

на значительном удалении от централизованных сетей электроснабжения, разработка и создание СВЧ-установок с автономным электроснабжением является актуальной задачей. В условиях развития фермерских хозяйств, находящихся на значительном удалении от централизованных сетей электроснабжения, применение ВИЭ является наиболее эффективным решением. В статье рассмотрены конструктивные особенности СВЧ-установки для сушки сыпучих материалов, в которой требуемый технологический режим обеспечивается за счет регулирования мощности магнетронных СВЧ-генераторов и скорости перемещения обрабатываемой продукции.

ӨОЖ 004.048
FTAXP 20.53.19

DOI 10.56339/2305-9397-2022-3-3-231-242

Криворотко О.И., физика-математика ғылымдарының кандидаты, **негізгі автор**, <https://orcid.org/0000-0003-0125-4988>

Ресей Ғылым академиясының Сібір бөлімшесінің есептеу математикасы және математикалық геофизика институты, Новосібір қ., Академик Лаврентьев көшесі 6, Ресей, krivorotko.olya@mail.ru

Кубегенова А.Д., техника ғылымдарының магистрі, <https://orcid.org/0000-0003-0156-7757>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Жәңгір хан көшесі, 51, Орал қ., 090009, Қазақстан Республикасы, aigul-03@mail.ru

Абуова Ж.М., техника және технология магистрі, <https://orcid.org/0000-0001-9456-9625>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Жәңгір хан көшесі, 51, Орал қ., 090009, Қазақстан Республикасы, Zhanargul81@mail.ru

Жанғалиева Е.С., экономика ғылымдарының магистрі, <https://orcid.org/0000-0003-0477-7952>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Жәңгір хан көшесі, 51, Орал қ., 090009, Қазақстан Республикасы, yelnaz@mail.ru

Krivorotko O.I., Candidate of Physical and Mathematical Sciences, **the main author**, <https://orcid.org/0000-0003-0125-4988>

Institute of Computational Mathematics and Mathematical Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Akademika Lavrentieva str. 6, Russia, krivorotko.olya@mail.ru

Kubegenova A.D., Master of Technical Sciences, <https://orcid.org/0000-0003-0156-7757>

NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, aigul-03@mail.ru

Abuova Zh.M., Master of Engineering and Technology, <https://orcid.org/0000-0001-9456-9625>

NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, Zhanargul81@mail.ru

Zhangalieva I.S., Master of Economics, <https://orcid.org/0000-0003-0477-7952>

NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, yelnaz@mail.ru

DATA MINING ТЕХНОЛОГИЯСЫН ҚОЛДАНУ АРҚЫЛЫ БІЛІМ АЛУШЫЛАРДЫҢ ОҚУ ҮЛГЕРІМІН ТАЛДАУ ANALYSIS OF STUDENTS' ACADEMIC PERFORMANCE USING DATA MINING TECHNOLOGY

Аннотация

Мақалада деректерді іздеу технологиясы, машиналық оқыту әдістері және кластерлік талдау негізінде есептерді сәйкестендіру, реттеу жүргізілді, университетіміздің білім алушылардың үлгерімі талдалып, алгоритмдері сипатталады.

Data Mining технологиясының көмегімен ақпараттық жүйелер және технологиялар білім беру бағдарламасының білім алушыларының үлгеріміне талдау жасалды. Data Mining

қолданбалы статистика, білімді тану, жасанды интеллект, мәліметтер базасының теориясы және т. б. сияқты ғылымдар негізінде пайда болған және дамитын көп салалы аймақ деп білеміз.

DataMining – бұл жасырын заңдылықтарды (ақпарат шаблондарын) деректерден іздеуге негізделген шешім қабылдауды қолдау процесі болып табылады. Статистикалық әдістердің көпшілігі деректерді талдау бойынша орташалау тұжырымдамасын пайдаланады ал Data Mining нақты мәндермен жұмыс жасайтынына көз жеткіздік.

Оқытуда білім деңгейін жоғарлату барысында үш белгі негізінде топтастыру үшін біз кластерлік талдауды қолдандық.

Білім алушылардың GPA балдарының көрсеткіші бойынша классификацияға бөлініп дендрограмма құрылды, кластерлерге бөлініп оқу үлгерімдері анықталып топтастырылып біріктірілді.

Кластерлік талдау-бұл объектілерді алдын ала анықталған және өлшенген критерийлер бойынша жұптасып салыстыру негізінде объектілерді салыстырмалы түрде біртекті сыныптарға (кластерлерге) ретке келтіру рәсімі ретінде білеміз.

ANNOTATION

The article identifies and corrects tasks based on data retrieval technology, machine learning methods and cluster analysis, analyzes academic performance and describes algorithms for the work of university students.

With the help of Data Mining technology, the analysis of the progress of students of the educational program information systems and technologies was carried out. We know that Data Mining is a multidisciplinary field that has emerged and is developing on the basis of sciences such as applied statistics, knowledge recognition, artificial intelligence, database theory, etc.

DataMining is a decision support process based on finding hidden patterns (patterns of information) from data. Most statistical methods use the concept of averaging data analysis, and we have seen that Data Mining works with specific values.

When increasing the level of knowledge in training, we used cluster analysis to group based on three features.

A dendrogram has been formed with a breakdown into classifications according to the indicators of students' GPA scores, distributed into clusters and grouped into groups with the definition of academic performance.

Cluster analysis is a procedure for ordering objects into relatively homogeneous classes (clusters) based on paired comparison of objects according to predefined and measured criteria.

Түйін сөздер: *Data Mining, деректерді талдау, кластерлік талдау, дендрограмма, статистикалық әдістер.*

Key words: *Data Mining, data analysis, cluster analysis, dendrogram, statistical methods.*

Кіріспе. Ақпараттық технологиялардың дамуы нәтижесінде электронды түрде жинақталған мәліметтер саны тез өсуде. Бұл деректер әртүрлі форматтарда бар: мәтіндер, суреттер, аудио, видео, гипермәтіндік құжаттар, реляциялық деректер базасы және т. б.

Алайда, қол жетімді ақпараттың басым бөлігі белгілі бір адамға ешқандай пайда әкелмейді, өйткені ол мұндай ақпаратты қайта өңдей алмайды. Пайдаланушыға пайдалы ақпаратты деректердің үлкен көлемінен алу мәселесі туындайды.

Сонымен қатар математика және есептеу техникасы бір орнында тұрмайды қарқынды дамып келеді. Деректерді сипаттап бөлу кезінде (дәстүрлі статистикалық әдістер талап ететіндей), белгілердің сызықтық емес тәуелділігі жағдайында да ақпаратты талдауға арналған әдістерге назар аудармауға мүмкіндік беретін жаңа алгоритмдер құрылуда. Қазіргі заманғы технологиялар жүздеген белгілердің ішінен ең ақпараттылығын таңдап, деректердің үлкен көлемін оңай өңдейді. Бұрын таңдалған белгілер сенімді, көптеген мыңдаған мүмкін нұсқалардың алгоритмдерін қолдана отырып жүзеге асырылады.

Қазіргі заманғы мәліметтерді сақтау мен өңдеу әдістерінің дамуы, жинақталған, талдауды керек ететін ақпараттардың тез өсуіне алып келуде. Жинақталған мәліметтердің соншалықты көптігі оны адам күшімен өңдеуге мүмкіндік бермейді, әрі бұл өңделмеген мәліметтердің ішінде, маңызды шешімдер қабылдауға керекті ақпараттар бар екені анық. Сол

себеппен, мәліметтерді автоматты талдау жасау үшін Data Mining технологиясын қолдану керек болады.

Data Mining қолданбалы статистика, білімді тану, жасанды интеллект, мәліметтер базасының теориясы және т. б. сияқты ғылымдар негізінде пайда болған және дамитын көп салалы аймақ деп білеміз.

Data Mining – бұл жасырын заңдылықтарды (ақпарат шаблондарын) деректерден іздеуге негізделген шешім қабылдауды қолдау процесі

Data Mining – ті анық емес, объективті және тәжірибеде пайдалы заңдылықтардың үлкен көлемін іздеуге арналған технология ретінде сипаттауға болады: – анық емес, өйткені табылған заңдылықтар ақпаратты өңдеудің стандартты әдістерімен немесе сараптамалық жолмен анықталмайды; – объективті, өйткені анықталған заңдылықтар әрдайым субъективті болып табылатын сараптамалық білімнен айырмашылығы шындыққа толығымен сәйкес келеді; – іс жүзінде пайдалы, өйткені қорытындылар практикалық қолдануға болатын нақты мағынаға ие.

Data Mining жүйелерінің нарығы өсіп келеді. Бұған ірі корпорациялардың қызметі ықпал етеді мысалға айтқанда: SAS, IBM, Microsoft, Oracle және тағы басқалары. 2027 жылға қарай кеңейтілген аналитиканың жаһандық нарығының көлемі 23,1% - ға өсіп, \$56,2 млрд белгісіне жетеді деп күтілуде.

Data Mining-тің соңғы тенденцияларына виртуалды және кеңейтілген шындық элементтері бар талдау әдістерін дамыту, оларды дерекқор жүйелерімен біріктіру, медицинадағы инновациялар үшін биологиялық деректерді өндіру, веб-майнинг өндірісі (интернеттегі деректерді талдау), нақты уақыттағы деректерді талдау және деректерді өндіру кезінде құпиялылықты қорғау шаралары кіреді. Сала жетекшілері болашақта майнинг деректер өндірісі корпоративті деректер қоймаларына салынатын ақылды қосымшаларда қолданылады деп санайды.

Мәліметтердегі заңдылықтарды анықтаудағы басты мәселе-ақпараттық массивтерді санау үшін қажет уақыт.

Белгілі әдістер мұндай сұрыптауды жасанды түрде шектейді немесе іздеу тиімділігін төмендететін шешімдердің бүтін ағаштарын салады. Бұл мәселені шешу Data Mining үшін өнім әзірлеушілердің басты мақсаты болып қала береді.

Data Mining технологиясын ғылыми зерттеулер үшін қолдану өте тиімді болып келеді, және жыл сайын адам өмірінің барлық салаларында зерттеудің өзекті бағытына айналууда. Data Mining білім саласында, инженерияда, медицинада, тағы басқа сала бағыттарында қолданылады.[1]

Data Mining-ке қандай міндеттеріне қою керектігіне қатысты бірыңғай пікір жоқ. Тапсырмалар өндірілетін ақпарат түрлеріне бөлінеді, бұл Data Mining есептерінің ең жалпы жіктелуі:

- жіктеу, (классификация)
- кластерлеу,
- болжау,
- қауымдастық, (ассоциация)
- визуализация.

Ағылшын тілінен аударғанда, Data Mining-бұл "деректерді өндіру". Бірақ іс жүзінде Data Mining термині деректерді немесе ақпаратты жинауды, талдауды және зерттеуді білдіреді. 1989 жылы Григорий Пятицкий-Шапиро семинар өткізді, онда келесі міндет қойылды: "жасырын білім" деп аталатын өте үлкен мәліметтер базасы бар, осы жасырын білімді бастапқы деректердің үлкен көлемінде анықтау әдістерін жасау қажет. Бұл тапсырма деректерді анықтау әдістеріне сілтеме жасау үшін қолданылатын Data Mining терминінің пайда болуына себеп болды.

Data mining мыналарды қамтиды:

- жасанды нейрондық желілерді, шешім ағаштарын, ассоциативті жадты, эволюциялық бағдарламалауды, анық емес логика мен генетикалық алгоритмдерді қолдану негізінде болжау, модельдеу және жіктеу әдістері;

• факторлық талдау, дисперсиялық талдау, сипаттамалық талдау, корреляциялық талдау, регрессиялық талдау, уақыттық серияларды талдау, компоненттік талдау сияқты статистикалық әдістер.

Сонымен, Data Mining технологиясының мәні деректердің үлкен көлемінде жасырылған анық емес, бірақ практикада пайдалы заңдылықтарды іздеу болып табылады. Data Mining белгілі бір заңдылықтарды білдіретін шаблондар тұжырымдамасына негізделген.[2]

Оларға Data Mining технологиясы шешетін міндеттер сәйкес келеді — жіктеу, болжау, ассоциация, кластерлеу, ауытқуларды анықтау және талдау, байланыс талдауы, бағалау, визуализация және қорытындылау. Жоғарыда аталған барлық тапсырмалар сипаттамалық және болжамды болып бөлінеді. Сипаттамалық тапсырмалар талданған деректерді түсінуді жақсартуға бағытталған, мұндай модельдердің басты сәті — адамның қабылдауының қарапайымдылығы. Болжау тапсырмалары екі кезеңде шешіледі: алдымен белгілі нәтижелері бар мәліметтер негізінде модель жасалады, содан кейін ол жаңа мәліметтер жиынтығы негізінде нәтижелерді болжау үшін қолданылады.[3]

Жоғарыда аталған мәселелердің әрқайсысын шешудің өзіндік әдістері бар, мысалы, жіктеу мәселесін шешу үшін байесиялық желілер, шешім ағаштарын индукциялау, жақын көршінің әдісі қолданылады, ал болжау есептерін шешу үшін математикалық статистика әдістері мен нейрондық желілер қолданылады. Сондай-ақ, көптеген тапсырмаларда қолданылатын Data Mining есептерін шешудің негізгі әдістері бар.

Деректерді өндірудің негізгі әдістеріне алгоритмдер жатады. Барлық нысандарды қарапайым санау $2 * N$ операцияны қажет етеді, мұндағы N — нысандар саны.

Деректердің үлкен көлемімен қарапайым есептеуді қолдану мүмкін емес екені анық, сондықтан жоғары есептеу қиындықтары бар тапсырмалар үшін әр түрлі жаңа шешу процесінің үлгілерін қолданамыз.

Мұндай әдістердің артықшылығы-түсіну мен іске асырудың қарапайымдылығы, ал кемшіліктері айтар болсақ зерттеу мен дамуға байланысты ресми теория мен күрделіліктің болмауы.

Сондай-ақ, Data Mining негізгі әдістеріне статистика теориясының элементтері жатады. Бұл корреляциялық, регрессиялық, дисперсиялық және статистикалық деректерді талдаудың басқа түрлері.[4]

Зерттеу материалдары мен әдістері. Зерттеу объектісі ретінде Жәңгір хан атындағы Батыс Қазақстан аграрлық техникалық университетінің «Ақпараттық жүйелер және технологиялар» мамандығының студенттерінің 3 жылдық оқу үлгерімі бойынша мәліметтер алынып Data Mining технологиясы арқылы талдау жүргізілді.

Деректерді талдау құралы ретінде Statistica колданбалы бағдарламалар пакеті қолданылды: Statistica Base, StatisticaAdvanced, деректерді өндіруге арналған құралдар Data Mining, SANN автоматтандырылған нейрондық желілер. StatisticaAdvanced бағдарламаны өндірген Stat. Soft Inc фирмасы (АҚШ), бұл фирма 1985 жылдан бастап статистикалық қосымшаларды шығарады. Бірінші коммерциялық программа Statistica Supplement Lotus 1 -2 -3 операциялық жүйеге арналып құрылды. 1991 жылы DOS операциялық жүйесіне арналған Statistica версиясы, 1992 жылы Macintosh-қа, 1994 жылы Windows-ға арналған версиялары шығарылған.[5]

Бағдарлама басқа графикалық қосымшалармен деректер алмасуда DDE технологиясын (деректерді динамикалық алмастыру) енгізді. Statistica дестесіндегі графикалық қосымшаларды OLE технологиясы арқылы (OLE объектілерін байланыстыру және ендіру) Word, Excel және тағы басқа қосымшаларға импорттаса болады, сонымен қатар деректерді алмастыру буфері арқылы да операциялар жүргізсе болады.

Statistica бағдарламасы статистикалық анализ және деректерді өңдеу жүйесі модульдер жиынтығынан тұрады. Әр модульде белгілі бір анализ топтары жинақталған. Программа үлкен графикалық мүмкіндіктермен қамтамасыздандырылған. Нәтижелік және бастапқы деректер ендірілген кестелерді пайдаланып жүздеген графиктер салса болады. Statistica бағдарламасы Statistics BASIC және SQL тілдері енгізілген, осыған байланысты тұтынушы керекті анализ әдісін таңдауға мүмкіндігі бар.

Кластерлеудің жаңа әдістері бір байланыс әдісіне (метод одиночной связи) негізделген. Графикалық формаларды қолдану арқылы, деректерді кластерлеу талдау уақытын қысқартуға, сондай-ақ білім алушының үлгерімін болжау алгоритмін әзірлеуге мүмкіндік берді.

Деректерді кластерлік талдауды қолданудың тәжірбиелік маңыздылығы мен өзектілігі зор, өйткені қазіргі ақпараттық қоғамда деректер мен оларды талдау нәтижелері үлкен рөл атқарады, ал кластерлеу бұл деректерді жақсы түсінуге мүмкіндік береді.[6]

Кластерлеу (немесе кластерлік талдау) міндеті — бұл көптеген нысандарды кластерлер деп аталатын топтарға бөлінеді. Әр топтың ішінде "ұқсас" нысандар пайда болады, ал әр топтың нысандары мүмкіндігінше әр түрлі болып келеді.

Кластерлеу мен жіктеудің басты айырмашылығы-топтардың тізімі нақты көрсетілмейді және алгоритм процесінде анықталады.[7]

Кластерлік талдауды қолдануды келесі кезеңде қарастыруға болады :

1. Кластерлеу үшін объектілерді іріктеу.
2. Іріктемедегі объектілер бағаланатын көптеген айнымалыларды анықтау және айнымалылардың мәндерін қалыпқа келтіру.(қажет еткен жағдайда)
3. Объектілер арасындағы ұқсастық өлшемінің мәндерін есептеу.
4. Ұқсас объектілер (кластерлер) топтарын құру үшін кластерлік талдау әдісін қолдану.
5. Талдау нәтижелерін ұсыну.[8]

Нәтижелерді алып талдағаннан кейін оңтайлы нәтиже алғанға дейін таңдалған метрика мен кластерлеу әдісін түзетуге болады.



Сурет 1 – Деректерді іздеу процесі

Қадамдарды толығырақ қарастырсақ:

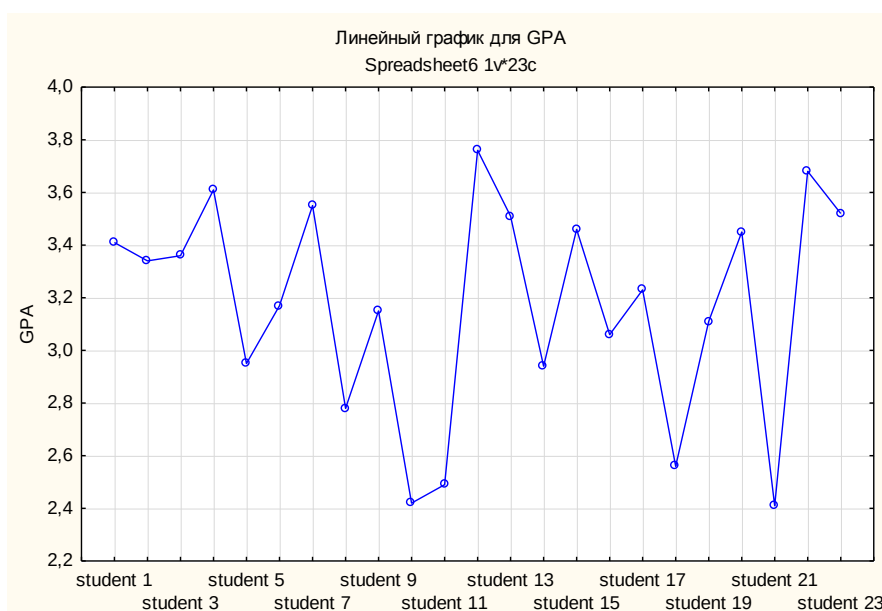
1. Деректерді жинау. Пайдалану және өзара қолдану туралы ақпарат деректер базасында сақталады. (Студенттерді тіркеу кеңсесі)

2. Деректерді алдын-ала өңдеу. Деректер алынып және талдау үшін тиісті форматқа түрлендіріледі.

3. Деректерді іздеуді қолдану. Деректерді іздеу алгоритмдері пайдаланушының қызығушылығын тудыратын білімді анықтайтын және жинақтайтын модельді құру және орындау үшін қолданылады. Осы мақсатқа жету үшін деректерді іздеу құралы қолданылады.

4. Нәтижелерді түсіндіру, бағалау және пайдалану. Алынған нәтижелер немесе модель түсіндіріледі және әрі қарайғы әрекеттер үшін қолданылады. Алынған модель түсіндіріледі және одан әрі талдау үшін қолданылады.

Зерттеу нәтижелері. Эксперименттік мәліметтерді өңдеу компьютерде статистикалық пакеттерде жүргізілді.



Сурет 2 – Барлық оқу кезеңіндегі GPA балдарының сызықтық графигі

GPA (Grade Point Average) балдарының сызықтық кестесі 3 жылдық оқу кезеңіндегі жиынтық деректерін ескере отырып құрылған.

Абсцисса осі бойынша студенттер тізімі саң бойынша белгіленген, координаталар осі бойынша - GPA балдары көрсетілген.

Диаграмма деректері көрсеткендей:

1. № 4,7,12,13,22,23 студенттердің көрсеткіштері жоғары 3,52 – 3,76 бал құрап отыр, бұл студенттер берілген тапсырмаларын уақытылы орындап сабаққа қатысып жүргенің білдіреді.

2. №1,2,3,6,9,15,16,17,19,20 студенттердің көрсеткіштері орташа 3,06 – 3,45 балды құрап отыр. Орташа бал жинап отырған студенттер топтын 40% көрсеткіші байқалады.

3. №5,8,10,11,18,21 студенттердің көрсеткіштері төмен 2,42 – 2,95 бал құрап отыр себебі бір студент денсаулық жағдайына байланысты академиялық демалыста болды, 4 студент емтиханда төмен бал жинап жазғы семестрге қалды.

Бақыланатын айнымалының іріктемелі орташа мәні формула бойынша анықталады (1):

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

мұндағы n -үлгінің көлемі (айнымалы x бақылауларының нақты саны).

Медиана - тапсырыс берілген мәндер қатарын жоғарыда да, төменде де осы мәндердің тең санымен жартысына бөледі. [9,10]

Мода - бұл мәліметтер жиынтығында жиі кездесетін мән.

Іріктемелі дисперсия айнымалының өзгергіштігін сипаттайды, формула бойынша есептеледі(2):

$$\overline{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (2)$$

мұндағы $\bar{x}$ - үлгінің орташа мәнін білдіреді.

Дисперсия 0-ден шексіздікке дейін өзгереді. Тұрақты айнымалылар - 0-дің экстремалды мәні өзгергіштіктің болмауын білдіреді.

Бастапқы деректер файлында Жәңгір хан атындағы Батыс Қазақстан аграрлық техникалық университетінің «Ақпараттық жүйелер және технологиялар» білім беру бағдарламасының білім алушылардың 3 жылдық оқу үлгерімі бойынша мәліметтер бар. Осы кластерлік талдаудың мақсаты кластерлерге бөлу және тәуекел топтарын анықтау үшін тиісті кластерді анықтау болып табылады. Бұл мәселені шешу үшін кластерлік талдауды қолдану негізгі тиімді және жаппай қолданылатын әдістердің бірі болып саналады.[25]

Жақындық шарасы ретінде Евклид арақашықтығын (Евклидово расстояние) пайдалана отырып, кластерлік талдаудың иерархиялық рәсімінің көмегімен GPA балдарының ерекшеліктері бойынша жіктеуді жүргіземіз, ал кластерлерді біріктіру үшін – толық байланыс әдісін (Метод полной связи) әдістерін қолданамыз. Осы әдістердің көмегімен екі кластерді бір-бірімен байланыстыруға болады.[11,12,13]

Кез-келген екі кластер бір-біріне жақындағанда, олар бір-біріне жақындай бастайды және байланыс қашықтығы мен ерекшеленеді. Тиісінше, біріктірілген кластерлер кездейсоқ қалған бөліктерден бөлек элементтерге айналады. Бұл құбылыс нысандарды бір-бірімен байланыстырады және кластерлерді құрайды. Алынған кластерлер ұзын тізбектермен ұсынылған. Кластерлердің табиғи санын анықтау аймақтарды кластерлерге біріктіру арқылы жүргізілді. [14,15]

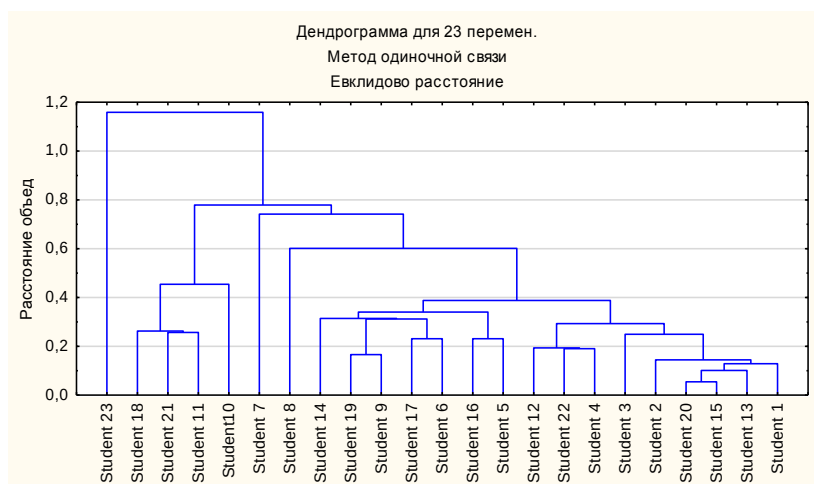
GPA балдарының көрсеткіші бойынша кластерлерге біріктіру тәртібі иерархиялық ағашта 3 суретте көрсетілген.

Бөлуді бағалау үшін төмендегі формула бойынша қарастыруға болады:

$$\pi = \frac{a_i + a_j}{2b_{ij}} \quad (3) \quad S = \frac{1}{k} \sum_{i=1}^{\bar{e}} \max \pi_{ij} \quad (4)$$

мұндағы a_i, a_j - кластардың орта кластерлік арақашықтығы; $i : j; b_{ij}$ сол кластерлік арасындағы орташа кластерлік қашықтық болып келеді.

Егер объектілер арасындағы барлық айырмашылықтар бір-біріне тең болса, онда бұл жағдайда олар 1-ге тең болады. Жоғарыда сипатталған алгоритмнің көмегімен алынған бөлімдер 1-ге тең немесе артық емес болып келеді. Нәтижесінде барлық нысандар бір кластерге біріктірілген кезде олар 1-ге тең болады деп айта аламыз.[16,17]



Сурет 3 – «Ақпараттық жүйелер және технологиялар» мамандығының студенттерінің GPA балдардың классификациясы.

Толық байланыс әдісі. Толық байланыс әдісі (ағылш. Complete-linkage clustering) - иерархиялық кластерлеу алгоритмдерінің бірі. Бастапқыда үлгінің әр элементі жеке кластер болып саналады. Содан кейін кластерлер барлық элементтер бір кластерге енгенше дәйекті

түрде біріктіріледі. Алгоритмнің әр қадамында екі кластер біріктіріледі, олардың арасындағы қашықтық минималды болып келеді.[18]

Объектілер жиынтығы бөлінген кластерлердің табиғи санын анықтау үшін иерархиялық кластерлеудің әр деңгейінде жиынтықты осы класстың санына бөлінді. Кластерлердің әр жұбымен олардың бір-бірімен ішкі байланысы бағаланды. Жәнеде әр кластер үшін орташа кластерлік қашықтық есептелді.

Байланыстылықты бағалау кезінде орташа кластерішілік қашықтықтың кластераралық қашықтыққа қатынасы қолданылады:

Дендрограммада объектілер кластерлерге біріктірілетін қашықтықтар (шартты бірліктерде) көлденеңінен белгіленеді. [19,20]

Көлденең ось бақылауларды білдіреді, тік-біріктіру қашықтығын білдіреді.

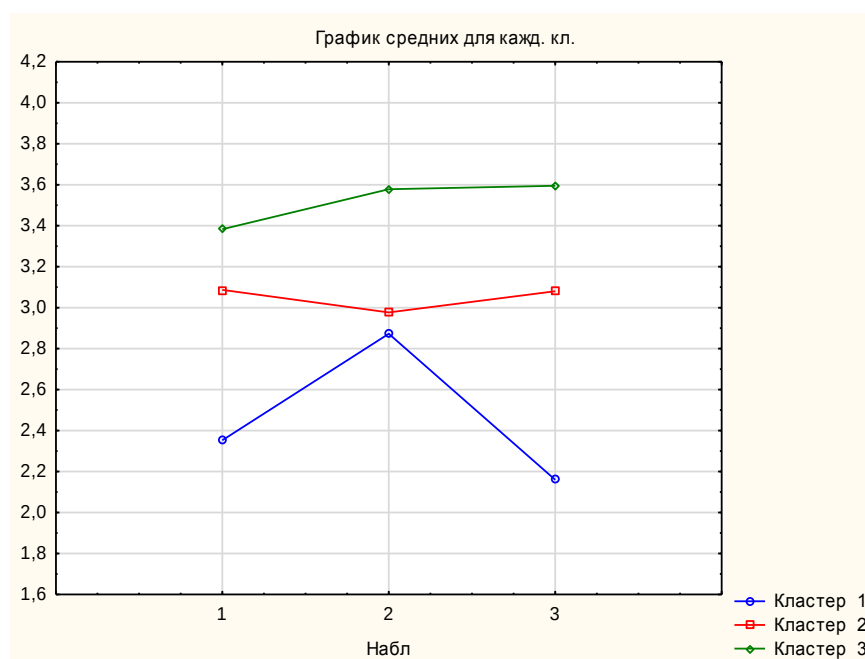
ГРА балдарының көрсеткіші бойынша классификациясының алғашқы қадамы келесі кластерлерден тұрады:

- 1) № 23,18,21,11,10,7,8 студенттердің көрсеткішінен;
- 2) № 14,19,9,17,6,16,5 студенттердің көрсеткішінен;
- 3) № 12,22,4,3,2,20,15,13,1 студенттердің көрсеткішінен құралды.

Бұдан әрі бір кластерге студенттердің көрсеткіштері № 23,18,21,11,10 және 8,16,15,12,3 нөмірлі студенттер бірігеді. Процесс барлық нысандарды бір кластерге біріктірумен аяқталады. Сонымен, дендограмм бойынша, бұл жағдайда үш кластерге бөлуге болады.

Case ID	Cluster Means (Spreadsheet3)		
	Cluster No. 1	Cluster No. 2	Cluster No. 3
C_1	11,50000	19,50000	4,000000
C_2	2,88375	3,10500	3,301429
C_3	3,20875	3,05375	3,422857
C_4	3,09250	3,04000	3,321429

Сурет 4 – Cluster Means электрондық кестесі



Сурет 5 – Әр кластер үшін орташа мәндер графигі

Алынған кластерлердің ерекшеліктерін талдап, білім алушылардың GPA балдары көрсеткіштерінің орташа мәнін салыстыра отырып, біз келесі нәтижелерге қол жеткіздік:

Бірінші кластер, әдетте жоғары деңгеймен ерекшеленеді бұнда білім алушылар сабаққа белсене қатысып тапсырмаларын уақытылы тапсырғандары көрсеткіштен көрініп тұр. Практикада қолдана алу дағдылары мен икемділігіне қатысты сипаттауға болады.

Екінші кластер үшін 1 кластермен салыстырғанда білім алушылардың GPA балдарының көрсеткіші төмен болып көрсетілген. Студенттердің бастапқы және базалық деңгейінің жоғары болғанын ескере кетуге болады.

Үшінші кластерде көрсетілген GPA көрсеткіш балы төмен болу себептері оқу барысында академиялық мдемамыс алған, емтиханда төмен балл жинап жазғы семестрге қалып сабақ оқып тапсырған студенттердің білім сапасы төмен болып келеді.

Қорытынды. Жұмыста Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университетінің «Ақпараттық жүйелер және технологиялар» білім беру бағдарламасының оқу қызметінің статистикалық талдау нәтижелері келтірілген.

Data Mining технологиясы арқылы жүргізілген талдау, студенттерінің 3 жылдық оқу жылы ішіндегі және жазғы семестрді қоса алғанда сабақ үлгерімінің нәтижелері бойынша GPA балдарының көрсеткішін кластерге бөліп білім деңгейін анықтауға мүмкіндік берді. Статистикалық талдау нәтижелері бойынша оқу үлгерімінің түріне қарамастан жақсы нәтиже беретіндей корреляциялық байланыс орнатылды.

Кластерлік жіктеу алгоритмі объектілер арасындағы ішкі байланысты анықтады және оқу іс-әрекетінде математикалық модельдің дұрыстығын көрсетті. Зерттеу нәтижелері студенттер үшін оқу процесі басқа мақсаттарға жетудің тәсілі ретінде әрекет етеді деген болжамды растайды.

Білім алушылардың оқу үлгерімін ескере отырып топқа бөлінген кластерлердің нәтижесі бойынша, аудиториялық сабақтарды келесі әдістерді қолдану арқылы өткізуге болады: жобалау, дифференциалды, интерактивті, интербелсенді.

Сонымен қатар жоғары оқу орнындағы білім алушының бойында коммуникативті дағды мен мүмкіндіктердің қалыптасу деңгейінің жоғары болуы, қарым – қатынас стилдері мен тәсілдерін игеріп, түрлі әлеуметтік рөлдерді тану арқылы, оны өз іс-тәжірбиесінде қолдана алу қабілеті бүгінгі күннің талабы болып келеді.

ӘДЕБИЕТТЕР ТІЗІМІ

1 Kubegenova A.D., Zhakhiena A.G., Baigubenova S.K., Utyasheva G.S., Omarov A.N. Clustering and data mining on the example of hiv-infected people data (2022) *Journal of Theoretical and Applied Information Technology*, 100 (13), pp. 5010-5018.

2 Romero C., Ventura S. (2020). Educational Data Mining and Learning Analytics: An Updated Survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Vol. 10, no. 3, pp. 1-21, doi: <https://doi.org/10.1002/widm.1355>

3 Cui, Zhihao and Yan Chaobing. «Deep Integration of Health Information Service System and Data Mining Analysis Technology» *Applied Mathematics and Nonlinear Sciences*, vol.5, no.2, 2020, pp.443-452. <https://doi.org/10.2478/amns.2020.2.00063>

4 Громыко Г.Л. Теория статистики: учеб. / Под ред. проф. Г. Л. Громыко. – 2-е изд., перераб. и доп. – М.: ИНФРА-М, 2006. – 476 с.

5 Нейский И.М. Характеристика технологий и процессов интеллектуального анализа данных // *Интеллектуальные системы и технологии: Сборник учебно-методических работ и статей аспирантов и студентов*. М., 2005. Вып. 7. С. 111– 122.

6 Kabanikhin S., Krivorotko O., Takuadina A., Andornaya D., Zhang. Sh. Geo-information system of tuberculosis spread based on inversion and prediction. *Journal of Inverse and Ill-posed Problems*, 2021, vol. 29, no. 1, pp. 65-79, doi: <https://doi.org/10.1515/jiip-2020-0022>.

7 Кубегенова А.Д. Интеллектуальный анализ и добыча данных в технологии Data Mining / А.Д. Кубегенова, Е.С. Кубегенов // *Современные инновации в технике и производстве: Сборник материалов I Международной научно-практической конференции*, Псков, 23–24 апреля 2020 года. – С. 130-135.

8 Кубегенова А.Д. Использование методов и основных алгоритмов в технологии Data Mining / А. Д. Кубегенова, К. Т. Искаков // *Исследования молодых ученых: материалы*

IV Международной научной конференции, Казань, 20–23 ноября 2019 года. Молодой ученый, 2019. – С. 8-11.

9 Кубегенова А.Д. Основополагающие моменты кластерного анализа/А.Д. Кубегенова, Е.С. Кубегенов // Инновационные идеи молодых исследователей: Сборник научных статей по материалам IV Международной научно-практической конференции, Уфа, 15 января 2021 года. Общество с ограниченной ответственностью «Научно-издательский центр «Вестник науки», 2021. – С. 53-59.

10 Козлов С.В. Перспективы внедрения методов интеллектуального анализа данных как цифровых средств поддержки процесса обучения. Труды III Всероссийской научно-практической конференции с международным участием, Брянск, 21–22 мая 2020 года.

11 Болюбаш Н.Н. Основные направления использования интеллектуального анализа данных в LMS-системах высших учебных заведений. Знание. – 2020. – № 1-1(77). – С. 78-85.

12 Казаринова Н.Л. Интеллектуальный анализ в оценке сформированности профессиональных компетенций // Актуальные научные исследования в современном мире. – 2021. – № 2-3(70). – С. 50-57.

13 Добрица В.П. Горюшкин Е.И. Управление качеством учебно-воспитательного процесса на основе анализа тестовых заданий с помощью искусственного интеллекта. Известия Юго-Западного государственного университета. 2018;22(6):175-182.
<https://doi.org/10.21869/2223-1560-2018-22-6-175-182>

14 Фомин В.В. Глава 2. Интеллектуальные технологии в решении образовательных задач // Интеллектуальные технологии в цифровой среде университета. – Санкт-Петербург: Центр научно-информационных технологий «Астерион», 2020. – С. 31-105.

15 Филоненко Е.М. Анализ академической успеваемости студентов в системе дистанционного обучения Moodle с помощью когнитивных карт .Актуальные научные исследования в современном мире. – 2018. – № 1-5(33). – С. 103-108.

16 Зыонг К.Х. Методы интеллектуального анализа данных в социальных сетях: систематическое мапирование. Сборник научных статей по итогам восьмой международной научной конференции, Казань, 30–31 августа 2020 года. – Казань: Общество с ограниченной ответственностью «КОНВЕРТ», 2020. – С. 146-157.

17 Танабаева М. Применение анализа образовательных данных как элемент образования будущего // Современное программирование: Материалы I Международной научно-практической конференции, Нижневартовск, 15–18 ноября 2018 года / Ответственный редактор Т.Б. Казиахмедов. – Нижневартовск: 2018. – С. 141-144.

18 Тонких А.П. Применение интеллектуального анализа данных для изучения образовательных моделей студентов в рамках подхода перевернутого класса при обучении программированию / А. П. Тонких // Информационные технологии в моделировании и управлении: подходы, методы, решения: III Всероссийская научная конференция с международным участием: сборник докладов, Тольятти, 21–23 апреля 2020 года. С. 430-438.

19 Жаткина К.Н. Применение методов интеллектуального анализа данных для цифровой образовательной платформы.// Системный анализ в науке и образовании. – 2020. – № 2. – С. 1-5. – DOI 10.37005/2071-9612-2020-2-1-5.

20 Оськин Д.А. Организация процесса интеллектуального анализа образовательных данных. // IT-технологии: развитие и приложение: Сборник докладов Международной научно-технической конференции, Владикавказ, 10 декабря 2019 года .Издательство «Терек», 2019. – С. 112-120.

REFERENCES

1 Kubegenova A.D., Zhahiena A.G., Bajgubenova S.K., Utyasheva G.S., Omarov A.N. Klasterizaciya i intellektual'nyj analiz dannyh na primere dannyh o vich-inficirovannyh lyudyah (2022) Zhurnal teoreticheskikh i prikladnyh informacionnyh tekhnologii, 100 (13), st. 5010-5018.

2 Romero K., Ventura S. (2020). Intellektualnyi analiz obrazovatelnyh dannyh i analitika obucheniya: Obnovlenniy opros. Mezhdisciplinarnye obzory Uajli: Intellektualinyj analiz dannyh i obnaruzhenie znanii. Tom 10, № 3, st. 1-21, doi: <https://doi.org/10.1002/widm.1355>

3 Cui, CHzhao i YAn', CHaobin. «Glubokaya integraciya Sistemy medicinskogo informacionnogo obsluzhivaniya i Tekhnologii intellektual'nogo analiza dannyh» Prikladnaya matematika i nelineinye nauki, tom 5, №2, 2020, st.443-452.

<https://doi.org/10.2478/amns.2020.2.00063>

4 Teoriya statistiki: ucheb. / Pod red. prof. G. L. Gromyko. – 2-e izd., pererab. i dop. – M.: INFRA-M, 2006. – 476 st.

5 Neiskii I. M. Harakteristika tekhnologii i processov intellektualnogo analiza dannyh // Intellektualnye sistemy i tekhnologii: Sbornik uchebno-metodicheskikh rabot i statej aspirantov i studentov. M., 2005. Vyp. 7. St. 111– 122.

6 Kabanihin S., Krivorotko O., Takuadina A., Andornaya D., Chzhan. Sh. Geoinformacionnaya sistema rasprostraneniya tuberkuleza, osnovannaya na inversii i prognozirovanii. Zhurnal obratnyh i nekorrektnykh zadach, 2021, tom 29, № 1, st. 65-79, doi.org/10.1515/jiip-2020-0022.

7 Kubegenova A. D. Intellektualnyi analiz i dobycha dannyh v tekhnologii Data Mining / A.D. Kubegenova, E.S. Kubegenov // Sovremennye innovacii v tekhnike i proizvodstve: Sbornik materialov I Mezhdunarodnoi nauchno-prakticheskoi konferencii, Pskov, 23–24 aprelya 2020 goda. – St. 130-135.

8 Kubegenova A. D. Ispo'zovanie metodov i osnovnykh algoritmov v tekhnologii Data Mining / A.D. Kubegenova, K.T. Iskakov // Issledovaniya molodykh uchenykh : materialy IV Mezhdunarodnoi nauchnoi konferencii, Kazan, 20–23 noyabrya 2019 goda. Molodoi uchenyi, 2019. – St. 8-11.

9 Kubegenova A.D. Osnovopolagayushchie momenty klasternogo analiza/ A.D. Kubegenova, E. S. Kubegenov // Innovacionnye idei molodykh issledovatelej: Sbornik nauchnykh statej po materialam IV Mezhdunarodnoi nauchno-prakticheskoi konferencii, Ufa, 15 yanvarya 2021 goda. Obshchestvo s ogranichennoy otvetstvennost'yu «Nauchno-izdatelskii centr «Vestnik nauki», 2021. – St. 53-59.

10 Kozlov S.V. Perspektivy vnedreniya metodov intellektualnogo analiza dannyh kak cifrovyykh sredstv podderzhki processa obucheniya. Trudy III Vserossiiskoi nauchno-prakticheskoi konferencii s mezhdunarodnym uchastie m, Bryansk, 21–22 maya 2020 g.

11 Bolyubash N. N. Osnovnye napravleniya ispolzovaniya intellektual'nogo analiza dannyh v LMS-sistemah vysshikh uchebnykh zavedenij. Znanie. – 2020. – № 1-1(77). – St. 78-85.

12 Kazarinova N. L. Intellektual'nyj analiz v ocenke sformirovannosti professionalnykh kompetencii // Aktualnye nauchnye issledovaniya v sovremennom mire. 2021. № 2-3(70). – St. 50-57.

13 Dobrica V.P., Goryushkin E.I. Upravlenie kachestvom uchebno-vospitatelnogo processa na osnove analiza testovykh zadaniy s pomoshchyu iskusstvennogo intellekta. Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. 2018; 22(6):175-182. <https://doi.org/10.21869/2223-1560-2018-22-6-175-182>

14 Fomin V.V. Glava 2. Intellektualnye tekhnologii v reshenii obrazovatelnykh zadach// Intellektualnye tekhnologii v cifrovoi srede universiteta. – Sankt-Peterburg: Centr nauchno-informacionnykh tekhnologii «Asterion», 2020. – Se. 31-105.

15 Filonenko E.M. Analiz akademicheskoy uspevaemosti studentov v sisteme distancionnogo obucheniya Moodle s pomoshchyu kognitivnykh kart. Aktualnye nauchnye issledovaniya v sovremennom mire. – 2018. – № 1-5(33). – St. 103-108.

16 Zyong K.H. Metody intellektual'nogo analiza dannyh v socialnykh setyah: sistematicheskoe mapirovanie. Sbornik nauchnykh statei po itogam vosmoi mezhdunarodnoi nauchnoi konferencii, Kazan, 30–31 avgusta 2020 goda. – Kazan: Obshchestvo s ogranichennoj otvetstvennost'yu «KONVERT», 2020. – St. 146-157.

17 Tanabaeva M. Primenenie analiza obrazovatelnykh dannyh kak element obrazovaniya budushchego // Sovremennoe programmirovaniye : Materialy I Mezhdunarodnoi nauchno-prakticheskoi konferencii, Nizhnevartovsk, 15–18 noyabrya 2018 goda / Otvetstvennyi redaktor T.B. Kaziahmedov. – Nizhnevartovsk: 2018. – St. 141-144.

18 Tonkih A. P. Primenenie intellektualnogo analiza dannyh dlya izucheniya obrazovatelnykh modelei studentov v ramkah podhoda perevernutogo klassa pri obuchenii programmirovaniyu/ A. P. Tonkih // Informacionnye tekhnologii v modelirovanii i upravlenii: podhody, metody, resheniya

: III Vserossiiskaya nauchnaya konferenciya s mezhdunarodnym uchastiem: sbornik dokladov, Tolyatti, 21–23 aprelya 2020 goda.– St. 430-438.

19 Zhatkina K.N. Primenenie metodov intellektual'nogo analiza dannyh dlya cifrovoj obrazovatel'noj platformy // Sistemnyi analiz v nauke i obrazovanii. – 2020. – № 2. – St. 1-5. – DOI 10.37005/2071-9612-2020-2-1-5.

20 Oskin, D. A. Organizatsiya processa intellektual'nogo analiza obrazovatel'nykh dannyh. // IT-tehnologii: razvitiye i prilozheniye: Sbornik dokladov Mezhdunarodnoi nauchno-tehnicheskoi konferencii, Vladikavkaz, 10 dekabrya 2019 goda .Izdatelstvo «Terek», 2019. – St. 112-120.

ӨОЖ 622.276

FTAXP 52.47.37

DOI 10.56339/2305-9397-2022-3-3-242-251

Казинский А.А., техника ғылымдарының кандидаты, доцент, негізгі автор, <https://orcid.org/0000-0001-7109-4145>

Ю.А. Гагарин атындағы СамГТУ машина жасау, материалтану және көлік институты, 443100, Самара қаласы, Молодогвардейская көш., 244, kozin202229@mail.ru

Купешова А.С., аға оқытушы, <https://orcid.org/0000-0002-6888-7619>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Жәңгір хан көшесі, 51, Орал қ., 090009, Қазақстан Республикасы, kupeshova.altynay@mail.ru

Ахметжан С.З., техника ғылымдарының кандидаты, қауымдастырылған профессор, <https://orcid.org/0000-0001-8319-6907>

«Батыс Қазақстан инновациялық және техникалық университеті», 090000, Орал қ. Ихсанов көш., 44, samal.zakey@mail.ru

Абдығалиева А.К., аға оқытушы, магистр, <https://orcid.org/0000-0002-2674-5268>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Жәңгір хан көшесі, 51, Орал қ., 090009, Қазақстан Республикасы, ainagul_123@mail.ru

Базарбаев А.А., магистрант, <https://orcid.org/0000-0003-0399-4816>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Жәңгір хан көшесі, 51, Орал қ., 090009, Қазақстан Республикасы, bazarbaev@mail.ru

Мергенев Р.А., магистрант, <https://orcid.org/0000-0003-3464-4464>

«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КеАҚ, Жәңгір хан көшесі, 51, Орал қ., 090009, Қазақстан Республикасы, roketa2000rm@gmail.com

Kazinsky A., candidate of Technical Sciences, associate professor <https://orcid.org/0000-0001-7109-4145>

«Institute of Mechanical Engineering, Materials Science and transport of Samgtu named after Yu.A.Gagarin», 443100, Samara, Molodogvardeiskaya STR., 244, kozin202229@mail.ru

Kupeshova A.S., Senior lecture, <https://orcid.org/0000-0002-6888-7619>

NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, kupeshova.altynay@mail.ru,

Akhmetzhan S.Z., candidate of technical sciences, ass. Professor, <https://orcid.org/0000-0001-8319-6907>

«West Kazakhstan University of Innovation and Technology», 09000044, Uralsk, Ihsanov Street 44, samal.zakey@mail.ru

Abdigalieva A.K., senior lecturer, master's degree, <https://orcid.org/0000-0002-2674-5268>

NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, ainagul_123@mail.ru

Bazarbayev A.A., master's student, <https://orcid.org/0000-0003-0399-4816>

NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, bazarbaev@mail.ru

Mergenev R.A., master's student, <https://orcid.org/0000-0003-3464-4464>

NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, roketa2000rm@gmail.com