

## **Үлкен көлемді деректерді жоғары оқу орындарында оқыту және тәжірибелік жүзеге асырылуы туралы (үлкен деректерге иерархиялық кластерлеуді қолдану)**

**Аңдатпа.** Бүгінде әртүрлі есептеу құрылғыларында сақталатын деректер ағыны көлемінің өте жоғары қарқынмен өсуі үлкен деректерді талдау қажеттілігін тудырып, әлеуметтік-экономикалық салаларда аталған бағыттағы құзыреттіліктері бар мамандарға деген сұранысты арттырып отыр. Үлкен деректер саласында мамандар даярлау жаңашыл платформаларды, аппараттық-бағдарламалық қамтамаларды, бағдарламалау орталарын жан-жақты зерттеп, білім мазмұнына ендіруді қажет етумен байланысты. Ұсынылып отырған зерттеу жұмысының нәтижесі ретінде R программалау ортасында үлкен көлемді деректерді сақтау, өңдеу мен талдау негізінде жоғары оқу орындарындағы білім алушылардың білімін арттыру, жаңа білік пен дағдысын қалыптастыру қажетті тәжірибелік жұмыстардан деректер ұсынылады.

**Түйін сөздер:** үлкен көлемді деректер, үлкен көлемді деректер туралы талдаулар, R программалау ортасы, кластерлеу, иерархиялық кластерлеу.

**DOI:** <https://doi.org/10.32523/2616-6895-2023-142-1-256-268>

### **Кіріспе**

Қазіргі таңда кез-келген білім беру жүйесінде оқу-әдістемелік және ғылыми-әдістемелік жұмысты ұйымдастыру үшін жоғары оқу орындарында білім мен ғылымның интеграциясын ескеру, оқытудың жаңа инновациялық технологияларын құрастыру мен ендіру, педагогтердің біліктілігін арттыруды қамтамасыз ету мақсатында оқу-әдістемелік және ғылыми-әдістемелік жұмыстармен қамтамасыз етілу қажет.

Жоғары оқу орындарының білім алушыларының кәсіби мамандықтар алуына сәйкес құрастырылатын білім беру бағдарламалары мемлекеттік жалпыға міндетті білім беру стандартының талаптарына сай жасалады.

Зерттеу жұмысымыздың мақсаты – жоғары оқу орнында үлкен көлемді деректер бойынша білім алушыларды даярлаудың теориялық негіздерін анықтап және тәжірибелік жүзеге асыру. Зерттеу жұмысының мақсатын жүзеге асыруда алға қойған міндеттердің бірі - жоғары оқу орындарының білім алушыларын үлкен көлемді деректерге оқытудың оқу-әдістемелік негіздерін анықтап, оқу үдерісінде тәжірибелік түрде жүзеге асыруды қамтамасыз ету. Қазақстандағы және әлемдік жоғары оқу орындарында үлкен көлемді деректер бойынша оқу курстарына шолу жасай келе, өз тәжірибемізге ендірілген оқу курсы жүргізуге арналған оқу-әдістемелік қамтама мазмұнынан ұсынамыз, соның ішінде оқу-әдістемелік қамтаманың мазмұнының

ішінде, иерархиялық кластерлер дендрограммасын, яғни ағаш құрылымы арқылы көрсетуге болатын және R бағдарламалық ортасында жүзеге асырылуын баяндайтын материалдар келтіріледі.

### **Әдебиеттерге шолу бөлімі**

Қазақстан және әлемдік жоғары оқу орындарында үлкен деректерді білім саласына ендіру бойынша деректер көздеріне тоқталсақ. С.Сейфуллин атындағы Қазақ агротехникалық университетінің «8D06101-Үлкен деректер аналитикасы» білім беру бағдарламасының мақсаты АКТ бағыты бойынша ғылыми-педагогикалық кадрларды және IT-компаниялар мен ірі өндірістік кәсіпорындарда сұранысқа ие менеджерлерді, талдаушыларды даярлауда үлкен көлемді деректерді талдау, деректерді оңтайлы жинау, жедел деректерді өңдеу процестерін құру, бизнес процестерді оңтайландыру, тұтынушылардың мінез-құлқын болжау, статистикалық көрсеткіштерді талдау, тәуекелдерді талдау, бизнес шешімдерін әзірлеу және т.б. компанияның тиімділігін арттыру сияқты сұрақтарды қарастырады. Деректанушы ғалымдардың негізгі жұмысы - жинақталған ақпарат жүйесіндегі логикалық байланыстарды байқау және соның негізінде белгілі іскерлік шешімдер мен модельдерді жасау. Бұл жаңа ғылыми жаңалықтарға, компанияның жұмыс нәтижелерін жақсартуға, кірістердің жаңа мүмкіндіктеріне, клиенттерге қызмет көрсетуді жақсартуға және т.б. әкеледі.

Авторлар жұмыстың мақсаты ретінде 2013-2018 жылдар аралығында білім берудегі үлкен деректерге арналған ғылыми мақалалардың шығарылуын талдау, сондай-ақ осы мақалаларда жиі қолданылатын кілттік сөздерді анықтауды қарастырған. Алдын ала анықталған критерийлерге негізделген іздеу алгоритмінің көмегімен тексерілген. Сандық процедураны қолдана отырып, соның ішінде мәтіндік өңдеу, білім берудегі үлкен деректер бойынша зерттеу жұмыстарын дайындаудың әр түрлі аспектілері талданған.

2015 жылдан бастап өндіріс көлемінің өсуіне қарай үлкен деректердің артуы да зерттеу нәтижелерінің білім беру саласына жаңаша мазмұн, ол өз кезегінде жоғары оқу орындарында оқытылатын пәндердің тенденциясының өзгеруіне әкелетінін көрсеткен [1].

Интеллектуалды жүйелерді қолдану арқылы Park Y.E. Сауд Арабиясы Эр-Риядтағы Ханзада Сұлтан университетінің іскери әкімшілік колледжінде үлкен деректерді қолданумен қатар, цифрлық трансформация қазіргі жоғары білім беруде терең өзгерістерге әкеліп отырғанын көрсеткен. Жұмыста үлкен деректерді және цифрлық технологияларды қолдану арқылы студенттердің қажеттіліктерін анағұрлым қанағаттандыру үшін білім берудің негізгі институционалдық құндылығын өзгертетіні қарастырылған [2]. Осы сала бойынша оқыту мен оқудағы жаңа тенденциялар мен бағыттарды қадағалауға, жоғары оқу орнында мәтінді интеллектуалды талдау әдісі көмегімен оқытуды негізге алынатыны айтылған. Жоғары оқу орындарында оқыту мен оқыту саласындағы 285 зерттеу мақалалары зерделенген, ол мақалалар 2018-2019 жылдары 2 жыл ішінде іздеу жүйелерін қолданатын бірнеше ірі деректер базасынан

(әр түрлі баспалардың веб-платформаларынан) жинақталған. Сонымен қатар адамның табиғи тілін өңдейтін семантикалық желілік талдаларды қарастырған, оқыту нәтижесін талдауда машиналық оқыту алгоритмі қолданылған [3]. Авторлардың жоғары оқу орнында үлкен көлемді деректерді оқу мен оқытуда оң нәтижелерге қол жеткізгені келтірілген.

Біз ұсынып отырған жұмыста да машиналық оқыту әдістері қолданылатынын атап өтеміз.

Бұлттық технологияларды, компьютерлік желілерді немесе жоғары технологияларды қолдану кезінде білім саласының дәстүрлі оқыту режимінен арылып келе жатқанын Du H., Zhu J., Sun K. өз еңбектерінде Қытай Халық Республикасының жоғары оқу орындары мен қарамақтарындағы оқу орындарының мысалдарында көрсетеді. Мақалада авторлар колледждер мен университеттердегі дене шынықтырудың қазіргі жағдайын статистикалық әдісті қолдана отырып, үлкен деректерді өңдеу негізінде таныстырады. Тәжірибе нәтижелері көрсеткендей, Қытайдағы негізгі университеттерде дене тәрбиесі бакалаврының даму тенденциясы айтарлықтай жақсарғанын, студенттердің дене шынықтыруға деген ынтасы 13% өскенін, ал дене дайындығы жақсы жаққа қарай дамығанын дәлелдейді [4]. Ал Юннань технологиялық және іскерлік университетінің зерттеушілері сонымен бірге үлкен деректердің пайда болуы қоғамның барлық салаларында революциялар тудырғанын атап өтеді. Ақпараттық симметрия және бірлескен алмасу дәстүрлі басқару моделіне әсер етіп қана қоймай, сонымен қатар жоғары білім беруді басқаруда күш беріп ынталандырғанын алға тартады. Университеттегі білім беруді басқарудың мақсаты мен мазмұны орасан зор өзгерістерге ұшырағанын, үлкен деректер дәуіріндегі жоғары өнімділік деректерді басқаруды және университетті басқаруда ғылыми және ұтымды жалпы жоспарлауды қолдануды талап ететінін айтып отыр.

Аталып отырған мақаланың мақсаты – үлкен деректер дәуірінің оң және теріс жақтарын және оның қазіргі кездегі университеттік білім беру жүйесіне әсерін терең талдау, сонымен қатар үлкен деректер дәуіріндегі университеттік білім беруді басқару жолын зерттеу. Мақалада 40 университеттердің қызметі туралы талдаулар жасалған. Үлкен деректер дәуірі жағдайында білім беру менеджменті туралы ақпаратты құру мен дамытуды жүзеге асыру үшін қазіргі үлкен деректер технологиясын интеграциялау керектігі және оны әр түрлі жоғары оқу орындарының менеджментіне қолдануын керек екендігі сараланады. Білім беруді басқаруда ақпараттың фрагментациясы мен асимметриясы 35,4% құрайтынын анықтаған. Бұл білім беру жүйесін инновациялауға және оңтайландыруға қол жеткізуге болатын жалғыз жол екендігі айтылған [5].

Болашақта бұлттық есептеу, веб-желілер және ақпараттық технологияларға негізделген оқытудың цифрлық ресурстарын басқару режимі прогрестің басты бағыты болады, яғни оқытудың ресурстарды басқарудың цифрлық жүйесін құру жобасы үлкен маңызға ие. Сондықтан Цзянси қаржы-экономикалық университетінің ғалымдары мақалада отандық және шетелдегі колледждер мен университеттердегі үлкен деректерді зерттеу және идеялық-саяси білім берудің жай-күйі көрсетілген.

Колледждер мен университеттердегі идеялық-саяси білімді алға жылжыту үшін үлкен деректерді қолдану, жоғары оқу орындарында оларды пайдалануға ықпал ету және нәтижелерін қорытындылау мен болашағын болжау қарастырылған [6].

Испанияның Малага университетінің Білім берудің теориясы мен тарихы және білім берудегі диагностика әдістері департаментінде, сонымен бірге мектептегі дидактика және ұйымдастыру департаментінде үлкен деректермен жұмыс істеу күнделікті жұмыстың біріне айналғанын айтады. Үлкен көлемді деректерді оқу процесінде оқытуда білім алушылардың бұл ұғымды әртүрлі қабылдайтыны туралы еңбектерінде атап өтеді. Мақаланың негізгі мақсаты - пайдаланушы профилдері мен олардың үлкен деректерді пайдалануға бейімділігін талдау және үлкен деректердің қолданылуына қандай факторлар әсер ететіндігін анықтау.

Жұмыста кластерлік талдауды реттік регрессиялық талдау және ағаштар арқылы сипаттаумен бірге жүргізген. Нәтижесінде студенттердің екі түрлі профильде, яғни үлкен деректерді қабылдау мен бағалау тұрғысынан және олардың білімдегі маңыздылығын бағалауымен бөлінетінін көрсеткен. Демек, үлкен деректерге оқыту мүмкіндіктері мен білім сапасын жақсарту тұрғысынан қарастыратын және қабылдайтын профильдегі студенттер жоғары балмен бағаланған. Зерттеу нәтижесінде университетте үлкен деректерге оқытуды дамытуда цифрлық технологиялар бойынша дағдыларын дамытуға ықпал ету қажеттілігін көрсетеді [7].

Ferrao M.E., Prata P., Alves M.T.G. Португалия және Бразилия елдерінің жоғары оқу орындарында білім беруді бағалау, бағалау және білім беруді зерттеу бойынша барлық цифрлық зерттеулер дерлік бірнеше жылдар бойы бір шешімсіз проблема болып келген толық емес деректер жиынтығына негізделгенін айтады. Үлкен анықталатын деректерді пайдалану жетіспейтін мәндерге қатысты жаңа қиындықтар тудырған. Зерттеу жұмысында кез-келген білім саласының зерттеушісіне толық емес болжамдарды тексеру үшін R процедураларын қолдану туралы қарапайым ұсыныстар береді, егер нәлдік гипотеза жоққа шығарылса, жетіспейтін деректерге ең дұрыс әдістердің бірі болып табылатын бірнеше импутацияны қалай жүзеге асыруға болатынын көрсетеді [8].

Үлкен деректерді білім мазмұнында қолдану туралы келесі жұмыста жалпы ғылым мен техниканың қарқынды дамуына байланысты «Интернет +», бұлттық есептеу, бұлттық қойма және жаңа тағы да басқа заманауи ақпараттық-коммуникациялық технологияларының дамуы ақпаратты кеңінен таратуға мүмкіндік береді, ал үлкен деректер қазіргі дәуірдің өнімі болып табылатынын айтады. Интеллектуалды білім берудің мәні қолданыстағы базалық білім беру моделін талдауда, белгілі бір білім беру формацияларының үйлесімінде, үлкен деректер мен білім берудің өзара интеграциясын талдауда, үлкен деректердің интеграциясы мен дамуын ынталандыруда жатқанын атап өтеді [9].

Көрсетіліп өткендей үлкен деректерге жоғары оқу орындарында оқыту қазіргі кезде кеңінен орын алуда. Аталып өткен әдістерді оқу процесінде қолдану біздің жұмысымызда да орын алды.

## Негізгі бөлім

Зерттеу жұмысымызда Л.Н.Гумилев атындағы Еуразия ұлттық университеті, С.Аманжолов атындағы Шығыс Қазақстан мемлекеттік университеті, Ы.Алтынсарин атындағы Арқалық педагогикалық институты жоғары оқу орындарының оқу процесінде үлкен көлемді деректерді басқаруға байланысты арнайы курсы ендіріліп, білім алушылардың осы мәселе бойынша білімін арттыру, жаңа білігі және дағдысын қалыптастыру мәселесі қарастырылды, атап айтқанда «6B01511-Информатика», «7M01511-Информатика», «7M01525-STEM білім беру» және «7M01524-Smart-технологиялар» білім беру бағдарламаларының білім алушыларының даярлықтарының үлкен деректерге байланысты критерийлер мен көрсеткіштер бойынша жетілдірілгенін атап өтеміз, атап айтқанда қарастырылып отырған мәселе - үлкен деректерге иерархиялық кластерлеуді қолдану әдісіне байланысты да жоғары нәтижелер көрсеткенін атап өтеміз. Зерттеу нәтижелері оң нәтижелерге қол жеткізгегін тәжірибелік-эксперименттік жұмыстар нәтижелерінен мынадай жағдайларды атап өтеміз:

- Білім беру бағдарламаларында студенттер мен магистранттардың басым бөлігі келесі оқу жылында өзіне қажетті пәндерді таңдау барысында үлкен көлемді деректерге байланысты пәндер таңдауы олардың үлкен көлемді деректерге деген қызығушылығы мен білімге деген құштарлығының артуын көрсетеді.

- Үлкен көлемді деректермен жұмыс істеуде университеттің локальды желісінде және ауқымды желіде жұмыс істеп, қашықтағы дискілік кеңістікте - бұлттық платформада деректер базасын құрады және сол деректер базасында үлкен көлемді деректерді ұйымдастырады. Бұл өз кезегінде студенттер мен магистранттардың білім мен білігінің артуына үлкен әсер етеді.

- Үлкен көлемді деректерге оқыту қазіргі кезде білім саласында кең қолдау тапқан STEM (Science, Technology, Engineering, Mathematics) оқытумен де байланысы бар.

Мұндай ойларды, осы салада қызмет жасап жүрген оқытушылардың ойларынан да байқаймыз. Атап айтсақ, білім алушылардың білімге деген ынтасын арттыру үшін оқытушылар білім беру процесін ұйымдастырудың түрлі тәсілдерін белсенді қолданады, соның ішінде HR-де қолданылатынын, адам ресурстарын басқаруға ұқсас Big Data аналитикасы оқушылардың қызығушылығын арттыруға және білім сапасын жақсартуға көмектесетінін айтады [10].

### **Эксперименттік жұмыстар бөлімі**

Эксперименттік жұмыстар Нұр-Сұлтан қаласындағы Л.Н.Гумилев атындағы Еуразия ұлттық университеті Ақпараттық технологиялар факультетінің «6B01511-Информатика», «7M01511-Информатика», «7M01525-STEM-білім беру», «7M01524-Smart-технологиялар» білім беру бағдарламаларында 5 кредит (1 лекция, 2 тәжірибелік сабақ, 2 өздік жұмыстар) көлемінде жүргізілді.

Оқу-әдістемелік қамтама мазмұнында қарастырылған кейбір тапсырмаларды

шешу жолдарына тоқталсақ.

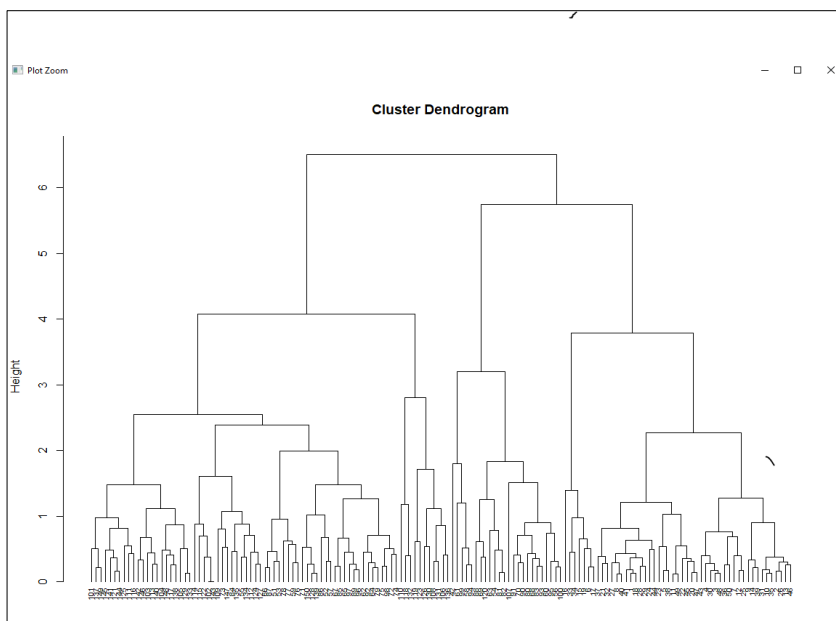
Үлкен деректерді сақтау және өңдеудің қуатты және негізгі технологиялары ретінде Hadoop фреймворгымен жұмыс істеуге арналды. Hadoop технологиясының негізі үлестірілген тораптарды біріктіру нәтижесінде кластер құрып, үлкен көлемдегі деректерді өңдеуге қол жеткізу болып табылады. Hadoop утилиталар жиынын қолдану үшін оны жергілікті желіде немесе бұлттық платформада жүзеге асыруға болады. Біздің жағдайда Hadoop арқылы үлестірілген тораптардан тұратын кластерді баптау GCP (good clinical practice) бұлттық платформасында жүзеге асырылды. Себебі, біздің жағдайда кластер үлкен көлемді есептеулер жүргізуге қажетті есептеу ортасы болып табылады [13]. Үлкен көлемді деректерді өңдеу барысында қайталау операторы арқылы деректерді талдау қиындық туғызады. Бұл жағдайда үлкен көлемді деректерді өңдеу үшін кластерлеу әдісі қолданылады. Тәжірибелік жұмыста кластерлеудің бір түрі иерархиялық кластерлеу түрін қарастырамыз.

Иерархиялық кластерлік талдау кластерлердің иерархиясы немесе берілген тәртібі бар кластерлеу әдісі болып табылады [11]. Иерархиялық кластерлерді дендрограмма деп аталатын ағаш құрылымы арқылы көрсетуге болады. Біздің жағдайда үлкен деректерді өңдеуде R ортасында иерархиялық кластерлеуді қарастырдық.

Тәжірибелік жұмыста объектілердің деректер жинағын, деректер нүктелері ретінде объектінің енін, ұзындығын қолдандық. Алдымен деректер жүктеліп, қалыпқа келтіріледі. Ұқсасыздық мәндерінің функциясы арқылы есептеуде иерархиялық кластерлеуді орындау үшін кластерлеу функцияларын қолдандық. Программа кодынан үзінді келтірсек:

```
Ең алдымен қажетті пакеттерді жүктейміз
// деректер жинағын қамту
library(datasets)
// алгоритмдерді кластерлеу
library(cluster)
// визуалдау
library(factoextra)
// map_dbl() функциясын қолдану
library(purrr)
# деректер жинағын жүктеп алдын ала өңдеу
df <- iris[, 1:4]
df <- na.omit(df)
df <- scale(df)
# айырмашылық матрицасы
d <- dist(df, method = "euclidean")
Алынған ұқсастық матрицасы алынып, дендрограмма құруға болады.
// Толық орналасуды(complete) пайдалану арқылы иерархиялық кластерлеу
hc1 <- hclust(d, method = "complete" )
// дендограмма тұрғызу
```

plot(hc1, cex = 0.6, hang = -1) (сурет 1)



Сурет 1. Кластер дендрограммасы

Сурет 1-дегі дендрограммада объектіні бақылауға сәйкес келетінін аңғаруға болады және ағашты жоғары жылжитқанда ұқсас бақылаулар жоғары биіктікте біріктірілді. Дендрограмманың биіктігі кластерлерді анықтайды. Кластерлерді анықтау үшін дендрограмманы cutree көмегімен қиюға болады. Содан кейін нәтижені шашыраңқы сызба ретінде визуалдауға болады.

```
# ағашты 3 топқа бөлу
sub_grps <- cutree(hc1, k = 3)
# Нүктелі диаграмма көмегімен визуалдау
fviz_cluster(list(data = df, cluster = sub_grps)) (сурет 2)
```

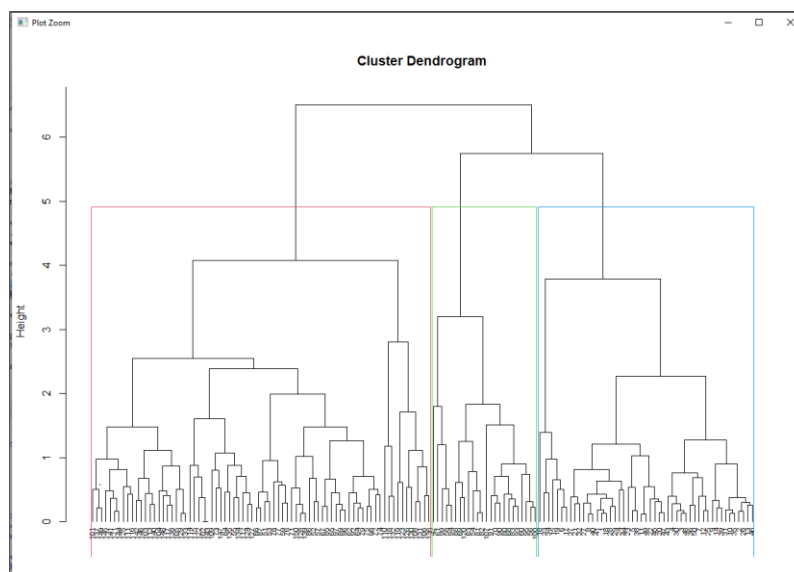


## Сурет 2. Нүктелі диаграмма

Төменде көрсетілгендей 3 кластердің айналасында дендрограмманың шекарасын да бере аламыз.

//Алынған дендрограмманы k кластер үшін тіктөртбұрыштың шекарасын пайдаланып сызамыз

```
plot(hc1, cex = 0.6, hang = -1)  
rect.hclust(hc1, k = 3, border = 2:4) (сурет 3)
```



Сурет 3. Кластерге бөлінген дендрограмма (қызыл және жасыл түспен ерекшеленген)

Иерархиялық кластерлеуді орындау үшін қажет функцияны пайдалана отырып, нәтижесін аламыз:

```
// агломерациялық бағалау әдістері  
m <- c("average", "single", "complete")  
names(m) <- c("average", "single", "complete")  
# иерархиялық есептеу функциясы  
# кластерлік коэффициент  
ac <- function(x) {  
  agnes(df, method = x)$ac  
}  
map_dbl(m, ac) (сурет 4)
```

average	single	complete
0.9035705	0.8023794	0.9438858



#### Сурет 4. Иерархиялық есептеу функциясының нәтижесі

Толық байланыс күшті кластер құрылымын береді. Сонымен, біз төменде көрсетілгендей қажет функция көмегімен иерархиялық кластерлеуді орындау үшін қажет әдісін қолдандық.

```
# иерархиялық кластерлеу
hc2 <- agnes(df, method = "complete")
# дендограмма салу
pltree(hc2, cex = 0.6, hang = -1,
       main = "Dendrogram of agnes")
...
// Бөлінген кластерлеуді жүзеге асыру
//бөлінген иерархиялық кластерлеуді есептеу
hc3 <- diana(df)
//бөлу коэффициенті
hc3$dc
// дендограмманы салу
pltree(hc3, cex = 0.6, hang = -1,
       main = "diana дендограммасы")
Нәтижесі
[1] 0.9397208
```

Сонымен, дендограммада иерархиялық кластерлеу нәтижелерін көрсету үшін визуализатор қолданылып, тиімді нәтиже алынды, яғни жекелеген объектілер мен кластерлердің жақындық дәрежесі көрсетілді, сонымен қатар олардың бөліну реттілігі графикалық түрде бейнеленді.

#### Қорытынды

Үлкен көлемді деректерді жоғары оқу орындарының оқу процесінде қолдануға байланысты еңбектерге талдаулар жасалып, біздің зерттеу жұмысымыздың теориялық және практикалық түрде жүзеге асырылуының нәтижелік жұмыстарынан деректер келтірілді. Зерттеу барысында қолданылған тәжірибелік нәтиже ретінде иерархиялық кластерлеуді, яғни кластерлеудің бұл түрінің басқа түрлеріне қарағанда қашықтық пен ережелердің ұқсастығын анықтауға оңай және сонымен қатар ерекшеліктерін атап өтуге болатынын көрсетті: кластер санын алдын ала орнатудың қажеті жоқ, иерархиялық класс қатынастарын анықтауға болады, сол сияқты басқа пішіндерге кластерлеуге болатын үлкен көлемді деректер курсындағы маңызды тақырып екендігін атап өтілді. Аталған бағытты деректерді өңдеудің әмбебап әдісі ретінде ұсынамыз және қарастырылған кодты қажеттілігіңізге сәйкес өзгерте алу мүмкіндігін де атап өтеміз. Теориялық ақпараттардың нақты баяндалған бағдарлама кодтары және

визуализдануы көрнекті сипатталған тәжірибелік тапсырмалармен бекітілуі, білім алушылардың тақырыптарды терең түсінуіне мүмкіндік беріп, қажетті білім мен дағдылардың қалыптасуына зор ықпал етпек. Баяндалған мәселе осы саланы зерттеуші оқытушылар мен студенттер, ғалымдар мен практик-оқытушыларға пайдалы және қажетті еңбек болады деген ойдамыз.

### Әдебиеттер тізімі

1. Terron A.M., Leiva J.J., Franco P.D. Big data irruption in education. Pixel-Bit, Revista de Medios y Educacion. – 2020. – Vol. 57. – P. 59-90.
2. Serik M., Nurbekova G., Mukhambetova M., Zulpykhar Z. The educational content and methods for big data courses including big data cluster analysis. World Transactions on Engineering and Technology Education. – 2022. –No.20(3). – P. 203-208.
3. Park Y.E. Uncovering trend-based research insights on teaching and learning in bigdata. Journal of Big Data. – 2020. – Vol. 7.
4. Du H., Zhu J., Sun K. Development Trend and Content Status of Physical Education Undergraduatein Normal Universities Based on Big Data // Advances in Intelligent Systems and Computing. – 2021. – Vol. 1303. – P. 488-493.
5. Yu W., Su C.The path of higher education management in the era of big data // Advances in Intelligent Systems and Computing. – 2021. – Vol. 1244. – P. 448-453.
6. Wu X. Research on the Innovation of Ideological and Political Education in Universities in the Era of Big Data // ACM International Conference Proceeding Series 16. – 2020. – P. 336-340.
7. Matas A., Leiva J.J., Negro C. Tendency to use big data in education based on its opportunities according to andalusian education students // Social Sciences. – 2020. – Vol. 9. – P. 205-211.
8. Ferrao M.E., Prata P., Alves M.T. Multiple imputation in big identifiable data for educational research: An example from the Brazilian Education assessment system. – 2020. –Vol. 28. – P. 599-621.
9. Jiang J., Chen T. Research on the Value of Smarter Education in the Era of Big Data.// 5th IEEE International Conference on Big Data Analytics. – 2020. – P. 42-45.
10. Вичугова А. Аналитика больших данных и Machine Learning в образовании: 5 кейсов из ВУЗов. [Электрон. ресурс]. – 2022. – URL: <https://www.bigdataschool.ru/blog/big-data-analytics-education-cases.html> (дата обращения: 01.02.2022).
11. Serik M., Nurbekova G., Mukhambetova M. Optimal organisation of a big data training course: Big data processing with BigQuery and setting up a data proc hadoop framework // World Transactions on Engineering and Technology Education. No.19(4). P. 417-422.

**М. Серік, Г.Ф. Нурбекова**

*Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан*

**О преподавании и практике внедрения больших данных в высших учебных заведениях  
(применение иерархической кластеризации к большим данным)**

**Аннотация.** В современном обществе стремительный рост объемов потоков данных, хранящихся в различных вычислительных устройствах, создает потребность в анализе больших данных и увеличивает спрос на специалистов с профессиональными компетенциями, так как подготовка будущих специалистов в области больших данных связана с необходимостью всестороннего изучения программно-аппаратных средств, сред программирования и интеграции их в образовательный процесс. Для этого в этой статье была предложена практическая работа, необходимые для повышения знаний студентов высших учебных заведений, формирования новых навыков и умений на основе хранения, обработки и анализа больших объемов данных в среде программирования R.

**Ключевые слова:** Big Data, анализ больших данных, язык программирования R, кластеризация, иерархическая кластеризация.

**M. Serik, G.F. Nurbekova**

*L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*

**On teaching and the practice of implementing big data in higher educational institutions  
(applying hierarchical clustering to big data)**

**Abstract.** In modern society, the rapid growth in the volume of data streams stored in various computing devices creates a need for big data analysis and increases the demand for specialists with professional competencies, since the training of future specialists in the field of big data is associated with the need for a comprehensive study of software and hardware, environments, programming and integrating them into the educational process. To do this, this article proposed practical work necessary to improve the knowledge of students of higher educational institutions, the formation of new skills and abilities based on the storage, processing and analysis of large amounts of data in the R programming environment.

**Keywords:** big data, big data analysis, R programming environment, clustering, hierarchical clustering.

## References

1. Terron A.M., Leiva J.J., Franco P.D. Big data irruption in education. Pixel-Bit, Revista de Medios y Educacion. 2020. Vol. 57. P. 59-90.
2. Serik M., Nurbekova G., Mukhambetova M., Zulpykhar Z. The educational content and methods for big data courses including big data cluster analysis. World Transactions on Engineering and Technology Education. 2022.No.20(3). P. 203-208.
3. Park Y.E. Uncovering trend-based research insights on teaching and learning in bigdata. Journal of Big Data. 2020. Vol. 7.
4. Du H., Zhu J., Sun K. Development Trend and Content Status of Physical Education Undergraduate in Normal Universities Based on Big Data. Advances in Intelligent Systems and Computing. 2021.Vol. 1303. P. 488-493.
5. Yu W., Su C. The path of higher education management in the era of big data. Advances in Intelligent Systems and Computing. 2021. Vol. 1244. P. 448-453.
6. Wu X. Research on the Innovation of Ideological and Political Education in Universities in the Era of Big Data. ACM International Conference Proceeding Series 16. 2020. P. 336-340.
7. Matas A., Leiva J.J., Negro C. Tendency to use big data in education based on its opportunities according to andalusian education students. Social Sciences. 2020. Vol. 9. P. 205-211.
8. Ferrao M.E., Prata P., Alves M.T. Multiple imputation in big identifiable data for educational research: An example from the Brazilian Education assessment system. 2020. Vol. 28. P. 599-621.
9. Jiang J., Chen T. Research on the Value of Smarter Education in the Era of Big Data. 5<sup>th</sup> IEEE International Conference on Big Data Analytics. 2020. P. 42-45.
10. Vichugova A. Analitika bol'shih dannyh i Machine Learning v obrazovanii: 5 kejsov iz VUZov. [Big Data Analytics and Machine Learning in Education: 5 Cases from Higher Education Institutions]. Available at: <https://www.bigdataschool.ru/blog/big-data-analytics-education-cases.html> [in Russian]. (.accessed 01.02.2022).
11. Serik M., Nurbekova G., Mukhambetova M. Optimal organisation of a big data training course: Big data processing with BigQuery and setting up a data proc hadoop framework. World Transactions on Engineering and Technology Education. No.19(4). P. 417-422.

### Авторлар туралы мәлімет:

*Serik M.* – педагогика ғылымдарының докторы, Информатика кафедрасының профессоры, Л.Н. Гумилев атындағы Еуразия ұлттық университеті,, Астана, Қазақстан.

*Nurbekova G.Ф.* – Информатика кафедрасының докторанты, Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан.

*Serik M.* – Doctor of Pedagogical Sciences, Professor of the Department of Computer

Science, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.

*Nurbekova G.F.* – Ph.D. student of the Department of Computer Science, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.