Information System for Predicting Personal Success Based on Open Data from Social Networks

Nagim Davletshin[1][0000-0002-5807-8527] and Konstantin Nikolaev[1][0000-0003-3204-238X]

[1] Kazan Federal University
davlet-9@yandex.ru, konnikolaeff@yandex.ru

Abstract. This paper presents the description and structure of an information and analytical system (IAS) aimed at collecting open quantitative and qualitative data of users of the social network on the Internet, as well as processing these data for use as input data for a complex neural network model for predicting professional and academic success. Methods for visualizing the results of the neural network model and methods for loading and unloading data from the IAS are also presented. Combining psychological and IT principles in the development of a neural network model and IAS offers a high-quality and reliable software product for predicting the success of users of the VK social network. The developed system can be used by HR companies to find the most suitable candidates, as well as by university admissions committees to find the most potentially successful applicants.

Keywords: success prediction, neural networks, web application, social networks, text processing.

1 Introduction

In this work, we present an interdisciplinary research study conducted at the intersection of two subject areas: psychology and information systems. The psychology of social networks is an intriguing field in modern psychology, dedicated to understanding the interaction between the I-real and I-virtual personalities.

A person's success consists of many separate types of success: professional, academic, social, and others. Within the framework of the project to determine various cognitive-behavioral predictors of personal success, professional and academic success were considered as the most indicative components of a personal success [1-2]. The criterion for assessing academic success was the students' grades, which were summarized into an integral index reflecting the average number of points for the entire period of their studies. The criterion for assessing professional success was an integral index, which was calculated on the basis of analyzing salary, work experience, length of stay at one place and knowledge of foreign languages.

As a data source for success predicting task, we used the data from the profiles of the VK social network users, assessments on subjects during training at the Kazan Fed-

eral University, as well as employment data (the profiles were obtained from the website hh.ru, which helps applicants in finding suitable vacancies). VK social network is a Russian–speaking social network, extremely popular among residents of Russia and countries of Commonwealth of Independent States (CIS) aged 15 to 45 years. This age range coincides well with the people in our research sample.

The purpose of this article is to present an information and analytical system that incorporates a neural network model for predicting academic and professional success based on users' activity in the VK social network [3-6]. By combining the principles of psychology and information systems by using advanced machine learning techniques, our information and analytical system offers a valuable tool for searching and analyzing VK users' profiles, extracting quantitative and qualitative user data, and providing forecasts for academic and professional success based on the developed neural network model.

2 Related work

The primary focus of our research specifically focused on predicting real-world behavior by analyzing users' virtual activities on social network [7-14].

Visual forms of digital content, such as posts, audio and video content, and photos, serve as makers within the personal profiles or social network users. By analyzing the digital content individuals generate on social media platforms, such as posts, audio and video content, and photos, researchers aim to predict real-life behaviors. However, the complexity of such studies arises from the need to collect and analyze large amounts of data from social networks, as well as the presence of nonlinear relationships between virtual activity indicators and real-life behavioral characteristics [15-17].

To overcome these limitations of traditional data processing and analysis methods, machine learning techniques, particularly artificial neural networks, offer promising solutions. Social network data mining is a rapidly evolving research field, and artificial intelligence methods have proven to be an effective in tasks such as classification, forecasting, anomaly detection, and clustering [18-19]. Intelligent social network analysis utilizes various types of neural networks, including perceptrons [20], recurrent neural networks (RNNs) and Gated Recurrent Units (GRUs) [21-22], and deep learning neural networks based on graphs (GCNs) [23]. Text analysis methods, such as text embedding and word2vec, based on dynamic neural networks (DNNs) and LSTM-based recurrent neural networks, are also extensively used for analyzing qualitative data such as posts and comments [18]. Moreover, machine learning and neural networks have been applied in studies focusing on extracting other characteristics from social network data, such as analyzing the risk of suicide based on Facebook posts [24-26].

In addition, there are a number of works aimed at predicting the academic performance of students using artificial neural networks. In [27], a model is proposed for organizing assistance to students in choosing areas of study in professional training. The right choice of directions increases the student's chances of successful development in the professional field and increases the chances of success in his career. The following data were used as data for training the neural network: personal interests, family

information, previous training, and others. The model showed 86% accuracy in predicting the choice of courses in higher education.

In [28], the neural network was used as one of the models considered to determine the factors influencing the choice of ICFES exams. The study showed that the gender of a person is the most important factor when choosing the direction of study, as well as the economic resources of the family. In [29], a neural network model was proposed to determine the most important parameters affecting a student's academic success. Among them there are the following parameters: reaction time, the place of employment of the mother and father, the educational institution in which the student studied before, and executive control (a collective measure characterizing a person's ability to manage and control cognitive processes).

In [30], a simple feedforward neural was trained to forecast evaluation test scores by taking into account the partial scores obtained during the course. Another study [31] uses the Cumulative Grade Point Average to predict academic performance specifically in the eighth semester also by using neural networks. The researchers in another investigation [32] compared two different neural network models, namely the Multilayer Perceptron and the Generalized Regression Neural Network, to determine which is the more effective in predicting students' academic performance. Lastly, the potential of neural networks in predicting learning outcomes was assessed and compared to a multivariate linear regression (LR) model within the field of medical education [33].

In most of the works, discussed above, the analysis of qualitative data of social networks is carried out without taking into account the psychological component, or only various academic indicators are predicted. Our work presents the results of a comprehensive study based on the analysis of various metrics of social networks (quantitative and qualitative), as well as a neural network based psychometric model of cognitive behavioral predictors of a person's activity.

3 Neural network model for predicting personal success

In this section, we provide a brief description of the neural network model aimed to prediction a person's success based on the data from social networks. A more detailed description of the system is given in [5-6].

The neural network model was developed by using Python programming language and an open API for obtaining data from VK social network. Initially, the subjects in the training collection were divided into five separate categories of success (academic and professional). For this purpose, the K-Means clustering algorithm on the basis of parameters such as academic grades for academic success and place of work, duration of stay in the same position, frequency of job changes, salary changes, etc. for professional success was used. A more detailed description of the parameters used and the adjustment of the coefficients of the algorithm are given in [34].

To predict personal success, we used forward neural network, and the best accuracy was obtained by using four-layered model with the following configuration: the first hidden layer contains 210 neurons, the second – 10 neurons, the third – 2 neurons, the output – one neuron. The number of neurons on the input layer is determined by a set

of input parameters used to predict personal success. Five unput parameters are quantitative parameters: the number of friends, subscribers, photos, videos, interesting pages. These parameters were selected during the statistical analysis of the relationship between all available metrics of the personal profile of the social network and the category of success of the subject. These parameters are scaled in the range [0; 1]. To prepare the remaining five parameters, sets of words were formed that are most often found in the texts of user posts from various categories of success (from 1 to 5). When analyzing the texts of the person under study, for each set of frequent words, the ratio of the words present in the texts of the posts of the subject to the total number of words in the set is calculated. One neuron in the output layer returns the success category number predicted by the neural network model.

For the practical application of the trained complex neural network model, a web application with user friendly interface was developed. This application performs the tasks of collecting data from the VK social network, preprocessing collected data and predicting professional and academic success.

4 IAS components description

The development of the web application (IAS) is based on Python language and its libraries, the server-side web framework Django, the Redis task queue system, and the Celery Python library. The user interface of the IAS was developed using a free set of HTML and CSS tools for creating websites and web applications, specifically the Bootstrap framework. The server-side of the IAS consists of a set of Django modules that handle the tasks of retrieving and preprocessing user data from VKontakte. The internal structure of the Django project follows the MVC (Model-View-Controller) pattern. For each web page accessible to the user, there is a controller file that fills the web document (HTML page) with the corresponding data.

The interaction with the database is performed through an Object-Relational Mapping (ORM) technology, which enables working with databases in an object-oriented programming paradigm.

Fig. 1 presents a database schema for storing information about users of the system, information downloaded from the VK social network, as well as the results of processing this data. In addition to the basic tables for storing information about users of the system, a subset of records from the Tasks table is linked to each record in the Users table. The Tasks table stores information about tasks running in the background for the Redis task queue system, which allows the user to run long lasting tasks (for example, downloading information from the VK user's wall). Upon successful completion of the user's task, all collected information is added to the corresponding data table, ready for further processing using a neural network model.

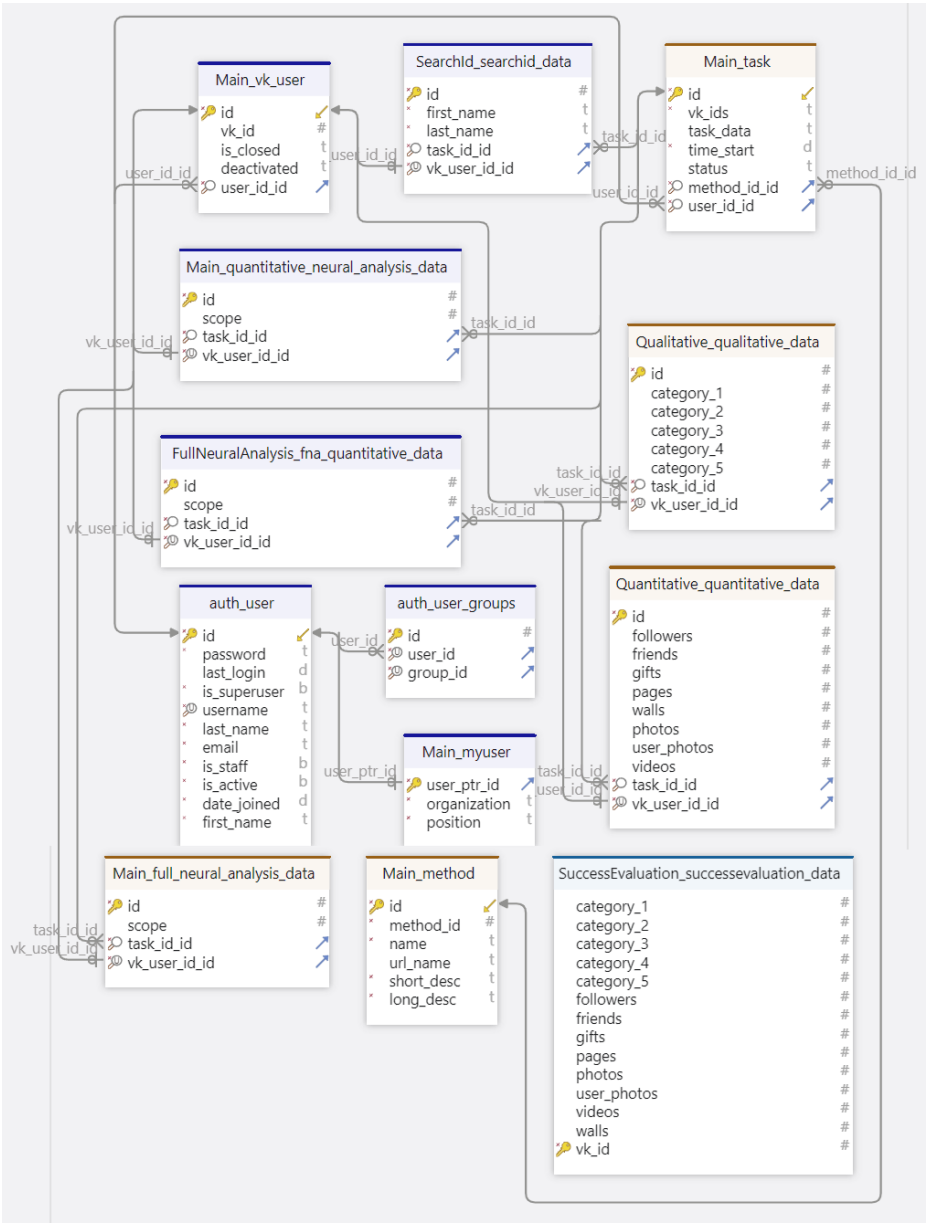

Fig. 1. IAS database scheme

Fig. 2 presents a block diagram of the IAS, reflecting the modules of the system, as well as their interaction with each other and with the database. The modular system was chosen because of the convenience of its implementation and support. In addition, IAS is aimed at the application of HR specialists and university admissions committees, so maximum structuring and ease of application was a priority.

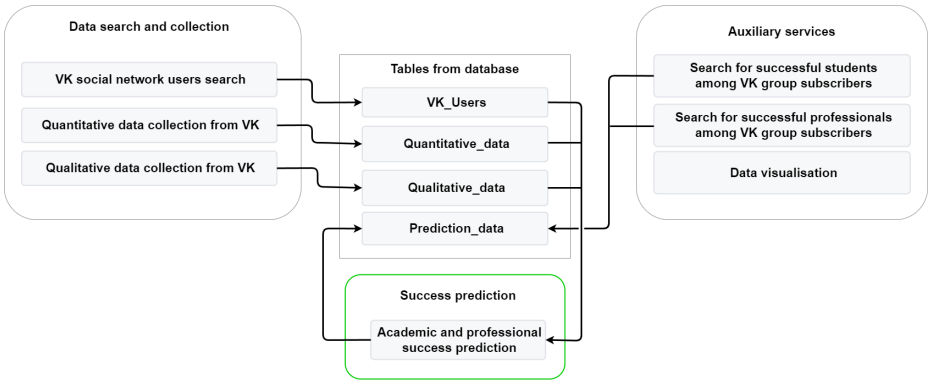

Fig. 2. IAS block diagram reflecting the system modules

IAS modules can be divided into three groups: data search and collection methods, a method for predicting success, and auxiliary methods. The module for searching and collecting data includes the methods "Search for people on the VK social network", "Collection of quantitative data on the VK social network", "Collection of qualitative data on the VK social network", "Collection of quantitative and qualitative data from the VK social network".

The "Search for people in the VK social network" method is used to quickly find the address of a personal VK page with main parameters as surname, first name, year of birth, city of residence. This method also supports batch processing of data to search for a large number of people in one operation. The found profiles are added to the VK_ID table and used in the data processing methods.

The method "Collecting quantitative data in the VK social network" is used to obtain quantitative characteristics of a user's personal page (for example, the number of friends, the number of entries on the page, the number of photos, date of birth, etc.). The obtained quantitative data is used as in input data for a neural network to predict academic and professional success.

The method "Collecting qualitative data in the social network VKontakte" receives qualitative information from users' personal pages (the text of posts and reposts of the user, the names of audio and video recordings, the names of interesting pages, as well as the texts of posts liked by the user). Based on the texts, a numerical metrics are formed that reflect the expression of the words most frequent for a particular category of success. As a result, a vector of 5 positive real numbers is formed for each user, in which the first element reflects the amount of frequency words from the texts of users of the 1st category of success, the second element – the 2nd category of success, etc. The method "Collecting quantitative and qualitative data from the social network VKontakte" is created for the convenience of the user, and it performs the work of two previous methods.

The method of predicting the subject's success - is the central method of the system. It uses quantitative and qualitative data obtained from the VK as parameters as input data for the neural network described in the previous section. The list of subjects avail-

able for processing reflects only subjects containing a full set of quantitative and qualitative data. The neural network uses as input quantitative parameters (the number of friends, subscribers, photos, videos, interesting pages) and metrics formed on the basis of the texts of posts and reposts of the user, reflecting the amount of frequency words from the corresponding categories of success. Each processed VK user is assigned a value from 1 to 5, reflecting the predicted success rate.

The third group of methods includes the method for searching of successful students or professionals among subscribers of the selected VK group, the methods of displaying information the data contained in the system, and the results of personal success prediction. Methods of searching for successful students / professionals allows us after input the URL of the VK group to start automatically the process of collecting all the necessary data, and predicting success for each subscriber of the group with an open profile. The method of displaying information in the system allows the IAS user to view and edit the results of the system (obtained quantitative and qualitative data, the results of success forecasting). In addition, the method allows to form a distribution of an arbitrary sample by categories of success and get a set of people from the category of interest.

5 Results

The result of this study is a full-fledged information system aimed at collecting quantitative and qualitative data of users of the VK social network, neural network analysis of this data, and visualization of the results obtained.

First of all, we present estimates of the accuracy of the neural network model. For training the neural network model of academic success prediction component, a dataset consisting of 33480 records was used. For training the neural network model of professional success prediction component, a dataset consisting of 61232 records was used. Both datasets were divided into a training and test collection in the ratio of 70% and 30%. The accuracy of the model is currently up to 83% for academic success and up to 78% for professional success.

As a result, the user obtains a set of records from the database describing quantitative and qualitative data obtained from VKontakte profiles, as well as the results obtained after their processing. The "Data visualization" method is used to visualize the generated data. In Fig. 3 is shown the examples of how the information is displayed in the system. This figure shows the table containing the internal ID of a personal VK page, and the predicted success value in the range [0;1]. The diagram at the right shows the distribution of the sample success values in five equal intervals (0-0.2, 0.2-0.4, etc.).

Fig. 4 shows an example of displaying the list of data processing tasks set by user. When clicking on the corresponding links, the system user can to obtain detailed information about a particular task: the list of processed VK profile identifiers, collected quantitative and qualitative data, success prediction results and other data.

In Table 1 we show a fragment of the merged table consisting of Quantitative_quantitative_data and Qualitative_qualitative_data tables. The obtained dataset of quantitative and qualitative data is used as an input data of the neural network model for predicting academic and professional success of a person.

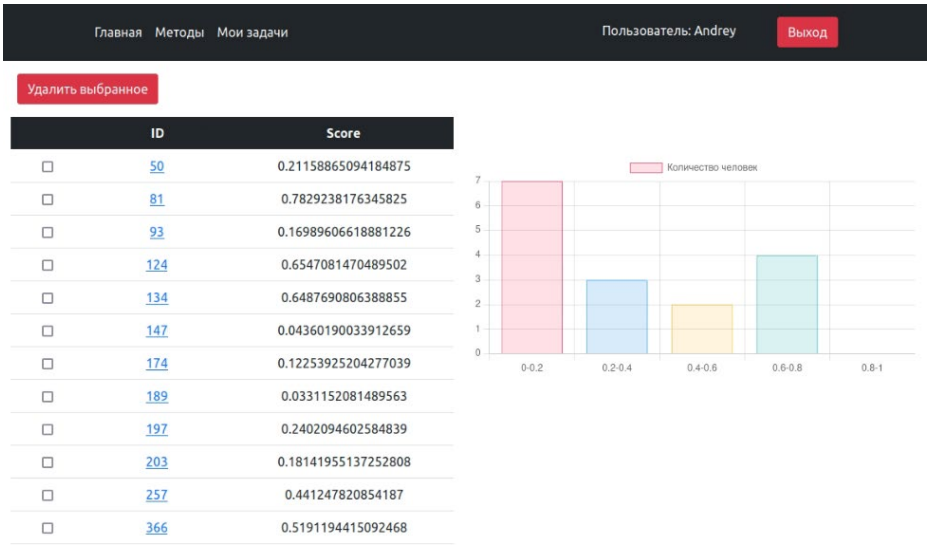

Fig. 3. The form of displaying the results of the personal success prediction and analysis

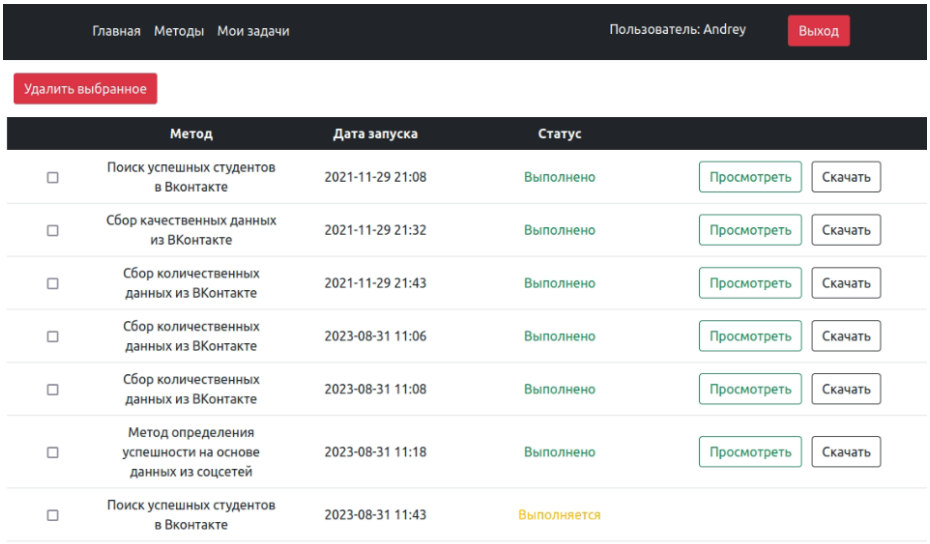

Fig. 4. Form for displaying user task's states

Table 1. Fragment of the result of merging Quantitative_quantitative_data and Qualitative_qualitative_data tables. The merged table contains quantitative and qualitative data of VKontakte users saved in the IAS.

id	7	8	9
followers	5941575	154	174
friends	0	568	566
gifts	0	0	26
pages	0	4	140
walls	296	496	852
photos	239	101	133
user_photos	0	0	0
videos	4	366	319
task_id_id	8	9	10
vk_user_id_id	18	24	20
category_1	0.3115942028985507	0.3115942028985507	0.31159
category_2	0.012254901960784314	0.012254901960784314	0.01225
category_3	0.025	0.025	0.025
category_4	0.02564102564102564	0.02564102564102564	0.02564
category_5	0.4803921568627451	0.4803921568627451	0.48039

6 Conclusion

This paper presents the system architecture and the description of informational and analytical system's main components. The informational and analytical system aimed to collecting and analyzing quantitative and qualitative data from VK user pages in order to predict academic and professional success of this social networks users. The methods for obtaining quantitative and qualitative data, for analyzing these data by using the neural network model, and for visualizing the results of the analysis are described. The developed system can be used by HR companies to find the most suitable candidates, as well as by university admissions committees to find the most potentially successful applicants.

In the future, it is planned to expand the neural network model of predicting success by adding a component focused to social success prediction. By the social success of a user of social networks, we mean the ability of a person in the course of virtual communication to create broad contacts, exchange information, show interest in other users, involve others in their interests, create their own groups, influence others. The list of possible metrics that will be included in the neural network model includes parameters of graphs of social relations between VK users: the number and density of connections, cliques, various centrality measures, clustering coefficients, and others.

We also plan to expand the ways of visualizing user data by introducing the methods for additional data processing and filtering the data by categories of academic and professional success. We also plan to publish this web-service on public domain.

Acknowledgement.

Funding: the study (all theoretical and empirical tasks of the research presented in this paper) was supported by a grant from the Russian Science Foundation (Project No. 19-18-00253, "Neural network psychometric model of cognitive-behavioral predictors of life activity of a person on the basis of social networks"), https://rscf.ru/project/19-18-00253/.

References

1. Ustin, P.N., Popov, L.M.: Problema prognozirovaniya zhiznennoj aktivnosti lichno-sti v social'nyh setyah. In: V Andreevskie chteniya: sovremennye koncepcii i tekhno-logii tvorcheskogo samorazvitiya lichnosti. pp. 457–461 (2020).
2. Ustin, P.N., Popov, L.M.: Kognitivno-povedencheskaya koncepciya i vozmozhnosti ee realizacii v zhiznennoj aktivnosti studentov. Psihologicheskij zhurnal. 42, 26–35 (2021).
3. Gafarov, F.M., Rudneva, Y.B., Sharifov, U.Y.: Predictive Modeling in Higher Education: Determining Factors of Academic Performance. Vyss. Obraz. v Ross. = High. Educ. Russ. 32, 51–70 (2023). https://doi.org/10.31992/0869-3617-2023-32-1-51-70.
4. Gafarov, F., Berdnikov, A., Ustin, P.: Predicting success of the online social network user by graph neural networks. Int. J. Adv. Intell. Informatics. 8, 285–298 (2022). https://doi.org/10.26555/ijain.v8i3.859.
5. Gafarov, F.M., Nikolaev, K.S., Ustin, P.N., Berdnikov, A.A., Zakharova, V.L., Reznichenko, S.A.: A Complex Neural Network Model for Predicting a Personal Success based on their Activity in Social Networks. Eurasia J. Math. Sci. Technol. Educ. 17, 1–9 (2021). https://doi.org/10.29333/ejmste/11175.
6. Gafarov, F.M., Rudneva, Y.B., Sharifov, U.Y., Trofimova, A.V., Bormotov, P.M.: Analysis of Students' Academic Performance by Using Machine Learning Tools. (2020). https://doi.org/10.2991/assehr.k.200509.104.
7. Hiranyachattada, T., Kusirirat, K.: The effect of social media and lecturer's direct instruction on students' digital sculpting skills. Eur. J. Sci. Math. Educ. 8, 163–169 (2020). https://doi.org/10.30935/scimath/9554.
8. Khandelwal, P., Gotlieb, M.R.: The Role of Perceived Utilitarian and Hedonic Value in Predicting Use of Location-Based Anonymous Social Networking Sites. Online J. Commun. Media Technol. 11, e202118 (2021). https://doi.org/10.30935/ojcmt/11114.
9. Levina, I.D., Ukolova, L.I., Lavrentyeva, E.Y., Akhilgova, M.T., Zharikov, Y.S., Popova, O. V., Semyanov, E. V., Malanov, I.A., Muskhanova, I. V., Magomeddibirova, Z.A., Bazaeva, F.U., Isaeva, M.A., Minasova, V.M.: Nursing home conditions for elderly people and its peculiarities of their adaptation. EurAsian J. Biosci. 13, 1549–1555 (2019).
10. Minakhmetova, A.Z., Skutelnik, O.A., Fedorchuk, Y.M., Shulga, T.I., Pavlushin, A.A., Shagiev, B. V.: Motivational and valuable orientation of teachers as a condition to form students' tolerant behavior. Man India. 97, 323–332 (2017).
11. Orekhovskaya, N.A., Chistyakov, A.A., Kryukova, N.I., Krokhina, J.A., Ospennikov, Y. V., Makarova, E. V.: Orthodoxy and modernity their contact facets in russian society. Eur. J. Sci. Theol. 15, 67–77 (2019).
12. Owan, V.J., Ekpe, M.B., Eneje, S.: Undergraduates' Utilization of Social Networking Media and Sexual Behaviours in Higher Education: A Case Study. Pedagog. Res. 5, (2020). https://doi.org/10.29333/pr/7940.
13. Piralova, O.F., Gerasimenko, S.A., Kuznetsov, V. V., Popova, O. V., Subbotin, G. V., Kolomyts, O.G., Mashkin, N.A.: Gaming industry trends in new generation specialist training in university environment. J. Environ. Treat. Tech. 8, 1132–1135 (2020).

14. Razumovskaya, M.I., Larionova, A.A., Zaitseva, N.A., Petrina, O.A., Vinogradova, M. V., Nagay, N.G., Takhumova, O. V.: Models of integrated interactions organization in the field of environmental education. J. Environ. Treat. Tech. 7, 576–580 (2019).
15. Du, L., Wang, Y., Song, G., Lu, Z., Wang, J.: Dynamic Network Embedding: An Extended Approach for Skip-gram based Network Embedding. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. pp. 2086–2092. International Joint Conferences on Artificial Intelligence Organization, California (2018). https://doi.org/10.24963/ijcai.2018/288.
16. Galchenko, N.A., Shatskaya, I.I., Makarova, E. V, Kulesh, E. V, Nizamutdinova, S.M., Yudina, A.M., Skutelnik, O.A.: Student hood spiritual needs in self-isolation period: Features and ways to meet them. EurAsian J. Biosci. 14, 2229–2234 (2020).
17. Tugun, V., Bayanova, A.R., Erdyneeva, K.G., Mashkin, N.A., Sakhipova, Z.M., Zasova, L. V.: The Opinions of Technology Supported Education of University Students. Int. J. Emerg. Technol. Learn. 15, 4 (2020). https://doi.org/10.3991/ijet.v15i23.18779.
18. Di Franco, G., Santurro, M.: Machine learning, artificial neural networks and social research. Qual. Quant. 55, 1007–1025 (2021). https://doi.org/10.1007/s11135-020-01037-y.
19. Young, T., Hazarika, D., Poria, S., Cambria, E.: Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. IEEE Comput. Intell. Mag. 13, 55–75 (2018). https://doi.org/10.1109/MCI.2018.2840738.
20. Jabłońska, M.R., Zajdel, R.: Artificial neural networks for predicting social comparison effects among female Instagram users. PLoS One. 15, e0229354 (2020). https://doi.org/10.1371/journal.pone.0229354.
21. Balakrishnan, M., T. V., G.: A neural network framework for predicting dynamic variations in heterogeneous social networks. PLoS One. 15, e0231842 (2020). https://doi.org/10.1371/journal.pone.0231842.
22. Yang, C., Sun, M., Zhao, W.X., Liu, Z., Chang, E.Y.: A Neural Network Approach to Jointly Modeling Social Networks and Mobile Trajectories. ACM Trans. Inf. Syst. 35, 1–28 (2017). https://doi.org/10.1145/3041658.
23. Tan, Q., Liu, N., Hu, X.: Deep Representation Learning for Social Network Analysis. Front. Big Data. 2, (2019). https://doi.org/10.3389/fdata.2019.00002.
24. Ophir, Y., Tikochinski, R., Asterhan, C.S.C., Sisso, I., Reichart, R.: Deep neural networks detect suicide risk from textual facebook posts. Sci. Rep. 10, 16685 (2020). https://doi.org/10.1038/s41598-020-73917-0.
25. Shatte, A.B.R., Hutchinson, D.M., Teague, S.J.: Machine learning in mental health: a scoping review of methods and applications. Psychol. Med. 49, 1426–1448 (2019). https://doi.org/10.1017/S0033291719000151.
26. Zheng, L., Wang, O., Hao, S., Ye, C., Liu, M., Xia, M., Sabo, A.N., Markovic, L., Stearns, F., Kanov, L., Sylvester, K.G., Widen, E., McElhinney, D.B., Zhang, W., Liao, J., Ling, X.B.: Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. Transl. Psychiatry. 10, 72 (2020). https://doi.org/10.1038/s41398-020-0684-2.
27. Sethi, K., Jaiswal, V., Ansari, M.D.: Machine Learning Based Support System for Students to Select Stream (Subject). Recent Adv. Comput. Sci. Commun. 13, 336–344 (2020). https://doi.org/10.2174/2213275912666181128120527.
28. Martínez Cervera, D.E., Salcedo Parra, O.J., Aguilera Prado, M.A.: Forecasting model with machine learning in higher education ICFES exams. Int. J. Electr. Comput. Eng. 11, 5402 (2021). https://doi.org/10.11591/ijece.v11i6.pp5402-5410.

29. Musso, M.F., Kyndt, E., Cascallar, E.C., Dochy, F.: Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. Front. Learn. Res. 1, (2013). https://doi.org/10.14786/flr.v1i1.13.
30. Gedeon, T.D., Turner, H.S.: Explaining student grades predicted by a neural network. Proc. Int. Jt. Conf. Neural Networks. 1, 609–612 (1993). https://doi.org/10.1109/ijcnn.1993.713989.
31. Arsad, P.M., Buniyamin, N., Manan, J.A.: A neural network students' performance prediction model (NNSPPM). In: 2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA). pp. 1–5. IEEE (2013). https://doi.org/10.1109/ICSIMA.2013.6717966.
32. R. Iyanda, A., D. Ninan, O., O. Ajayi, A., G. Anyabolu, O.: Predicting Student Academic Performance in Computer Science Courses: A Comparison of Neural Network Models. Int. J. Mod. Educ. Comput. Sci. 10, 1–9 (2018). https://doi.org/10.5815/ijmecs.2018.06.01.
33. Dharmasaroja, P., Kingkaew, N.: Application of artificial neural networks for prediction of learning performances. In: 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). pp. 745–751. IEEE (2016). https://doi.org/10.1109/FSKD.2016.7603268.
34. Vakhitov, G., Enikeeva, Z., Yangirova, N., Shavalieva, A., Ustin, P.: Identification of the Clusters of Social Network Communities for Users with a Specific Characteristic. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE). pp. 140–146. IEEE (2019). https://doi.org/10.1109/DeSE.2019.00035.