

**«6D075100 – Информатика, есептеу техникасы және басқару»
мамандығының PhD докторанты
Даркенбаев Даурен Кадыровичтің
«Үлкен өлшемді деректі өңдеуге арналған сандық модельдеу және
бағдарламалық қамтама құру»**

тақырыбындағы диссертациялық жұмысына

АҢДАТПА

Зерттеу тақырыбының өзектілігі. Деректердің өсу қарқыны соңғы онжылдықта айтарлықтай біліне бастады. Зерттеулер көрсеткендей, соңғы екі онжылдықта әр екі жыл сайын деректер көлемі шамамен он есе өсті. Процессорлардың қуатын екі есе көбейтетін Мур заңынан асып кетті. Әр секунд сайын отыз мың гигабайтқа жуық дерек жинақталады және оларды өңдеу жылдамдықты қажет етеді. Әлеуметтік желілерде қолданушылары видеоларын, фотоларын және хаттарын жүктеулері үлкен өлшемді құрылымданбаған деректердің көптеп жинақталуына әкеледі. Бізге әртүрлі форматтағы деректермен жұмыс істеуге тура келеді және жақсы нәтижелерге қол жеткізу үшін деректерді даярлаумен қатар, өңдеулерді моделдеу қажет. Диссертациялық жұмыста жүргізілген зерттеулерден үлкен өлшемді деректерді өңдеуде модель мен алгоритмдер құру өзекті екенін көрсетіп отыр. Ақпараттық ағындар жыл сайын арта түсетіні және осыған байланысты үлкен өлшемді деректерді сақтау мен өңдеу мәселелерін шешу өзекті болатындығы сөзсіз. Диссертация тақырыбының өзектілігі, қазіргі қоғамның көптеген салаларының цифрлануы және кәсіби қызметтердің онлайн режимге ауысуларының артуына байланысты болып отыр.

Диссертациялық жұмыста үлкен өлшемді деректерді өңдеудің моделін құру, ипотекалық несиені беру мәселелерін шешуде Data Mining және машиналық оқыту әдістерін қолдану арқылы ипотекалық несиені алатын жеке тұлғалардың төлем қабілеттерін анықтау және болжау бойынша зерттеулердің нәтижелері берілген. Диссертациялық жұмыс банк жүйесінің өзекті мәселелерінің бірі - ипотекалық несиелендіру жұмыстарын шешуге арналған. Негізгі мәселе деректерді интеллектуалды талдау әдісін қолдана отырып, ұзақ мерзімге ипотекалық несиені алушылардың төлем қабілеттерін болжау болып отыр. Негізгі міндет - ипотекалық несиені алушылардың төлем қабілеттерін анықтайтын жүйені негізінде үлкен өлшемді деректерді өңдеу үдерісін жүзеге асыру. Қазіргі уақытта ұзақ мерзімді ипотекалық несиелендіру қарқыны жыл сайын артып келеді, диссертациялық жұмыста ипотекалық несиелендіру жүйесінің моделін құру бойынша зерттеулер өте өзекті және уақытылы жүргізілуі баспана алғысы келетін кез-келген адамның төлем қабілетін нақты болжап, тиісті шешімдер қабылдауға ықпал етеді.

Ипотекалық несиелендіру мәселесін таңдау себебіміз, қазіргі кезде Қазақстан Республикасында ипотекалық несиелендіру бағдарламалары жүзеге асырылып жатыр, бұл өз кезегінде ұзақ уақыт бойы ипотекалық несиені алушылардың

төлем қабілеттерін талдау, анықтау және болжау жүйесін дамытуды талап ететіні анық. Жеке тұлғалар туралы деректер мен олардың атрибуттарының көптігі, үлкен өлшемді деректерді өңдеу қажеттілігін туындатып отыр. Диссертациялық жұмыста ипотекалық несиелеу мәселелерін оңтайлы шешу үшін бағдарламалық қамтама құрылды, бағдарламалық қамтамада нейронды желі алгоритмдерінің қолданылуы және салмақ коэффициенттері уақытылы жаңарып отыруы деректерді талдау жұмыстарын айтарлықтай жеңілдетті. Мысалы, ипотекалық несие беруші қаржы ұйымдарының жұмыстарына құрылған бағдарламалық қамтаманы енгізу қызмет көрсету сапасын арттырады, басқару жүйесін жаңартады, қауіпсіздік және т.б. мәселелерді де толық шешеді.

Көптеген несие беруші ұйымдар, атап айтқанда банктер және шағын несие беруші ұйымдар өз клиенттеріне несие беру немесе бермеу туралы шешім қабылдауда автоматтандырылған жүйе жұмысына жүгінеді, бұл жүйе клиенттердің басындағы сипаттамаларға негізделіп, несиелік балды есептеп сәйкесінше оң немесе теріс нәтиже береді. Бұл жүйе орталық несиелік бюроға сұрау салып, несие алушының деректерін тексеріп, несиелік тарихын тексеру арқылы жүзеге асады. Егер, несие алып уақытында түрлі жағдайларға байланысты қайтарым жасай алмаған клиент үшін ұзақ мерзімге ипотекалық несие алу мүмкіндігі жоқ деп айтуға да болады. Бұл жұмыс ипотекалық несие алып баспаналы болғысы келген адамның арманы мен мүмкіндігін шектегендік болады. Сондықтан, диссертациялық жұмыста кез келген баспаналы болғысы келген адамның төлем қабілетін нақты болжап, сәйкесінше шешім қабылдайтын ипотекалық несие беру жүйесінің моделі құрылды. Әрбір адамның яғни ипотека беруші ұйымдардың клиенттерінің деректерін өңдеу заманауи технология мен сапалы техникалық жабдықтарды талап ететіні анық. Егер осы жүйе біздің елде қолданысқа енсе, көптеген жұмыстар жеңілдемек. Атап айтсақ, ипотекалық несие алушының деректері орталық дерекқордан алынып, нақты уақыт режимінде өңделіп, шешім қабылданатын болады. Тікелей ипотекалық несие беруші ұйымның қызметкері мен тұтынушы арасында байланыс болмағандықтан, түрлі заң бұзушылыққа жол берілмей, шынымен баспаналы болғысы келетін және төлем қабілеті сәйкес келетін азаматтар пәтерге ие болатын еді. Сонымен қатар, қашықтан өтінім беру - клиенттердің үйден шықпай, уақыт үнемдеумен қатар, қазіргі таңда ғасыр дерті болып отырған әлемдік пандемияның таралмауына да өз септігін тигізетіні анық. DataMining әдістерін пайдаланып, құрылатын бағдарламалардың барлығы заман талабынан туындайтын ойлар негізінде жүзеге асады. Ипотекалық несие беруші үлкен ұйымдарда клиенттерінің деректері туралы жазбалар миллиондап сақталып тұр, ал жыл сайын клиенттері туралы жаңа деректерді сақтап, өңдейді. Бұл өз кезегінде шынайы деректерге бай жана жүйелік модель құруға үлкен септігін тигізеді. Ипотекалық несие алушының деректерін салыстырып өңдеп, нақты нәтиже алуымызға бұл таптырмас құрал деп айтуымызға негіз бар.

Жыл сайынға қаржы ұйымдарының арасындағы бәсекелестіктің артуы өздерінің тұтынушыларының деректерін жылдам өңдеп, қысқа уақытта сәйкесінше шешім қабылдауды қажет етіп отыр. Қолданыстағы бар жүйелер қаржы ұйымдарының талаптарын толық қанағаттандырып отырған жоқ. Диссертациялық жұмыста жүргізілген үлкен өлшемді деректерді өңдеуді моделдеу және ипотекалық несие беру жүйесін құру жұмыстары, осы мәселелерді шешуге бағытталған өзекті зерттеу жұмыстары болып табылады.

Жұмыстың зерттелу деңгейі. Зерттеу жұмысының өзектілігі әлем ғалымдарының зерттеу тақырыбына байланысты жазған еңбектерінен айқындалады. Зерттеу тақырыбына ұқсас зерттеулер жүргізген алыс шетел ғалымдары Chung, H.M., Joyce Jackson, Srinivasan V., Kim Yong, Henley W. E., Desai V. S., Conway D. G., Crook J. еңбектеріне шолу жасалды. Ресей ғалымдары Н.В. Бабина, А.А. Земцов және Т.Ю. Осипова, В.Расторгуев секілді ғалымдардың еңбектеріне шолу жасалды. Отандық ғалымдар Калимолдаев М.Н., Амиргалиев Е.Н., Балакаева Г.Т., Мамырбаев Ө.Ж. және т.б. отандық ғалымдардың еңбектеріне шолу жасалды.

Диссертациялық жұмыстың мақсаты. Үлкен өлшемді деректерді өңдеудің моделі мен алгоритмдерін құру, ипотекалық несие берудің қолданбалы міндеттерін шешуде талдау жүргізу және болжам жасау.

Зерттеудің міндеттері. Алға қойған мақсатқа жету үшін төмендегі міндеттерді шешу қажет:

1. Деректерді өңдеуге арналған әдістер мен өңдеу жүйелеріне сараптама жүргізу;

2. Data Mining әдістері: сызықты регрессия, логистикалық регрессия, көпқабатты нейронды желі негізінде үлкен өлшемді деректерді өңдеудің алгоритмдері мен моделін құру;

3. Үлкен өлшемді деректерді өңдеу жүйесінің жұмыс сапасын бағалау, деректерді тестілеу;

4. Құрылымданбаған үлкен өлшемді деректерді өңдеу негізінде жеке тұлғалардың төлем қабілеттеріне талдау жүргізу және болжам жасау;

Зерттеу нысаны. Ипотекалық несие алушылардың төлем қабілеттеріне талдау жүргізу және болжам жасау үшін үлкен өлшемді деректерді өңдеу жүйесін құру.

Зерттеу пәні. Үлкен өлшемді деректерді өңдеудің әдістері мен алгоритмдері.

Зерттеу әдісі. BigData теориясы және технологиялары, Data Mining әдістері: сызықты регрессия, логистикалық регрессия, нейрондық желілер. NoSQL технологиялары, машиналық оқыту алгоритмдері, бағдарламалық қамтаманы жобалау.

Жұмыстың ғылыми жаңалығы.

1. Үлкен өлшемді құрылымданбаған деректерді өңдеудің алгоритмдері құрылды;

2. Үлкен өлшемді құрылымданбаған деректерді өңдеудің сандық моделі құрылды;

3. Машиналық оқыту алгоритмдеріне шешілетін міндеттің форматына сай модификация жасалып, ипотекалық несие алушы жеке тұлғалардың төлем қабілеттеріне талдау жүргізу және болжам жасау механизмдері құрылды;

Жұмыстың теориялық және практикалық маңызы. Алынған нәтижелерді теориялық және практикалық тұрғыдан ипотекалық несие беруші қаржылық ұйымдардың жұмысын автоматтандыру үшін қолдануға болады.

Дүние жүзінде таралып отырған пандемия кезінде қаржы ұйымының қызметкерлері ипотекалық несие алушы азаматтардың төлем қабілеттерін қашықтан болжап, шешімін айта алады. Сонымен қатар, диссертациялық жұмыста жүргізілген зерттеулер ұзақ мерзімге ипотекалық несие алушыларға ипотекалық несие берілуі немесе берілмеуі туралы талдау жүргізіп, шешім қабылдауда өзінің септігін тигізеді. Құрылған жаңа жүйені азаматтардың төлем қабілеттерін анықтауға ғана емес, ғылымның басқа салаларына да қолдануға болады. Нақтырақ айтқанда, медицинада, геоинформатикада қолдануға болады, сонымен қатар, білім алушылардың оқу үлгерімдеріне болжам жасауға болады.

Қорғауға шығарылған негізгі тұжырым. Заманауи технологиялар (NoSQL, MongoDB) негізінде DataMining әдістерін қолдана отырып, үлкен өлшемді құрылымданбаған деректерді өңдеу жүйесінің моделі құрылды. Жеке тұлғалардың төлем қабілеттерін анықтауда машиналық оқыту алгоритмдеріне модификация жасалды. Дерекқорда тіркелген ипотекалық несие алушы жеке тұлғалардың деректеріне талдау жүргізіліп, төлем қабілеттеріне болжам жасалды. Үлкен өлшемді құрылымданбаған деректерді өңдеуде көпқабатты нейронды желіні қолдану, жоғары нәтиже беретіні анықталды.

Зерттеушінің жеке үлесі. Ізденуші диссертациялық жұмыстың барлық міндеттерін шешті. DataMining, MongoDB және т.б. технологияларды зерттеді және мәселелерді шешуде қолданды. Ипотекалық несие беру мәселелерін шешуде талдау мен болжам жасау үшін, үлкен өлшемді құрылымданбаған деректерді өңдеу жүйесі құрылды.

Диссертацияның құрылымы мен көлемі. Диссертациялық жұмыс кіріспе, 3 тарау, қорытынды және пайдаланылған әдебиеттерден тұрады. Диссертацияның толық көлемі: 101 бет жазба мәтіні, соның ішінде 34 сурет, 10 кесте, 83 пайдаланылған әдебиеттер тізімінен және 3 қосымшадан тұрады.

Кіріспеде тақырыптың өзектілігі, диссертациялық жұмыстың мақсаты, міндеті, зерттеу әдістері айқын жазылған. Алынған нәтижелері, ғылыми жаңалығы мен маңызы сипатталған. Сонымен қатар аталмыш диссертациялық жұмыс бойынша жазылған мақалалар тізімі берілген.

Диссертациялық жұмыстың **бірінші тарауында** ипотекалық несие алушы жеке тұлғалардың төлем қабілетін анықтау мақсатында ғылыми еңбектерге шолу жасалып, үлкен көлемді деректерге яғни BigData-ға толық анықтама берілді. Сонымен қатар үлкен көлемді деректерді өңдеу және сақтау үшін қолданылып жүрген құрылғыларға шолу жасалды.

Құрылымданбаған деректерді сақтауға арналған бірінғай дерекқор ретінде MongoDB дерекқоры таңдалып алынып оның қолданысқа енгізілуі туралы мәселелер қарастырылған.

Екінші тарауда ұзақ мерзімге ипотекалық несие алушы жеке тұлғалардың төлем қабілетін анықтайтын жүйе моделін құру әдістері талқыланады. Негізгі кезеңдер толық қарастырылды. Деректерді өңдеу жүйесінің моделін құруда Data Mining әдістерінің қалай қолданылғаны көрсетілді. Жеке тұлғалардың төлем қабілетін анықтау мақсатындағы жүйе құрудың негізгі міндеттері мен кездесетін қиындықтары анықталды. Диссертациялық жұмыстың жалпылама міндеттері ашылып жазылып, шешу әдістері ұсынылды.

Үшінші тарауда модификация жасалынған машиналық оқыту алгоритмдері жүзеге асырылды және DataMining әдістері негізінде бағдарламалық қамтама құрылған. Диссертациялық жұмыстың эксперименталды бөлімі көрсетілген. Алынған барлық нәтижелер кесте түрінде және суреттермен берілген.

Қорытындыда негізгі нәтижелер мен диссертациялық зерттеулердің қорытындылары берілген.

Зерттеу нәтижелерінің апробациясы. Ғылыми зерттеу жұмысының нәтижелері әл-Фараби атындағы Қазақ ұлттық университетінің информатика кафедрасының ғылыми семинарларында талқыланды және келесі халықаралық конференцияларда баяндамалар жасалды:

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

– V 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)

Диссертация тақырыбы бойынша 13 мақала, авторлық куәлік және өндіріске енгізілгені туралы акт алынды:

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.Қ. 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.

Авторлық куәлік және өндіріске енгізілгені туралы акт алынды:

1. ЭВМ-ге арналған бағдарлама «NoSQL технологияларын және нейрондық желілерді қолданып үлкен көлемді деректерді өңдеу» авторлық куәлік № 8459 от «28» ақпан 2020 жыл.
2. Диссертациялық жұмыстың нәтижелерінің өндіріске енгізілгені туралы акт.