

Socio-Economic Problems and the State

journal home page: <http://sepd.tntu.edu.ua>



Chala, N., Voropai, O. & Pichyk, K. (2021) Using data mining to create innovations in education. *Sotsialno-ekonomichni problemy i derzhava* [Socio-Economic Problems and the State] (electronic journal), Vol. 25, no. 2, pp. 21-28. Available at: <http://sepd.tntu.edu.ua/images/stories/pdf/2021/21cndiie.pdf>



USING DATA MINING TO CREATE INNOVATIONS IN EDUCATION

Nina CHALA

Skovoroda Str., 2, Kyiv, 04070, Ukraine
e-mail: n.chala@ukma.edu.ua
ORCID ID: <http://orcid.org/0000-0002-0356-9003>

Olga VOROPAI

National University of Kyiv-Mohyla Academy
Volos'ka Str., 10, Kyiv, 04655, Ukraine
e-mail: olga.voropai@ukma.edu.ua
ORCID ID: <http://orcid.org/0000-0001-5257-7619>

Kateryna PICHYK

Skovoroda Str., 2, Kyiv, 04070, Ukraine
e-mail: pichykv@ukma.edu.ua
ORCID ID: <http://orcid.org/0000-0003-1161-270X>



Article history:

Received: October, 2021
1st Revision: November, 2021
Accepted: November, 2021

JEL classification:

C10
C13
C89
I20
I29

UDC:

004.8:378

DOI:

<https://doi.org/10.33108/sepd.2022.02.021>

Abstract. *The article substantiates the need for educational institutions to use Data Mining technology as a key to successful management decisions in modern realities. The study focuses on working with social media data. The authors emphasized the lack of attention to this issue among both foreign and Ukrainian scientists.*

The article outlines the algorithm for collecting and transmitting primary data obtained as a result of monitoring the activity of educational institutions in social networks to form models of various types of their actions.

The model presented by the authors includes four stages. Stages one and two provide the list of factors / metrics that can be included in the model. These factors require an appropriate and high-quality data collection process. At the next stage, the authors propose data clustering as the most important process for the future use of social network data. It is emphasized that the formation of clusters will depend on the tasks facing the management teams of the educational market. The authors give several examples of such clustering but point out that the list is not exhaustive and can be significantly expanded.

An important aspect of the availability of such databases is access to information not only for teachers, but also for all interested university staff. At the same time, each user (students, teachers, staff, administration) will receive data relevant to their requests and needs.

The developed methodology will help increase the efficiency of management decision-making and implementation and provide an opportunity to justify the parameters of successful innovation in educational institutions in many respects, including the development of educational programs, implementation of new certification programs and disciplines, other services, etc.

Key words: *Data Mining technology, social networks, clusters, educational institutions, educational services, innovations in education.*



Чала Н. Using data mining to create innovations in education [Електронний ресурс] / Ніна Чала, Ольга Воропай, Катерина Пічик // Соціально-економічні проблеми і держава. — 2021. — Вип. 2 (25). — С. 21-28. — Режим доступу: <http://sepd.tntu.edu.ua/images/stories/pdf/2021/21cndiie.pdf>



This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

1. Statement of the Problem.

Data Mining technologies are a powerful tool for business analytics and data research to identify hidden patterns and build models. The key point is that large data sets are devoid of structure and connections and the purpose of technology is to identify these structures and find patterns where, at first glance, chaos reigns. Companies are already highly active in using this tool to analyze almost all available financial and behavioral indicators, because Data Mining is not based on theoretical considerations but actual data, making justification and management decisions faster and more accurate. At the same time, research has testified that universities are just beginning to use Data Mining technologies in their activities. Most Ukrainian higher education institutions are just thinking about this issue.

Being in the grip of state regulation, Ukrainian educational institutions focus more on the compliance of educational programs with licensing requirements than on the introduction of innovations or the development of effective interaction with stakeholders. The development of the Data mining and Machine Learning systems will provide an opportunity to develop the most personalized educational offer, which will meet, on the one hand, the licensing requirements, and on the other - the needs of stakeholders (employers, students, etc.).

2. Analysis of Recent Research and Publications.

Data Mining approaches have been widely used in marketing, in particular in social media data analysis. The methodology of analyzing important characteristics of publications in social networks is covered in [6]. The authors examined the official pages of 14 brands on Facebook. According to the results, they conclude that the format and subject matter of the publication have the most significant effect on the interaction with the brand page. The date and time of publication did not significantly increase user interaction with the page. Moro, Rita and Vala [4] show the use of Data Mining to analyze the brand page in social networks. In this paper, the authors investigated the importance of each of the publication factors to predict the success of the post.

Researchers from India have shown in their work [7] the importance of taking into account the activity of students in social networks in the university's development strategy: their opinions posted, reactions, discussions. Indian researchers also show how to understand students' behavior with this type of information better.

Other authors [9] consider the processes of co-creating value for universities with other stakeholders in social networks. Understanding the value universities create for their users, university managers will develop innovative products that their customers expect.

3. Unsolved Aspects of the Problem.

Educational institutions can use the activity of potential customers in social networks, particularly the preferences of messages, dissemination of news, comments, etc., to create innovative programs. However, the collection and processing of primary analytical data on social media activity by universities is currently underused. This situation is especially true for universities that receive state funding, which does not depend on the university's activity in the social network. Not all universities have their own social media pages. And those who do, do not always plan to communicate systematically and in advance. Basically, performance

appraisal comes down to simply counting the number of subscribers per page, likes, comments, and distributions. The presented work shows the importance and usefulness for universities to work with information obtained from social networks to develop innovative educational products.

4. Task Formulation.

Our study aims to offer an effective mechanism for collecting and transmitting a set of data obtained from monitoring social media activities. University management of different levels may further use this data to form effective models of university activities.

We offer a cyclic model of extracting, primary processing, and transmitting social network data for the innovative activities of a company providing educational services.

5. Main Research Results.

Figure 1 demonstrates the idea of a cyclic model of working with social media data within universities or other companies, providing educational services.

Working with the data will involve four main stages (Fig. 1): extracting primary data, forming additional secondary data with AI tools, data clustering, and data transfer for further analysis of employees' requests.



Fig. 1. Using social intelligence for Edu market innovations

Stage 1. Primary SM Data: extracting internal and external data from SM.

Internal and external data from social networks should be used when implementing in-depth analytics. Internal data includes page usage statistics (e.g., Facebook, Instagram, etc.) available for every SM page owner. In addition to the available analytics, the model can also use external data. For example, using Google Trends Semrush may contribute to determining the popularity of search queries of the relevant educational institution. Web-page traffic information will provide insights about the geographical location of the users, used devices, time spent on the platform, etc.

In the model, one of the following factors can be selected as the dependent variable. Accordingly, others will act as independent factors of the model (one can choose not all factors from the list for the modeling process):

- M1. Post time.
- M2. Like, Share, Comment.
- M3. Type of content.
- M4. Promotion.
- M5. Post Length.
- M6. Post Type.
- M7. Frequency.
- M8. Google Trends.
- M9. Page Traffic.
- M10. New Page Likes / Non-Likes.
- M11. Page likes reach.
- M12. Engagement Rate, ER (%)
- M13. Language.

Stage 2. Secondary SM Data: analyzing primary data with AI tools.

The second stage involves forming the additional secondary data sets based on the collected primary data using modeling tools and artificial intelligence (social intelligence). The main types of secondary data include the following metrics:

M14. Sentiment analysis. This metric allows for an initial analysis of publications and comments in the SM to determine users' positive, negative or neutral attitudes towards the topic under discussion.

M15. Influence makes it possible to identify SM participants who form other participants' attitudes and are opinion leaders on the analyzed topic.

M16. Network measures allow you to determine the structure of interaction, the strength of ties, and cluster SM information and participants.

Stage 3. Data Clusters: Creating data clusters for various marketing innovation problems.

The clustering stage is the most important in the process of using social networks for innovative activities in the educational market. The formation of clusters depends on the typical tasks facing the management team of educational organizations.

These tasks may require information about the socio-demographic data of the target groups, their lifestyle and the parameters influencing the choice of educational institutions or educational programs. Research testifies that entrants to master's programs actively use social networks to search for information about student life at the university and make the final decision about admission based on this data [2]. At the same time, the analysis of the SM use by Canadian universities showed a distorted coverage of aspects related to student life, particularly in Tweeter [8]. The authors point to the need to reconsider this approach to avoid misleading students.

S. Desai and M. Hang [1] consider the possibility of using artificial social intelligence (social intelligence) to analyze university courses and student attitudes towards them. It should also be supplemented by advanced analytics to identify trends in the early stages [5], which will make it possible for universities to plan a pool of research topics, develop a portfolio of educational programs and more.

Below is an indicative list of possible use cases.

1. New employees - search for new teachers, guest lecturers (experts) for educational programs. Search for partners for project implementation, student internships, etc.
2. Students/applicants - identify the most motivated applicants who share corporate values and university culture and aim to apply for this particular university.
3. Student leisure. Formation and support of team spirit among students, organization of extracurricular activities. Identifying the most exciting activities and informing participants about the upcoming events and regular activities.
4. New educational products - new programs, dual education, interdisciplinary, international, and double degree programs.
5. Advanced training and adult education - identifying requests for short-term or long-term advanced training courses, certificate programs, master classes. Creating and launching the university's innovative proposals, defining market demand for new programs and courses.
6. Reading list - the analysis of social networks to create recommendation lists of the up-to-date and relevant literature on a specific subject. Identifying relevant SM discussions will allow, for example, libraries to form opinion-leaders clusters, researchers interested in the topic and suggest these results for department heads, lecturers and students. The latter may use such recommendations to improve their courses with up-to-date literature, invite industry experts in class, ask for expert advice for a graduate thesis.
7. Research topics (who discusses, who supports) - monitoring current discussions in social networks identifying the tone of these discussions and the level of public interest. Such information will form thematic clusters for the distribution of university research resources, setting up research priorities. It will allow tracking opinion leaders and experts on selected topics, etc.

Stage 4. Data Users: using data by company employees in their innovation decision-making process.

The fourth stage is the last in the algorithm of transferring raw data. In the future, users, upon request, will use the appropriate data cluster to model processes, reactions and make management decisions. These decisions can relate to changes at the operational level (e.g., expanding the library stock based on SM data) and at the strategic level (e.g., deciding to launch new programs for a new target audience).

Each of the clusters can have its data users (see Fig. 2). The strategic department can use the Going Global (New International Program) set to make decisions about entering new geographic / language markets. In turn, the Department of International Cooperation based on the same cluster will supplement this decision with a list of possible partners for cooperation under this program. This cluster will allow heads of departments to take into account the feedback in the SM, the tone of information, etc., to develop the content of the new program and its promotion plan.

However, one user can access different data clusters to solve various problems. Thus, the department head can turn to other data clusters to solve other educational, scientific and managerial issues (Fig. 2).

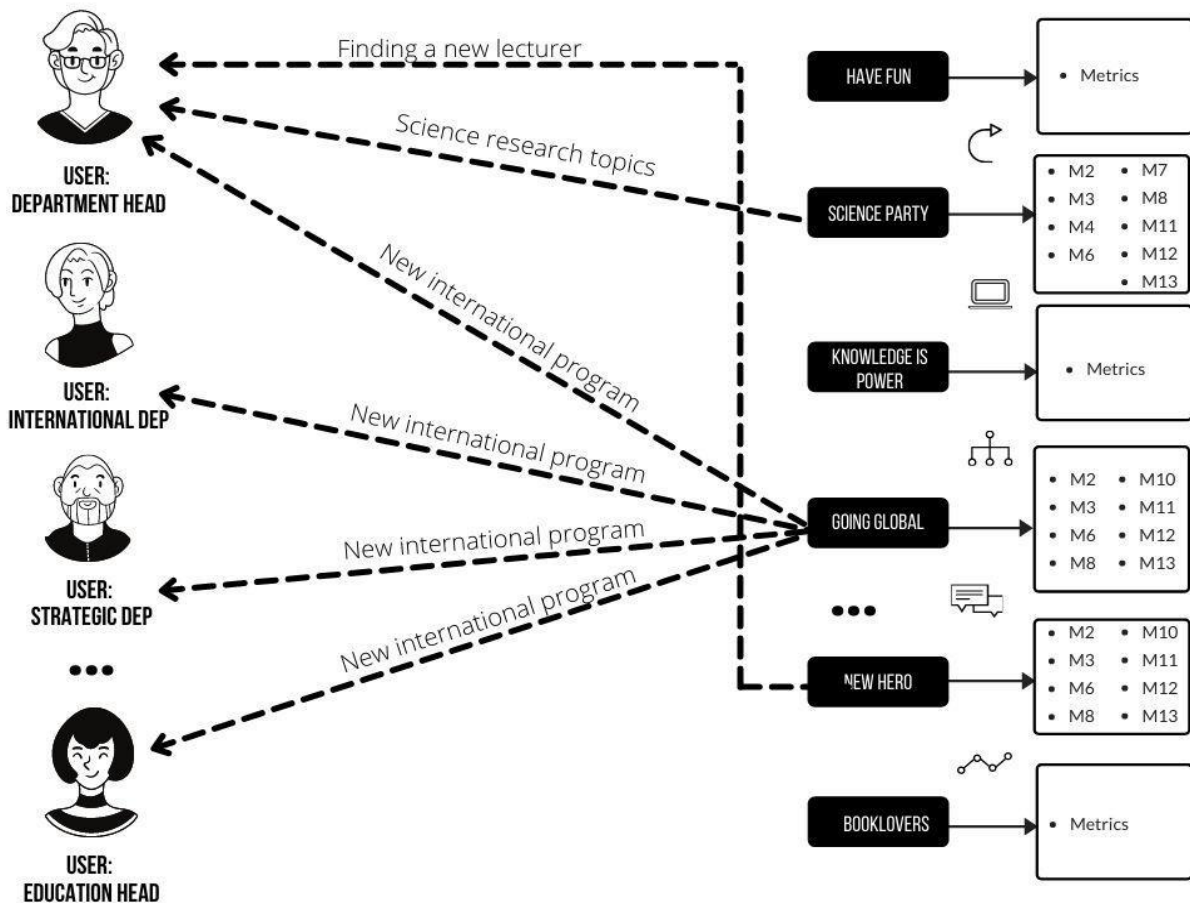


Fig. 2. Using an appropriate data cluster.

The choice of research topics will be more reasonable considering the cluster of social data "Science Party" (Research topics). Finding a new professor to teach the courses will require working with the New Hero cluster.

6. Conclusions and Further Research Implications.

Further successful use of data obtained from social networks has several obstacles caused by the specific characteristics of the educational market. Most of its players use analytical data for standardized reporting to government institutions. At the same time, internal data processing is limited to the unit that participated in data collection [3]. That is why practical work with data requires developing internal policies that will provide access to information for all stakeholders. At the same time, the efficiency of working with data obtained from social networks will be higher if they are pre-clustered at the stage of data transfer to users. This aggregation will involve grouping the data into sets needed to solve standardized tasks. The availability of ready-made clusters will automate the requests of various users of educational institutions (from students to senior management).

The proposed methodology will provide an opportunity to understand the parameters of successful implementation of innovation in educational institutions while developing educational programs, courses, training and other services, defining pricing policy, research areas, selection of the most motivated students, inviting appropriate lecturers and more.

Author details (in Ukrainian)**ВИКОРИСТАННЯ ТЕХНОЛОГІЙ DATA MINING ДЛЯ РОЗРОБКИ ІННОВАЦІЙНИХ ПРОДУКТІВ В СФЕРІ ОСВІТИ****Ніна ЧАЛА**

Національний університет "Кієво-Могилянська Академія"
вул. Г. Сковороди, 2, м. Київ, 04070,
Україна
e-mail: n.chala@ukma.edu.ua
ORCID ID: <http://orcid.org/0000-0002-0356-9003>

Ольга ВОРОПАЙ

Національний університет "Кієво-Могилянська Академія"
вул. Волоська, 10, м. Київ, 04655,
Україна
e-mail: olga.voropai@ukma.edu.ua
ORCID ID: <http://orcid.org/0000-0001-5257-7619>

Катерина ПІЧИК

Національний університет "Кієво-Могилянська Академія"
вул. Г. Сковороди, 2, м. Київ, 04070,
Україна
e-mail: pichykv@ukma.edu.ua
ORCID ID: <http://orcid.org/0000-0003-1161-270X>

Анотація. В статті обґрунтовано необхідність використання закладами освіти технології Data Mining, адже в сучасних реаліях це ключ для успішного прийняття управлінських рішень. Особлива увага дослідження зосереджена на роботі з даними соціальних мереж. Авторами вивчено роботи, пов'язані із даною тематикою, та наголошено на недостатній увазі до відзначеного питання як серед іноземних, так і українських вчених.

У статті окреслено алгоритм збору та передачі первинних даних, отриманих внаслідок моніторингу активності закладів освіти у соціальних мережах, для формування моделей різноманітних видів їх діяльності.

Наведена авторами модель містить чотири етапи. За першими двома представлено фактори/метрики, які можуть бути включені в модель, і відповідно потребують збору необхідних даних. На наступному етапі авторами запропоновано кластеризацію даних, як найважливіший процес для майбутнього використання даних соціальних мереж. Наголошено, що формування кластерів буде залежати від завдань, які стоять перед керівництвом щодо провадження діяльності на ринку освітніх послуг. Автори наводять декілька прикладів такої кластеризації, але вказують, що перелік не вичерпується запропонованим списком і може бути значно розширений.

Важливим аспектом наявності таких баз даних, є можливість доступу до інформації всіх зацікавлених співробітників університетів. У той же час, кожен з користувачів (студенти, викладачі, співробітники, адміністрація) зможуть отримувати дані, релевантні своїм запитам і потребам.

Розроблена методика сприятиме підвищенню ефективності прийняття та впровадження управлінських рішень і надасть можливість обґрунтувати параметри успішного впровадження інновацій в закладах освіти щодо багатьох аспектів, зокрема, при розробці освітніх програм, впровадженні нових сертифікатних програм та дисциплін, пропозиції інших послуг тощо.

Ключові слова: технологія Data Mining, соціальні мережі, кластери, заклади освіти, освітні послуги, інновації в освіті.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://sepd.tntu.edu.ua/images/stories/pdf/2021/21cndiie.pdf>

Funding

The authors received no direct funding for this research.

Citation information

Chala, N., Voropai, O. & Pichyk, K. (2021) Using data mining to create innovations in education. *Sotsialno-ekonomichni problemy i derzhava* [Socio-Economic Problems and the State] (electronic journal), Vol. 25, no. 2, pp. 21-28. Available at: <http://sepd.tntu.edu.ua/images/stories/pdf/2021/21cndiie.pdf>

References

1. Desai S., Meng H. (2019) Social Media Content Analytics beyond the Text: A Case Study of University Branding in Instagram. 94-101. DOI: [10.1145/3299815.3314441](https://doi.org/10.1145/3299815.3314441)
2. Galan M., Lawley M., Clements, M. (2015) Social media's use in postgraduate students' decision-making journey: An exploratory study. *Journal of Marketing for Higher Education*, 25, 287–312 DOI: [10.1080/08841241.2015.1083512](https://doi.org/10.1080/08841241.2015.1083512)

3. Krawitz M., Law J., Litman S. (2018) How higher-education institutions can transform themselves using advanced analytics. *McKinsey and Company*, 1-6. Available at: <https://www.mckinsey.com/industries/public-and-social-sector/our-insights/how-higher-education-institutions-can-transform-themselves-using-advanced-analytics?cid=app> (accessed 19 August 2021)
4. Moro S., Rita P., Vala B. (2016) Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach, *Journal of Business Research*, Volume 69, Issue 9, 2016, Pages 3341-3351, ISSN 0148-2963, <https://doi.org/10.1016/j.jbusres.2016.02.010>
5. Mulligan Ch., Northcote N., Röder T., Vesuvala S. (2021) The strategy-analytics revolution. *McKinsey and Company*, 1-6. Available at: <https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/the-strategy-analytics-revolution?cid=other-eml-alt-mip-mck&hdpid=d040c36b-ab38-4318-8f3e-e01884025bba&hctky=3072032&hlkid=3a7d50730d574ff0bbf82bf62bb74d3d> (accessed 4 September 2021)
6. Pletikosa I., Spiegler E., Michahelles F. (2011). The Effect of Post Type, Category and Posting Day on User Interaction Level on Facebook. 810-813. 10.1109/PASSAT/SocialCom.2011.21. Available at: https://www.researchgate.net/publication/220876232_The_Effect_of_Post_Type_Category_and_Posting_Day_on_User_Interaction_Level_on_Facebook (accessed 4 September 2021)
7. Singh, A. (2017) Mining of Social Media data of University students. *Education and Information Technologies* 22,1515–1526. <https://doi.org/10.1007/s10639-016-9501-1>
8. Veletsianos G., Kimmons R., Shaw A., Pasquini L., Woodward S. (2017) Selective openness, branding, broadcasting, and promotion: Twitter use in Canada's public universities, *Educational Media International*, 54:1, 1-19, DOI: [10.1080/09523987.2017.1324363](https://doi.org/10.1080/09523987.2017.1324363)
9. Voropai, O., Pichyk, K., Chala, N. (2019). Increasing competitiveness of higher education in Ukraine through value co-creation strategy. *Economics and Sociology*, 2019, 12(4), стр. 214–226. DOI: [10.14254/2071-789X.2019/12-4/14](https://doi.org/10.14254/2071-789X.2019/12-4/14)



© 2021 Socio-Economic Problems and the State. All rights reserved.
 This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.
 You are free to:
 Share — copy and redistribute the material in any medium or format Adapt — remix, transform, and build upon the material for any purpose, even commercially.
 The licensor cannot revoke these freedoms as long as you follow the license terms.
 Under the following terms:
 Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
 You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
 No additional restrictions
 You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Socio-Economic Problems and the State (ISSN: 2223-3822) is published by Academy of Social Management (ASM) and Ternopil Ivan Puluj National Technical University (TNTU), Ukraine, Europe.

Publishing with SEPS ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the SEPS website
- Rapid publication
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a SEPS journal at <http://sepd.tntu.edu.ua>

