

Uşaqların fərdi məlumatları əsasında İnternet asılılığın müəyyən olunması üsulu

Fərqanə Abdullayeva¹, Sabirə Ocaqverdiyeva²

^{1,2}AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
a_farqana@mail.ru¹, allahverdiyevsabira@gmail.com²

Xülasə— Məqalədə İnternet mühitində uşaqların fəaliyyətləri ilə əlaqədar təhlillər aparılmışdır. Loq-fayllar və İnternet asılılıq problemi araşdırılmışdır. Uşaq və yeniyetmələrin İnternet asılılığının dərəcələrinin müəyyən olunması üçün açıq verilənlər bazası üzərində məşin təlimi üsulları tətbiq edilmişdir. Üsulların effektivliyi qiymətləndirmə metrikaları ilə analiz edilib.

Açar sözlər— loq-fayllar; uşaq və yeniyetmələr; İnternet asılılıq; məşin təlimi

I. GİRİŞ

İnformasiya cəmiyyətinin formalaşdırılması, İnternet şəbəkəsinin sürətli inkişafı və insanlara təqdim etdiyi xidmətlərin kifayət qədər populyarlıq qazanması informasiya fəzalarında böyük həcmli məlumatın formalaşmasına səbəb olmuşdur. Bu isə informasiya təhlükəsizliyi ilə bağlı müəyyən problemlər yaradır. Həmin problemlərin həlli və təhlükəsizliyin təmin olunması üçün yeni metodların, mexanizmlərin işlənməsi və tətbiqinə ehtiyac vardır[1].

Kompüter texnologiyalarının inkişafı istifadəçilərin şəbəkədə olan davranışları və onların fəaliyyətləri ilə əlaqədar müxtəlif təhlillərin aparılmasına imkan yaradır.

İnternet istifadəçilərinin bir hissəsi uşaq və yeniyetmələrdir. Uşaqlar virtual mühitdə aktivdirlər. Müasir şəraitdə informasiya texnologiyalarının inkişaf etdiyi bir dövrdə sosiallaşma amili kimi İnternet şəbəkəsi əlverişli vasitədir. Eyni zamanda istifadəçi maraqlarını izah etmək, texniki bacarıqları inkişaf etdirmək və özünü ifadə etməyin yeni formaları ilə təcrübə aparmaq üçün bir platformadır. İnternetdən istifadənin faydalı tərəfləri olduğu kimi, informasiya bolluğu şəraitində mövcud olan təhlükələrdən xəbərsiz uşaq və yeniyetmələrin bilməyərək müxtəlif təhlükələrlə rastlaşmasına səbəb ola bilər [2].

Hal-hazırda loq məlumatları analiz etməklə bir sıra məsələlər həll edilir. Şəbəkə aktivliyinin yoxlanması, resursların və haker proqramlarının izlənməsi, təhdidlərin aşkarlanması üçün metod və hesablama vasitələrinin (alqoritm və proqramlarının) yaradılmasına xüsusi diqqət yetirilir. Loq-faylların analizi informasiya təhlükəsizliyinin təmin olunması sahəsində effektiv həllərin, metod və mexanizmlərin işlənilməsinə imkan verir.

Məqalədə uşaq və yeniyetmələrin loq-fayllarını analiz etməklə, onların İnternet mühitində fəaliyyətləri ilə əlaqədar təhlillər aparılır və şəbəkədən asılılıq dərəcəsi müəyyən edilir.

II. İNTERNET ASILILIĞI

İnternet asılılığı fenomeninin psixoloji problem kimi elmi cəhətdən öyrənilməsi iki amerika mütəxəssisi – psixiatr İ.Qoldberq və kliniki psixoloq K.Yanqın adı ilə bağlıdır. İnternet-asılılıq termini ilk dəfə elmi ədəbiyyata 1996-cı ildə İ.Qoldberq tərəfindən daxil edilmişdir [3].

Psixoloqlar, həm də elmi tədqiqatla məşğul olan digər peşə sahibləri tərəfindən araşdırılan bir problemdir. İnternet asılılıq sindromu insanın informasiyaya olan həvəs və tələbatından yaranır. İnternet məlumatla zəngin bir mühit olduğundan, istifadəçi daha çox informasiya əldə edir, paylaşır və nəticə olaraq özünün informasiyaya olan tələbatını ödəyir.

İnternet asılılıq (ing. Internet addiction) – şəxsin psixoloji durumuna, təhsil və ya peşəsinə, o cümlədən sosial əlaqələrinə ziyan vurulması ilə nəticələnən İnternetdən həddindən artıq istifadə olunmasıdır [4]. Uşaq və yeniyetmələr vaxtlarının müəyyən hissəsini İnternetdə keçirir, axtarış edir, şəbəkə oyunları oynayır, dərslərini hazırlayırlar. Ətrafda baş verənləri öyrənmə həvəsi, şəbəkə oyunlarına olan maraqları onların saatlarla İnternetdən istifadə etməsinə səbəb olur. Araşdırmalara görə İnternet asılılıq yeniyetmələr arasında daha çox müşahidə olunur [5].

Uşaqların İnternet asılılıq sindromunun qarşısını almaq üçün bəzi ölkələrdə müxtəlif tədbirlər həyata keçirilir. Məs., Cənubi Koreya dövləti İnternetdən asılılıq üzrə qarşılıqlı tədbirləri həyata keçirir və hətta bu məqsədlə məktəbəqədər uşaqlardan başlayaraq plan hazırlamışdır [6].

III. LOQ ANALİZİN TƏTBİQLƏRİ VƏ PROBLEMLƏRİ

Veb-serverdə toplanan qeydiyyat faylları (ing. log files) Veb-qovşaqdan keçən trafik və istifadəçilərin davranışları haqqında əsas informasiya mənbəyi hesab olunur və əhəmiyyətli məlumatları özündə ehtiva edir. Loq-fayllar istifadəçilərin fəaliyyətləri ilə bağlı müəyyən prosesləri, hərəkət ardıcılıqlarını və istifadəçi maraqlarını ifadə edir. Loq-

faylların təhlili əməliyyat sistemləri və qurğular tərəfindən qeyd alınan hadisələri nəzərdən keçirməklə təhdidləri aşkarlamaq üçün effektiv bir yoldur [7]. Loq sistemin vəziyyəti, istifadəçi davranışı, sistem əməliyyatları və s. ilə bağlı başlıca informasiya mənbəyidir [8].

Kompüterlərin, şəbəkələrin və başqa İT sistemlərin generasiya etdiyi yazılar audit yazıları və ya sistemin davranışını sənədləşdirən yazılar adlanır. Loq-faylların analizi – loq yazıların qiymətləndirilməsidir. Təşkilatlar tərəfindən müxtəlif rüblərin qarşısının alınması və normativ tələblərə cavab vermək üçün istifadə edilir.

Loq analizinin tətbiqləri aşağıdakılardır:

- təhlükəsizlik insidentlərinə cavabvermə;
- sistemlərdə, kompüterlərdə, şəbəkələrdə baş verən problemlərin aradan qaldırılması, istifadəçilərin davranışının öyrənilməsi;
- sistemin proqram tətbiqlərinin əlçatanlığını və keyfiyyət göstəricilərinin səviyyəsinin monitorinq edilməsi;
- müştərilərin hər hansı bir məhsula olan marağının modelləşdirilməsi;
- uşaqların internetdən asılılığının modelləşdirilməsi.

Loq-fayllarda olan məlumatları analiz edərək obyektlərin parametrlərinə (vaxt, məzmun və s.) görə klassifikasiya etməklə uşaqların İnternetdən nə dərəcədə asılı olması müəyyən edilə bilər və onların valideynləri, müəllimləri bu haqda məlumatlandırıla bilərlər. Təqdim olunan işdə loq-faylların analizinin məqsədi uşaqların İnternetdən asılılıq dərəcəsini müəyyən etməkdir.

IV. KLASSİFİKATORLAR

Loq-faylların analizində müxtəlif metod və alqoritmlərdən istifadə olunur. İşdə aşağıdakı metodlar tətbiq edilmişdir [9-12]:

Dayaq vektorlar metodu (Support vector machine, SVM) – təsnifat və reqressiya problemləri üçün çox dəqiq və məşhur metoddur, nöqtələr fəzasını optimal hiperxətt vasitəsilə kateqoriyalara ayırmağa əsaslanır. Xətti və qeyri-xətti, bir çox praktik məsələlərin həllində istifadə oluna bilər. Bu üsuldə əlamət fəzasının ölçüsü sərbəstdir və çoxparametrlilik problemi yoxdur, lakin nöqtələrin sayı artdıqca onları ayırmaq üçün optimal xəttin tapılması mürəkkəbləşir.

Loqistik reqressiya (Logistic regression) – bütün reqressiya analizləri kimi, loqistik reqressiya proqnozlaşdırıcı bir təhlildir. Loqistik reqressiya məlumatları təsvir etmək üçün istifadə olunur. Bir asılı ikili dəyişən və bir və ya daha çox nominal, nizamlı, interval və ya nisbət səviyyəsində sərbəst dəyişənlər arasındakı münasibətləri izah etmək üçün əsas şəkildə ikili asılı dəyişən modelləşdirmək üçün loqistik bir funksiyadan istifadə edən statistik bir modeldir.

MLPClassifier (Multi-layer Perceptron, MLP) – neyron şəbəkənin klassik bir növüdür. Bir məlumat birdən çox siniflə

əlaqələndirilə bilər. Onlar bir və ya daha çox neyron təbəqədən ibarətdir. Məlumatlar giriş qatına daxil olur. Abstraksiya səviyyəsini təmin edən bir və ya daha çox gizli təbəqədən ibarət ola bilər. Proqnozlar çıxış səviyyəsində də aparılır. Proqnozlaşdırma məsələlərində istifadəsi əlverişlidir.

BernoulliNB (Bernoulli Naive Bayes) – bu metod yalnız ikili xüsusiyyətlər üçündür. Bunun üçün tipik bir nümunə (Bernoulli vəziyyətində) və ya bir müddətin ehtimalını ifadə edən (çoxnomial vəziyyətdə) təsnifatıdır. Multinomial model kimi, bu model verilənlərin təsnifatlaşdırılmasında populyardır.

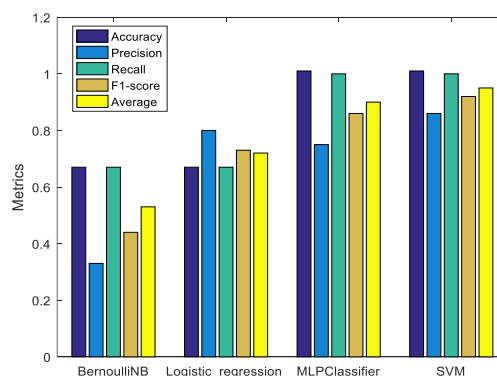
Yuxarıda qeyd olunan üsullar vasitəsilə müxtəlif eksperimentlər aparılmış və bu üsulların üstünlükləri və çatışmazlıqları aşkarlanmışdır.

V. EKSPERİMENTLƏR

Eksperimentlər İnternet istifadəçilərinin İnternet asılılığı ilə bağlı toplanmış loq-faylları üzərində aparılmışdır [13]. İstifadə olunan verilənlər bazası 105 əlamətdən (cins, yaş, psixoloji vəziyyət, təhsil səviyyəsi, İnternetdə keçirilən vaxt və s.) ibarətdir. Təqdim edilən eksperimentlər “İnternet Time” parametrinə əsasən aparılmışdır. Burada “İnternet Time” parametri uşaqların İnternetdə keçirdiyi vaxtın müddətini göstərir. Verilənlər bazasında uşaqların bir həftə ərzində İnternetdən asılılıq kateqoriyaları təyin edilmiş və “çox asılı” sinif kimi İnternetdə ən çox fəaliyyət müddəti 51-120 saat intervalı götürülmüşdür. Bu kateqoriyalar aşağıdakı kimi müəyyən edilmişdir:

- normal – 1-15 saat;
- az asılı – 16-35 saat;
- orta asılı – 36-50 saat;
- çox asılı – 51-120 saat;

Eksperimentlər Python proqram paketində aparılmışdır. Eksperimentlərin aparılmasında BernoulliNB, Loqistik reqressiya, MLPClassifier, və SVM maşın təlimi üsulları istifadə edilmişdir. Metodlar Accuracy, Precision, Recall, F1-score və Average metrikaları əsasında qiymətləndirilmişdir. Metodların effektivliyinin İnternet asılılıq verilənlər bazası üzərində müqayisəli analizinin nəticələri vizual olaraq Şəkil 1-də verilmişdir.



Şəkil 1. Metodların effektivliyinin İnternet asılılıq verilənlər bazası üzərində müqayisəsi

Şəkil 1-dən göründüyü kimi MLPClassifier və SVM klassifikatorları digər metodlarla müqayisədə daha yaxşı nəticələr göstərmişdir. Belə ki, BernoulliNB klassifikatoru “Çox asılı” sinfini 0.53, Loqistik reqressiya klassifikatoru 0.72, MLPClassifier 0.90, SVM 0.95 dəqiqliyi ilə klassifikasiya edə bilmişdir.

NƏTİCƏ

İnformasiya bolluğu şəraitində və Big data probleminin olduğu bir dövrdə uşaqların İnternetdə mövcud olan təhlükələrdən qorunması vacibdir. İnformasiya təhlükəsizliyinin vəziyyətinin monitorinqinin (statistikasının) aparılması üçün loq-faylların analizi və səmərəli qərarların qəbul edilməsi çox əhəmiyyətlidir.

Təklif olunan yanaşma İnternet asılılığının təyində və İnternet təhlükələrindən qorunma problemlərinin həll edilməsində faydalı ola bilər. Loq-faylları analiz etməklə, İnternet asılılığının təyini və oradan alınan qiymətləndirmələr təhsil müəsisələri, valideynlər və ya əlaqədar şəxslərin müəyyən qərarların qəbulu zamanı faydalı ola bilər.

İSTİNADLAR

- [1] F. Yusifov, “Loq-faylların analizi əsasında informasiya təhlükəsizliyinin təmin edilməsi”, İnformasiya texnologiyaları problemləri, №1(7), 31-37, 2013.
- [2] R. Alguliyev and S. Ojagverdiyeva, “Conceptual Model of National Intellectual System for Children Safety in Internet Environment”, International Journal of Computer Network and Information Security(IJCNIS), № 3, pp. 40-47, 2019.
- [3] R. Əliquliyev, R. Mahmudov, “İnternet asılılığı və onunla mübarizə məsələləri”, İnformasiya cəmiyyəti problemləri, № 2, s. 7-18, 2012.
- [4] K. Beard and E.Wolf, “Modification in the proposed diagnostic criteria for Internet addiction”, Cyber-Psychology & Behavior, 4(3), pp.377–383, 2001.
- [5] D. Kuss, M.Griffiths, L.Karila and J. Billieux, “Internet addiction: A systematic review of epidemiological research for the last decade”, Current Pharmaceutical Design, 20(25), pp.4026–4052, 2014.

- [6] J.Romano, “Prevention in the twenty-first century: Promoting health and well-being in education and psychology”, Asia Pacific Education Review, 15(3), pp. 417–426, 2014.
- [7] T.Qin, YL. Gao, LY. Wei, ZL. Liu and CX.Wang, “Potential threats mining methods based on correlation analysis of multi-type logs”, IET Networks, vol.7, pp. 299-305, 2018.
- [8] H.Hemant and İ. Rasika, “An approach for MapReduce based Log analysis using Hadoop”, IEEE Sponsored 2’nd international conference on electronics and communication systems, pp.1264- 1268, 2015.
- [9] P.Rushikesh, “Support Vector Machines(SVM) — An Overview”, <https://towardsdatascience.com/https-medium-com-pupalrushikesh-svm-f4b42800e989>, 2018.
- [10] J.Tolles and W.Meurer, “Logistic Regression Relating Patient Characteristics to Outcomes”, JAMA, vol. 316, № 5, pp.533-534, 2016.
- [11] Когда использовать нейронные сети MLP, CNN и RNN. <https://www.machinelearningmastery.ru/when-to-use-mlp-cnn-and-rnn-neural-networks>, 2018.
- [12] R.Sebastian, “Naive Bayes and Text Classification I - Introduction and Theory”, (this version, v4), <https://arxiv.org/abs/1410.5329>, 2017.
- [13] C. Wu, M. Lee, S.Liao and L.Chang, “Risk Factors of Internet Addiction among Internet Users: An Online Questionnaire Survey”. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4603790/> (Data set: “20150610_Internet Addiction dataset.sav”), 2015.

AN IDENTIFICATION METHOD OF INTERNET ADDICTION BASED ON CHILDREN’S PERSONAL DATA

Fargana Abdullayeva¹, Sabira Ojagverdiyeva²

^{1,2}Institute of Information Technology of ANAS, Baku, Azerbaijan

¹a_fargana@mail.ru, ²allahverdiyevabasira@gmail.com

Abstract— The analysis on the activities of children in the Internet environment have been conducted in the article. Log files and the Internet addiction problem have been investigated. Machine learning techniques have been applied on open databases to determine the degree of Internet dependence of children and adolescents. The effectiveness of methods has been analyzed by evaluation metrics.

Keywords— *log files; children and adolescents; Internet addiction; machine learning*