Торайғыров университетінің
ҒЫЛЫМИ ЖУРНАЛ

НАУЧНЫЙ ЖУРНАЛ
Торайгыров университета

ТОРАЙҒЫРОВ УНИВЕРСИТЕТІНІҢ ХАБАРШЫСЫ

Физика, математика және компьютерлік ғылымдар сериясы
1997 жылдан бастап шығады

TORAIGHYROV UNIVERSITY

ВЕСТНИК ТОРАЙГЫРОВ УНИВЕРСИТЕТА

Серия: Физика, математика и компьютерные науки
Издается с 1997 года

ISSN 2959-068X

№ 2 (2022)
Павлодар

НАУЧНЫЙ ЖУРНАЛ
ТОРАЙГЫРОВ УНИВЕРСИТЕТА

Серия: Физика, математика и компьютерные науки
выходит 4 раза в год

СВИДЕТЕЛЬСТВО
о постановке на переучет периодического печатного издания,
информационного агентства и сетевого издания
№ KZ91VPY00046988
выдано
Министерством информации и общественного развития
Республики Казахстан

Тематическая направленность
публикация материалов в области физики, математики,
механики и информатики

Подписной индекс – 76208

https://doi.org/ 10.48081/JVDF6096

СЕКЦИЯ «КОМПЬЮТЕРНЫЕ НАУКИ»

SRSTI 20.51.15

https://doi.org//10.48081/DVJY9039

M. B. Zhadrenova
Kazakh British Technical University, Republic of Kazakhstan, Nur-Sultan

PREDICTING AND MAKING DECISIONS REGARDING PROVIDING PUBLIC SERVICES FOR CITIZENS PROACTIVELY BY IMPLEMENTING BIG DATA ANALYTICS TECHNOLOGIES VIA E-GOVERNMENT OF KAZAKHSTAN

One of the examples of successful big data analysis projects is the E-government of Kazakhstan. This project contains various information systems, governmental databases, and technological solutions for providing public services to citizens. The availability of public services in an online format became possible thanks to the provision of electronic digital signatures to citizens free of charge. Progressively, E-government became a single mechanism for the interaction of the state and citizens and government agencies with each other, ensuring their consistency with the help of information technology. It was this mechanism that made it possible to reduce the queues to state bodies and to simplify and speed up the receipt of certificates, certificates, permits, and much more. It was an unknown study and question of whether big data analysis technologies could expand new functions that then would be used for further service progresses. This article will conduct the chances of data lake implementation and machine learning technologies for further predictive analytics and computerized decisions to provide public services proactively.

Keywords: information systems, big analysis, databases, proactive services, machine learning algorithms, technologies, data lake, implementations.

Introduction

Citizens of Kazakhstan have an opportunity to receive offline services by visiting citizens' service centers or online via the Internet. Any citizen of Kazakhstan has the opportunity to receive offline services. Egov.kz web-platform is the website for providing electronic services to citizens. For any service people

are authorized in the Egov.kz portal then they attach needed documents and the system sends this request with the user's personal data from databases to information systems of government bodies. The workers of agencies reconsider such applications for further provisions or rejections timely. So, the e-government of Kazakhstan is not only Egov.kz portal, it is a variety of information systems and databases, which stands behind the Egov.kz web-platform and keeps personal data about all Kazakhstan citizens (Ospanova & Baimakhanova, 2018). The government of Kazakhstan is on the way to implementing proactive services, which means without citizens' request and by the initiative of state bodies. For instance, according to Egov.kz, the first proactive service of «Childbirth» was automated in 2019, the government sends a message with congratulations and proposes to parents several services in one application without going to governmental agencies. This proactive service includes formalization of governmental payment for childcare, registration of a name for a child, receiving a birth document, activating a request for future kindergarten, and registering a child's living address. Before automating this proactive service, people should collect documents and go to governmental agencies for receiving these services. This one life case was analyzed by Egov.kz project's developers manually without implementing computerized decisions, but in the future could be revealed by implementing big data analysis technologies, particularly data lake technological solutions for data storing, processing, and machine learning algorithms for data analysis.

In this literature review different researches on big data analysis technologies, namely the data lake concept would be analyzed. It could optimize data storage, and studies of how machine learning algorithms could be a great tool for data analysis, particularly for making predictions and computer decisions. Through some studies will be given definitions, main properties, and architecture of existing data lake solutions. Also, the overview in this paper experiments shows that machine learning algorithms were created and successfully implemented for making judges' decisions, predicting parking availability places, and potential cardiovascular diseases.

The purpose of this literature review is to examine how to apply big data analysis technologies in order to provide proactive public services via e-government infrastructure in Kazakhstan. These findings potentially could make easier the ways of receiving online public services for citizens, improve the performance of e-government, and as a result increase positions of Kazakhstani e-government on worldwide ratings.

E-government of Kazakhstan

The E-government of Kazakhstan could be used for implementing a successful big analysis project for providing proactive public services. According to Ospanova and Baimakhanova (2018), Kazakhstani e-government includes

a variety of information systems, governmental databases and technological solutions for providing public services to citizens via one window of Egov.kz web-portal. Particularly the majority of all public services are available at egov.kz, which therefore process by governmental agencies via their information systems without direct interaction with citizens, thus it excludes possibilities for bureaucracy and corruption. Since the creation of e-government in 2006 thousands of state services in healthcare, education, labor, finance, entrepreneurship, and others were automated, and now only by using Egov.kz mobile application people can receive about 80 online services (Ospanova & Baimakhanova, 2018). The Kazakhstani e-government project took 33rd place at the top countries worldwide in «E-government development» ranking by the United Nations in 2016 (Muslimova & Alzhanova, 2015). Regarding the study of Amanbek et al (2020) in 2009 e-government had 10960 users, and in comparison, in 2018 the number was raised dramatically to 8.6 million citizens, and moreover in 2020 quarantine situation forced people to become online public services receivers due to restrictions for offline visiting state centers. The sum of payments and transactions via e-gov.kz platform was 18,5 billion tenge and the total number of provided services was 27,5 million in the context of 235 different types (Petrenko & Shevyakova, 2019). This statistical data shows that the e-government of Kazakhstan daily consumes and processes vast amounts of citizens' personal data, which could be analyzed for making predictions in order to provide proactive public services and assistance to people in crisis cases. E-government systems and governmental databases could be integrated with each other, so this mixed data collection with adoption of predictive analytics technology will lead to enhancement of interactions between government and citizens (Dawes et al., 2016). Thus, given above statistics could be the proof of existing big data on e-government databases and that it has great potential for making big data analysis.

Data lake concept

Many studies advocate that the data lake concept is the core architecture for big data analysis projects because it could collect and match various types of data immediately. As cited in Miloslavskaya and Tolstoy (2016), Dixon (2010) described a data lake as a reservoir, which is filled with different types of water, such as natural water, waterfall, glacial water, and everyone is able to taste, swim or examine the chosen kind of water. According to Miloslavskaya and Tolstoy (2016) data, a lake is a repository, which holds an immense volume of raw data, which could be processed and analyzed in real-time. Khine & Wang (2018) and many other studies divided big data into three main categories, such as structured, unstructured, and semi-structured. Structured data includes database records, which are mostly in written format. The next unstructured data are datasets, which mostly raw data, which means it should be cleaned and prepared by

algorithms, for instance, images and videos are good examples of unstructured data. The last category is semi-structured, combining raw and read-on data, which needs computer processing, for instance, it could be history from web pages. Miloslavskaya and Tolstoy (2016) also defined a separate category of fast data, which is data changing in real-time, such as for instance, information from sensors, geo-location, and social networks. Moreover, the authors argued that for the most effective and accurate predictive analysis fast data should be collected.

Many researchers support the view that the data lake concept is the most appropriate architectural solution for gathering data, storing datasets, and integrating diverse data sources. According to Mehmood et al (2020) conducted to enhance the living standards of citizens by implementing disruptive technologies in four smart cities. For their Data lake authors suggested using and analyzing diverse types of data, such as data from traffic sensors, cameras, social networks, policemen's systems, geo-data, and weather conditions. The architecture and description of fusing such varied data types are provided and could be used for analog projects. As in the study of Hai, Geisler, and Quix (2016), they proposed the data lake model named as «Constance», which according to authors could be adopted into future projects.«Constance» data lake includes three stages of pre-processing data, which are required for qualitative big data analysis. These three processes of data enrichment includes requesting insufficient values, cleaning incorrect ones and matching overlapping values. Liu et al (2020) proposed their architecture of the data lake concept as a model for future projects. Also, in their research authors conducted an overview of existing technologies for implementing the data lake concept (names given in Figure 1). Additionally, Figure 1 fully represents the architectural model of how they integrated technological components with each other. Their model of data lake was aimed at the future inventory of vehicles based on real-time information from car sensors, transportation systems, and social networks. They stated that the model is scalable, which means the data sources could be changed and added.

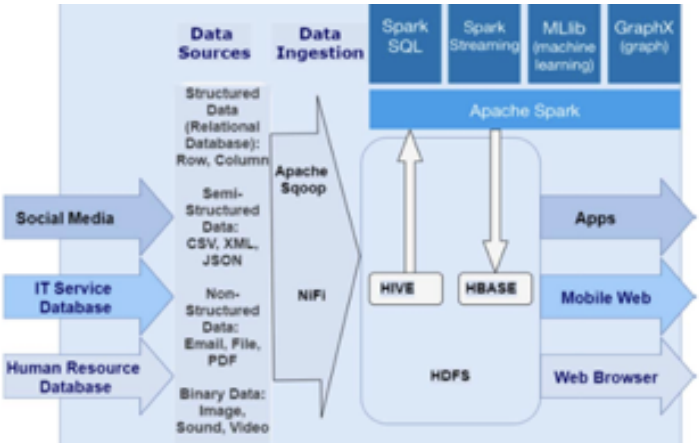

Figure 1 – The architecture represents the data lake model.
IEEE. https://ieeexplore.ieee.org/abstract/document/9245460/

Machine learning technology

While previous research works support the view that the data lake concept should be implemented for essential collecting, storing, and processing data, others consider that machine learning algorithms could be implemented for big data analysis, particularly for making computerized unbiased decisions and predictions. Kleinberg et al (2018) proposed the model for making judges decisions regarding being in jail or home for defendants pending a final court decision. Such millions of human decisions are taken annually by judges, who potentially could have personal preferences, although according to law it should be in consideration with a peoples' actions in a release. The model is based on initial triggers, such as the severity of violence, absence of the previous crimes in a person's past, random assignment between judges and others, also this model could be improved for future implementation. Study results were based on model simulations and show that the crime rate could be decreased to 24.8 % without changes in the number of prisoners and the same level of crime will be with decreasing prisoners to 42 %. Even more, their machine learning model supposes that reducing the number of African Americans and Hispanics in prison will lead to a decrease in all types of crime rates.

Materials and methods

Theoretical analysis of programs; network and artifact analysis; focus groups; action research; observations; data orchestration processes and tools to centralize the data in standard or relational database e-government; experiments; analysis and synthesis of research papers; predictive modeling; research method of machine learning algorithms in addressing classification and prediction problems.

Results and discussion

In Mehmood et al (2020) research of predicting parking availability places, authors propose their «Cutler» named project, which is based on machine learning algorithms coded on «Spark» programming language. The results of their research showed that the problem of lack of parking places could be solved by machine learning predictive analytics, which was based on traffic sensors, cameras, systems, geo-data, and weather conditions. Regarding the research of Mainiet al (2018) aimed at prevention of deaths from cardiovascular diseases by making predictions. For this, the authors adopted their machine learning model, which gathered data from patient's medical history and clinical data and based on that, assisted doctors to pay attention to such people. Mostly were used such factors as blood test, age, blood pressure, cholesterol level and genetic predisposition. The study did not provide the results of probabilities figures of correct diagnosis. Reviewed studies confirm the hypothesis that machine learning technology could be implemented in real cases of big data analysis for making predictions and automated decisions.

Conclusion

It is important to conduct research to determine whether big data analysis technologies, such as data lake and machine learning could appropriately store, analyze data and make predictions for further providing public services in advance. Implementing big data analysis technologies from the perspectives of overviewed different studies, these disruptive technologies could be effective tools, particularly the data lake concept for collecting, storing, preparing data and machine learning algorithms for making analysis, predictions. There was no research in this area within Kazakhstan, but it is clear from the reviewed statistics that the e-government process a huge amount of data there will be demand from Kazakhstani citizens. Even only one proactive service of «Childbirth», which is implemented manually, shows that it dramatically improves citizens' experience with e-government, although citizens have thousands of such life cases. Thus, based on the research given above, it seems possible to implement data lake and machine learning technologies for further predictive analytics and computerized decisions to provide public services proactively. Such implementations will improve citizens' experience with e-government, which will lead to empowering the status of Kazakhstani e-government in the world ranking.

REFERENCES

1 **Amanbek, Y., Balgayev, I., Batyrkhanov, K., & Tan, M.** (2020) Adoption of e-government in the Republic of Kazakhstan. Journal of Open Innovation: Technology, Market, and Complexity, 6(3), 46. [Electronic resource]. – https://www.mdpi.com/2199-8531/6/3/46

2 **Dawes, S. S., Vidiasova, L., & Parkhimovich, O.** (2016) Planning and designing open government data programs: An ecosystem approach. Government Information Quarterly, 33(1), 15–27. [Electronic resource]. – Planning and designing open government data programs: An ecosystem approach

3 **Hai, R., Geisler, S., & Quix, C.** (2016, June) Constance: An intelligent data lake system. In Proceedings of the 2016 International Conference on Management of Data (pp. 2097–2100). [Electronic resource]. – https://dl.acm.org/doi/abs/10.1145/2882903.2899389

4 **Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S.** (2018) Human decisions and machine predictions. The quarterly journal of economics, 133(1), pp. 237–293. [Electronic resource]. – https://www.nber.org/system/files/working_papers/w23180/w23180.pdf

5 **Liu, R., Isah, H., & Zulkernine, F.** (2020, September) A Big Data Lake for Multilevel Streaming Analytics. In 2020 1st International Conference on Big Data Analytics and Practices (IBDAP) (pp. 1–6). IEEE. [Electronic resource]. – https://ieeexplore.ieee.org/abstract/document/9245460/

6 **Maini, E., Venkateswarlu, B., & Gupta, A.** Data Lake-An Optimum Solution for Storage andAnalytics of Big Data in Cardiovascular Disease Prediction System, 2018. [Electronic resource]. – http://ijcem.org/papers112018/ijcem_112018_05.pdf

7 **Mehmood, H., Gilman, E., Cortes, M., Kostakos, P., Byrne, A., Valta, K., ... & Riekki, J.** Implementing big data lake for heterogeneous data sources. In 2019 ieee 35th international conference on data engineering workshops (ICDEW) (pp. 37–44), 2019, April. IEEE. [Electronic resource]. – https://ieeexplore.ieee.org/abstract/document/8750951/

8 **Miloslavskaya, N., & Tolstoy, A.** Big data, fast data and data lake concepts. Procedia Computer Science, 88, 300–305, 2016. [Electronic resource]. – https://www.sciencedirect.com/science/article/pii/S1877050916316957

9 **Muslimova, K., Alzhanova, A.** Electronic government of the Republic of Kazakhstan: the history of formation and development. Bulletin of the Karaganda University, 2015. [Electronic resource]. – https://articlekz.com/en/article/14398#google_vignette

10 **Ospanova, D., & Baimakhanova, D.** Administrative and legal support of digitalization of public services in the context of the development of cyber space in the Republic of Kazakhstan. News of the national academy of sciences of the Republic of Kazakhstan, 5 (321), 152–159, 2018. – DOI: 10.32014/2018. 2224-5294.26

11 **Petrenko, E. S., & Shevyakova, A. L.** Features and perspectives of digitization in Kazakhstan. In Ubiquitous Computing and the Internet of Things: Prerequisites for the Development of ICT (pp. 889–899), 2019. Springer, Cham. [Electronic resource]. – https://link.springer.com/chapter/10.1007/978-3-030-13397-9_91

Material received on 07.06.22.

**М. Б. Жадренова*
Қазақстан-Британ техникалық университеті,
Қазақстан Республикасы, Нұр-Сұлтан қ.
Материал баспаға 07.06.22 түсті.

ҚАЗАҚСТАННЫҢ ЭЛЕКТРОНДЫҚ ҮКІМЕТІ АРҚЫЛЫ BIG DATA ANALYTICS ТЕХНОЛОГИЯЛАРЫН ЕНГІЗУ АРҚЫЛЫ АЗАМАТТАРҒА ПРОАКТИВТІ МЕМЛЕКЕТТІК ҚЫЗМЕТ КӨРСЕТУ ЖӨНІНДЕ БОЛЖАУ ЖӘНЕ ШЕШІМДЕР ҚАБЫЛДАУ

Үлкен деректерді талдау бойынша сәтті жобалардың мысалдарының бірі Қазақстанның электрондық үкіметі болып табылады. Бұл жобада әртүрлі ақпараттық жүйелер, мемлекеттік деректер базалары, азаматтарға мемлекеттік қызмет көрсетудің технологиялық шешімдері бар. Мемлекеттік қызметтердің онлайн форматында қолжетімді болуы азаматтарға электрондық цифрлық қолтаңбаны тегін ұсынудың арқасында мүмкін болды. Прогрессивті түрде «Электрондық үкімет» мемлекет пен азаматтардың және мемлекеттік органдардың бір-бірімен өзара әрекеттесуінің, ақпараттық технологиялардың көмегімен олардың жүйелілігін қамтамасыз ететін біртұтас механизмге айналды. Дәл осы механизм мемлекеттік органдарға кезектерді азайтуға және анықтамаларды, рұқсаттарды және т.б. алуды жеңілдетуге және жеделдетуге мүмкіндік берді. Бұл белгісіз зерттеу және үлкен деректерді талдау технологиялары жаңа функцияларды кеңейте ала ма деген сұрақ болды, содан кейін олар одан әрі қызмет көрсету прогресі үшін пайдаланылады. Бұл мақалада мемлекеттік қызметтерді проактивті түрде көрсету үшін одан әрі болжамды талдау және компьютерлендірілген шешімдер үшін деректер көлі мен машиналық оқыту технологияларын енгізу мүмкіндіктері қарастырылады.

Кілтті сөздер: ақпараттық жүйелер, үлкен талдау, деректер қоры, проактивті қызметтер, машиналық оқыту алгоритмдері, технологиялар, деректер көлі, іске асыру.

***М. Б. Жадренова**
Казахско-Британский технический университет,
Республика Казахстан, г. Нур-Султан.
Материал поступил в редакцию 07.06.22.

ПРОГНОЗИРОВАНИЕ И ПРИНЯТИЕ РЕШЕНИЙ ПО ПРЕДОСТАВЛЕНИЮ ГОСУДАРСТВЕННЫХ УСЛУГ ГРАЖДАНАМ ПРОАКТИВНО ПУТЕМ ВНЕДРЕНИЯ ТЕХНОЛОГИЙ АНАЛИТИКИ БОЛЬШИХ ДАННЫХ ЧЕРЕЗ ЭЛЕКТРОННОЕ ПРАВИТЕЛЬСТВО КАЗАХСТАНА

Одним из примеров успешных проектов по анализу больших данных является Электронное правительство Казахстана. Этот проект содержит различные информационные системы, государственные базы данных и технологические решения для оказания государственных услуг гражданам. Доступность государственных услуг в онлайн-формате стала возможной благодаря предоставлению гражданам ЭЦП на безвозмездной основе. Постепенно Электронное правительство стало единым механизмом взаимодействия государства и граждан и государственных органов друг с другом, обеспечив их согласованность с помощью информационных технологий. Именно этот механизм позволил сократить очереди в госорганы и упростить и ускорить получение справок, разрешений и многого другого. Это было неизвестное исследование и вопрос о том, могут ли технологии анализа больших данных расширить новые функции, которые затем будут использоваться для дальнейшего развития обслуживания. В этой статье будут рассмотрены возможности внедрения озер данных и технологий машинного обучения для дальнейшей прогнозной аналитики и компьютеризированных решений для упреждающего предоставления государственных услуг.

Ключевые слова: информационные системы, большой анализ, базы данных, проактивные сервисы, алгоритмы машинного обучения, технологии, озеро данных, реализации.

Теруге 07.06.2022 ж. жіберілді. Басуға 30.06.2022 ж. қол қойылды.
Электрондық баспа
8,05 Mb RAM
Шартты баспа табағы 8,57. Таралымы 300 дана. Бағасы келісім бойынша.
Компьютерде беттеген: З. С. Искакова
Корректор: А. Р. Омарова
Тапсырыс № 3960

Сдано в набор 07.06.2022 г. Подписано в печать 30.06.2022 г.
Электронное издание
8,05 Mb RAM
Усл.печ.л. 8,57. Тираж 300 экз. Цена договорная.
Компьютерная верстка З. С. Искакова
Корректор: А. Р. Омарова
Заказ № 3960

«Toraighyrov University» баспасынан басылып шығарылған
«Торайғыров университеті» КЕ АҚ
140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы
«Торайғыров университеті» КЕ АҚ
140008, Павлодар қ., Ломов к., 64, 137 каб.
67-36-69

e-mail: kereku@tou.edu.kz
www.vestnik.tou.edu.kz
https://vestnik-pm.tou.edu.kz/