UDC 519.6
DOI: 10.31891/CSIT-2021-4-11

OLHA KUZMENKO, KONSTANTIN GRITSENKO,
HANNA YAROVENKO, OLEKSANDR KUSHNEROV, AVGUSTA HRYTSENKO
Sumy State University, Sumy, Ukraine

PREDICTIVE MODELING OF THE OUTCOMES OF CYBER SPORT MATCHES USING DATA MINING TECHNOLOGIES

The e-sports industry is a highly developed interdisciplinary field in which machine learning and artificial intelligence technologies are widely used. The article examines the application of Data Mining technologies in order to predict the results of e-sports matches. The technique of predictive modeling is considered. Based on the correlation analysis, the input variables for predicting the results of the matches of the online game League of Legends are determined. The SAS Enterprise Miner package builds predictive models in the form of a decision tree, logistic regression and neural network. Based on the analysis of the quality indicators of forecast models, the use of the neural network as a predictor of the results of e-sports matches is substantiated. The purpose of this predictor is to give players advice on how to change the strategy of the game in which they are defeated.

Keywords: e-sports match, online game, forecasting results, decision tree, logistic regression, neural network.

ОЛЬГА КУЗЬМЕНКО, КЩСТЯНТИН ГРИЦЕНКО,
АННА ЯРОВЕНКО, ОЛЕКСАНДР КУШНЕРЬОВ, АВГУСТА ГРИЦЕНКО
Сумський державний університет

ПРОГНОЗНЕ МОДЕЛЮВАННЯ РЕЗУЛЬТАТІВ КІБЕРСПОРТИВНИХ МАТЧІВ ІЗ ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ DATA MINING

Останнім часом в Україні активно розвивається індустрія кіберспорту. Кіберспорт сприяє розвитку розумових здібностей у використанні інформаційних технологій. Кіберспорт – це змагання з комп'ютерних ігор. Вона охоплює велику кількість населення та широкий спектр професій. Індустрія кіберспорту включає не лише гравців, а й розробників ігор, менеджерів команд, організаторів турнірів, маркетологів, стрімінгових компаній, численних спонсорів та державних установ. У багатьох країнах світу кіберспорт має державну підтримку, змагання високого рівня та відповідні освітні програми. У 2020 році кіберспорт був визнаний офіційним видом спорту в Україні, а в серпні 2021 року Україна вперше прийняла Чемпіонат Європи з кіберспорту. Особливо слід відзначити роль штучного інтелекту та технологій машинного навчання у розробці комп'ютерних онлайн-ігор.

Кіберспортивна галузь є високорозвиненою міждисциплінарною сферою, в якій широко використовуються технології машинного навчання та штучного інтелекту. В статті досліджено питання застосування технологій Data Mining з метою прогнозування результатів кіберспортивних матчів. Розглянута методика прогнозного моделювання. На основі кореляційного аналізу визначено вхідні змінні для прогнозування результатів матчів онлайн гри League of Legends. В пакеті SAS Enterprise Miner побудовано прогнозні моделі у вигляді дерева рішень, логістичної регресії та нейронної мережі. На основі аналізу показників якості прогнозних моделей обгрунтовано використання нейронної мережі в складі предиктора результатів кіберспортивних матчів. Метою даного предиктора є надання гравцям порад про те, яким чином змінити стратегію гри, в якій вони зазнають поразки.

Ключові слова: кіберспортивний матч, онлайн гра, прогнозування результатів, дерево рішень, логістична регресія, нейронна мережа.

Introduction. Recently, the e-sports industry has been actively developing in Ukraine. E-sports promotes the development of mental abilities in the use of information technology. E-sports is a competition in computer games. It covers a large population and a wide range of professions. The e-sports industry involves not only players, but also game developers, team managers, tournament organizers, marketers, streaming companies, numerous sponsors, and government agencies. In many countries around the world, e-sports has government support, high-level competitions, and relevant educational programs [1]. In 2020, e-sports was recognized as an official sport in Ukraine, and in August 2021, Ukraine hosted the European E-Sports Championship for the first time. The role of artificial intelligence and machine learning technologies in the development of computer online games should be especially noted.

The purpose of the research is to create a high-quality model for predicting the results of e-sports matches using Data Mining technologies. To achieve this goal, it is necessary to solve some tasks: developing a predictive modeling methodology, building predictive models based on the most common Data Mining technologies: decision trees, logistic regression, a neural network, choosing the best model in terms of quality indicators for the purpose of further use as part of the predictor of the results of e-sports matches.

Related works. The relevance of research in the field of cybergaming with the use of machine learning tools, which includes Data Maning technologies, is confirmed by a significant number of publications. So, for the query "computer gaming and machine learning" in the Scopus database were found 1802 documents published by 4766 scientists over the past five years. Watson B., Sput J., Kim J., Listman J., Kim, S., Wimmer R., Lee B., Putrino D. [0] described the competitiveness of e-sports compared to traditional sports. Sanauja-Paris G., Camacho M., Balado-Albiol M. [7] considered the consolidation of e-sports in a pandemic. Imas Ye., Petrovska T., Hanaha O. [0]

who studied it from the point of view of a cultural phenomenon in Ukraine, can be singled out among domestic scientists engaged in e-sports research. Zhmai O.V. [4] studied the investment attractiveness of this industry in Ukraine. Gorova K.O., Goroviy D.A., Kiporenko O.V. [5] studied the main trends in the e-sports market. Lazneva I.O., Tsaranenko D.I. [6] studied the impact of e-sports on changing the structure of the global computer game market. Despite the current advances in e-sports, a number of unresolved issues remain regarding the specifics of the use of machine learning technologies and artificial intelligence in e-sports, which require further in-depth research.

Summary. The research was conducted on a dataset describing 7620 professional matches from the online computer game League of Legends (LoL), obtained from the analytical resource Kaggle (kaggle.com). League of Legends is a multiplayer online battle arena (MOBA) developed by Riot Games and released in 2009. Today, this computer online game is one of the most popular in the world both in terms of the number of players and the number of views of matches on the Twitch streaming service. In this game, two teams of 5 players each, each controlling a champion, compete to destroy the Nexus object inside the enemy base. A typical map of the League of Legends game is shown in Figure 1.

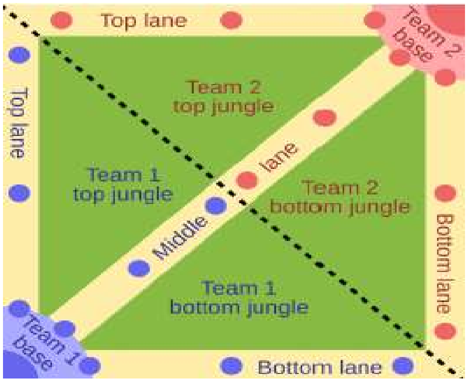

Fig. 1. A typical map of the online computer game League of Legends

The map of the game includes three roads on which the minions (servants) of each team move: Top lane, Middle lane and Bottom lane. Roads on the map are shown in yellow. The game map also includes the jungle, which is depicted in green. They contain powerful monsters that heroes can kill to get gold. Each character has unique skills that can be used against other heroes, minions or monsters. The map also contains towers that protect the team's territory from enemy attacks. In fig. 1 they are depicted with blue and red dots.

The minions regularly move in waves from the base of each team in the direction of the enemy and pass on all three roads until they find the minions of the other team to confront them. Heroes get gold by killing enemies. Gold can be used to buy items that improve the skills of the heroes and give them a competitive advantage. To increase the strength of the heroes, you need to defeat certain characters in the game, such as dragon, baron and herald.

The main goal of the game is to destroy the Nexus object inside the base of each team. But in order for this to be possible, all the towers and the ingibitor on the way to the enemy base must be destroyed in advance. Towers are structures that cause damage to enemies within their range. Inhibitors are structures that, when destroyed by an enemy team, make enemy minions more powerful and allow you to destroy a Nexus object.

When choosing a hero for themselves, players should consider what role he/she will play in the team. Different heroes can play one or more roles, but most of them are better at fulfilling exactly the main role for which they were created. The player must also be aware of the current metagame [8]. Metagames are dictated by the type of heroes, high-quality ones are adequate for each position on the map, as well as the heroes are the most powerful in the game. Due to the constant updates of the game, which are performed with patches, the metagame tends to change. An example of a metagame is the most common team in the *League of Legends*: usually two heroes go to the Bottom lane, one *Support* and one *Attack Damage Carry* (ADC). *Mage* or *Assassin* enters the Middle lane. In the jungle comes *Tank* or *Fighter*. On the upper road, the hero can act as an *Initiator*, *Front-liner* or *Tank*.

The following performance indicators of *blue* and *red* teams were used as input variables of the e-sports match of the online game League of Legends, respectively:

blue_end_gold, red_end_gold. Gold is earned by killing monsters, other players and objects. It can be spent on items that strengthen the hero.

blue_kills, red_kills. Killing players of the opposing team gives extra money and also slows down the progress of the opposing team.

blue_towers, red_towers. Destroying the towers that are placed on each road map, allows your team to move forward.

blue_inhibs, red_inhibs. Destroying the inhibitors that are placed at each team base is necessary for your team to capture the inner towers and win the game.

blue_dragons, red_dragons. Killing a dragon, a powerful monster gives your team gold, which allows you to gain a competitive advantage over another team.

blue_barons, red_barons. Killing a baron, the most powerful monster gives your team a competitive advantage over another team.

blue_heralds, red_heralds. Killing the Rift Herald is the third most powerful monster gives your team gold, which allows you to gain a competitive advantage over another team.

The resulting binary variable *winner* contains the number of the team that won the match (1 is *blue*, 2 is *red*). The results of correlation analysis of input variables using a thermal map are presented in Figure 2.

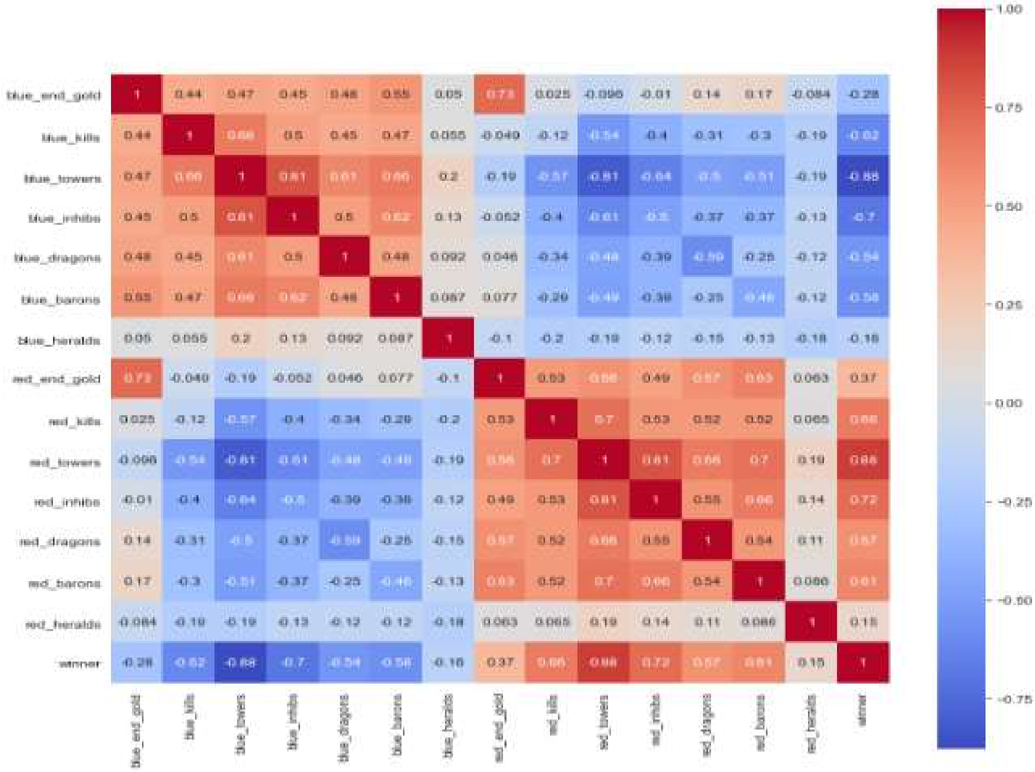

Fig. 2. The results of correlation analysis of input variables

The correlation between the output variable *winner* and the input variables *blue_towers* and *red_towers* is high (0.88). The correlation between the output variable *winner* and the input variables *blue_inhibs* and *red_ inhibs* is also high, but lower (0.72). Thus, destroying the towers in the game is extremely important, even more so than destroying the inhibitors. A fairly high level of correlation is also observed between the original variable winner and variables such as: *blue_kills* and red_kills (0.66); *blue_barons*, *red_barons* (0.61); *blue_dragons*, *red_dragons* (0.57).

For predictive modeling of e-sports results using Data Mining technologies, the SAS Enterprise Miner package was chosen, which is an integrated component of the SAS data mining system, designed to detect in large data sets the information needed for decision-making. To build models, we used the tools of decision trees, regression analysis and neural networks (Figure 3).

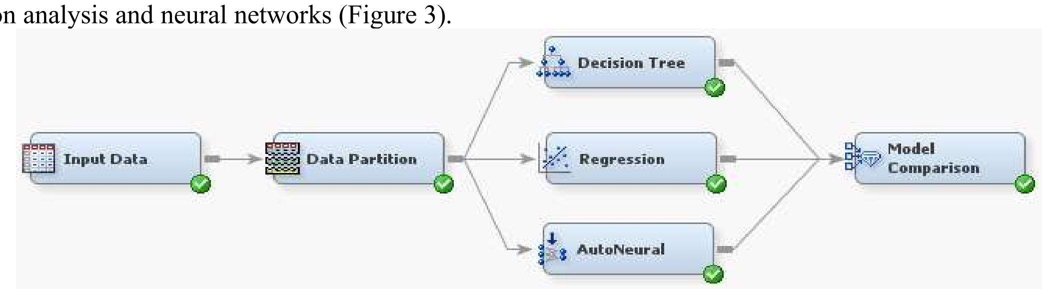

Fig. 3. ETL-diagram of the process of modeling the results of e-sports matches using Data-Mining technologies in the SAS Enterprize Miner package

The *Input Data* node of the ETL process diagram contains a set of input data consisting of 14 interval input variables and 1 binary output variable. The number of observations is 7620. The *Data Partition* node of the ETL process diagram uses the tool Data Partition, with which the entire input data set (100% - 7620 matches) is randomly divided into two parts while maintaining the proportion of response distribution of the target variable (*winner*): 60% (4571) is the training data on which the model is based; 40% (3049) is the validation data, which checks the quality of possible variants of the model specification and selects the best of them. The distribution of data as a percentage is presented in Table 1.

Table 1.

Partitioning the input data set using the Data Partition tool in the SAS Enterprise Minner package

Data	Number of observations with team winning 1 (*blue*)		Number of observations with team winning 2 (*red*)	
	units	%	units	%
Primary	4146	54,41 %	3474	45,59 %
Training	2487	54,41 %	2084	45,59 %
Validation	1659	54,41 %	1390	45,59 %

Thus, the models were built and tested on equivalent data sets.

In the second stage, a *Decision Tree* node, *Regression* node and *AutoNeural* node were built. In the third stage, using the Model Comparison tool, a comparative analysis of the constructed models was performed and the best one was selected. The constructed decision tree with 6 leaves is presented in Figure 4.

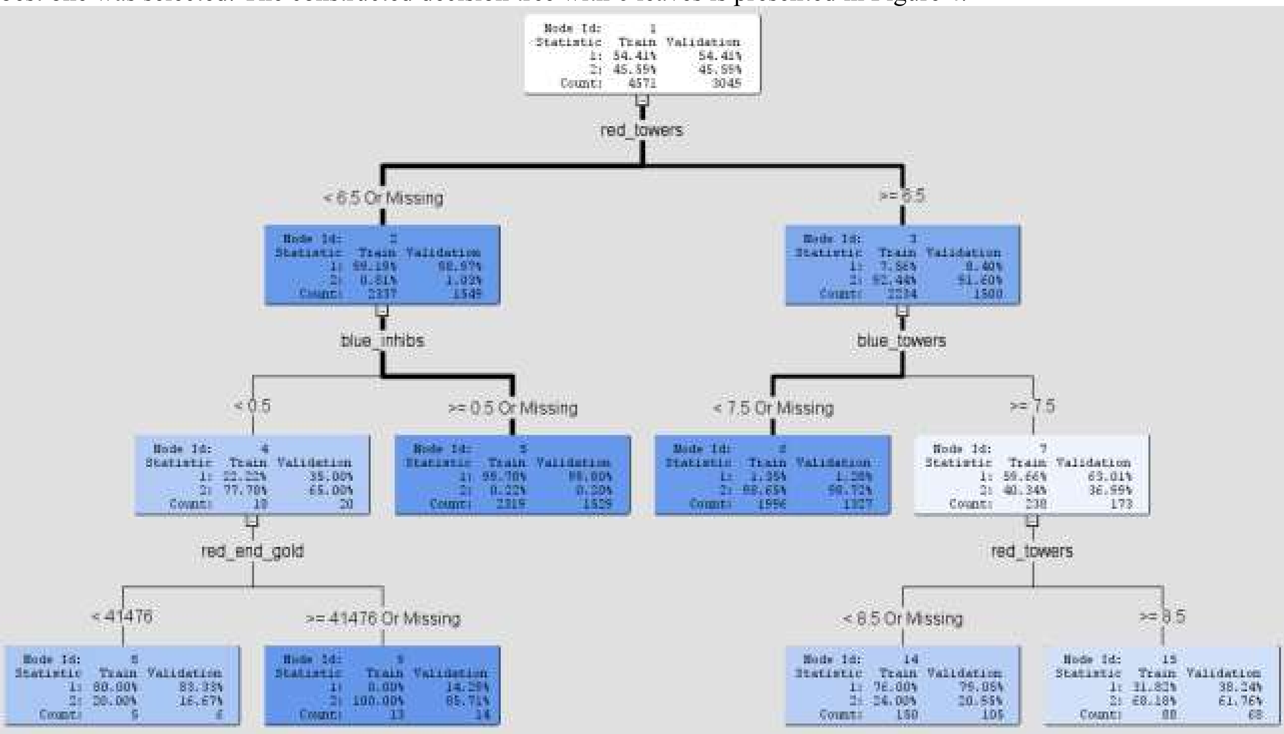

Fig. 4. Screenshot with the results of predictive modeling of the results of e-sports matches using the decision tree in the SAS Enterprize Miner package

The use of input variables in the process of constructing branches corresponds to the degree of their correlation with the output variable (see Figure 2) and is confirmed by the results of assessing their significance in the SAS Enterprise Miner package, presented in Figure 5.

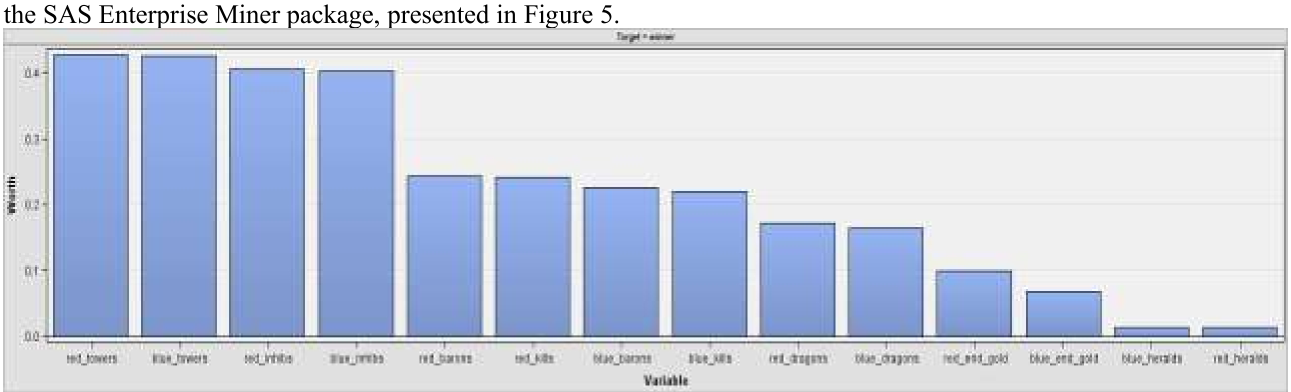

Fig. 5. Screenshot with the results of assessing the significance of input variables in building a decision tree in the SAS Enterprize Miner package

To optimize the model of logistic regression, the method of stepwise exclusion of insignificant factors (*Stepwise*) was chosen, the significance of which was determined by the statistical criterion of Chi-Square. Figure 6 shows the calculated values of the Chi-Square criterion for the selected factors (*Score Chi-Square* column).

Summary of Stepwise Selection

Step	Effect Entered	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Validation Misclassification Rate
1	red_towers	1	1	3583.5358		<.0001	0.0574
2	blue_towers	1	2	503.6451		<.0001	0.0216
3	red_kills	1	3	62.3712		<.0001	0.0223
4	blue_kills	1	4	68.0320		<.0001	0.0167
5	blue_inhibs	1	5	12.5208		0.0004	0.0167
6	blue_heralds	1	6	6.7336		0.0095	0.0151
7	red_inhibs	1	7	5.4592		0.0195	0.0161
8	red_heralds	1	8	4.5606		0.0327	0.0154

Fig. 6. Screenshot with the results of assessing the significance of input variables in the construction of logistic regression in the package SAS Enterprize Miner

Following the *Score Chi-Square* column, the *Pr> Chi-Square* column demonstrates how significant the selected factor is, namely, the smaller the value in it, the more significant the factor in the model. Variables with a calculated value less than 0.0001 have a high statistical significance (in this case, the factors *blue_towers* and *red_towers*, *blue kills* and *red_kills*). The factors *blue_inhibs* and *red_inhibs*, *blue_heralds* and *red_heralds* are also statistically significant (the corresponding calculated value is less than 0.05). Estimates of the odds ratio, showing how the selected variables of the logit model affect the target variable, are presented in Figure 7.

Odds Ratio Estimates

Effect	Point Estimate
blue_heralds	1.896
blue_inhibs	1.528
blue_kills	0.794
blue_towers	0.273
red_heralds	0.505
red_inhibs	0.737
red_kills	1.333
red_towers	3.038

Fig. 7. Screenshot with the odds ratio estimates in the construction of logistic regression in the SAS Enterprize Miner package

When constructing the neural network, the activation function was chosen as the hyperbolic tangent, and the type of architecture with one hidden layer.

The best model was selected on the basis of the Misclassification Rate, the Average Squared Error and the Gini Coefficient. The lowest values of the Misclassification Rate, the Average Squared Error and the highest values of the Gini Coefficient are characterized by the neural network (Figure 8). In second place is the logistic regression. The last place is occupied by the decision tree.

Statistics	AutoNeural	Reg	Tree
Valid: Kolmogorov-Smirnov Statistic	0.97	0.97	0.95
Valid: Average Squared Error	0.01	0.01	0.02
Valid: Roc Index	1.00	1.00	0.99
Valid: Average Error Function	0.04	0.04	.
Valid: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff	0.78	0.82	0.83
Valid: Cumulative Percent Captured Response	21.94	21.94	21.53
Valid: Percent Captured Response	10.94	10.94	10.80
Valid: Divisor for VASE	6098.00	6098.00	6098.00
Valid: Error Function	268.16	263.22	.
Valid: Gain	119.35	119.35	115.23
Valid: Gini Coefficient	1.00	1.00	0.98
Valid: Bin-Based Two-Way Kolmogorov-Smirnov Statistic	0.97	0.96	0.95
Valid: Kolmogorov-Smirnov Probability Cutoff	0.48	0.32	0.24
Valid: Cumulative Lift	2.19	2.19	2.15
Valid: Lift	2.19	2.19	2.17
Valid: Maximum Absolute Error	1.00	1.00	1.00
Valid: Misclassification Rate	0.01	0.02	0.02
Valid: Mean Squared Error	0.01	0.01	.
Valid: Sum of Frequencies	3049.00	3049.00	3049.00
Valid: Root Average Squared Error	0.11	0.11	0.14
Valid: Cumulative Percent Response	100.00	100.00	98.12
Valid: Percent Response	100.00	100.00	98.72
Valid: Root Mean Squared Error	0.11	0.11	.
Valid: Sum of Squared Errors	71.87	71.72	112.89
Valid: Sum of Case Weights Times Freq	6098.00	6098.00	.
Valid: Number of Wrong Classifications	42.00	.	.

Fig. 8. Screenshot with the results of quality assessment of decision tree models, logistic regression and neural network in the SAS Enterprize Miner package

Conclusions. This research paper demonstrates the practical implementation of the most common Data Mining technologies predict the results of e-sports matches, namely: decision trees, logistic regression, and neural network. According to the results of the study, it can be conclude that the best Data Mining technology that can be recommended for the practical implementation of the predictor of the results of e-sports matches is the technology of neural networks. The second in terms of forecasting quality is logistic regression. Based on the obtained models, real-time predictors can be built, which are able to provide players with information about what priorities should be during the game, for example, to change the strategy of the game in which they are defeated.

References

1. Jenny, S., Gawrysiak, J., & Besombes, N. (2021). Esports.edu: An Inventory and Analysis of Global Higher Education Esports Academic Programming and Curricula. International Journal of Esports, 1(1). Retrieved from https://www.ijesports.org/article/59/html

2. Watson, B., Spjut, J., Kim, J., Listman, J., Kim, S., Wimmer, R., Lee, B. (2021). Esports and high performance HCI. Paper presented at the Conference on Human Factors in Computing Systems. Proceedings, doi:10.1145/3411763.3441313. Retrieved from www.scopus.com

3. Imas Ye., Petrovska T., Hanaha O. (2021) Kibersport v Ukraini yak suchasnyi kulturnyi fenomen. Teoriia i metodyka fizychnoho vykhovannia i sportu. № 1. S. 75–81.

4. Zhmai O.V. (2021) Kibersport yak investytsiino pryvablyva haluz dlia Ukrainy. Stan, dosiahnennia ta perspektyvy informatsiinykh system i tekhnolohii: Materialy konf. Odesa. S. 214–215. URL: http://dspace.onu.edu.ua:8080/handle/123456789/31438 (accessed at 20.06.2021)

5. Gorova K.O., Goroviy D.A., Kiporenko O.V. (2016). Osnovni tendentsii rozvytku rynku kibersportu. Problemy i perspektyvy rozvytku pidpryiemnytstva. № 4 (2). S. 51–55. URL: http://nbuv.gov.ua/UJRN/piprp_2016_4%282%29__12 (accessed at 20.06.2021)

6. Lazneva I.O., Tsaranenko D.I. Kibersport ta yoho vplyv na zminu struktury svitovoho rynku kompiuternykh ihor (2018). Naukovyi visnyk Uzhhorodskoho natsionalnoho universytetu. Seriia: Mizhnarodni ekonomichni vidnosyny ta svitove hospodarstvo. S. 63-66.

7. Sanahuja-Peris, G., Camacho, M. M., & Balado-Albiol, M. (2021). El afianzamiento de los e-sports en pandemia, una oportunidad para las marcas. [El afianzamiento de los e-sports en pandemia, una oportunidad para las marcas] Prisma Social, 34, 165-186. Retrieved from www.scopus.com

8. Kokkinakis, A., York, P., Moni Sagarika Patra, Justus Robertson, Ben Kirman, Alistair Coates, Alan P. Chitayat, Simon Demediuk, Anders Drachen, Jonathan Hook, Isabelle Nolle, Oluseyi Olarewaju, Daniel Slawson, Marian Ursu, & Florian Oliver Block. (2021). Metagaming and metagames in Esports. International Journal of Esports, 1(1). Retrieved from https://www.ijesports.org/article/51/html

Olha Kuzmenko **Ольга Кузьменко**	D.Sc., Professor, head of the Economic Cybernetics Department, Sumy State University, Sumy, Ukraine e-mail: o.kuzmenko@biem.sumdu.edu.ua orcid.org/0000-0001-8520-2266, Scopus Author ID: 57214957739, Researcher ID: W-5504-2019 https://scholar.google.com/citations?user=Gpd0708AAAAJ&hl=ru	доктор економічних наук, професор кафедри економічної кібернетики, Сумський державний університет, Суми, Україна
Konstantin Gritsenko **Костянтин Гриценко**	Ph.D, Docent, Associate Professor of the Economic Cybernetics Department, Sumy State University, Sumy, Ukraine e-mail: k.hrytsenko@uabs.sumdu.edu.ua orcid.org/0000-0002-7855-691X, Scopus Author ID: 6601977173 https://scholar.google.com/citations?hl=ru&user=CrvuGuwAAAAJ	кандидат технічних наук, доцент кафедри економічної кібернетики, Сумський державний університет, Суми, Україна
Hanna Yarovenko **Ганна Яровенко**	D.Sc., Docent, Associate Professor of the Economic Cybernetics Department, Sumy State University, Sumy, Ukraine e-mail: a.yarovenko@uabs.sumdu.edu.ua orcid.org/0000-0002-8760-6835, Scopus Author ID: 57210605480, Researcher ID: P-3734-2014 https://scholar.google.com.ua/citations?hl=en&user=kk67yQUAAAAJ&view_op=list_works&sortby=pubdate	доктор економічних наук, доцент кафедри економічної кібернетики, Сумський державний університет, Суми, Україна
Oleksandr Kushnerov **Олександр Кушнерьов**	PhD Student, Assistant of the Economic Cybernetics Department, Sumy State University, Sumy, Ukraine e-mail: o.kushnerov@uabs.sumdu.edu.ua orcid.org/0000-0001-8253-5698 https://scholar.google.com/citations?user=qbbAxbUAAAAJ&hl=ru	аспірант, асистент кафедри економічної кібернетики, Сумський державний університет, Суми, Україна
Avhusta Hrytsenko **Августа Гриценко**	applicant, Sumy State University, Sumy, Ukraine e-mail: a.hrytsenko@student.sumdu.edu.ua orcid.org/0000-0002-3641-637X	здобувач освіти,, Сумський державний університет, Суми, Україна