

Log-File Analysis to Identify Internet-addiction in Children

Rasim M. Alguliyev

Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan
E-mail: r.alguliyev@gmail.com

Fargana J. Abdullayeva

Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan
E-mail: a_farqana@mail.ru

Sabira S. Ojagverdiyeva

Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan
E-mail: allahverdiyevsabira@gmail.com

Received: 03 December 2020; Revised: 26 December 2020; Accepted: 18 January 2021; Published: 08 October 2021

Abstract: The problem of the Internet addiction (IA) arose after the rise of the Internet. Some of the Internet users include children and teenagers and they are active in a virtual environment. Most minor users are not well aware of the dangers posed by information abundance. One of these dangers is the IA. Excessive use of the Internet is addictive, and some users experience a high risk of addiction. IA can negatively affect the children's health, psychology, socialization and other activities. There is a great need to the development of forecasting programs and various technological approaches for the identification of IA among Internet users, especially children and adolescents. This article uses machine-learning techniques to detect IA. Activities of children in the Internet environment is analyzed. The log-files of children and their IA problem are explored. To determine the degree of IA among children and adolescents an experiment is conducted on public dataset. The effectiveness of the methods is analyzed by various evaluation metrics and promising results are obtained. The results show better performance of Weighted SVM, compared to BernoulliNB, Logistic Regression, MLPClassifier, SVM classifiers. Acquired results of the research provide kids information security. To evaluate a kids IA helps to identify their psychological conditions, and it creates a better situation for parents, teachers, and other related people to communicate with children and teenagers better way.

Index Terms: Internet addiction, child protection, log file, machine learning, weighted support vector machine

1. Introduction

The formation of the information society, the rapid development of the Internet, and the popularity of the services it offers have contributed to the generation of large amounts of information in the information space. Users have great interest on the Internet, they try to solve their problems in a virtual environment, build relationships and spend most of their time online. Over time, the number of Internet users is growing, real relationships are being integrated into the virtual environment, and new forms of human relationships are emerging.

The benefits of using the Internet are obvious, and it helps to increase productivity and efficiency at work, save time, and solve problems (online banking, online shopping, social networking, education, virtual relationships, entertainment, etc.) without leaving the work place. However, it poses certain problems with information security and causes the problems of IA. The scientific study of the IA phenomenon as a psychological problem is related to the names of two American experts - psychiatrist I. Goldberg and clinical psychologist K. Yang. I. Goldberg first introduced the term IA to the scientific literature in 1995 [1]. In 1994, psychologist K. Yang designed a questionnaire to detect the Internet addiction, posted it on the site, and received about 500 responses [2]. According to survey results, approximately 5% of internet users have this kind of problem. While characterizing IA K. Yang shows that people more frequently use chat (37%), multiplayer games (28%), teleconferences (15%), e-mails (13%), other websites (7%), other network services (2%) [1]. IA appears in virtual dating and communication, repeated web-surveys, cybersexuality, financial dependency [3].

Goldberg commented on this problem not from the medical and psychological point of view, but from the social status of the personality [4]. IA leads to negative situations in the lives of people. Moreover, it also causes many social

problems as divorce, financial difficulties, loss of jobs, suicides and so forth [5]. IA does not cover a particular age group, there is no restriction, and it can be represented in any age group and profession. Children and adolescents, as a more sensitive and less experienced people in society, are more vulnerable to this problem. Obviously, a part of the Internet users includes children and teenagers. They are extremely active in the global network. Children surf the Internet, do their homework, play networking games and find relevant engagements [6]. In the age of digital technology, it is common for children to spend time online and access information through the Internet. However, after a while, these activities become habitual and they sit in front of the computer for hours unaware of how time passes. Curiosity towards the network is associated to the psychology, health and behavior of children and adolescents, and attitudes towards their surroundings. The Internet has a serious impact on the lifestyle of these children, radicalizing them.

New methods are being used by secure Internet centers, researchers, etc. to diagnose, assess, and treat Internet addiction. One of the most used methods to measure IA is the Internet addiction test (IAT). [7] uses IAT in his/her research to measure IA among Croatian teenagers. In the research 15-20 years old, 352 teenagers participated. The sample chosen among Croatian high school students observed contingency was higher (.91). It should be noted that the approach applied among random samples, and if done by the automatic method it might give different results.

In another research [8] is investigating addiction to computer technologies and the internet among university students. In the research, the role of smartphones is mentioned among the main factors for IA. But there might be other reasons for IA. In this approach, exact prediction methods were not shown.

According to studies, the rate of IA varies by country and even by province. For example, IA among Taiwanese university students accounts for 15.3%. In Hunan province, China, this figure varies between 2.4% -5.5%, whereas in Shanxi province, China, it accounts for 6.4% [9]. Some countries are taking steps to solve the problem of IA. [10] conducts a survey on children living in Australia. During the period from November 2010 to November 2011, a considerable amount of information was collected within the survey. As a result, based on the monitoring of 400 children aged 9-16 and their families, a real database was created that represented the Internet usage of children.

There are different approaches to the identification of IA in children. The rate of addiction among children is determined by the time they spend on the Internet. Specifies the frequency and duration of the Internet use as a risk factor. The author hopes that if teenagers spend less time online, their risk of being addicted is much lower. Otherwise, the user must have the symptoms of addiction to claim that he/she is addicted. Children and adolescents should be targeted when designing preventive measures on the Internet [11]. Because the highest rate of IA is observed among children and adolescents [12]. Therefore, the preventive programs designed for the use of the Internet within proper time intervals should be implemented in the school environment and should be applied from the elementary classes [11].

IA is described as a serious problem throughout the world. The social solution of this problem is not enough yet. Prevention of the threats posed by the Internet usage and ensuring information security of the users necessitate the development and implementation of new technological tools, techniques and mechanisms [6].

Nowadays, advanced computer technology allows the experts to monitor the users' online behaviors, surfing, activities, and so forth to be analyzed.

Due to there are no exact criteria for the diagnostic methods for internet addiction, it is not diagnosed among international psychiatrists. This research was done in a limited frame; hence it was addressed to the internet addictive group.

The article proposes a new method for the identification and evaluation of IA among children by analyzing their log-files on the Internet. The purpose of this study was to determine the behavior of children and adolescents in the network, walks, time spent, etc. to determine the degree of dependence of children on the Internet by conducting analysis on logs.

This evaluation will be significant to treat, show psychological support and eliminate symptoms of addiction among kids and teenagers. Also, it will help, parents, schools, related people, specialists for their work in this direction.

The sections of the paper are organized as follows: Section II is about the Log file. Section III provides an overview of related work. Section IV provides information on the classification algorithms used in the study. Section V presents the Proposed Weighted Support Vectors Method. Section VI presents the results of experiments conducted to verify the accuracy of the proposed method. Finally, Section VII contains the result.

2. About Log-files

Log files collected on the Web server is the main source of data about the traffic system status, user behavior, system operations, etc. that transmitted via the web-node and contains important information [13]. Log files represent certain processes related to the user activities, behavior sequences and user interests. Log files are generated by network devices, applications, operating systems, programmable or smart devices [14]. Once the log files are collected, processed and structured, they can be analyzed and used to identify the images and anomalies such as network interventions [15, 16]. Log files can be collected in every field: CDN traffic, database queries, server breaks, errors, and so forth. The records generated by computers, networks and other information technology systems are called audit records or system behavior documentation records. Log files analysis is an assessment of log files and used by organizations to prevent different kinds of risks and meet regulatory requirements. Based on the query all log files of

the resources entered into the server were analyzed.

In this way, it is possible to detect the threats by reviewing the accidents recorded by operating systems and devices [17]. Applications of log analysis:

- Responding to security incidents;
- Troubleshooting in systems, computers, networks, and studying user behavior;
- Monitoring the accessibility and quality indicators of system software applications;
- Modeling the customers' interest in any product;
- Modeling the IA in children.

Traditional problems of log analysis:

Large volumes of exponential data storage. Due to the increased size and complexity of distributed systems, the volume of log files can be quite large. As a result, storing huge amount of data in traditional systems becomes impossible. Because storage in traditional systems is limited to a system, and the data is growing rapidly. Therefore, for common methods, analyzing the system log files on a single node is not considered effective.

Processing of data with complex structures (structured, unstructured, semi-structured). The main challenge for traditional methods is to handle the heterogeneity of the data. A single system, however, is not perfect or ineffective to maintain the diversity of these data generated from different sources.

Large amounts of data causes bottleneck at the input of computing node. Even if the system's memory capacity is increased, the transmission capacity of the drive and access rate does not increase analogically. Diagnostics of large volumes of log files generated by large-scale distributed systems at only one node is a time-consuming process.

3.Related Works

A number of VMSTAT, NETSTAT, and other monitoring systems have been developed so far. VMSTAT (Virtual memory statistics) is a computer monitoring system that collects and displays information about operating systems, memory, processes and interruptions [18]. NESTAT is a command line-based network tool that represents the network connections for TCP (Transmission Control Protocol) routing schedules. NESTAT is used to identify the network problems and to determine the amount of traffic as a quality indicator in the network [19]. These systems are not distributed, and although the log file size is large enough, such systems cannot be effectively used for the log files analysis in a distributed environment. Log files analysis is applied for the determination of the number of visits to each IP address, and detection (finding) of the most popular and least popular file on the website, and so on. [20] proposes a MapReduce approach that minimizes the amount of time spent on operation by log files clustering with the use of K -means algorithm. It considers the excessive use of the Internet as the IA [21]. IA is the excessive use of the Internet, which can damage the psychological state, education or profession, as well as social relationships [22]. IA is more common among adolescents. [23] analyses the relevant literature sources published in the field of IA prevention.

South Korean government developed an action plan to prevent the IA. The implementation of this plan is scheduled to start even at preschool age [24]. Intensive Internet access is assessed as a risk factor for the IA. Other risk factors for IA include psychopathological factors (depressive disorders and anxiety and social phobia), obsessive compulsive symptoms, personal symptoms (hyperactivity and impulsiveness, high rate of curiosity and low reward dependence, introversion, low conscience and consent, and high neurotic/low emotional stability, hostility, low self-control and self-regulation), and physiological features (high pulse rate, low body temperature) [25].

[26] performs experimental evaluations using classification algorithms such as Naive Bayes, JRip, ZeroR, J48, RepTree, and WEKA to identify the Internet-addicted people. As a database, 100 questionnaires with 7 attributes are filled in by students from different universities. The classification accuracy is measured by Root Mean-Squared Error (RMSE), Time and Mean Absolute Error. Compared to the other four algorithms, JRip performs best results.

[27] proposes an automated method and machine learning approach to predict the IA using Bayesian network. The Bayesian network model is built using the IAT features as domain knowledge. Using the unique features of Bayesian network, the symptoms that lead to the probability of the IA are identified. The presence or absence of the IA measured by six parameters of the IAT (Salience, Excessive use, Neglect work, Anticipation, Lack of control, Neglect Social Life). According to the results, out of the six parameters Excessive use and Salience plays a significant role in increasing the risk of IA.

[28] assesses the global distribution of the IA using a multinational meta-analysis. The IA test data set is used, which includes 164 distribution rules from 80 reports involving 89281 participants from 31 nations in 7 regions of the world. The highest prevalence of the IA is observed in the Middle East 10.9% and the lowest prevalence is observed in Northern and Western Europe 2.6%. IA is higher in the countries with low life satisfaction index, high pollution index, high traffic hours index and low gross domestic product per capita.

In [29], the authors evaluate the IA by gender. The study uses 84 female and 84 male samples. Two types of data analysis are applied using JASP 0.8.5.1 program: Classical Test Theory Independent Sample T-Test and Bayesian Independent Sample T-Test. The Open Science Framework (osf.io / fmeah) is used as a database.

[30] explores the relationship between the Internet use, IA and excessive Internet use of 1,034 students (aged 18-27) and their dissociative symptoms. Sociodemographic survey (by age, gender, Internet usage rate, etc.) consisting of 14 Articles is prepared. Sociodemographic questionnaire is used when collecting the data about the IA scale, Dissociative Practice and size. Pearson correlation analysis, t-test and χ^2 test are used for data analysis. Excessive use of the Internet is mostly associated with the dissociative symptoms. According to IAS, 9.7% of the sample is estimated to be addicted to the Internet. According to the results of this study, IA is common for the young college students in Turkey.

In [31], IA is discovered using machine learning techniques. The database uses the data of 2,397 Chinese college students (Age: 19.17 ± 0.70 , Male: 64.17%). T-test and 2-linear normalization are used for the feature selection of student. In some researches, one of the ways to end IA is shown family support [32].

4. Classification Algorithms

The main principle of the classification is machine learning-based data generation strategy. Classification refers to predefined classes.

Different methods and algorithms are used for log files analysis [32-34]. The following methods are used in this work: Logistic regression, as all regression analysis, is a predictive analysis. It is used to describe the data. It is a statistical model that uses a logistic function to model a binary dependent variable as a basic to explain the relationship between one or more nominal, ordinal, interval or relative independent variables.

MLPClassifier (Multi-layer Perceptron, MLP) is a classic type of neural network. One data can be linked to more than one class. They consist of one or more neural layers. The data enters input layer. It may include one or more hidden layers that provide the abstraction level. The forecasts are also performed at the output level. It is beneficial for forecasting issues.

Bernoulli Naive Bayes is for binary features only. A typical example is the classification (in case of Bernoulli) that expresses the probability of a period (in a multinomial case). As a multinomial model, this model is most commonly used in the data classification.

Support vector machine (SVM) is one of the most popular machine learning algorithms. It is a very accurate method for solving classification and regression problems. It is based on the categorization of the pointsspace by optimal hyper line. It can be used to solve both linear and nonlinear practical issues. In this method, the size of the pointsspace is arbitrary and there is no problem of multi-parameter, but as the number of points increases, finding the optimal line for their separation becomes difficult.

A training data set T and N points with input data, $\{(x_i, y_i)\}_{i=1}^N$ $x_i \in R^d$ and output data $y_i \in \{-1, 1\}$ is given. Assume that a set of training data is mapped into a higher dimensional feature space by a nonlinear function $\varphi(x)$, later it is linearly divided. Support vector machine finds a hyperplane $\Phi(x) = \omega^T \varphi(x) + b$ by utilizing the soft margin loss function. The following quadratic programming problem (QPP for short) is solved to obtain the nonlinear SVM:

$$\begin{aligned} \min_{\omega, b, \xi} \quad & \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^N s_i \xi_i \\ \text{s.t.} \quad & y_i ((\omega \bullet \varphi(x_i)) + b) \geq 1 - \xi_i, \quad i = 1, \dots, N, \quad \xi_i \geq 0, \quad i = 1, \dots, N \end{aligned} \quad (1)$$

Here, C denotes a predefined parameter. It weights the compromise the fitting errors and flatness of the classification function, $\xi_i = (\xi_1, \dots, \xi_N)^T$ denotes the slack vectors.

The dual problem of QPP (1) is set as follows by presenting the Lagrange multipliers $\alpha = (\alpha_1, \dots, \alpha_N)$:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j (\varphi(x_i) \bullet \varphi(x_j)) - \sum_{i=1}^N \alpha_i, \\ \text{s.t.} \quad & \sum_{i=1}^N y_i \alpha_i = 0 \\ & 0 \leq \alpha \leq C, \quad i = 1, \dots, N \end{aligned} \quad (2)$$

The unique solution $\alpha^* = (\alpha_1^*, \dots, \alpha_N^*)$ can be achieved effortlessly since it is a convex quadratic programming problem. After resolving the QPP (2), its solution ω^* and the threshold b^* can be completed. In this case, the segmentation hyperplane will be as follows:

$$\Phi(x) = \sum_{i=1}^N \alpha_i^* y_i K(x_i, x) + b^* \quad (3)$$

where, $K(x_i, x) = (\varphi(x_i) \bullet \varphi(x))$ denotes the kernel function, through which the dot product in the high dimensional feature space is given.

The decision function is found as follows after the optimal (ω^*, b^*) is identified:

$$f(x) = \text{sign}(\omega^{*T} \varphi(x) + b^*) = \text{sign}\left(\sum_{i=1}^N \alpha_i^* y_i K(x_i, x) + b^*\right) \quad (4)$$

5. Proposed Weighted Support Vectors Method

Standard SVM model takes the importance rate of all training equally. However, in the case of unbalanced training data, the classification effect of the method is low when using the SVM.

This article presents an approach based on weighted SVM to give different weights to separate training data. The weighted SVM method was first proposed in [35, 36]. The main idea of this method is to assign different weights in accordance with the relative importance of each data point in the class, which is important for the study of the decision surface of different data points.

WSVM (Weighted Support Vectors Method) complicates the penalties to reduce the effect of less important data points (for example, noise and external noise). The constrained optimization problem is designed as follows

Assume that $S = \{(x_i, y_i, s_i)\}_{i=1}^N$ is labeled training data. Here, x_i is the i -th training data, y_i - corresponding sign of data class, and s_i - weight of corresponding class. The weighted optimization problem statement in SVM is as follows:

$$\min_{\omega, b, \xi} \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^N s_i \xi_i \quad (5)$$

Restrictions:

$$y_i (\omega \cdot \varphi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, \dots, N$$

$$\xi_i \geq 0, \quad i = 1, \dots, N$$

6. Experiments

Experiments are implemented on the Internet users' log files associated with the IA [37]. The database used contains 105 features (gender, age, psychological state, education level, time spent on the Internet, etc.). The experiments presented are based on the Internet Time setting. The Internet Time setting here shows how long the children spend on the Internet. Database categories the addiction rate of the children for one week, and the maximum presence period on the Internet denoted as a "highly addicted" class is between 51 and 120 hours. These categories are defined as:

- normal - 1-15 hours;
- low addiction - 16-35 hours;
- medium addiction - 36-50 hours;
- high addiction - 51-120 hours.

Experiments are performed in Python software package. BernoulliNB, Logistic regression, MLPClassifier, SVM and Weighted SVM machine learning techniques are used in the experiments. In the suggested drawn SVM model important class points were valued with higher numbers, non-important class numbers were valued with lower numbers. Valuation of the points means; training algorithm classifies the valued points more exactly. In the experiment for the (Normal (0), Low Addiction (1), Medium Addiction (2), High Addiction (3)) classes value was indexed as a 0:3, 1:5, 2:80, 3:9 accordingly. Methods are evaluated for Accuracy, Precision, Recall, F1-score and Average metrics. The results obtained from the comparative analysis of the classifiers are presented in Table 1.

Table 1. Evaluation of the effectiveness of classification algorithms

Metrics Classifiers	Child addiction class	Accuracy	Precision	Recall	F1-score	Average
BernoulliNB	Normal	0.87	0.87	0.83	0.85	0.86
	Low Addiction	0.31	0.62	0.31	0.41	0.41
	Medium Addiction	0.51	0.10	0.50	0.17	0.32
	High Addiction	0.67	0.33	0.67	0.44	0.53
Logistic regression	Normal	0.94	0.97	0.94	0.95	0.95
	Low Addiction	0.79	0.69	0.79	0.73	0.75
	Medium Addiction	0.17	0.17	0.17	0.17	0.17
	High Addiction	0.67	0.80	0.67	0.73	0.72
MLPClassifier	Normal	0.96	0.98	0.96	0.97	0.97
	Low Addiction	0.88	0.86	0.88	0.87	0.87
	Medium Addiction	0.67	0.80	0.67	0.73	0.72
	High Addiction	1.01	0.75	1.00	0.86	0.90
SVM	Normal	0.98	0.96	0.98	0.97	0.97
	Low Addiction	0.81	0.81	0.81	0.81	0.81
	Medium Addiction	0.17	0.33	0.17	0.22	0.22
	High Addiction	1.01	0.86	1.00	0.92	0.95
Weighted SVM	Normal	0.99	0.99	0.99	0.99	0.99
	Low Addiction	0.98	1.00	0.98	0.99	0.99
	Medium Addiction	0.83	1.00	0.83	0.91	0.99
	High Addiction	1.00	0.75	1.00	0.86	0.99

According to Table 2, the weighted SVM method performs more accurate predictions than others. It recognized 217 points out of 220, and cannot recognize only 3 points.

Table 2. Accuracy of data points recognition in the database

Classifier	Correct predictions	False predictions	Total number of points
BernoulliNB	157	63	220
Logistic regression	194	26	
MLPClassifier	207	13	
SVM	203	17	
Proposed Weighted SVM	217	3	

The results of the comparative analysis of the effectiveness of the methods on the IA database are visually depicted in Fig.1.

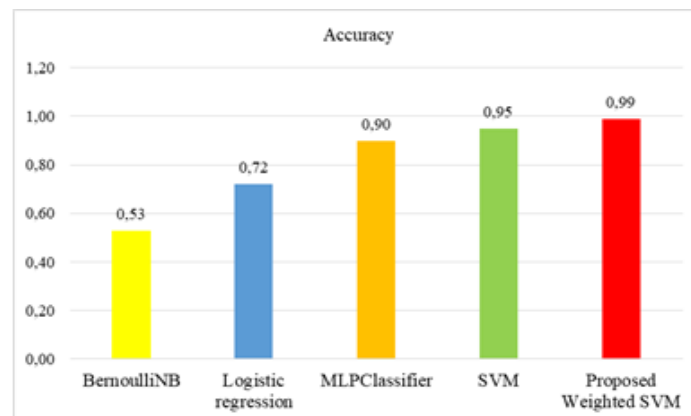


Fig. 1. Comparison of Method Efficiency on the IA Database

As shown in Fig.1. the MLPClassifier and SVM classifiers perform better than the other methods. Thus, the BernoulliNB classifier is able to classify the "High Addiction" class with an accuracy of 0.53, Logistic Regression Classifier - 0.72, MLPClassifier - 0.90, SVM - 0.95, and Weighted SVM classifier - 0.99.

7. Conclusion

Modern children are growing up in a digital environment and they are significantly interested in technology. The excessive use of the Internet and becoming addicted to it has led to serious problems related to the lives and health of children, adolescents and young people.

The purpose of categorizing children's Internet dependence in different classes is to enable appropriate safeguards to be identified for each classified class.

This article used machine-learning techniques to identify IA by analyzing the log files of children and adolescents. Five different classification algorithms were explored during the experiments. Among these algorithms, the Weighted SVM classifier performed higher accuracy when determining the Internet addiction. As a result, this classifier was able to classify the points with the accuracy of 0.99, showing the best recognition.

Acquired results in the research let us determine kids' and teenager's IA categories and make predictions for their psychological situations. To determine kids and teenager's IA is also providing higher efficiency for teachers and other centers working with kids (safe internet centers, state security organizations, child security organizations, and so on).

References

- [1] Alguliyev R.M. & Mahmudov R.Sh., "Internet addiction and issues on the struggle against it", *Problems of Information Technologies*, No.2, pp.7-18, 2012.
- [2] Young K.S. & Rodgers R.C., "The Relationship Between Depression and Internet Addiction", *CyberPsychology & Behavior*, vol.1, No.1, pp. 25-28, 1998.
- [3] Young K.S., "Internet addiction: Evaluation and treatment," *Student British Medical Journal*, 1999, 7, 351-393.
- [4] Young K.S., "Internet addiction: The emergence of a new clinical disorder", *Cyberpsychology & Behavior*, vol.1, No.3, pp.237-244, 1998.
- [5] Whang LS-M., Lee, S., & Chang, G., "Internet over-users' psychological profiles: a behavior sampling analysis on Internet addiction", *CyberPsychology & Behavior*, vol.6, No.2, pp.143-150, 2003.
- [6] Alguliyev, R.M., & Ojagverdieva, S.S., "Conceptual Model of National Intellectual System for Children Safety in Internet Environment", *International Journal of Computer Network and Information Security (IJCNIS)*, vol.11, No.3, pp.40-47, 2019.
- [7] Černja, I., Vejmelka, L., & Rajter, M., Internet addiction test: Croatian preliminary study. *BMC Psychiatry* 19, 388 (2019).
- [8] Mahbobor R., "Addiction of Information and Communication Technology (ICT) and Internet by the Bangladeshi University Students and Its Impact on Their Future", *International Journal of Information Technology and Computer Science(IJITCS)*, vol.10, No.8, pp.56-68, 2018.
- [9] Mei, S., Yau, Y.H., Chai, J., Guo, J., & Potenza, M.N., "Problematic internet use, well-being, self-esteem and self-control: Data from a high-school survey in China", *Addictive Behaviors*, vol.61, pp.74-79, October 2016.
- [10] Green, L., Brady, D., Ólafsson, K., Hartley, J., & Lumby, C., "Risks and safety for Australian children on the Internet", *Cultural science*, vol. 4, No. 1, p.1-75, 2011.
- [11] Jang, M.H., & Ji, E.S., "Gender differences in associations between parental problem drinking and early adolescents", *Internet addiction. Journal for Specialists in Pediatric Nursing*, vol.17, No.4, pp.288-300, 2012.
- [12] Šmahel, D., Vondráčková, P.L., Blinka, S., & Etcheverry, G., "Comparing addictive behavior on the Internet in the Czech Republic, Chile and Sweden", *World Wide Internet: Changing societies, economies and cultures, China: University of Macau*, pp.544-582, 2009.
- [13] Hemant, H., & Rasika, I., "An approach for MapReduce based Log analysis using Hadoop", *In IEEE Sponsored 2'nd International Conference On Electronics And Communication Systems*, IEEE, pp.1264-1268, 2015.
- [14] Qin, T., Gao, Y.L., Wei, L.Y., Liu, Z.L., & Wang, C.X., "Potential threats mining methods based on correlation analysis of multi-type logs", *IET Networks*, vol.7, pp. 299-305, 2018.
- [15] Azizi, Y., Azizi, M., & M.Elboukhari, "Log files Analysis Using MapReduce to Improve Security", *Procedia Computer Science*, vol.148, pp. 37-44, 2019.
- [16] Fu, Q., Lou, J.G., Wang, Y., & Li, J., "Execution Anomaly Detection in Distributed Systems through Unstructured Log Analysis", *ICDM '09: Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, In ninth IEEE International Conference on Data Mining (ICDM)*, pp.149-158, December 2009.
- [17] Qin T., Gao Y.L., Wei L.Y., Liu Z.L., & Wang, C.X., Potential threats mining methods based on correlation analysis of multi-type logs/IET NETWORKS, 2018, vol. 7, pp. 299-305.
- [18] Displaying Virtual Memory Statistics (vmstat), 2010, <https://docs.oracle.com/cd/E19455-01/805-7229/6j6q8svh5/index.html>
- [19] Vigna, G. & Kemmerer, R.A., "NetSTAT: A network-based intrusion detection approach", *In 14th Annual Computer Security Applications Conference (Cat. No.98EX217)*, IEEE., pp. 1-10, 1998.
- [20] Cao, N., Qiao, G., Liu, Y., & Pan, W., "System anomaly detection in distributed systems through MapReduce-based Log Analysis", *In 3rd International conference on advanced computer theory and engineering (ICACTE)*, IEEE, pp.410-413, 2010.
- [21] Vondráčková, P., & R.Gabriel k, "Prevention of Internet addiction: A systematic review", *Journal of Behavioral Addictions*, vol.5, No.4, pp.568-579, 2016.

- [22] Beard, K.W., & Wolf, E.M., "Modification in the proposed diagnostic criteria for Internet addiction", *Cyber-Psychology & Behavior*, vol.4, No.3, pp. 377-383, 2001.
- [23] Romano, J.L., "Prevention in the twenty-first century: Promoting health and well-being in education and psychology", *Asia Pacific Education Review*, vol.15, No.3, pp. 417-426, 2014.
- [24] Kuss, D.J., Rooij, A.J., Shorter, G.W., Griffiths, M.D., & Mheen, D., "Internet addiction in adolescents: Prevalence and risk factors", *Computers in Human Behavior*, vol.29, Issue 5, pp 1987-1996, 2013.
- [25] Nandhini, C., & Krishnaveni, K., "Evaluation of Internet Addiction Disorder among Students", *Indian Journal of Science and Technology*, vol.9, Issue: 19, pp.1-5, 2016.
- [26] Singh, A., & Babbar, S., "Detecting Internet Addiction Disorder Using Bayesian Networks", In: Panda B., Sharma S., Roy N. (eds) *Data Science and Analytics. REDSET 2017. Communications in Computer and Information Science*, vol.799, pp.80-95, 2018, Springer, Singapore.
- [27] Cheng, C., & Li, A.Y., "Internet Addiction Prevalence and Quality of (Real) Life: A Meta-Analysis of 31 Nations Across Seven World Regions", *Cyberpsychology, Behavior, And Social Networking*, vol.17, No. 12, pp.755-760, 2014.
- [28] Ifdil, I., Putri, Y.E., Fadli, R.P., Erwindi, L., Suranata, K., Ardi, Z., Fitria, L., Churnia, E., Zola, N., Barriyah, K., & Rangka, I.B., Measuring internet addiction: comparative studies based on gender using Bayesian analysis, *Journal of Physics: Conf. Series*, (1114), pp.1-8, 2018,
- [29] Canan, F., Ataoglu, A., Ozcetin, A., & Icmeli, C., "The association between Internet addiction and dissociation among Turkish college students", *Comprehensive Psychiatry*, vol.53, №5, pp.422-426, 2012.
- [30] Di, Z., Gong, X., Shi, J., Ahmed, H.O., & Nandi, A.K., "Internet addiction disorder detection of Chinese college students using several personality questionnaire data and support vector machine", *Addictive Behaviors Reports*, vol.10, pp.1-9, 2019,
- [31] Chen, C., & Lee, H., (2013). Discussion on Adolescent Internet Addiction Counseling Strategies through DEMATEL. *I. J. Modern Education and Computer Science*, 6, 9–16.
- [32] Rushikesh P., "Support Vector Machines (SVM) — An Overview", <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>, 2018.
- [33] Tolles, J., & Meurer, W., "Logistic Regression Relating Patient Characteristics to Outcomes", *JAMA*, vol. 316, № 5, pp.533-534, 2016.
- [34] Sebastian R., "Naive Bayes and Text Classification I - Introduction and Theory", (this version, v4), <https://arxiv.org/abs/1410.5329>, 2017.
- [35] Batuwita, R., & Palade, V., "FSVM-CIL: Fuzzy Support Vector Machines for Class Imbalance Learning", *IEEE Trans. Fuzzy Syst.*, vol. 18, №3, pp. 558–571, Jun. 2010.
- [36] He, H., & Ma, Y., "Imbalanced Learning: Foundations, Algorithms, and Applications. Hoboken", NJ, USA: John Wiley & Sons, Inc., 2013.
- [37] Wu, C., Lee, M., Liao, S., & Chang, L., "Risk Factors of Internet Addiction among Internet Users: An Online Questionnaire Survey", *Plos one*, vol.10, No.10, pp. 1-10, 2015.

Authors' Profiles



Rasim M. Alguliyev is a Director of the Institute of Information Technology of the Azerbaijan National Academy of Sciences (ANAS) and the Academician-Secretary of the ANAS. He is a full member of ANAS and a Full Professor. He received his BSc and MSc in Electronic Computing Machines from the Azerbaijan Technical University in 1979. He received his PhD and Doctor of Science (higher degree after PhD) in Computer Science in 1995 and 2003, respectively. His research interests include information security, e-government, data mining, big data, online social network analysis, cloud computing, evolutionary and swarm computation and scientometrics. He is the author of more than 600 papers, four monographs, four patents and several books.



Fargana J. Abdullayeva graduated from the 'Computer Techniques and Technologies' faculty of the Sumgait State University in 2002. In 2003, she was accepted for employment at the Institute of Information Technology of the ANAS, where she addressed issues of information security provision. In 2004, as a respondent, she commenced his work on the subject of 'Development of methods and algorithms for providing information security of population and migration system'. In 2010, defense of the thesis took place. At present, she conducts researches on cloud computing security, anomaly detection, load balancing and task scheduling.



Sabira S. Ojagverdiyeva graduated from Applied Mathematics faculty of Baku State University (BSU). Since the same year, she began working Institute of Information Technology of ANAS. Her area of interest includes information security, on Internet child protection, data sanitization and Data Mining technologies. She carries out scientific research on "Ensuring children's safety on Internet" in the field of information security.

How to cite this paper: Rasim M. Alguliyev, Fargana J. Abdullayeva, Sabira S. Ojagverdiyeva, " Log-File Analysis to Identify Internet-addiction in Children", International Journal of Modern Education and Computer Science(IJMECS), Vol.13, No.5, pp. 23-31, 2021.DOI: 10.5815/ijmecs.2021.05.03