Iranian J Publ Health, Vol. 39, No.4, 2010, pp.51-63

Original Article

The Analysis of Internet Addiction Scale Using Multivariate Adaptive Regression Splines

**M Kayri*

Dept. of Computer and Instructional Technology Education, Educational Faculty, Yuzuncu Yil University, Van, Turkey

(Received 3 May 2010; accepted 3 Oct 2010)

Abstract

Background: Determining real effects on internet dependency is too crucial with unbiased and robust statistical method. MARS is a new non-parametric method in use in the literature for parameter estimations of cause and effect based research. MARS can both obtain legible model curves and make unbiased parametric predictions.

Methods: In order to examine the performance of MARS, MARS findings will be compared to Classification and Regression Tree (C&RT) findings, which are considered in the literature to be efficient in revealing correlations between variables. The data set for the study is taken from "The Internet Addiction Scale" (IAS), which attempts to reveal addiction levels of individuals. The population of the study consists of 754 secondary school students (301 female, 443 male students with 10 missing data). MARS 2.0 trial version is used for analysis by MARS method and C&RT analysis was done by SPSS.

Results: MARS obtained six base functions of the model. As a common result of these six functions, regression equation of the model was found. Over the predicted variable, MARS showed that the predictors of daily Internet-use time on average, the purpose of Internet- use, grade of students and occupations of mothers had a significant effect ($P< 0.05$). In this comparative study, MARS obtained different findings from C&RT in dependency level prediction.

Conclusion: The fact that MARS revealed extent to which the variable, which was considered significant, changes the character of the model was observed in this study.

Keywords: *MARS, Piecewise function, Internet addiction, Linear correlation*

Introduction

Whether the correlations between variables in the research design are linear defines the preferable regression method. Accurate modeling of cause-effect relationship becomes harder as the number of predicted and predictive variables increase. In other words, with the increase in the number of variables in the model and interactions between variables, parameter estimations might be biased (1, 2). In this context, regression methods in accordance with the structure of variables in the research design during application could have different functions. Regression might be linear, non-linear or mixed according to models to show correlations between variables. Linear correlation between variables is considered as an important assumption in order to apply parametric regression methods. In case of non-linear correlation between variables, the model is fitted into non-parametric methods. However, it is occasionally observed that distribution curve or regression curve is too rough for non-parametric methods and therefore the curve becomes difficult to interpret (3). The aim of such smoothing functions or additive algorithm process is to get readable curves and lower mean square error (MSE) for unbiased parameter estimations (4).

One of the latest non-parametric methods is Multivariate Adaptive Regression Spline (MARS), which involves a number of variables in the model in an independent or in an interaction fashion and enables unbiased parameter estimations with strong algorithms. MARS could be viewed as generalization of repeated discriminate method and stepwise linear regression to improve the performance of a given regression set (5). MARS creates a new regression equation for each linear region in model. Obtained each linear region is called as

***Correspondence:** Tel: +90 507 9842982, Fax: 0432 2251368, E-mail: mkayri@yyu.edu.tr

"knot". This method first divides data space (pile) into areas and then forms a regression equation for each, which highlights MARS as an applicable solution to multivariate regression problems that might cause multidimensionality for other methods. MARS uses both forward and backward progresses for robust and unbiased parameter estimations. At first, MARS maximizes all the possible effects of predictive variables in the forward model and then removes the least effective functions in the backward model using Ordinary Least Squares method. In general terms, regression methods get a single regression equation in the model, whereas MARS, unlike these methods, a lot of piecewise regression equations in the model. With these qualities, MARS could make a coherent, unbiased estimation for the predicted variable/s, which can be continuous, ordinal, and nominal in the research design (6).

MARS method was developed by Friedman (1991) to smooth rough regression curves and to bring correlation between the predicted variable and predictive variables into conformity (7). The main principle of MARS is based on revealing the effect of predictive variables partial, without breaking linearity. In other words, every point where linearity breaks is taken as a knot and predictive variables, which are influential up to that point, are modeled using a new regression equation. Then, separate regression equations are obtained for each knot defined by the examination of the other dimensions of the curve. Thus, the study design is constantly examined within a linear relationship. The number of regression equations is the same as the number of knots defined in the process and the effect of predictive variables, which are influential on each knot, is clear. MARS reveals the final model, taking the obtained combination of basis functions into account (regression equation for each knot; Basis Function-BF).

Generally, MARS model calculates predictive variables whose effect on a single predicted variable is being examined in the model as in Equation 1 (8).

$$Y = \sum_{k=1}^{M} \beta_k BF_k(X) + \varepsilon \qquad [1]$$

Y in Equation [1] represents the predicted variable, while X shows predictive variable set. Term BF_k in the model refers to k_{th} basis function (BF) in every linear knot. As mentioned above, BF divides a correlation structure, which is not linear into linear segments and expresses regression equation obtained for each linear segment. Term *M* gives the number of basis functions in the final model and βk refers to the estimated value of lowered Mean Square Error (MSE).

Model selection in MARS is calculated using Generalized Cross–Validation (GCV) (9).

$$GCV = \frac{1}{N} \frac{\sum_{i}^{N} [Y_i - f_M(X_i)]^2}{\left[1 - \frac{C(M)}{N}\right]^2} \qquad [2]$$

In Equation [2], N calculates observation number; $\sum_{i}^{N} [Y_i - f_M(X_i)]^2$ calculates lack-of-fit of the sum of squared residuals ($f_M(X_i)$) of BF in M number which is found for data set. In $\left[1 - \frac{C(M)}{N}\right]^2$ is the penalty term to apply to M number of BF. The penalty term is applied for reducing the number of BFs, which tend to increase in the model and for restricting the ideal model number. Finally, the ideal MARS model is represented by an equation estimated by the lowest GCV obtained from Equation [2].

MARS, developed to fill that deficiency of non-parametric methods, both obtains readable regression curves and makes unbiased estimation using split method and a solution approach (10-12). Although the performance of MARS depends on the structure of variables in data set (13), it is generally accepted as a preferable method because of accurate estimation and fast calculation as well as ease of interpretation (14).

In the literature, it is seen that there are various methods, which model the effect of predictive variables on the predicted variable. Logistic Regression (LR), Classification and Regression Tree (CART), Principal Component Regression (PCR) and Generalized Additive Models (GAM) are commonly used methods. In many studies, it is

reported that MARS gives more effective results than these methods (15-17). Nevertheless, it is also stated the predictive performance of MARS lowers when the sample size is insufficient (18). Hence, it is essential to pay attention to obtain data set to which MARS is applied from big samples. Moreover, Briand and et al. (2007) suggest multicollinearities might occur since MARS gets interaction between predictive variables involved in the model (19). In this context, a study to determine whether there is multicollinearity in the model is needed.

The aim of the present study was to show practically a general introduction of non-parametric MARS method, which could efficiently model variables of mixed structure in the study design. Applicability of MARS method will be shown over a data set compiled from a scale study, which was developed to reveal Internet dependency profile in Turkey. The study aims at presenting an alternative point of view about predictive power of variables by a different regression method. It was also aimed to reveal the main factors on internet dependency. Revealing the power of the MARS, the data set will be analyzed with Classification and Regression Tree method. The findings will be discussed comparatively.

Materials and Methods

Data Gathering Instrument

With using MARS method, internet addiction term and internet addiction level were examined. The factors, which affect the dependency, were scrutinized with MARS. Addiction is defined as being unable to give up or control certain behavior or substance abuse (20, 21). Internet dependency was first used in an e-mail sent by Dr. Ivan Goldberg as a joke in 1996 (22). It might be suggested that most Internet dependency sufferers are male and young, although it is the common problem of individuals from every social group and age group (23). The concept was first used following the term "Internet dependency" in the literature and was later called in different ways by various researchers and clinicians. These terms are "Internet dependency" (24), "pathological Internet use" (25, 26), "problematic Internet use" (27, 28), "excessive Internet use" (29), "Internet abuse" (30), "Internet dependency disorder" (31, 32) etc. Furthermore, in some studies, we come across the term "cyber-addiction", which is used to mean on-line or off-line dependency (33). In short, it might be suggested that these terms express undesirable cases, particularly caused by excessive Internet use.

Data set to which MARS method is applied is taken from the Internet Dependency Scale (IDS), developed by Günüç (2009). The sample consists of 754 secondary school students (301 female, 443 male with 10 missing data). IDS is a five-item Likert scale and attitude levels are as follows: "I totally agree", "I agree", "I am not sure", "I disagree" and "I totally disagree". The Internet Dependency Scale is a measure tool, which was developed to define Internet dependency levels. The measure tool of dependency level consists of 35 items and the minimum score is 35, whereas the maximum score is 175. High scores in IDS show stronger dependency. Cronbach Alpha for Internal consistency for the tool is found as 0.944. The scale has four sub-dimensions and the overall explained variance of the scale is calculated as 47.463%. It consists of the following sub-dimensions: "Deprivation", "Control Difficulty", "Functional Decline" and "Social Isolation". The contribution of Deprivation to the overall explained variance is calculated as 15.084%. The percentage is 11.911% for Control Difficulty, 10.553% for Functional Decline and finally 9.915% for Social Isolation. Reliability coefficients of the measure tool sub-dimensions are respectively 0.877, 0.855, 0.827 and 0.791. Factor load values of all the items in the measure tool range from 0.40 to 0.702. Correlation coefficients of the items in IDS range from 0.590 to 0.800. In the light of the given data, IDS is claimed to be ideally reliable and valid.

Process

For the students included in the study design, "Internet dependency level" is taken as the predicted variable, whereas "grade (grades 9, 10, 11 and 12)",

"gender", "age", "the number of siblings (sibling)", "educational background of father (father educational background)", "educational background of mother (mother educational background)", "occupation of father (father occupation)", "occupation of mother (mother occupation)", "cigarette smoking (smoking)", "monthly economic income level (income)", "the most common intended use of the Internet (intended use)" and "daily Internet use time on average (time)" are assigned as predictive variables in the model. The literature is referred to in order to define predictive variables. The literature shows that variables such as Internet use time (34-37), age and income could be influential on Internet dependency levels. In this context, most predictive variables are included in the model in reference to the literature. The predicted variable in the study design is obtained by an interval scale and is continuous. The following predictive variables are nominal scale type: "grade", "gender", "occupation of father", "occupation of mother", "cigarette smoking" and "intended use" The following predictive variables included in the model are ordinal scale type: "age", "the number of siblings", "educational background of father", "educational background of mother", "income level" and "time" Regression equation of the predictive vector under examination which affects the predicted variable is as follows:

Internet dependency level(Y)= $\beta_0+\beta_1$*grade+β_2* gender+β_3*father_educational background+β_4* mother_educational background+β_5*father_occupation +β_6*mother_occupation+β_7*cigarettesmoking+β_8*income+β_9*intended use+β_{10}*age+β_{11}*siblings+ β_{12} *time+ε

Regression equation, which consists of various scales (continuous, nominal, ordinal) and, a group of predictive variable data set will be examined by MARS method. Testing multicollinearity is considered essential for a coherent estimation by MARS, since there are many variables included in the model (19). In the literature, whether there are high correlations between predictive variables are tested by different criteria. Collinearity diagnostics is examined to assess whether there is multicollinearity between predictive variables. In the table, there are eigenvalues, case indexes and variance ratios for each variable and a higher eigenvalue of a variable than others means multicollinearity (38). Tolerance and Variance Inflation Factor (VIF) might be examined in order to determine multicollinearity (38). Menard (1995) suggests there is considerable multicollinearity when tolerance value is <.1 (39). Another method to check multicollinearity is the assessment of standard errors of unstandardized regression coefficients (β) (40). If standard errors of all variables are <2, it is decided that there is no multicollinearity. Because of the analysis, standard error mean of all the variables is calculated as 1.204. According to that criterion, there is no multicollinearity between the variables. When tolerance values of the variables are examined, it is observed that tolerance values range from 0.318 to 0.959. When this method is used, it is seen that there is no multicollinearity in the model. When collinearity diagnostics is checked, it is seen that eigenvalues of all the predictive variables are close to one another and therefore there is no multicollinearity.

In order to examine the performance of MARS, MARS findings will be compared to Classification and Regression Tree (C&RT) findings, which are considered in the literature to be efficient in revealing correlations between variables. C&RT, defined as Classification Tree in the event that the predicted variable was obtained by classified scale, and as Regression Tree in case of continuous variable, is known as a classification technique which is not held responsible for parametric regression technique assumptions and defines correlations between dependent variable(s) and independent variables in its own population, without any inference with data set values (41, 42). Since the dependent variable in this study is a continuous variable, Regression Tree (RT) will be applied to the data set. Although the fact that C&RT has certain advantages is highlighted in the literature, it cannot provide parametre prediction in cases where linearity is broken as in MARS. C&RT considers the data set as a whole, but it takes the effect of sub-levels of independent variables into account. MARS, which examines the extent to which parametres are effective in cases where linearity is

broken, has a superior advantage in this respect. In the research, the performance of MARS will be examined in comparison with RT. The analysis of regression equation built in the study by MARS is realized using Mars 2.0 Trial evaluation version.

Results

First, descriptive statistics about the variables in the model are given. The continuous variables in the model include data on age and daily Internet use time on average. The average age of the sample is found as 15.83±1.2. It is observed in the sample that the minimum age is 14 and the maximum age is 22. Daily Internet use time on average by the students is calculated as 2.65±2.08. The finding is that the individuals included in the sample use the Internet at least one hour a day and the maximum Internet use time is ten hours. In the study, the Internet Dependency Scale gives Internet dependency level in the sample (the predicted variable) as 73.057±25.476. The lowest dependency level in the sample is 36, while the highest level is 170. The other variables in the sample are nominal and ordinal, and data on the demographic variables is presented in Table 1.

MARS uses both forward and backward progress algorithms for strong, robust parameter estimations. Because of many testing, algorithm obtains the model function (regression equation), in case of the lowest Generalized Cross Validation (GCV) value. GCV reaches its lowest value where the quantity of error is most minimized in the model. In other words, GCV in the model functions as the controller of Mean Square Error (MSE). The case of function representing the model shown by GCV is presented in Figure 1.

When Fig. 1 is examined, it is seen that the point where GCV most minimizes error is the area where the model is expressed by six separate regression equations. Here, the model might be expressed by six basis functions in terms of variable structure and the effect of variables (independent and interactive) on the predicted variable in the model. Findings about the estimated number of functions in the model produced by GCV during analysis are given in Table 2.

Table 2 shows that the lowest value is 484.897 obtained by GCV for the model and this value overlaps the lowest MSE in the model. Following partial linear function number of GCV representing the model, MARS presents other important findings of the model. MARS could express the whole model where one predicted and 12 predictive variables are included by 0.348% ($R^2 = 0.348$).

MARS defines the significant variables in the model respectively as follows: daily Internet use time on average, the most common intended use of the Internet, grade of students and occupation of mother. MARS is sensitive to missing values of variables in data set and differs from many other regression methods. Taking missing values of variables into account, the finding about the level of significance of MARS predictive variables in the model is presented in Table 3.

When Table 3 is examined, it is clear that missing values of the variables in the model are important at certain points. MARS, in this context, sensitively functions and clearly shows the character of the study design. In fact, MARS implies it is essential to pay attention to the related variable while revealing the level of significance of missing values of that predictive variable.

MARS shows predictive variables, which affect Internet dependency level, are expressed by six basis functions (BF). Level of significance of the obtained functions is listed in Table 4. In Table 4, MARS summarizes the effect of predictive variables on the predicted variable, taking BFs in the model into account, with the following regression model:

Internet dependency Level (Y)= 60.517+17.052 *BF1+3.192 *BF3-9.938 *BF4-13.623 *BF7+9.892 *BF11-14.331 *BF15

Internet dependency Level (Y)= BF1 BF3 BF4 BF7 BF11 BF15 (Regression equation in MARS modeling is shown in this way)

MARS summarizes BFs in the regression equation in the model as follows:

BF1= (HOW MANY_HOURS ne.*);
BF3= max (0, HOW MANY_HOURS-2)* BF1;
BF4= max (0, 2–HOW MANY_HOURS)* BF1;
BF5= (THE_MOST ne.);
BF7= (THE_MOST in (3, 1) * BF5;
BF9= (GRADE ne.) * BF5;
BF11= (GRADE in (1)) * BF9;
BF13= (MOTHER_MSL ne.) * BF5;
BF15= (MOTHER_MSL in (4)) *BF13;*: nested

Of basis functions, BF1 expresses missing values of the variable of daily Internet use time on average (time) and "ne." means that missing values are nested in the observed "time" variable. Function BF3 considers the two-hour Internet use time important, by taking function BF1 into account (BF3 is formed by BF1) and creates a knot at that point. As it will be remembered, MARS produced a function for the obtained linear model after leaving a knot at the reached point when linearity tended to break. Like BF3, BF4 is shaped by BF1 and creates a separate unrestricted linear area for the two-hour-Internet use time. Nonetheless, what is remarkable in BF4 here is that it brings a minus value into the regression model, contrary to BF3 (-9.938). In other words, the daily average two-hour-Internet use time in the sample causes an increase in Internet dependency levels of some of the users (+3.192*BF3), and the same case (the two-hour-Internet use) lowers Internet dependency levels of some of the individuals (-9.938*BF4). In this context, BF3 and BF4 shaped by BF1 are considered as two separate linear areas (knots). However, all individuals are represented by a single function in other classical linear or non-linear regression methods. As such, MARS process is considered more realistic. BF5 obtained in the model represents missing values about the intended use of the Internet and shows that missing values are covered by the observed variables. Influenced by BF5, BF7 produces three separate linear areas and finds those who use the Internet for "research" (Table 1)" significant. BF9 and BF11 are affected by function BF5 and unlike what is expected, use of the Internet for "research" for 9 graders increases dependency level (+9.892*BF11). Nevertheless, for the others, use of the Internet for "research" is a reducing factor of dependency (-13.623*BF7). Finally, similar to "grade", BF13 reveals that the intended use of the Internet by those whose mothers' occupations are coded as "4 (free-lance; self-employment)" differs from that of the others. Here, in the linear function obtained within the unrestricted area, it is observed that Internet dependency levels of those whose mothers' occupations are coded as "free-lance" are lowered (-14.331*BF15). It is also seen that six functions are used in the regression model, although MARS produces nine BFs (BF1, BF3, BF4, BF7, BF11, BF15). It is accepted that functions BF5, BF9 and BF13, removed from the regression model by MARS, have an indirect effect and they especially create BFs assigned in the model. MARS produces a single regression equation, taking the six BFs obtained in the model and the overwhelming effect of each function into account.

When correlation between the dependent variable and the independent variable was examined by Regression Tree, findings different from MARS were obtained. As it is clear from Table 5, it was certain that the most influential predictor on Internet dependency was "the most common intended use of the Internet".

However, in MARS, the most influential variable on the dependency level was "daily Internet use time on average". RT, which emphasizes the intended use of the Internet, varies according to occupation of father, provided data different from MARS. As it is mentioned above, MARS, unlike RT, highlighted educational background of mother was influential on the dependency level, not occupation of father. In addition, MARS found daily Internet use time the most influential variable on the dependency level, while RT considered this variable as secondarily significant. RT did not find grade influential on dependency, whereas MARS considered both grade and interaction between grade and occupation of mother significant in the model. Particularly, it was interesting that RT found occupation of father influential on depend-

ency level, but not occupation of mother. On the other hand, MARS considered educational background of mother significant in the model. In addition, the variables in the model were found significant neither in the MARS model nor in the RT model. Briefly, MARS found grade and occupation of mother significant in the model, while RT didn't consider these variables significant. Also, occupation of father was found significant by RT, not by MARS. In this context, it was observed that both techniques obtained different models. Besides, the extent to which both techniques found the independent variables significant was rather different. As it is clear from Table 6, MARS found missing values that belonged to daily internet use time 100% significant, whereas RT found the most common intended use of the Internet 100% significant. MARS considered daily internet use time 90.1% significant, while RT obtained 74.21% significance. See Table 3 and 6 for the significant levels of the other variables. As it is obviously seen, the significant levels of the variables shown by the two methods were different.

Table 1: Descriptive statistics for nominal and ordinal predictive variables

Variable	Level	f	%
Grade	1. Class 9	356	47.2
	2. Class 10	160	21.2
	3. Class 11	135	17.9
	4. Class 12	101	13.4
	Missing observation	2	0.3
Gender	1. Female	301	39.9
	2. Male	443	58.8
	Missing observation	10	1.3
Income of Family	1. 0 - 500 TL	102	13.53
	2. 501 - 1000 TL	289	38.3
	3. 1001 - 1500 TL	174	23.1
	4. 1501 – 2000 TL	68	9.0
	5. 2001 – 2500 TL	24	3.2
	6. Upper 2501	25	3.3
	Missing observation	72	9.5
Mother Educational Background	1. Unliterated	122	16.2
	2. Literated	43	5.7
	3. Primary School	330	43.8
	4. Elemantary	107	14.2
	5. High School	116	15.1
	6. University	18	2.4
	Missing observation	18	2.4
Father Educational Background	1. Unliterated	28	3.7
	2. Literated	22	2.9
	3. Primary School	284	37.7
	4. Elemantary	163	21.6
	5. High School	177	23.5
	6. University	72	9.5
	Missing observation	8	1.1
Mother’s occupation	1. Housewife	639	84.7
	2. Officer	21	2.8
	3. Worker	28	3.7
	4. Free-lance(self-employment)	36	4.8
	Missing observation	30	4.0
Father’s occupation	1. Unemployed	42	5.5
	2. Worker	138	18.3
	3. Free-lance(self-employment)	320	42.4
	4. Retired	81	10.7
	5. Officer	114	15.4
	Missing observation	59	7.9

Smoking	1. Yes	51	6.8
	2. No	653	86.6
	Missing observation	22	2.9
The most intended internet use	1. Research	356	47.2
	2. Chat	116	15.4
	3. News	25	3.3
	4. Music-film	70	9.3
	5. Game	54	7.2
	6. Pornography	13	1.7
	7. On-line shopping	1	0.1
	8. Web_sorf	38	5.0
	Missing observation	81	10.7
Siblings number	1. 0 – 2 siblings	309	41.0
	2. 3 – 5 siblings	307	40.7
	3. 6 – 8 siblings	83	11.0
	4. upper 9 siblings	36	4.8
	Missing observation	*19*	2.5

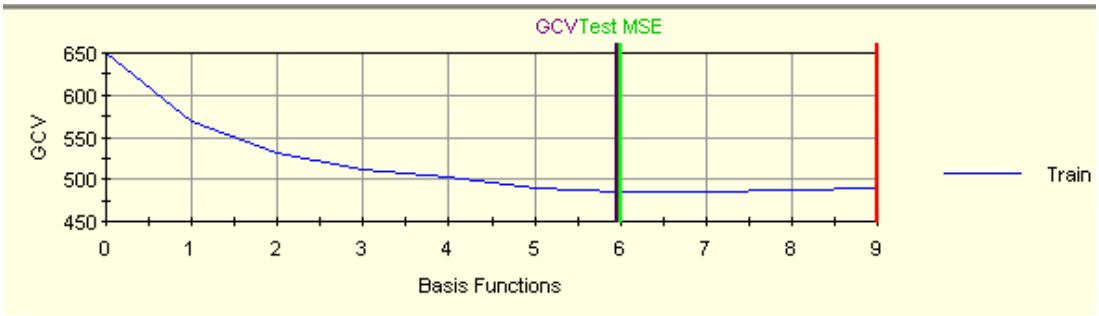

Fig. 1: Minimized MSE of GCV

Table 2: GCV findings related to model

Basis Functions	Number of Predictive Variables	Values of GCV	MSE
9	12	490.1	473.3
8	12	488.0	473.0
7	12	486.3	473.1
****6**	**12**	**484.9**	**473.0**
5	9	490.6	480.9
4	9	503.3	495.3
3	9	511.8	505.5
2	6	530.4	525.8
1	3	568.5	565.7
0	0	649.911	648.189

Table 3: The significance level of meaningful explanatory variables in the model

The Meaningful predictors in the model	Significance (%)
Hour (time)_missing observation	100
Hour (time)	90.1
Intended internet use_missing observation	88.2
Intended internet use	72.2
Grade_missing observation	51.9
Grade	51.9
Mother's educational background_missing observation	26.7
Mother's educational background	26.7

Table 4: Parameter estimations for Basis Functions

Function	Parameter	Estimation Value	Std. Error	T – Ratio	*P*
-	intercept	60.5	3.2	18.6	0.000
1	BF 1	17.1	3.4	5.1	0.000
2	BF 3	3.2	0.5	6.5	0.000
3	BF 4	-9.9	1.9	-5.3	0.000
4	BF 7	-13.6	1.6	-8.3	0.000
5	BF 11	9.9	1.6	6.1	0.000
6	BF 15	-14.3	4.0	-3.5	0.000

Table 5: Meaning explanatory variables with RT

Node	Mean	Std. Dev.	n	%	Mean	Primary Independent Variable	
						Variable	Improvement
0	73.05	25.47	754	100.0	73.05		
1	66.56	22.99	449	59.5	66.56	Intended*	73.522
2	82.61	25.96	305	40.5	82.61	Intended	73.522
3	58.86	18.73	162	21.5	58.86	Hour	25.456
4	70.90	24.04	287	38.1	70.90	Hour	25.456
5	78.51	24.60	241	32.0	78.51	Hour	24.382
6	98.09	25.31	64	8.5	98.09	Hour	24.382
7	74.93	24.03	186	24.7	74.93	Father_Occupation	14.602
8	90.61	22.76	55	7.3	90.61	Father_Occupation	14.602

*: Intended internet use

Table 6: The significance level of meaningful explanatory variables in the model with RT

Independent Variable Importance		
Independent Variable	**Importance**	**Normalized Importance**
Intended*	100	100.0
Hour	74.21	73.8
Father_occ.	25.61	25.5
Income	9.93	9.9
Mother_occ.	9.71	9.7
Father_educ.	8.75	8.7
smoking	7.37	7.3
Mother_educ	6.89	6.9
sibling	6.70	6.7

Discussion

In the light of the research model obtained by MARS, daily Internet use time on average, the intended use of the Internet might be considered as the main triggering factors for Internet dependence for Turkish secondary school students. In the literature, these results were supported and it was concluded that Internet use time on average caused Internet dependence (23, 34-36). Besides, as in other studies (24, 32, 35, 37), the present study showed that the intended use of Internet was an important indicator of Internet dependence, as well as the Internet use time on average.

The present study attempts to introduce MARS, one of the non-parametric regression methods, in theory and practice. Unlike the well-known clas-

sical methods, MARS does not obtain a generalizing function (regression equation) for the population or all the individuals in the sample, but splits the whole model into linear regions and produces separate functions (BF) for the each generated linear area (knot). Afterwards, it obtains a single regression equation, which represents all the BFs, taking the overwhelming effects of the defined BFs into account.

In the literature, it is highlighted that regression equations obtained by MARS method make robust and coherent parameter estimations (15-17). However, it is also pointed out that MARS could make biased estimations in case of multicollinearities in the model (19). In this context, it is essential to test whether there are high correlations between variables in the study design. Similarly, it is also stated the predictive performance of MARS is adversely affected when the sample size is small. Hence, it is essential for researchers to pay attention to such cases (multicollinearity and ideal sample size) in cause-effect based research to which MARS is applied.

In this comparative study, MARS obtained different findings from RT in dependency level prediction. Similarly, MARS obtained significant levels of the variables in the model differently. In this respect, as it was shown by similar research, MARS was considered more efficient in model estimation (15-17). Without any confusion in cause-effect relationship of the advantage of MARS, the case could be explained by linear knots. In addition, MARS calculates the effect of missing values of independent variables, while RT ignores missing values. It could be thought that MARS as linear knots estimates parameters more effectively. Significant levels of the predicted variables, which were considered as functions in linear knots, were found different by the MARS model. Based on the other studies in the literature, it could be suggested that findings obtained by MARS are more robust.

Data set to which MARS is applied is taken from a study on Internet dependency levels, as an example for future epidemiological studies. In this context, the study mainly attempts to introduce MARS and secondarily reveal factors, which affect Internet dependency levels of those included in the sample. The model is built on 12 predictive variables (ten nominal/ordinal and two continuous) and a dependent variable, Internet dependency level (the predicted). MARS shows that the predictive variables (daily Internet use time on average, the intended use of the Internet, grade of students and occupation of mother) have a significant effect on the predicted variable ($P< 0.05$). MARS takes missing values of the variables, which are considered significant into account, and shows missing values of many variables, which are considered significant in the model, are significant, as well. It is striking that the factor, which is found significant by MARS in the model reveals there are missing values of the Internet use time variable. Moreover, the fact that MARS reveals the extent to which the variable, which is, considered significant, changes the character of the model could be viewed as an advantage of MARS. It is known in the literature that the piecewise function model obtained by MARS is the solution to the complexity of regression curves (7). Whereas, it is occasionally observed that distribution curve or regression curve is too rough for non-parametric methods and therefore the curve becomes difficult to interpret.

It is concluded that MARS method, which does not set the conditions for scales (continuous-discrete) from which the predicted and the predictive variables are obtained, will make unbiased predictions, since it gets interactions between variables involved in the model and obtains regression equation that represents the model from many partial functions. It is advisable for all qualitative study researchers in general and for epidemiologists in particular to benefit from MARS as a method in eligible study designs.

Ethical Considerations

All ethical issues including plagiarism, Informed Consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc have been completely observed by the author.

Acknowledgements

The authors declare that they have no conflicts of interest.

References

1. Hardle W, Müler M, Sperlich S, Werwatz A (2004). *Nonparametric and Semiparametric Models*. Springer Series in Statistics: Springer Press, Germany.
2. Kayri M (2007). Two-step clustering analysis in researches: A case study. *Eurasian Journal of Educational Research*, 7(28): 89-99.
3. Yatchew A (2003). *Semiparametric Regression for the Applied Econometrician*. Cambridge University Press, England.
4. Kayri M, Zırhlıoğlu G (2009). Kernel smoothing function and choosing bandwidth for non-parametric regression methods. *Ozean Journal of Applied Science*, 2: 49-54.
5. Akyol M (2007). Genç sporcuların performansının MARS ile kestirilmesi. *Cumhuriyet University Medicine Faculty X. Biostatistics Congress*, 5-8 September, Sivas, Turkey.
6. Stephon P (2001). Forecasting recessions: can we do beter on MARS^{TM}? *Fed Res Bank St. Louis Rev*, 83: 39-49.
7. Frieadman JH (1991). Multivariate additive regression splines. *Ann. Of Stat.*, 19: 1-67.
8. Xiong R, Meullenet JF (2004). Application of multivariate adaptive regression splines (MARS) to the preference mapping of cheese sticks. *Sensory and Nutritive Qualities of Food*, 69: 131 – 140.
9. Friedman JH, Roosen CB (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research*, 4: 197-217.
10. Deconinckb E, Zhang MH, Petitet F, Dubus E, Ijjali I, Coomans D, Vander Heyden Y (2008). Boosted regression trees, multivariate adaptive regression splines and their two-step combinations with multiple linear regression or partial least squares to predict blood–brain barrier passage: a case study. *Analytica Chimica Acta*, 609: 13-23.
11. Doksum K, Peterson D, Samarov A (2000). On variable bandwidth selection in local polynomial regression. *Journal of the Royal Statistical Society Series B,* 62: 432-48.
12. York TP, Eaves LJ, Van Den Oord G (2006). Multivariate adaptive regression splines: a powerful method for detecting disease–risk relationship differences among subgroups. *Statist. Med.*, 25: 1355-67.
13. Türe M, Kurt İ, Kurum AT, Özdamar K (2005). Comparing classification techniques for predicting essential hypertension. *Expert Systems with App.*, 29: 583-88.
14. Leathwick JR, Elith J, Hastie T (2006). Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modeling of species distributions. *Ecol Mod*, 199: 188-96.
15. Kuhnert PM, Do K, McClure R (2000). Combining non-parametric models with logistic regression: an application to motor vehicle injury data. *Comp Stat and Data Anal*, 34: 371-86.
16. Mukkamala SA, Sung H, Abraham A, Ramos V (2004). Intrusion detection systems using adaptive regression splines. *Proceedings of the 6th International Conference on Enterprise Information Systems* (April 14-17, Porto), Proceedings, p.26-33, Porto.
17. Muñoz J, Fellicisimo AM (2004). Comparison of statistical methods commonly used in predictive modeling. *J of Vegetation Sci*, 15: 285-92.
18. Jin R, Chen W, Simpson TW (2000). Comparative studies of metamodeling techniques under multiple modeling criteria. *American Institute of Aeronautics and Astronautics*, Inc 2000; AIAA-2000-4801.
19. Briand LC, Freimut B, Vollei F (2007). Using multiple adaptive regression splines to understand trends in inspection data and identify optimal inspection rates. ISERN TR 00-07 2007. Avaliable from:

http://www.salfordsystems.com.

20. Egger O, Rauterberg M (1996). Internet behaviour and addiction [Master thesis]. Work & Or-ganisational Psychology Unit (IfAP) Swiss Federal Institute of Technology, Zurih.
21. Henderson EC (2001). *Understanding Addiction*. University Press of Mississippi, Mississippi.
22. Günüç S (2009). Development of internet addiction scale and scrutinizing the relations between the internet addiction and some demographic variables [Master thesis]. Institute of Social Sciences, Yuzuncu Yil University, Van.
23. Gonzalez NA (2002). Internet addiction disorder and its relation to impulse control [Master thesis]. Texas A&M University, Kingsville.
24. Tvedt H (2007). Internet use and related factors among fifth-graders [Master thesis]. Umeå University Department of Psychology, Umea.
25. Davis RA (2001). A cognitive-behavioral model of pathological internet use. *Computers in Human Behavior,* 17: 187-95.
26. Young KS (2004). Internet addiction: A new clinical phenomenon and its consequences. *American Behavioral Scientist,* 48: 402-15.
27. Caplan SE (2002). Problematic internet use and psychosocial well-being: Development of a theory-based cognitive-behavioural measurement instrument. *Computers in Human Behavior,* 18: 553–75.
28. 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: 89–96.
29. Yang CK, Choe BM, Baity M, Lee JH, Cho JS (2005). Scl-90-r and 16pf profiles of senior high school students with excessive internet use. *Canadian Journal of Psychiatry,* 50: 407.
30. Young KS, Case CJ (2004). Internet abuse in the workplace: New trends in risk management. *Cyberpsychology & Behavior*, 7: 105-11.
31. Kiralla LV. Internet addiction disorder: A descriptive study of college counselors in four-year institutions [PhD thesis]. Department of Organizational Leadership, University of La Verne, LA Verne; 2005.
32. Thatcher A, Goolam S (2005). Development and psychometric properties of the problematic internet use questionnaire. *South African Journal of Psychology*, 35: 793-809.
33. Chou C, Condron L, Belland JC (2005). A review of the research on internet addiction. *Educational Psychology Review*, 17: 363-88.
34. Cao F, Su L (2007). Internet addiction among chinese adolescents: prevalence and psychological features. *Child: Care, Health & Development,* 33: 275-281.
35. Everhard RA (2000). Characteristics of pathological internet users: An examination of on-line gamers [PhD thesis]. The Department of Psychology, Spalding University, Louiseville.
36. Hardie E, Tee MY (2007). Excessive internet use: The role of personality, loneliness and social support networks in internet addiction. *Australian Journal of Emerging Technologies and Society,* 5: 34-47.
37. Chen K, Chen I, Paul H (2001). Explaining online bahavioral differences: An Internet dependency perspective. *The Journal of Computer Information Systems,* 41: 59.
38. Field A (2005). *Discovering Statistics Using Spss.* 2nd ed. Sage Publication, London.
39. Menard S (1995). *Applied Logistic Regression Analysis.* Sage university paper series on quantitative applications in the social sciences, Thousand Oaks, CA, Sage.
40. Mertler CA, Vannatta RA (2005). *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation.* 3rd ed. Pyrczak Publishing, Glendale.
41. Chang LY, Wang HW (2006). Analysis of traffic injury: Anna application of non-para-

metric classification tree techniques. *Accident Analysis Prevention*, 38: 1019-27.

42. Kayri M, Boysan M (2008). Assesment of relation between cognitive vulnerability and depression's level by using classification and regression tree analysis. *Hacettepe Universitesi Egitim Fakultesi Dergisi-Hacettepe University Journal of Education*, 34: 168-77.