk, kontrol güçlüğü, işlevsellikte bozulma ve sosyal izolasyon alt kategorilerinin etkilerini incelemektir. Bu alt kategorilerin her iki ülkede de üniversite öğrencileri arasında internet bağımlılığı ile anlamlı bir ilişkisi olduğu tespit edilmiştir. MARS tarafından oluşturulan modelde hem Türkiyeli hem de Iraklı öğrencilerin internet bağımlılık düzeyini etkileyen en önemli faktörün “günlük internet kullanım süresi” olduğu tespit edilmiştir. Çalışmaya ait detaylar ise tezin Bulgular ve Sonuç bölümlerinde yer almaktadır. MARS yöntemi ile internet bağımlılık düzeyini etkileyen faktörlere ilişkin bulguların literatüre katkı sağlayacağını diliyor ve bu tür veri madenciliği yöntemlerinin teorisi ile yaşama dönük verilere uygulamalı olarak yansıması tavsiye edilmektedir.

CURRICULUM VITAE

Personal Information

Name Surname : Hewa Ghafor HASSAN

E-mail :

Education Information

Undergraduate

University : Salahaddin University – Hawler

Faculty : Administration and Economics

Department : Statistics

Graduation Year : 2008

Master's degree

University : The University of Nottingham

Faculty : Mathematical Science

Department : Statistics

Graduation Year : 2012

Publications:

Hassan, H., Kayri, M., Şevgin, H., (2023). Examine The Internet Addiction Scale Of Students In Turkey And Iraq Comparatively With The Multivariate Adaptive Regression Splines (MARS) Method. *Journal of Namibian Studies*. https://doi.org/10.59670/jns.v34i.1792.

Hassan, H., Qader, J., (2020). Characterizing the Multivariate Exponential Distributions. *Journal of Business Management*. https://doi.org/10.53555/m.v6i1.3379.

Hassan, H., Qader, J., (2018). International Supply Chain: An Application Study in Oil and Gas Industry. *Journal of Business Management*. https://doi.org/10.53555/bm.v4i3.1899.

Hassan, H. (2018). Adaptation Talent Management to Enhance Organisation's Business Strategy, Erbil International Airport as a Case. *International Journal of Social Science and Economic Research*, 3(3), 918-927.

VAN YÜZÜNCÜ YIL UNIVERSITY
THE INSTITUTE OF NATURAL AND APPLIED SCIENCES
THESIS ORIGINALITY REPORT

Date 15/08/2023

Thesis Title: EXAMINE THE INTERNET ADDICTION LEVELS OF STUDENTS IN TÜRKİYE AND IRAQ COMPARATIVELY WITH THE MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) METHOD

The title of the mentioned my thesis, above having total ………..(...........) pages with cover page, introduction, main parts and conclusion, has been checked by me/my supervisor for originality by Turnitin computer program on the date of 15/08/2023 and its detected similary rate was 18% (Eighteen) according to the following specified filtering.

The applied filters are given below:

- Excluding the acceptance and approval page,
- Excluding Acknowledgments,
- Excluding table of contents,
- Excluding symbols and abbreviations,
- Excluding materials and methods,
- Excluding references,
- Excluding quotations,
- Excluding publications derived from the thesis
- Excluding the text parts less than 7 words (Limit match size to 7 words)

I read the Thesis Originality Report Instruction of Van Yuzuncu Yil University for Obtaining and Using Similarity Rate for the thesis and I declare my thesis does not contain any plagiarism according to the maximum similarity rates specified in this directive; otherwise, I accept all kinds of legal responsibility for any dispute arising in situations which are likely to be detected and I declare that the information I have given above is correct.

I submit it to your information.

Date and Signature

Name and Surname: Hewa Ghafor HASSAN
Student ID: 19910001255
Department: Statistics
Program: Statistics
Status: ( ) M.Sc (X) Ph.D.

SUPERVISOR	**APPROVAL OF THE INSTITUTE**
SUITABLE	SUITABLE