

Даркенбаев Даурен Кадырович

# ҮЛКЕН КӨЛЕМДІ ДЕРЕКТЕРДІ ӨНДЕУ ӘДІСТЕРІ

Монография



Алматы  
2024

ӘОЖ 004.45  
КБЖ32.973.26-018.2  
Д18

*ҚР ҒЖБМ ҒК Ақпараттық және есептеу технологиялары  
институты Ғылыми кеңесімен ұсынылған  
Хаттама №4,15 сәуір 2024жыл*

**Рецензенттер:**

**Г.Т.Балакаева**, физика-математика ғылымдарының докторы,  
профессор

**Н.С.Заурбеков**, техника ғылымдарының докторы, профессор

Монография ҚР ҒЖБМ бағдарламалық-мақсатты қаржыландыруы  
аясында жүзеге асырылған №BR 24992907 (2024-2026жж.)  
«Қазақстан қалаларының жылумен жабдықтау жүйесінің жұмысын  
Цифрландыру және оңтайландыруға арналған бағдарламалық өнімді  
әзірлеу» жобасының қолдауымен жарық көрді.

**Даркенбаев Д.К.**

Д18 Үлкен көлемді деректерді өңдеу әдістері: монография/

Д.К. Даркенбаев. – Алматы: Дарын, 2024. – 151 б.

**ISBN978-601-80492-2-4**

Монографияда үлкен көлемді деректерді өңдеу әдістері зерттелген.  
Зерттеу негізінде ипотекалық несие алушы жеке тұлғалардың деректері  
өңделіп, сәйкесінше болжамдар жасалған.

Монографияда үлкен көлемді деректерді өңдеу мәселелерімен қатар,  
NoSQL және DataMining технологиялары сәтті интеграцияланған.  
Сонымен қатар, үлкен көлемді деректерді өңдеу жүйесінің моделі  
құрылған. Жеке тұлғалардың төлем қабілеттерін бағалаумен қатар,  
классификациялық алгоритмдерді салыстыра және дәлдігін арттыра  
отырып, заманауи есептеу тәжірибелерімен негізделген. Деректерді оқыту  
және тестілеу процестерінде алынған өңдеу нәтижелерінің сапасы  
айқын көрсетілген.

Монографияны ақпараттық технологиялар саласында үлкен көлемді  
деректерді өңдеу әдістерін зерттеп жүрген мамандар, ЖОО оқытушылары  
мен студенттері мен көмекші құрал ретінде қолдануларына болады.

ӘОЖ 004.45  
КБЖ32.973.26-018.2

**ISBN978-601-80492-2-4**

©Даркенбаев Д.К. 2024  
©«Дарын баспасы», 2024

# МАЗМҰНЫ

|                                                                                                                       |            |
|-----------------------------------------------------------------------------------------------------------------------|------------|
| <b>БЕЛГІЛЕУЛЕРМЕН ҚЫСҚАРТУЛАР .....</b>                                                                               | <b>4</b>   |
| <b>АНЫҚТАМАЛАР .....</b>                                                                                              | <b>5</b>   |
| <b>КІРІСПЕ .....</b>                                                                                                  | <b>6</b>   |
| <b>1 ҮЛКЕН КӨЛЕМДІ ДЕРЕКТЕРДІ ӨНДЕУДІҢ ЗАМАНАУИ<br/>ЖАҒДАЙЫ ЖӘНЕМ ӘСЕЛЕРІ .....</b>                                   | <b>14</b>  |
| Үлкен көлемді деректерді өңдеудің рөлі мен маңызы .....                                                               | 14         |
| Үлкен көлемді деректерді сақтауға және өңдеуге арналған<br>қолданыста бар өнімдер .....                               | 27         |
| Үлкен көлемді деректердің аналитикалық өңдеу технологиялары .....                                                     | 38         |
| NoSQL дерек қорының түрлері .....                                                                                     | 43         |
| Үлкен көлемді деректерді сақтауда MongoDB қолданылуы .....                                                            | 51         |
| MonetDB .....                                                                                                         | 53         |
| CouchDB .....                                                                                                         | 58         |
| <b>2 DATA MINING ӘДІСТЕРІНІҢ ДЕРЕКТЕРДІ ӨНДЕУДЕ<br/>ҚОЛДАНЫЛУЫ .....</b>                                              | <b>61</b>  |
| Ұзақ мерзімге ипотекалық несие беру жүйесінің математикалық моделі<br>64                                              |            |
| Сызықты регрессия әдісінің қолданылуы .....                                                                           | 74         |
| Логистикалық регрессия әдісінің қолданылуы .....                                                                      | 82         |
| Регрессия теңдеуін түрлендірудің қажеттілігі .....                                                                    | 82         |
| Логистикалық белсендіру функциясының салмақ векторын<br>анықтауда ең кіші квадраттар әдісінің қолданылуы .....        | 91         |
| Максималдық тигімалдық және логистикалық регрессия .....                                                              | 92         |
| ROS талдау .....                                                                                                      | 104        |
| ROS қисықтың канондық алгоритмін құру .....                                                                           | 106        |
| Көпқабатты нейрондық желілерді қолданып ипотекалық несие<br>алушылардың төлем қабілеттерін анықтау .....              | 108        |
| <b>3 ҮЛКЕН КӨЛЕМДІ ДЕРЕКТЕРДІ ӨНДЕУДІҢ КОМПЬЮТЕРЛІК<br/>МОДЕЛІН ҚҰРУ .....</b>                                        | <b>115</b> |
| Бағдарламалау тілінің көмегімен бағдарламалық қамтамасыз ету .....                                                    | 115        |
| Деректерді аярау .....                                                                                                | 117        |
| Деректерді нормализациялау .....                                                                                      | 122        |
| Бағдарламалық қамтамасыз ету архитектурасы .....                                                                      | 125        |
| Деректерді өңдеу алгоритмдерінің жүзеге асырылуы және өңдеу<br>моделінің құрылуы .....                                | 130        |
| Ипотекалық несие алушылардың төлем қабілеттерін анықтауда<br>Data Mining әдістерін қолданып эксперимент жүргізу ..... | 134        |
| Қолданылған әдістердің нәтижелерін салыстыру .....                                                                    | 139        |
| Деректерді өңдеу нәтижелері .....                                                                                     | 141        |
| <b>ҚОРЫТЫНДЫ .....</b>                                                                                                | <b>144</b> |
| <b>ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ .....</b>                                                                          | <b>146</b> |

## БЕЛГІЛЕУЛЕРМЕНҚЫСҚАРТУЛАР

РДҚБЖ–Реляциялық деректер қорын басқару жүйесі ДҚБЖ–

Деректер қорын басқару жүйесі

NoSQL–NotonlyStructuredQueryLanguage

SQL – Structured Query Language

MongoDB–MongoDatabase

ЕКӨ–Еңкішкік квадраттар әдісі

IBS – Information Business Systems

IBM–International Business Machines

SNA – Systems Network Architecture

IMDB – Internet Movie Database

RAID–Redundant Array of Independent Disks

MIMD–Multiple Instruction stream, Multiple Data stream MISD –

Multiple Instruction, Single Data

SIMD–Single Instruction, Multiple Data

SISD – Single Instruction, Single Data LR

– Linear Regression

LogR–Logistic Regression NN

– Neural Networks

HDFS–Hadoop Distribution File System

ROC – receiver operating characteristic

EDWH – Enterprise Data Warehouse

HTTP – HyperText Transfer Protocol

## АНЫҚТАМАЛАР

Деректерді өңдеу – бастапқы деректерді құнды бизнес ақпаратына түрлендіруді қамтиды. Әдетте деректер ғалымдары деректерді өңдеуді, соның ішінде жинауды, ұйымдастыруды, тазалауды, тексеруді, талдауды және графиктер немесе құжаттар сияқты оқылатын пішімдерге түрлендіруді қосады. Деректерді өңдеу үш әдіспен жүзеге асырылуы мүмкін: қолмен, механикалық және электронды.

DataMining–деректерді қазбалау, іздеу табу әдісі

Дефолт – несие бойынша міндеттемелердің орындалмау жағдайы

BigData–үлкен көлемді деректер

Ипотекалық несие – жылжымайтын мүлік кепілімен берілетін несиенің санында қаржы ресурстарын тарту тәсілі, яғни ипотекалық, коммерциялық банкілер, сақтандыру және басқа несие-қаржы мекемелері жылжымайтын мүлік (жер, құрылыс, тағы басқа) кепілімен беретін несие.

REST – желідегі таратылған қолданбалы компоненттердің өзара әрекеттесуіне арналған архитектуралық стиль. Басқаша айтқанда, REST – бұл барлық жүйелер деректермен оңай алмасуға және қолданбаны масштабтауға болатындай етіп, бағдарламалаушының серверлік қолданба үшін жазу кодын қалай ұйымдастыруға болатынына қатысты ережелер жиынтығы.

## КІРІСПЕ

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 [1]. Әр секунд сайын отыз мың гигабайтқа жуық дерек жинақталады және оларды өңдеу жылдамдықты қажет етеді. Әлеуметтік желілерде қолданушылары видеоларын, фотоларын және хаттарын жүктеулері үлкен көлемді құрылымданбағандеректердің көптеп жинақталуына әкеледі [2].

Бізге әртүрлі форматтағы деректермен жұмыс істеуге тура келеді және жақсы нәтижелерге қол жеткізу үшін деректерді даярлаумен қатар, өңдеулерді модельдеу қажет [3]. Монографияда жүргізілген зерттеулерден үлкен көлемді 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 [4]. Монография тақырыбының өзектілігі, қазіргі қоғамның көптеген салаларының цифрлануы және кәсіби қызметтердің онлайн режимге ауысуларының артуына байланысты болып отыр.

Монографияда үлкен көлемді деректерді өңдеудің моделін құру, ипотекалық несие беру мәселелерін шешуде Data Mining және машиналық оқыту әдістерін қолдану арқылы ипотекалық несие алатын жеке тұлғалардың төлем қабілеттерін анықтау және болжау бойынша зерттеулердің нәтижелері берілген. Монография банк жүйесінің өзекті мәселелерінің бірі – ипотекалық несиелендіру жұмыстарын шешуге арналған. Негізгі мәселе деректерді интеллектуалды талдау әдісін қолдана отырып, ұзақ мерзімге ипотекалық несие алушылардың төлем қабілеттерін болжау болып отыр. Негізгі міндет – ипотекалық несие алушылардың төлем қабілеттерін анықтайтын жүйе негізінде үлкен өлшемді деректерді өңдеу үдерісін жүзеге асыру. Қазіргі уақытта ұзақ мерзімді ипотекалық несиелендіру

- Ауытқуды талдау.

Data Mining табиғаты бойынша көп салалы, өйткені ол сандық әдістердің элементтерін, математикалық статистика мен ықтималдықтар теориясын, ақпарат теориясы мен математикалық логиканы, жасанды интеллект пен машиналық оқытуды қамтиды.

Бизнесті талдау тапсырмалары әртүрлі тәсілдермен тұжырымдалған, бірақ олардың көпшілігінің шешімі бір немесе басқа Data Mining тапсырмасына немесе олардың комбинациясына келеді. Мысалы, тәуекелді бағалау – регрессия немесе жіктеу мәселесінің шешімі, нарықты сегменттеу – кластерлеу, сұранысты ынталандыру – ассоциация ережелері. Шын мәнінде, Data Mining тапсырмалары көптеген нақты бизнес мәселелерінің шешімін «жинақтауға» болатын элементтер болып табылады.

Жоғарыда аталған мәселелерді шешу үшін әртүрлі Data Mining әдістері мен алгоритмдері қолданылады. Data Mining математикалық статистика, ақпарат теориясы, машиналық оқыту және мәліметтер базасы сияқты пәндердің қиылысында дамып, дамып келе жатқандығына байланысты, Data Mining алгоритмдері мен әдістерінің көпшілігі осы пәндердің әртүрлі әдістеріне негізделген әзірленгені табиғи нәрсе. . Мысалы, k-means кластерлеу алгоритмі статистикадан алынған.

Data Mining-те келесі әдістер өтетанымал болды: нейрондық желілер, шешім ағаштары, кластерлік алгоритмдер, соның ішінде масштабталатындар, оқиғалар арасындағы ассоциативті байланыстарды анықтау алгоритмдері және т.б.

Data Mining – бұл деректерді қазбалау, талдау, белгісіз деректерді табу технологиясы. Адам өміріндегі түрлі салаларда шешім қабылдауда өз ықпалын тигізеді. Басты міндеттерінің бірі-деректерді классификациялау, болжам жасау болып табылады.

Бүгінде әлемде деректерді өңдеудің әртүрлі әдістері бар. Олардың барлығының өзіндік қиындықтары мен ерекшеліктері бар.

Тиісті алгоритмдерге мыналар жатады:

1. Кластер типін талдау;
2. Факторлық талдау;

## ҚОРЫТЫНДЫ

Монографияда үлкен көлемді деректерді өңдеу моделі мен бағдарламалық қамтама құрылды. Келесідей міндеттер шешілді:

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

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

3. Үлкен өлшемді деректерді өңдеу жүйесінің жұмыс сапасы бағаланып, деректер тестілеуден өтті;

4. Құрылымданбаған үлкен көлемді деректерді өңдеу негізінде жеке тұлғалардың төлем қабілеттеріне болжам жасалды, бағдарламалық қамтама құрылды, ипотекалық несие беру ұйымының шешім шығару жүйесіне енгізілді. .

Бағдарламалық кешен келесідей алгоритмдермен жүзеге асырылды:

1. Деректерді даярлау:

- csv форматтағы үлкен өлшемді құрылымданбаған деректерді MongoDB деректер қорына трансформациялау жүргізілді.

2. Оқыту деректер импортталады.

3. Импортталған деректерді нормализациялау жүргізілді.

4. Өңделетін деректер алдын-ала белгілі жауаппен салыстырылды. (futurematrix, target  $y_i$ )

5. Өңделетін деректер DataMining әдістері: сызықты регрессия, логистикалық регрессия, көп қабатты нейрондық желілер бойынша өңделіп, нәтижелердің дәлдіктері анықталды.

6. Егер модель көп қателіктерді анықтаса, цикл қайталанып MongoDB дерекқорынан қашанғы қателік пайызы азайғанша деректерді алып өңдеуін тоқтатпайды. Қателік азайғанға дейін цикл қайталанып салмақ коэффициенті өзгеріп отырады. Қателікті қолмен енгізумен қатар, машиналық оқыту алгоритмдері сызықты регрессияны, логистикалық регрессияны, көпқабатты нейронды желіні қолдану кезінде қайта оқыту болмауыңқадағалауымыз керек, олушінарбір итерация үшін

график алып бақылауда ұстауымыз маңызды. Егер қайта оқыту болған жағдайда, ол модельдің шынайылығы күмән туғызатын болады және оны жарамсыз деп айтуымызға негіз бар.

Бұл құрылған жүйе қаржы ұйымының ипотекалық несиеберу кезіндегі жұмыстардәстілеуден сәтті өтті. Үлкен көлемді құрылымданбаған деректерді өңдеуге арналған бағдарламалық қамтама жеке тұлғалардың төлем қабілеттерін анықтауға, болжам жасауға үлкен септігін тигізеді. Барлық шынайы деректер қаржы ұйымдарынан алынды, құрылған модель сапасының жоғары екеніне көз жеткізілді.

# ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ

- 1 Хашковский В.В., Шкурко А.Н. Современные подходы организации систем обработки больших объемов данных. Известия Южного федерального университета. Технические науки. 2014. – № 8 (157). – С. 241–250.
- 2 Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP. – 2 изд. – СПб.: БХВ-Петербург, 2007. – 384 с.
- 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 №1 (97) 2018. – P.120-126.
- 4 Даркенбаев Д.К. Үлкен көлемді деректерді сақтау және талдау әдістері. «Заманауи зерттеулердің өзекті мәселелері» II Халықаралық ғылыми-практикалық интернет конференциясы. – Нұр-Сұлтан, 2019. – Б.120-124. ISBN 978-601-275-886-4.
- 5 Thomas L.C., Edelman D.B., Crook, J.N. Credit Scoring and Its Applications // SIAM Monographs on Mathematical Modeling and Computation. – Philadelphia: SIAM, 2002. – 263 p.
- 6 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. – No 13. – P. –2659-2670. ISSN:1992-8645.
- 7 Большие Данные. Толковый словарь на Академике. 2014. <https://dic.academic.ru/dic.nsf/ruwiki/1422719>. (қаралған күні: 18.04.2019).
- 8 Рубанов В.А. Между стандартами управления и информационной стихией // Технологический Прогноз. – 2010. – № 3.
- 9 Tom White Hadoop: The Definitive Guide, 3rd Edition. – O'Reilly Media, 2012, – 688 p.
- 10 Большие данные // <https://ru.wikipedia.org>. (қаралған күні: 20.07.2019).
- 11 Нурлыбаева К.К., Балакаева Г.Т. Алгоритмизация процесса построения скоринговых моделей // Вестник КазНТУ. Серия Технические науки. – 2014. – №6(106). – С.195-200.
- 12 Артемов С. Big Data: новые возможности для растущего бизнеса // Инфосистемы Джет. URL: <http://www.pcweek.ru>. (қаралған күні: 20.08.2018).
- 13 Doug L. 3D Data Management: Controlling Data Volume, Velocity and Variety // Meta Delta. – 2001. – P.949-951.
- 14 Pettey C., Goasduff L. Gartner Says Solving Big Data Challenge Involves More Than Just Managing Volumes of Data. <http://www.gartner.com>. (қаралған күні: 21.11.2019).
- 15 Doug L. 3D Data Management: Controlling Data Volume, Velocity and Variety // Meta Delta. – 2001. – P. 949-951.

16 Pettey C., Goasduff L. Gartner Says Solving Big Data Challenge Involves More Than Just Managing Volumes of Data. URL: <http://www.gartner.com>. (қаралған күні: 27.06.2020).

17 Петухов Д. BigData. Проблема и решения. <http://www.codeinstinct.pro> (қаралған күні: 11.08.2019).

18 HDFS Architecture Guide  
[https://hadoop.apache.org/docs/r1.2.1/hdfs\\_design.html](https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html). (дата обращения: 15.04.2019).

19 Apache Cassandra. <http://cassandra.apache.org>. (дата обращения: 15.04.2019).

20 Семенов Ю. А. Большие объемы данных (bigdata). <http://book.itep.ru>. (дата обращения: 21.04.2020).

21 Jacobs A. The pathologies of big data // Communications of the ACM. 2009. – Vol. 52. – Iss. 8. – P. 36–44.

22 Tsvetkov V. Ya., Lobanov A. A. Big Data as Information Barrier // European researcher. – Series A. 2014. – Vol. 78. – Iss. 7-1. – P. 1237–1242.

23 Баженов Р. И. Интеллектуальные информационные технологии. – Биробиджан: ПГУ им. Шолом-Алейхема, 2011. – 176 с.

24 Lockwood G. K. Conceptual Overview of Map/Reduce and Hadoop. <http://www.glennklockwood.com>. (қаралған күні: 28.06.2020).

25 Аншина М. Методы работы с большими данными и их эффективность // Конференция Big Data: возможность или необходимость, 26 марта. – Москва, 2013. – 312 с.

26 Шипунов А. Б., Балдин Е. М. Анализ данных с R. <http://www.inp.nsk.su>. (қаралған күні: 05.01.2018).

27 Abadi D. J., Madden S., Hachem N. Column Stores vs. Row Stores: How Different Are They Really? // Proceedings of the ACM SIGMOD International Conference on Management of Data. – Vancouver. – 2008. – 475 p.

28 Гаврилов Д. Аналитическая платформа – что это такое? <http://www.abc.org.ru> (қаралған күні: 28.10.2020).

29 Fernández A., Río S., López V., Bawakid A., del Jesus M., Benítez J., Herrera F. Big data with cloud computing: an insight on the computing environment, MapReduce and programming framework. WIREs Data Min Knowl Discov 4(5). – P. 380–409.

30 Котенко И. В., Саенко И. Б. Построение системы интеллектуальных сервисов для защиты информации в условиях кибернетического противоборства // Труды СПИИРАН. 2012. – № 3(22). – С. 84–100.

31 Balakayeva G. T., Phillips C., Darkenbayev D. K., Turdaliyev M. Using NoSQL for processing unstructured Big Data. News of the National Academy of sciences of the Republic of Kazakhstan. ISSN 2224-5278. – Volume 6, Number 438, 2019. – P. 12–21.

32 Balakayeva G. T., Darkenbayev D. K., Chris Phillips. Investigation of technologies of processing of Big Data. International Journal of Mathematics and Physics 8, – №2, (13) 2017. – P. 13–18.

33 Зобова Е.В., Самойлова С.С. Управление кредитным риском в коммерческих банках // Социально-экономические явления и процессы. – Тамбов, 2012. – № 12 (046). – С. 74-81.

34 Международная конвергенция измерения ка-питала и стандартов капитала: уточненные рамочные подходы (Базель 2). – М., 2004.

35 Черкашенко В.Н. Управление рисками кредитования малого и среднего бизнеса. <http://bankir.ru>. (каралган күні: 31.07.2018).

36 Ворошилова И.В., Сурина И.В. К вопросу о совершенствовании механизма оценки кредитоспособности индивидуальных заемщиков <http://ej.kubagro.ru>. (каралган күні: 03.08.2020).

37 Усачев С. Кредитный скоринг // Банки и технологии. – № 04. 2008. – С.128.

38 Усачёв С. Кредитный скоринг: решения desktop или enterprise // Банки и технологии. – 2008. – № 04. – С.50-54.

39 Evgueni Solojentsev D., Scenario Logic and Probabilistic Management of Risk in Business and Engineering. Springer Science Business Media Inc 2005. Printed in USA springeronline.com

40 Abdou HAN, Pointon J. Credit scoring, statistical techniques and evaluation criteria: Are view of the literature This version is available at: <http://usir.salford.ac.uk/id/eprint/16518/> 2011. (каралган күні: 14.06.2020).

41 Буракова В.А. Проблемы применения скоринга в российской банковской практике. <http://bizness-gruppa.ru>. (каралган күні: 28.01.2019).

42 Нурлыбаева К.К., Балакаева Г.Т. Анализ больших объемов данных для принятия решения // Сборник тезисов Международной конференции студентов и молодых ученых. – Алматы, 2014. – С.133.

43 Скажи, кто твой заемщик, и скажут твои. <http://www.credits.ru>. (каралган күні: 05.11.2019).

44 Darkenbayev D.K. Building a linear regression model for processing Big Data in the definition of solvency of citizens. Материалы международной научно-технической конференции «100-летие чтений Бойко Ф.К.», посвященной 100-летию Бойко Ф.К. – Павлода, 2020. – С. 23-29.

45 Darkenbayev D.K. Numerical solution of the regression model for analysis and processing of Big Data. Vestnik KazNRTU, Technical sciences series – № 6(130). – 2018. – Р. 132-139.

46 Balakayev G.T., Darkenbayev D.K. Correlation and regression analysis for Big Data processing. Vestnik KazNRTU, Technical sciences series – № 1(131), 2019. – Р. 338-345.

47 Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – 2-е изд. – М.: Финансы и статистика, (перевод с английского) 1986. – 296 с.

48 Теория вероятностей и математическая статистика / В.Е. Гмурман. – 9-е изд. – М.: Высшая школа, 2003. – 126 с.

49 Теория вероятностей / Чернова Н.И. – Новосибирск: Новосибирский государственный университет, 2007. – 318 с.

50 Бизнес-аналитика: от данных к знаниям / Паклин Н.Б., Орешков В.И. – 2-е изд. – Санкт-Петербург: Питер, 2013. – 312 с.

- 51 Data Science Наука о данных с нуля /ДжозелГрас.– Санкт-Петербург: БХВ Петербург, 2017. – 308 с.
- 52 Практическая статистика для специалистов Data Science/П. Брюс, Э. Брюс. – Санкт-Петербург: БХВ Петербург, 2018. – 127 с.
- 53 Барсегян А.А., Куприянов М.С. и др. Методы модели анализа данных: OLAP и Data Mining. – СПб.: БХВ-Петербург, 2004. – 116 с.
- 54 Data Mining в системе управления знаниями. <http://www.smart-edu.com/upravlenieznaniami/izvlechenie-znaniy-data-mining-vsisteme-upravleniya-znaniyami>. (каралған күні: 14.08.2020).
- 55 Барсегян А.А. Анализ данных и процессов: учеб. пособие – 3-е изд. – СПб.: БХВ-Петербург, 2009. – 121 с.
- 56 Berson A., Smith S. J. Data Warehousing, Data Mining & OLAP. McGraw Hill, 1997. – P. 15-16
- 57 Беляков Д.Е., Решение Задач Классификации с Помощью Деревьев Решений. – Москва, 2016. – 473 с.
- 58 Aronson J.E., Leing T.-P., Turban E., "Neural Network in Data Mining," in Decision Support System, (Prentice-Hall of India Private Limited, 2007). – 18p.
- 59 Anderson J.A., Silverstein J.W., Ritz and S.A. Jones R.S., "Distinctive features, categorical perception, and probability learning: Some applications of a neural model", –vol. 84, 1977. – 413 p.
- 60 Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers // Kluwer Academic Publishers. 2004. – 127 p.
- 61 Zweig M.H., Campbell G. ROC Plots: A Fundamental Evaluation Tool in Clinical Medicine // Clinical Chemistry. – Vol. 39, – No. 4, 1993. – P. 1227-1233.
- 62 Linear Regression Using Least Squares. <https://towardsdatascience.com/linear-regression-using-least-squares-a4c3456e8570>. (дата обращения: 04.09.2020).
- 63 Solutions of problems on the topic: The Least Square Method (Processing Experimental Data). MatBuro – Solving problems in mathematics, statistics, economics and programming. [www.matburo.ru](http://www.matburo.ru). ( дата обращения: 17.07.2020).
- 64 Newsom I. Data Analysis II: Logistic Regression, 2015. – P. 573-580
- 65 Chitra K., and B. Subashini. "Data Mining Techniques and its Applications in Banking Sector." International Journal of Emerging Technology and Advanced Engineering 3.8, 2013. – P. 219-226
- 66 Логистическая регрессия и ROC-анализ – математический аппарат. <https://loginom.ru/blog/logistic-regression-roc-auc>. ( дата обращения: 14.11.2019).
- 67 C. Drummond, R. C. Holte, "Cost curves: An improved method for visualizing classifier performance", Mach. Learn. 65, 2006. – P. 96-98.
- 68 Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: классификация и снижение размерности. – М.: Финансы и статистика, 1989. – С. 42-60.
- 69 David W. Hosmer, Stanley Lemeshow. Applied Logistic Regression, 2nd ed. – New York, Chichester, Wiley. 2002. – 392 p. ISBN 0-471-35632-8.
- 70 Fausett Laurene, Fundamentals of Neural Networks: Architectures, Algorithms and Applications. – Prentice-Hall, New Jersey, USA. Int., Inc., 1994 – P. 264-265.

- 71 Aronson J.E., Leing T.-P., Turban E., "Neural Network in Data Mining," in Decision Support System, (Prentice-Hall of India Private Limited, 2007). – P. 18
- 72 Ковалев М., Корженевская В. Методика построения банковской скоринговой модели для оценки кредитоспособности физических лиц // Банки Казахстана. – 2008. – №1. – С. 43-48.
- 73 Base Group Labs Нейронные сети. // <http://www.basegroup.ru>. (дата обращения: 21.02.2020).
- 74 Anderson J.A., Silverstein J.W., Ritzand S.A., Jones R. S., "Distinctive features, categorical perception, and probability learning: Some applications of a neural model", – vol. 84, 1977. – 413 p.
- 75 Boyle M., Crook J.N., Hamilton R., Thomas L.C. Methods applied to slow payers. – Oxford: Oxford University Press, 1992. – № 65 (1). – P. 58–70.
- 76 Henley W.E. Statistical Aspects of Credit Scoring. – Milton Keynes: Open University, 1995. – №1 (45). – P. 77-95.
- 77 Desai V.S., Conway D.G., Crook J., Overstreet G. Credit-scoring models in the credit union environment using neural networks and genetic algorithms // IMA Journal of Mathematics Applied in Business and Industry. – 1997. – № 8(4). – P. 323-346.
- 78 West D. Neural Network Credit Scoring Models // Computers and Operations Research. – 2000. – №27. – P. 1131-1152.
- 79 Baesens B. Developing Intelligent Systems for Credit Scoring Using Machine Learning Techniques: Thesis Ph.D., Katholieke Universiteit. – Leuven, 2003. – 180 p.
- 80 Srinivasan V., Kim Yong H. Credit Granting: A Comparative Analysis. Of Classification Procedures // the Journal of Finance. – 1987. – №42 (3). – P. 665–681.
- 81 Yobas M.B., Crook J.N., Ross P. Credit Scoring Using Evolutionary Techniques // IMA Journal of Mathematics Applied in Business and Industry. – 2000. – №11. – P. 111-125.
- 82 Malhotra R., Malhotra D.K. Evaluating consumer loans using Neural Networks // Omega the International Journal of Management Science. – 2003. – №31(2). – P. 83-96.
- 83 Lee T.S., Chen N.J. Investigating the Information Content of Non-Cash-Trading Index Futures Using Neural Networks // Expert Systems with Applications. – 2002. – № 22 (3). – P. 225-234.

Ғылыми басылым

Даркенбаев Даурен Кадырович

**ҮЛКЕН КӨЛЕМДІ ДЕРЕКТЕРДІ  
ӨНДЕУ ӘДІСТЕРІ**

Монография

Басылуға 22.05.2024 қол қойылды.

Пішімі 60х84/16. Көлемі 9,4 б.т.

Таралымы 500 дана

«Дарын баспасы» ЖШС Алматы

қ., Жамбылқ-сі, 114/85, Бизнес-

СитиБО, Литер Д, 510 каб.

e-mail: [darynbasp2022@mail.ru](mailto:darynbasp2022@mail.ru),

+7747150 1177

«Дарын баспасы» ЖШС баспаханасында басылды