

АННОТАЦИЯ

диссертационной работы Даркенбаева Даурена Кадыровича на тему: «**Численное моделирование и разработка программного комплекса для обработки большого объема данных**», представленной на соискание степени доктора философии (PhD) по специальности «6D075100-Информатика, вычислительная техника и управление»

Актуальность темы исследования. Темпы роста данных значительно возросли в последнее десятилетие. Исследования показали, что за последние два десятилетия объем данных увеличивается примерно в десять раз каждые два года - это превысило закон Мура, который удваивает мощность процессоров. Каждую секунду накапливается около тридцати тысяч гигабайт данных, и их обработка требует увеличения эффективности обработки данных. Загрузка видео, фотографий и писем пользователей в социальных сетях приводит к накоплению большого объема данных, в том числе неструктурированных. Это приводит к необходимости работать с большими данными разных форматов, которые должны быть определенным образом подготовлены к дальнейшей работе в целях получения результатов моделирования и вычислений. В связи с вышеизложенным проведенные в диссертационной работе исследования обработки больших данных, разработка модели и алгоритмов для решения весьма актуальны. Несомненно, с каждым годом информационные потоки будут нарастать и в этой связи актуально решение проблем хранения и обработки больших объемов данных и кроме того, актуальность диссертационной темы обусловлена растущей цифровизацией, возрастающим переходом на профессиональную деятельность в режиме онлайн во многих сферах жизни современного общества.

Диссертационная работа содержит результаты исследований по разработке модели системы обработки больших данных, анализа и проведения прогноза с использованием методов Data Mining и машинного обучения в решении задачи ипотечного кредитования, конкретно для анализа, прогнозирования и определения платежеспособности физических лиц, получающих ипотечный кредит. В диссертационной работе рассмотрено решение одной из актуальных проблем банковской системы - ипотечное кредитование. Основная проблема заключается в прогнозировании платежеспособности ипотечных заемщиков на длительный срок с использованием метода интеллектуального анализа данных. Главная задача - реализовать процесс обработки больших данных на основе разработанной системы, определяющей платежеспособность ипотечных заемщиков. В настоящее время темпы долгосрочного ипотечного кредитования ежегодно растут, вследствие этого своевременность проведенных в диссертационной работе исследований по разработке модели системы ипотечного

кредитования, которая четко прогнозирует платежеспособность любого желающего получить жилье и принимает соответствующие решения, весьма актуальна.

Выбор проблемы ипотечного кредитования обусловлен тем, что в настоящее время в Республике Казахстан реализуются программы ипотечного кредитования, что, в свою очередь, требует разработки системы для анализа, определения и прогнозирования платежеспособности получателей ипотечных кредитов на длительный срок. В связи с большим количеством данных по потребителям, их атрибутам необходима обработка большого объема данных. В диссертационной работе разработан эффективный программный комплекс для ипотечного кредитования, в котором предусмотрено моделирование с весовым коэффициентом, который обновляется в соответствии со временем с использованием алгоритмов нейронной сети, что значительно упрощает работу по анализу данных. Например, внедрение разработанного программного комплекса финансовыми кредитующими организациями значительно упрощает такие вопросы, как качество обслуживания, составление новых программ, сфера управления, обеспечение безопасности и др.

Многие кредитные учреждения, в частности банки и микрокредитные организации, используют автоматизированную систему, чтобы решить, предоставлять или нет кредит своим клиентам, система рассчитывает кредитный рейтинг на основе характеристик клиента и дает положительный или отрицательный результат соответственно. Эта система реализуется путем запроса в центральное кредитное бюро, проверки данных заемщика и проверки кредитной истории. Следует отметить, что получить долгосрочный ипотечный кредит невозможно для клиента, который не смог погасить вовремя предыдущий кредит по различным обстоятельствам, что в свою очередь станет ограничением возможности получения жилья. В связи с вышеизложенным в диссертационной работе разработана модель системы ипотечного кредитования, которая достаточно точно прогнозирует платежеспособность любого человека, желающего иметь жилье и позволяет принимать решение по выдаче кредита. Очевидно, что обработка данных каждого клиента требует современных технологий и качественного технического оборудования. Разработанная в диссертационной работе система позволяет решить многие проблемы по выдаче ипотеки, в частности, данные заемщика из центральной базы данных, будут обработаны в режиме реального времени и принято соответствующее решение. Проблема состоит в том, чтобы быстро обработать данные заемщиков и создать систему, которая меняется в зависимости от факторов макроэкономической среды. Есть только одно решение этой проблемы - создание автоматизированной системы ипотечного кредитования на основе результатов обработки, анализа и сравнения имеющихся у нас данных. В банках хранятся миллионы записей о клиентах, и каждый год они обрабатывают новые данные о клиентах. Это, в свою очередь, внесет большой вклад в создание новой модели системы, богатой реальными

данными. Подача заявки на получение ипотеки может производиться удаленно и узнать решение финансовой организации можно, не выходя из дома, что существенно поможет экономить время многим клиентам и немаловажно в настоящих условиях пандемии.

Ежегодный рост конкуренции между различными финансовыми организациями диктует необходимость быстрой обработки данных клиентов и принятия соответствующих решений в кратчайшие сроки. Существующие устройства и системы не удовлетворяют требованиям финансовых организаций. В этой связи проведенные в диссертации исследования по моделированию обработки больших данных, созданию системы ипотечного кредитования являются весьма актуальными.

Степень изученности темы исследования. Актуальность выбранного исследования в мире подтверждается исследованиями зарубежных и отечественных авторов. Аналогичные исследования проводились в трудах ученых дальнего зарубежья Chung, H.M., Joyce Jackson, Srinivasan V., KimYong, Henley W. E., Desai V. S., Conway D. G., Crook J. российских ученых Н.В. Бабиной, А.А. Земцова, Т.Ю. Осипова, В.Расторгуева, отечественных ученых Калимолдаева М.Н., Амиргалиева Е.Н., Балакаевой Г.Т., Мамырбаева О.Ж. и др.

Цель диссертационной работы: Разработка модели и алгоритмов обработки больших данных, анализ и прогнозирование в реализации прикладной задачи ипотечного кредитования.

Задачи исследования:

1. Анализ методов и систем обработки больших данных;
2. Разработка алгоритмов и моделей для обработки больших объемов данных на основе методов Data Mining: линейная регрессия, логистическая регрессия, многослойная нейронная сеть;
3. Анализ и тестирование разработанной системы обработки больших данных;
4. Анализ и прогнозирование платежеспособности физических лиц для ипотечного кредитования .

Объект исследования. Разработка системы обработки больших данных для анализа и прогнозирования платежеспособности в ипотечном кредитовании.

Предмет исследования: Методы и алгоритмы обработки больших данных.

Методы исследования: Теория и технологии BigData, Методы Data Mining: линейная регрессия, логистическая регрессия, нейронные сети. Технологии NoSQL, алгоритмы машинного обучения, проектирование программного комплекса.

Научная новизна работы:

1. Разработаны алгоритмы обработки больших данных, в том числе неструктурированных;
2. Разработана численная модель обработки больших неструктурированных данных;

3. Модифицированы алгоритмы машинного обучения в соответствии с форматом решаемой задачи, разработаны механизмы анализа и прогноза платежеспособности физических лиц, получающих ипотечные кредиты.

Теоретическая и практическая значимость работы. Полученные результаты могут быть использованы в теории и на практике для автоматизации работы финансовых организаций ипотечного кредитования. Во время глобальной пандемии ипотечным заемщикам можно будет удаленно, онлайн подать заявку на получение квартиры и получить решение финансового учреждения. Таким образом, проведенные в диссертационной работе исследования позволяют провести анализ и эффективно решить, предоставлять или нет ипотечный кредит заемщикам на длительный срок. Разработанную систему обработки больших данных можно использовать не только для расчета платежеспособности граждан, но и в других областях, например для диагностики заболеваний в области медицины, в геоинформатике, в сфере образования и других областях.

Основные положения, выносимые на защиту. На основе современных технологий (NoSQL, MongoDB) разработана модель системы обработки больших неструктурированных данных с использованием методов DataMining. Алгоритмы машинного обучения модифицированы для определения платежеспособности физических лиц. Проанализированы данные физических лиц зарегистрированных в базе данных прогнозированы платежеспособности физических лиц получающих ипотечные кредиты. Доказано, что использование многослойных нейронных сетей в обработке больших неструктурированных данных является высокоэффективным.

Личный вклад исследователя. Соискатель самостоятельно решил все задачи диссертационной работы. Диссертантом изучены и использованы современные технологии Data Mining, MongoDB и др., разработана численная модель системы обработки больших неструктурированных данных создан программный комплекс для анализа и прогноза в ипотечном кредитовании.

Объем и структура работы. Диссертация состоит из введения, трех разделов, заключения и списка литературы. Общий объем диссертации: 101 страниц письменного текста, в том числе 34 рисунка, 10 таблиц, список литературы из 83 источников, 3 приложения.

Во введении была определена актуальность работы и показаны проблемы, связанные с темой. Описана идея работы, цель и задачи исследования, научная новизна и практическая ценность исследования, методы исследования.

В первой главе диссертации, приводится обзор научных работ и подробный анализ технологий Big Data. Также был представлен обзор устройств, используемых для обработки и хранения больших объемов данных. В качестве единой базы данных для хранения неструктурированных данных был выбран Mongo DB и рассмотрены методы реализации.

Во второй главе рассматриваются методы создания модели системы, определяющей платежеспособность физических лиц, получающих долгосрочные ипотечные кредиты. Подробно рассмотрены основные этапы. Было показано, как методы Data Mining использовались для создания модели системы. Выявлены основные задачи и сложности создания системы определения платежеспособности физических лиц. Описаны общие задачи диссертации и предложены методы решения. Был проведен анализ для проверки качества разработанной модели системы.

В третьей главе реализованы модифицированные алгоритмы модели системы на основе методов Data Mining и машинного обучения, представлены результаты разработки программного комплекса. Представлена экспериментальная часть диссертации, полученные результаты представлены в виде таблиц, графиков и скрин-шотов.

В заключении изложены основные результаты и выводы диссертационного исследования.

Уровень достоверности и результаты апробации. Результаты исследования обсуждались на научных семинарах кафедры информатики КазНУ им. аль-Фараби и докладывались на следующих международных конференциях:

– XIV Miedzynarodowej naukowo-praktycznej konferencji «Naukowa przestrzeń Europy – 2018» (Прага, Чехия);

– Студенттер және жас ғалымдардың «Фараби әлемі» атты халықаралық ғылыми конференциясы (2018, Алматы, Қазақстан);

– Международная конференция «Актуальные проблемы вычислительной и прикладной математики», «Марчуковские чтения – 2019» (Академгородок, Новосибирск, Россия);

– II Халықаралық ғылыми-практикалық интернет конференциясы «Заманауи зерттеулердің өзекті мәселелері» (2019, Нұр-Сұлтан, Қазақстан);

– Ф.К.Бойконың 100 жылдығына арналған «Ф.К.Бойко I мерейтойлық оқулары» атты халықаралық ғылыми-техникалық конференциясы (2020, Павлодар, Қазақстан)

По теме диссертации опубликованы 13 статей, получено авторское свидетельство и акт о внедрении:

1. Даркенбаев Д.Қ. Big Data. Үлкен көлемді деректермен жұмыс істеу қағидалары // ҚазҰПУ хабаршысы. – 2017. -№ 3 (59). – Б. 211-214.

2. Balakayeva G.T., Darkenbayev D.K., Chris Phillips. Investigation of technologies of processing of Big Data // Internation Journal of Mathematics and Physics. – 2017. – Vol.8. No.2. – P.13-18.

3. Balakayeva G.T., Darkenbayev D.K. Modeling the processing of a large amount of data// Al-Farabi Kazakh National University. Journal of Mathematics, Mechanics and Computer Science. – 2018. -Vol.1(97). – P.120 – 126.

4. Балақаева Г.Т., Даркенбаев Д.Қ. Үлкен өлшемді деректерді өңдеу үдерісін моделдеу // ҚазҰПУ хабаршысы. – 2018. -№ 1(61). – Б. 248-252.

5. Darkenbayev D.K. Numerical solution of the regression model for analysis and processing of Big Data//Vestnik KazNRTU. – 2018. – № 6(130).–P.132 – 139.

6. Balakayeva G.T., Darkenbayev D.K. Correlation and regression analysis for Big Data processing // Vestnik KazNRTU. – 2019. – № 1(131). – P.338 – 345.

7. Balakayeva G.T., Chris Phillips, Darkenbayev D.K., Turdaliyev M. Using NoSQL for processing unstructured Big Data // News of the National Academy of sciences of the Republic of Kazakhstan. – 2019. –Vol.6.No.438. – P. 12 – 21.

8. G. Balakayeva, D. Darkenbayev. The solution to the problem of processing Big Data using the example of assessing the solvency of borrowers // Journal of Theoretical and Applied Information Technology. – 2020. – Vol.98. No13.– P. 2659-2670. (Scopus).

9. Darkenbayev D.K. Increasing the efficiency of processing large-size data using Big SQL technology//Materialy XIV Miedzynarodowej naukowo-praktycznej konferencji, «Naukowa przestrzeń Europy - 2018».– Vol.10. – P. 50-55.

10. Даркенбаев Д.К. Повышение эффективности и применение новых технологий для обработки больших объемов данных //V Международные Фарабиевские чтения. – Алматы, 2018. – С. 215.

11. D. K. Darkenbayev, G. T.Balakayeva. Modeling big data processing using regression analysis // Марчуковские научные чтения – 2019. – Академгородок, Новосибирск, Россия. – С. 135.

12. Даркенбаев Д.Қ. Үлкен көлемді деректерді сақтау және талдау әдістері // Заманауи зерттеулердің өзекті мәселелері» II Халықаралық ғылыми –практикалық интернет конференциясы. – Нұр-Сұлтан, 2019. – Б.120-124.

13. Darkenbayev D.K. Building a linear regression model for processing Big Data in the definition of solvency of citizens// Материалы международной научно-технической конференции «I юбилейные чтения Бойко Ф. К», посвященной 100-летию Бойко Ф. К. – Павлодар, 2020. – С. 23-29.

Получено Авторское свидетельство и акт о внедрении:

1. Программа для ЭВМ «NoSQL технологияларын және нейрондық желілерді қолданып үлкен көлемді деректерді өңдеу» авторское свидетельство № 8459 от «28» февраля 2020 года.

2. Акт о внедрении результатов диссертационной работы.