ANNOTATION
of Mukasheva Assel's
PhD thesis on the specialty 6D070300 – «Information Systems»
on the topic «Research and development of an information system for diagnosing diabetes based on BigData technology tools»

Relevance of the topic. The prevalence of diabetes in Kazakhstan is epidemic and leads to large financial costs. The analysis of the current state of research on predicting patient growth has shown insufficient elaboration in this subject area. To improve this issue, it was proposed to use real statistical data, which can be used to identify the most effective method for predicting the growth of patients with diabetes mellitus against the background of passive detection to further determine the amount of purchased insulin by region. The volume of the predicted amount of insulin, in turn, will allow you to plan financial costs in order to avoid unnecessary costs.

The relevance of the work is emphasized by the high level of disease prevalence and the lack of problem-oriented information technologies for diagnosis and development of recommendations for the treatment of diabetes. In this regard, the need for research and development of an information support system for an endocrinologist for the diagnosis of diabetes mellitus based on BigData technology tools is high, since the developed system will allow doctors to use unstructured data, on the basis of which they can make an optimal decision. The use of modern information technologies in the healthcare system will improve the quality of medical services provided. Therefore, the study of these issues determines the relevance of the topic of this dissertation.

The purpose of the study. Research and development of a method for predicting the growth of patients with diabetes in the Republic of Kazakhstan as well as the development of an information support system for an endocrinologist for the diagnosis of diabetes on the basis of BigData technology tools.

The goal defined the main objectives of the dissertation.

1. Analysis of the current state of application of forecasting methods and the use of BigData technologies in the field of medical care.
2. Development of a method for predicting patient growth using regression analysis methods and using the scikit-learn library to determine the amount of purchased insulin and other medications.
3. Development of a system based on algorithmic and software for the diagnosis and maintenance of diabetes treatment.
4. Development of the platform of the information system of information resources on diabetes in the BigData environment, as well as the collection of initial information.
5. Conducting exploratory analysis, processing of large data using summary statistics and graphical representations of data.

6. Execution of experimental studies, analysis of obtained results and formulation of conclusions on the work as a whole.

Thus, the goal and objectives of the study on predicting the possible number of patients are formulated, which will allow doctors and medical organizations to plan in advance the amount of purchased insulin and other means of controlling diabetes. In the future it is planned to launch this information system in open access for practical training in the methods of diagnosing diabetes based on the theoretical knowledge obtained by students of medical universities.

The objects of study are the processes of predicting patients with diabetes for the next years in order to prepare the procedure for purchasing the necessary amount of insulin and the information system for diagnosing patients in medical institutions based on the entered data using BigData technology tools.

The scientific novelty of the thesis is:

– based on a set of acceptable solutions to statistical problems, the results obtained allowed to predict the number of patients;

– distributed computing that provides efficient execution of the developed algorithms;

– a conceptual model of the processes of searching, extracting, processing and analyzing data in large data sets has been developed;

– developed and tested on the basis of a systematic approach information system for the diagnosis of diabetes on the basis of tools BigData technologies.

Research methods. The theoretical basis of the thesis is based on the methods of mathematical statistics, machine learning, and mathematical analysis. The practical implementation of the algorithms was based on the methods of object-oriented programming and object-oriented databases.

The practical significance of the work. The use of methods in the system for predicting the growth of patients with diabetes in the following years will allow all doctors and medical organizations to plan in advance the amount of purchased insulin and other diabetes control products. The developed information system for the diagnosis of diabetes has allowed to improve the quality of diagnosis of the disease on the basis of algorithmic and software.

The results of research of the dissertation thesis will facilitate the work of endocrinologists, because they can use not only their own knowledge,but also the world's knowledge base resources about diabetes. Higher educational institutions that teach, prepare and train endocrinologists will also benefit from the use of this system. The practical value of the dissertation research is confirmed by the received act on the implementation of the results of the work.

Testing the results of the thesis. The main provisions and results of the dissertation research were reported and discussed at scientific seminars of the department «Cybersecurity, processing and storage of information» KazNRTU named after K.I. Satpayev, as well as at scientific conferences of various levels.

Publications. According to the results of the dissertation thesis 14 works were published, including 5 articles recommended by the Committee for Control in the Field of Education and Science of the Republic of Kazakhstan; 1 article - in a foreign journal, which is included in the ISI Web of Knowledge (IF = 2,827) and Scopus (76 percentile) databases, 8 abstracts in international conferences.

The structure and scope of the thesis. The thesis work consists of 4 sections, conclusion, list of references from 139 titles. The work is presented on 117 pages, includes 88 figures and 6 tables.

The introduction reveals the relevance, specifies the problems associated with the topic under study. The purpose and objectives of the study, the scientific novelty and practical value of the work, research methods.

The first chapter of the dissertation presents current problems of diagnosis and prediction of diabetes in the modern world. The advantages and disadvantages of current modern medical information systems for diagnostics are analyzed. The goals and objectives of the study are presented.

In the second chapter of the dissertation, the methods of applying Big Data technology in the field of global health are studied. Information technologies that solve medical problems are considered and analyzed.

The third chapter describes forecasting models based on regression analysis methods. As a result of experiments for comparative analysis between the performed calculations, the capabilities of the scikit-learn library for solving machine learning problems were further considered, which also demonstrated the growth of patients. A research analysis of big data processing using summary statistics and graphical data representations was performed.

In the fourth chapter, based on the software implementation of the information system, the results of research on the diagnosis of diabetes using BigData technology tools are considered. An information model for diagnosing diabetes with the help of a database has been developed, and a list of standard medical tests that can be used to diagnose diabetes is shown based on the mathematical apparatus.

In conclusion, the main results and conclusions of the thesis are reflected.

According to the topic of the thesis 14 publications that were published are followings:

1. А.К. Мукашева, Н.П. Сапарходжаев. «Қазақстан Республикасының аумағында BigData технологиясы негізінде науқастарға қызмет көрсету жүйесін әзірлеу». Вестник КазНИТУ, №6 (124) ноябрь 2017г. – Алматы: КазНИТУ имени Сатпаева, 2017, стр. 193-197.

2. Н. П. Сапарходжаев, Г. К. Балбаев, А.К. Мукашева. «Разработка информационной системы на основе технологий BigData для диагностики и лечения диабета». Вестник АУЭС, №4 (6) (43) 2018, Алматы

3. Н.П. Сапарходжаев, А.К. Мукашева. «Параметрлерді және ұсыныстарды бағалау үшін ақпараттық ресурстарды пайдалануға және қалыптастыруға арналған ақпарат жүйесін құрастыру: BigData құралдары

негізінде MongoDB-мен жұмыс». Вестник КазНИТУ, №6 (130) 2018г. – Алматы: КазНИТУ имени Сатпаева, 2018, стр. 198-204.

4. Н.П. Сапарходжаев, А.К. Мукашева. «Прогнозирование распространенности диабета в республике Казахстан на основе методов регрессионного анализа». Вестник АУЭС, №5 2019, Алматы

5. А.К. Мукашева, Н.П. Сапарходжаев. «Прогнозирование распространенности диабета в республике Казахстан на основе методов регрессионного анализа». Вестник КазНИТУ, №5 () 2019г. – Алматы: КазНИТУ имени Сатпаева, 2019.

6. A.K. Mukasheva, N. Saparkhojayev. «The concept of monetization of IoT-based project: case of Medical System in Kazakhstan». The 15th International conference information technologies and management 2017 April 27-28, 2017, ISMA University, Riga, Latvia.

7. А.К. Мукашева, Н.П. Сапарходжаев. «Разработка системы обслуживания пациентов на основе технологии BigData на территории Республики Казахстан». МНПК «Математические методы и информационные технологии макроэкономического анализа и экономической политики» КазНИТУ им. К.И.Сатпаева, стр. 182-187, 11-12 апреля 2017г. – с.182-187

8. Н.П. Сапарходжаев., А.К. Мукашева, «Анализ системы для диагностики сахарного диабета на основе технологии BigData» ISBN 978-601-323-111-2. Труды Международных Сатпаевских чтений «Инновационные решения традиционных проблем: инженерия и технологии», 12 апреля 2018г. И66 – Алматы: КазНИТУ имени Сатпаева, 2018.-1254-1256.

9. N. Saparkhojayev, A.K. Mukasheva, «Introduction to BigData technology for diagnosis of diabetes». Information Technologies and Management, 2018 April 26 – 27, “Information Technologies, Management and Society”, ISMA University, Riga, Latvia.

10. N. Saparkhojayev, A.K. Mukasheva, «The development of information system of formation and use of information resources for evaluation of parameters and evaluation of recommendations based on BigData technology tools: work with MongoDB». International Conference on Cyber Security and Computer Science (ICONCS’18), Oct 18-20, 2018 Safranbolu, Turkey

11. A. Mukasheva, N. Saparkhojayev, Z. Akanov, A. Algazieva, «Prevalence of diabetes in the republic of Kazakhstan based on regression analysis methods». ISBER. International Conference on Research in E-Learning & Distance Education, Social Sciences, Economics and Management. July 13-14, 2019 / Turkey.

12. Mukasheva A., Saparkhojayev N., Akanov Z. et al. Forecasting the Prevalence of Diabetes Mellitus Using Econometric Models. Diabetes Therapy (2019). https://doi.org/10.1007/s13300-019-00684-1.

13. Мукашева А. К., Н. П. Сапарходжаев, И. А. Зимин. «Разработка информационной системы на основе технологии BigData». «Научное творчество: научно-исследовательская практика и приоритетные направления» /

Материалы XXII республиканской научной конференции студентов и молодых ученых, посвященной 1150-летию Абу Насыр аль-Фараби: Том I. (23 апреля 2020 г), стр. 527-532, ISBN 978-601-285-096-3.

14. А. А. Куандыков, А.К. Мукашева, И. А. Зимин, Обработка журналов веб серверов для получения статистики сайтов с помощью инструмента Apache Spark. МНПК «Модели инновационных решений повышения конкурентоспособности отечественной науки» (4 июня 2020 г, г. Челябинск). В 2 ч. Ч. 2 / - Уфа: OMEGA SCIENCE, 2020. – стр. 79-84, ISBN 978-5-907347-21-2.