

**БОШҚАРУВ ТИЗИМИГА КИРУВЧИ МАЪЛУМОТЛАР
МАССИВЛАРИГА ИНТЕЛЛЕКТУАЛ ИШЛОВ БЕРИШДА
КЛАСТЕРЛАШ УСУЛИДАН ФОЙДАЛАНИШ (СОЛИҚ ОРГАНЛАРИ
МИСОЛИДА)**

Ҳакимжон Насириддинович ЗАЙНИДИНОВ

профессор

техника фанлари доктори

Муҳаммад ал-Хоразмий номидаги ТАТУ

Тошкент, Ўзбекистон

Одилбек Улугбек ўғли АСҚАРАЛИЕВ

техника фанлари бўйича (PhD) фалсафа доктори

PERFECT UNIVERSITY

Тошкент, Ўзбекистон

oasqaraliyev77@gmail.com

Малика Норкуловна НОРМАТОВА

таянч докторант

Гулистон давлат университети

Тошкент, Ўзбекистон

normatova.2989@gmail.com

Аннотация

Ушбу мақолада маълумотлар массиви, уларнинг таҳлили, турли бошқарув тизимларига кирувчи ахборот оқимларига ишлов бериш ва бошқариш масалалари кўриб ўтилган. Олинган хулосалар асосида кластерлаш усуллари таҳлили амалга оширилган. Энг оптимал усуллар асосида бошқарув тизимларига кирувчи маълумотлар массивларини муҳимлилик тоифалари бўйича кластеризациялашнинг оптимал алгоритми ишлаб чиқилган.

Таянч сўзлар: ахборот оқими, кластеризациялаш, интеллектуал тизим, эксперт тизими, бошқарув тизими, алгоритм, математик модел.

**ИСПОЛЬЗОВАНИЕ МЕТОДОВ КЛАСТЕРИЗАЦИИ ПРИ
ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКЕ ДАННЫХ, ПОСТУПАЮЩИХ
В СИСТЕМУ УПРАВЛЕНИЯ (НА ПРИМЕРЕ НАЛОГОВЫХ
ОРГАНОВ)**

Ҳакимжон Насириддинович ЗАЙНИДИНОВ

профессор

Доктор технических наук

ТУИТ имени Мухаммада аль Хоразми

Ташкент, Узбекистан

Одилбек Улугбек углы АСҚАРАЛИЕВ

Доктор философии (PhD) по техническим наукам

PERFECT UNIVERSITY

Ташкент, Узбекистан

oasqaraliyev77@gmail.com

Малика Норкуловна НОРМАТОВА

базовый докторант
Гулистанский государственный университет
Ташкент, Узбекистан
normatova.2989@gmail.com

Аннотация

В данной статье рассматриваются массивы данных, их анализ, обработка и управление информационными потоками, поступающими в различные системы управления. На основании полученных выводов был проведен анализ методов кластеризации. На основе наиболее оптимальных методов разработан оптимальный алгоритм кластеризации массивов данных, входящих в системы управления, по категориям важности.

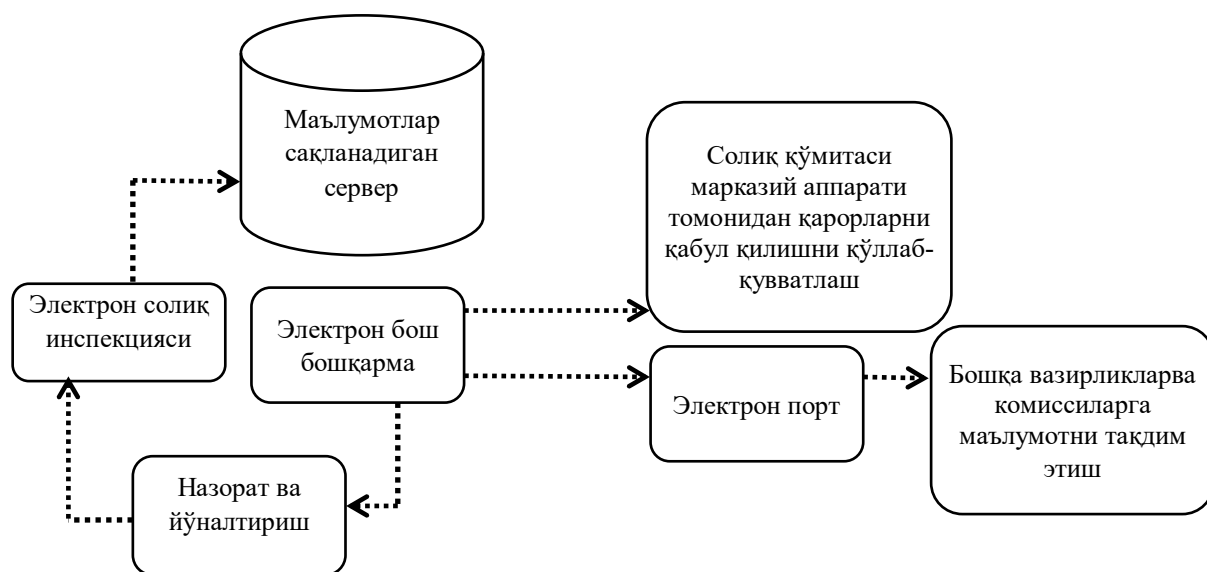
Ключевые слова: информационный поток, кластеризация, интеллектуальная система, экспертная система, система управления, алгоритм, математическая модель.

Жаҳонда корпоратив интеграллашган бошқарув тизимларининг катта массивли маълумотларга ишлов бериш, таҳлил қилиш, прогнозлаш (башорат қилиш) ва тадқиқ қилинаётган вазиятларни баҳолаш масалаларини ечиш учун замонавий интеллектуал корпоратив ахборот-коммуникация тизимларини (Intelligence Enterprise Information Systems) шакллантиришнинг илмий асослари яратилмоқда. Ушбу соҳада, жумладан, қарорларни қабул қилишни қўллаб-қувватлаш тизимлари бизнес-стратегиялари ва ахборот технологияларини бирлаштириш, функционал компонентларнинг билимлар базаси учун қоидалар ва далилларни қуришнинг усул ҳамда алгоритмлари интеллектуал тизимларнинг ишлаш самарадорлигини оширишда муҳим вазифалардан ҳисобланади. Шу билан бирга бошқарув тизимлари қуйи поғоналаридан динамик кировчи ахборотларни интеллектуал қайта ишлашнинг билимлар базаси асосида ҳулосалаш ва қарор қабул қилиш механизмини ишлаб чиқиш зарур ҳисобланади.

Бугунги кунда барча соҳаларда бўлгани каби автоматлаштирилган интеллектуал ахборот тизимлари, уларни ишлаб чиқиш ва интеграллашган ҳолда ишловчи бошқарув тизимларини амалиётга жорий қилиш жараёнлари жадал ривожланиб бормоқда. Шу жумладан интернет тармоқ технологиялари, билимлар базасини ишлаб чиқиш, сақлаш ва тақдим этиш

усуллари, сунъий интеллектга асосланган дастурий воситалар, дастурлаш тиллари ва аппарат-тузилмавий дастурий таъминот, ашёлар интернет (IoT-Internet of Things) ва катта ҳажмли маълумотлар (Big Data) каби замонавий соҳаларда интеллектуаллашган бошқарув модуллари асосида тизимлар ва жараёнларни ишлатиш долзарб аҳамият касб этади.

Бошқарув тизимида турли маълумотларни тақдим этиш, бошқа ташкилотлар ва ҳамкор корхоналар билан ўзаро интеграллашган режимда ишлашда маълумотларни узатиш функционал жараёнлар орқали амалга оширилади. Бу жараённи амалга оширишда ички ва ташқи талаблар инобатга олинади ва тегишли мулоқот тизимлари асосида ахборот оқими жараёни содир бўлади. Ушбу жараённинг функционал схемаси қуйидаги 1-расмда келтирилиб ўтилган.



1-расм. Бошқарув тизимида маълумотларни узатишнинг функционал схемаси.

Умуман олганда мукаккаб-интеграллашган бошқарув тизимларида ахборотни интеллектуал қайта ишлашни тегишли йўналишларга мос муаммоларни илмий усуллар ёрдамида ечиш орқали амалга ошириш мақсадга мувофиқ бўлади. Расмда тасвирланган бошқарув интеллектуал ахборот тизими ҳам бугунги кундаги мавжуд ҳолатни ўзида акс эттиради.

Субъектив ва объектив характериға эға бўлган ва турли хусусиятли тизим муаммоларини ҳал қилиш билан боғлиқ бўлган катта ҳажмдаги

маълумотларни таҳлил қилиш зарурати ортиб бораётганлиги билан янги илмий йўналишларнинг фаол ўсиши талаб қилинди, шу жумладан маълумотларни таҳлил қилиш, қайта ишлаш масалалари ҳам тобора долзарб бўлиб бормоқда.

Маълумотларга интеллектуал ишлов бериш усуллари орасида юқори ўринни кластерлаш усуллари эгаллайди, улар кўп ўлчовли жараёнлар ва ходисаларни ўрганиш учун энг истиқболли ва энг муқобил танлов ҳисобланади.

Кластерлашнинг вазифаси ўрганилаётган X маълумот тармоғи шаклида ифодаланган бир нечта кластерларга (кичик тўпламларга) ажратишдан иборат бўлиб, улардаги объектлар бошқа кластерлар объектларига қараганда бир-бирига “ўхшаш” ва метрик фазодаги “ўхшашлик” масофаси аниқланади.

Кластерлаш алгоритмларида ҳар бир X объект ўзининг характеристикалари $X_i = (x_1, \dots, x_d)$ вектори билан аниқланади, бу ерда $x_i, i = 1, \dots, d$ объектларнинг алоҳида характеристикалари сифатида тақдим этилади ва d орқали характерланувчи майдоннинг ўлчами аниқланади.

Кичик тўплам $M = (X_1, \dots, X_n)$ хусусиятлар векторларидан иборат бўлиб, $X_i = (X_{i1}, \dots, X_{id})$ объектлар бир-бирига яқин бўлган кластер ҳисобланади улар орасидаги масофа $D(X_i, X_j)$, бу ерда X_i ва X_j кластерлаш объектлари сифатида белгиланади.

Кластерларни аниқлашга қўйиладиган талаблар қуйидагилар:

- ҳар бир кластерда объектлар ўхшаш, хосса ва атрибутларининг қийматлари бир хил бўлиши керак;
- объектлар тўплами барча кластерлар бўйича тақсимланиши керак;
- кластерлар бир-бирини истисно қилиши, бошқача айтганда ҳар бир объект бир вақтнинг ўзида иккита кластерга тегишли бўлмаслиги керак.

Кўп ўлчовли маълумотлар тўпламлари кластерли таҳлил қилиш алгоритмлари ва усуллари учун дастлабки таҳлилий восита ҳисобланади. Шу

сабабли тадқиқот ишида интеграллашган бошқарув тизимларига алоқадор бўлган кирувчи маълумотлар (солиқ тўловчилар) муҳимлилик тоифалари бўйича тақсимлаш учун кластер таҳлил усули танланди. Айни пайтда барча турдаги кластерлаш масалаларини ҳал этувчи ягона стандарт ечим мавжуд эмас. Шунинг учун кластерлаш вазифалари ўрганилаётган соҳанинг ўзига хос хусусиятлари билан тавсифланади.

Тадқиқот давомида кластерлаш усуллари таснифлаш амалга оширилган ва унинг натижалари қуйида келтириб ўтилган.

Маълумотларни қайта ишлаш усуллари	
<i>Иерархик:</i> <u>Алгоритмик:</u> CURE, ROCK, CHAMELON <u>Бўлинувчи:</u> BIRCH, MTS	<i>Ноиерархик, итератив:</i> к-средных (k-means) PAM (k-means+k-medions) CLOPE
Маълумотларни таҳлил қилиш усуллари	
Аниқ	Ноаниқ
Кластерлаш алгоритмларининг поғонали қўлланилиш усуллари	
Бир этапли кластеризация	Кўп этапли кластеризация
Қайта ишланган маълумотлар ҳажмини катталаштириш усуллари	
Масштабланувчи	Масштабланмайдиган
Кластеризация бажарилишини вақтга боғлиқлик усуллари	
Потокли (on-line)	Потоксиз (off-line)

2-расм. Кластеризациялаш усуллари классификацияси.

Норавшан кластерлаш усуллари орасида биз c-means (Fuzzy c-means - FCM) алгоритмини танлаб оламиз ва тадқиқот ишида маълумотлар массивларини кластеризациялашда фойдаланамиз.

1-жадвал

Кластерлаш усуллари афзалликлари ва камчиликлари

Кластеризация усуллари ному	Афзалликлари	Чекловлари
CURE	турли ўлчамдаги ва зичликдаги шарсимон бўлмаган шаклдаги кластерларни ҳосил қилади	объектларни зичлиги ва кластерлар сони учун баъзи чегараларни ўрнатишни талаб этади ва бу ҳар доим ҳам мақбул эмас
ROCK	ўлчовларни аралаш турларига эга	кўп сонли объектлар учун

	бўшлиқларни кластерлаш учун объектлар орасидаги боғланиш даражаси тушунчасини қабул қилади ва уларнинг умумий қўшнилари сони, кластерлаш сифатини баҳолаш функцияси билан белгиланади	кенгайтириб бўлмайди
CHAMELON	иккита кластерни бирлаштириш учун динамик симуляциялардан фойдаланилади, бу табиий ва бир хил кластерларни ўрганишни осонлаштиради	бу жараён қийин ва вақтга боғлиқ
BIRCH	чекланган хотира ҳажмидан фойдаланган ҳолда катта ҳажмдаги маълумотларни қайта ишлайди, бунда маълумотларнинг бир хил тақсимланмаганлиги фактига асосланади, шунингдек зичлиги юқори бўлган майдонни ягона кластер сифатида қайта ишлайди	фақат рақамли қийматларни қайта ишлашга имкон беради, бўртма ёки шарсимон шаклга эга бўлган кластерларни ажратиб кўрсатади шунингдек чегара қийматларини ўрнатиш зарурати мавжуд
MTS	ихтиёрий шаклдаги кластерларни, шу жумладан тушириш ва конкав шакллар кластерларини танлаш имконини беради ва бир нечта оптимал ечимлардан энг мақбулини танлайди	эмиссияларга жуда сезгир
k- средних (k-means)	оддий, ишлатиш учун қулай, шаффоф ва тушунарли	катта маълумотлар базасида ҳаддан ташқари сезгир, бир-бирини устига чиқадиغان кластерда қўллашнинг мумкин эмаслиги ва кластерлаш сонини белгилаш зарурати
CLOPE	юқори масштаблилиқ ва иш тезлиги, кластер сонини автоматик танлаш қулайлиги	эмиссияларга жуда сезгир
PAM	фойдаланиш осон, аммо алгоритмнинг бошқаларига нисбатан сезгирлиги камроқ	кластерлар сонини ўрнатиш ва катта маълумотлар базаларида ишлашни секинлаштириш зарурати мавжуд
Ўз-ўзидан ташкил этиладиган Кохонен карталари	усулни амалга ошириш жуда осон	фақат рақамли маълумотлар билан ишлаш имконини беради ва тармоқ ҳажмини минималлаштиришни талаб этади
FCM c-средних (c-means)	микдорий ва сифат характеридаги, танловларга нисбатан кам сезгирликдаги маълумотлар билан ишлаш имконини беради	кластер сонини белгилаш зарурати; ҳисоблаш мураккаблиги

Кластер таҳлилининг вазифаси A солиқ тўловчилар тўпламини бири-бирига ўхшаш объектларга эга кластерлар (тоифалар) бўйича тақсимлашдан иборат. Норавадан кластерлаш усули ҳар бир объектга бир вақтнинг ўзида бир нечта ёки барча кластерларга тегишли бўлишига имкон беради, аммо тегишлилик даражаси ҳар хил кўринишда бўлиши мумкин. Кластерлар сони c олдиндан маълум деб тахмин қилинади.

Кластерлаш объектлари тўплами $A = \{a_1, a_2, \dots, a_n\}$ элементларнинг чекланган тўплами бўлиб, бу ерда a_i солиқ тўловчи ва хусусиятларнинг чекланган топламидир. $P = \{p_1, p_2, \dots, p_q\}$ уларнинг ҳар бири миқдорий жиҳатдан ўрганилаётган ҳудуднинг характерли элементлари. Бунда солиқ тўловчиларнинг умумий сони натурал n ни, солиқ тўловчиларнинг ўлчанган белгиларининг умумий сони эса натурал q ни белгилайди.

Ҳар бир солиқ тўловчи учун барча P хусусиятлари маълум миқдорий шкала бўйича ўлчанади. Шундай қилиб, $x_i = (x_1^i, x_2^i, \dots, x_q^i)$ элементларнинг ҳар бирига $a_i \in A$ маълум бир вектор тайинланади, бу ерда x_j^i маълумотлар объекти учун $a_i \in A$ хусусиятнинг $p_j \in P$ миқдорий қиймати. Биз $x_i = (x_1^i, x_2^i, \dots, x_q^i)$ хусусият қийматлари векторларини X ўлчамли $(n \times q)$ маълумотлар матрицаси кўринишида ифодалаймиз, уларнинг ҳар бир сатри x_i векторининг қийматига тенг.

Кластер структураси $(c \times n)$ ўлчамли μ аъзолик матрицаси билан берилган, a_i – чи элементнинг k – кластердаги $\mu_{A_k}(a_i)$ – аъзолик даражаси, c – кластерлар сони.

Бундан ташқари сўровлар шаклида кирувчи солиқ назорати жараёнларини бошқариш шартлари учун аъзолик матрицаси қуйидаги чекловларга жавоб бериши керак:

$$a) \mu_{A_k}(a_i) \in [0,1], k = \overline{1,c}, i = \overline{1,n};$$

б) $0 < \sum_{i=1}^c \mu_{A_k}(a_i) = 1$ – ҳар бир солиқ тўловчи барча кластерлар ўртасида

тақсимланиши керак;

в) $0 < \sum_{i=1}^n \mu_{A_k}(a_i) < 1$ – ҳеч бир кластер бўш бўлмаслиги, яъни барча солиқ

тўловчиларни ўз ичига олиши керак.

Кейинчалик барча кластерларнинг ҳар бир хусусият учун ва қуйидаги формула бўйича ҳисобланган керакли кластер учун v_k марказларини ҳисобга оламиз.

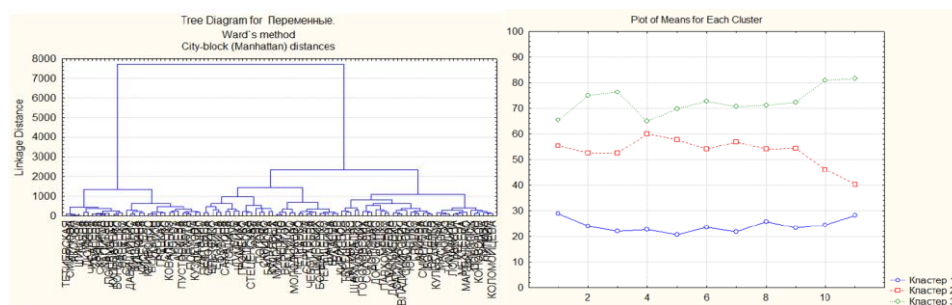
$$v_j^k = \frac{\sum_{i=1}^n (\mu_{A_k}(a_i))^m \cdot x_j^i}{\sum_{i=1}^n (\mu_{A_k}(a_i))^m}, j = \overline{1, q},$$

Бу ерда $m \in (1, \infty)$ – экспоненциал оғирлик, $(m > 1)$ ҳақиқий сонга тенг.

Солиқ тўловчини муҳимлилик тоифаларига бўлиш сифатини баҳолаш учун биз объектлардан тегишли аъзолик даражасига эга бўлган кластерлар марказларигача бўлган масофалар йиғиндисини кўрсатадиган тарқалиш мезонидан фойдаланамиз.

$$f(A_k, v_j^k) = \sum_{i=1}^n \sum_{k=1}^c (\mu_{A_k}(a_i))^m \sum_{j=1}^q (x_j^i - v_j^k)^2.$$

Кластерлар сонини аниқлаш учун биз киришда кластерлар сонини аниқлашни талаб қилмайдиган ва уларнинг сонини аниқлаш имконини берадиган бошқа усуллар билан кластерлашни амалга оширамиз. Солиқ тўловчилар фаолияти тўғрисидаги дастлабки маълумотларга асосланиб, кластерлаш “Statistica” пакетида бажарилган “к-средних” усулидан фойдаланган ҳолда амалга оширилди, натижалар 3-расмда келтирилган ва ушбу танлов учун фақат учта кластер, яни тоифалар мавжудлигини кўрсатади.

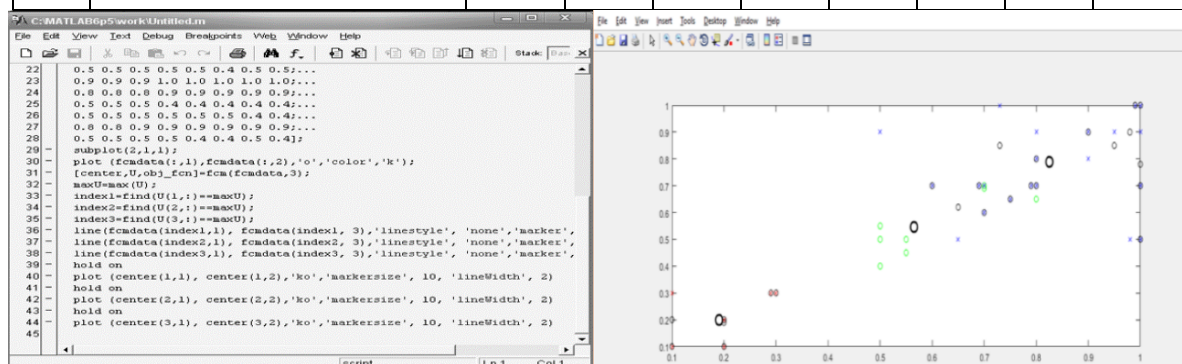


3-расм. “Statistica” пакетида бажарилган “к-средных” алгоритмини амалга ошириш натижалари.

“с-средных” алгоритмини амалга ошириш учун эса математиканинг классик ва замонавий соҳаларини қамраб олувчи, сонли ҳисоб-китобларни бажариш учун интеграллашган дастурий муҳит бўлган ва норавшан моделлаштириш учун мўлжалланган махсус MATLAB компьютер дастуридан Statistica пакетидан фойдаланамиз.

Қўйилган масаланинг реализация жараёнлари 4-расмда кўрсатилган бўлиб, унда дастлабки маълумотлар, кластерлаш дастур кодлари ва олинган натижаларнинг визуализацияси акс эттирилган.

№ 1	Солиқ тўловчининг идентификация рақами	Қаралувчи асосий омиллар							
		1	2	3	4	5	6	7	8
1	199785432	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
2	234678987	1	1	1	1	0,9	0,9	1	1
3	675646456	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
...	097865342	...	...	...	...	...	...	...	...
100	543797864	0,5	0,5	0,5	0,5	0,4	0,4	0,5	0,4



4-расм - “с-средных” алгоритми бўйича амалга оширилган солиқ тўловчилар ҳақидаги маълумотларни кластерлаш натижалари.

“с-средных” алгоритмида объект ва кластер маркази орасидаги масофа Эвклид нормаси орқали аниқланади ва барчасининг шакли бир хил бўлади.

Кластерлаш алгоритмлари маълумотларни ғайриоддий тузилишга эга бўлишга мажбур қилади, бу эса келажакда субоптимал натижаларга олиб келиши мумкин. Ушбу камчиликни бартараф этиш учун биз тадқиқот доирасида ҳар бир кластер учун мослашув нормасидан фойдаланадиган Густафсон-Кессел алгоритмини қўллаб кўрамиз, натижада ҳар бир i – кластер ўз нормаларини яратувчи матрицага эга. Натижада норавшан бўлиниш матрицалари ва кластерлар марказлари координаталари, шунингдек барча кластерларнинг меъёр ҳосил қилувчи матрицалари оптималлаштирилди, бу эса турли геометрик шакллардаги кластерларни олиш имконини беради.

Густафсон-Кессел алгоритми қуйидаги такрорий босқичлардан иборат.

1-қadam. Норавшан ахборот манбааларига асосланган маълумотлар матричасини яратиш.

- керакли кластерлар сони ёки муҳимлилик тоифалари сони $c = 3$;
- алгоритмни такрорлашнинг максимал сони $s = 100$;
- алгоритмнинг яқинлашув параметри $\varepsilon = 0,00001$;
- мақсад функцияси ва кластер марказларини ҳисоблашнинг экспоненциал оғирлиги $m = 2$.

Мутахассис аъзолик функциялари тўплами билан тавсифланган кластерларга ноаниқ бўлинишга қараб X маълумотлар матрицаси $\mu'_k(a_i)$ ни ўрнатади.

Солиқ тўловчининг муҳимлиликнинг қайси тоифасига тегишли эканлигини аниқлашда мутахассис учун кўплаб омиллар мавжуд. Ўтказилган тадқиқотларга асосан биз улардан саккизтасини ажратиб оламиз, улар энг юқори даражада солиқ тўловчининг фаолиятини тавсифлайди (3-жадвал).

3-жадвал

Муҳимлилик тоифасини аниқлашда асосий бўлган омиллар

№	Омиллар
1.	Солиқ тўловчи давлат рўйхатидан ўтказилган манзилда мавжуд бўлмаганда

2.	Солиқ тўловчининг манзили манзил-маълумот бюроси маълумотлари базасида мавжуд бўлмаганда
3.	Раҳбарнинг ёки раҳбарлик вазифаларини ёки молиявий бошқарув вазифаларини амалга ошираётган шахснинг шахси аниқланмаса, ёхуд аризада кўрсатилган шахслар билан тузилган меҳнат шартномалари Ягона миллий меҳнат тизимида тасдиқланмаганда
4.	Таъсисчи тўғрисида маълумотлар солиқ органининг маълумотлар базасида мавжуд бўлмаганда
5.	Ўз кўчмас мулкига эга бўлмай туриб ижарага олинган ёки текин фойдаланилаётган кўчмас мулки бўлмаганда
6.	Электрон ҳисоб-варақ фактуралар маълумоти асосида солиқ тўловчининг кирим ва чиқим операцияларида товарлар номутаносиблиги аниқланганда
7.	Ариза тақдим этилган пайтда солиқ тўловчининг солиқ қарзи мавжуд бўлганда, шунингдек, улуши 50 фоиз ва ундан кўп бўлган таъсисчиларининг бошқа юридик шахсларида солиқ қарзи мавжуд бўлганда
8.	Солиқ тўловчига солиқ қарзи мавжуд ёки солиқ хавфи юқори бўлган корхона қўшилганда

Мутахассис-эксперт олиб борилган тадқиқотлар бўйича ушбу омиллар асосида ҳар бир солиқ тўловчини таҳлил қилади, ҳар бир солиқ тўловчига мос равишда $x_i = (x_1^i, x_2^i, \dots, x_q^i)$ вектор келтиради, бу ерда x_j^i – хусусиятнинг миқдорий қиймати сифатида ифодаланган. Эксперт баҳолаш натижалари эса 4-жадвалда келтириб ўтилган.

4-жадвал

Солиқ тўловчиларни эксперт баҳолашнинг бошланғич матрицаси

№	Кластерланаётган кирувчи ахборот манбаалари (солиқ тўловчилар)	Омиллар							
		1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8	9	10
1	MCHJ "XМ AGROFOOD"	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
2	MCHJ "GUVEN MAKINA GROUP"	1	1	1	1	0,9	0,9	1	1
3	MCHJ "SINEKTA SALES GROUP"	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
4	MCHJ "SERGELI FREE ZONE"	0,7	0,7	0,7	0,7	0,7	0,6	0,5	0,5
5	MCHJ "MAROQAND AVIA TOUR"	0,5	0,5	0,5	0,5	0,5	0,4	0,4	0,4
...	...	...	...	...	...	...	...	...	...
94	MCHJ "YEUROMED SYSTEMS"	0,4	0,5	0,5	0,7	0,5	0,5	0,8	0,5
95	MCHJ "SAID INVEST GROUP"	1	1	1	1	0,9	0,8	1	0,9
96	XK "AUTOMOTORS"	0,1	0,1	0,1	0,1	0,29	0,2	0,29	0,2
97	MCHJ "ZAFAR LES GOLD"	0,7	0,69	0,6	0,60	0,7	0,6	0,5	0,5
98	MCHJ "MEGA IDEAL AVTO"	0,79	0,7	0,7	0,7	0,8	0,9	0,8	0,69
99	MCHJ "CITY BRAND"	0,29	0,3	0,3	0,3	0,3	0,2	0,2	0,1
100	MCHJ "RUSH STONE"	0,2	0,19	0,2	0,2	0,2	0,2	0,2	0,19

Кластер тузилмаси аъзолик матрицаси μ билан белгиланади, бу ерда $\mu_{A_k}(a_i)$ – солиқ тўловчининг k – кластерга мансублик даражаси. Шу билан бир қаторда эксперт солиқ тўловчиларни муҳимлилик тоифалари бўйича тақсимлашни ҳам амалга оширади. Эксперт баҳолаш натижалари 5-жадвалда келтирилган.

5-жадвал

Солиқ тўловчиларнинг эксперт тизими баҳоси матрицаси

№	Кластерланаётган кирувчи ахборот манбаалари (солиқ тўловчилар)	Муҳимлилик тоифалари		
		Яшил	Сариқ	Қизил
1.	MCHJ "HM AGROFOOD"	0,0500	0,9000	0,0500
2.	MCHJ "GUVEN MAKINA GROUP"	0,9000	0,0500	0,0500
3.	MCHJ "SINEKTA SALES GROUP"	0,1000	0,2000	0,7000
4.	MCHJ "SERGELI FREE ZONE"	0,5500	0,3500	0,1000
5.	MCHJ "MAROQAND AVIA TOUR"	0,1000	0,8000	0,1000
...	...	...	...	...
94.	MCHJ "YEUROMED SYSTEMS"	0,0100	0,0900	0,1900
95.	MCHJ "SAID INVEST GROUP"	0,0200	0,04500	0,1200
96.	XK "AUTOMOTORS"	0,0400	0,1000	0,6000
97.	MCHJ "ZAFAR LES GOLD"	0,7500	0,0800	0,1000
98.	MCHJ "MEGA IDEAL AVTO"	0,7500	0,2000	0,0500
99.	MCHJ "CITY BRAND"	0,0500	0,0500	0,9000
100.	MCHJ "RUSH STONE"	0,1000	0,2000	0,7000

2-қadam. Кластер марказларини ҳисоблаш.

Кейинги босқичда 2.1-формуладан фойдаланиб, кластерларнинг марказларини ҳисоблаймиз. Ҳисоблашлар натижалари 6-жадвалда кўрсатилган.

6-жадвал

Кластерлар марказлари

Муҳимлилик тоифалари	Омиллар							
	1	2	3	4	5	6	7	8
Яшил	0,23564	0,24754	0,26316	0,25478	0,2679	0,21784	0,28221	0,20789
Сариқ	0,78414	0,76809	0,76387	0,80704	0,81388	0,80441	0,79142	0,77872
Қизил	0,52152	0,50933	0,50497	0,487	0,5043	0,47473	0,45326	0,44784

3-қadam. k – кластер учун ковариация матрицасини аниқлаш

$$B_j^k = \frac{\sum_{i=1}^n (\mu_{A_k}(a_i))^m \cdot (x_j^i - v_j^k)^T \cdot (x_j^i - v_j^k)}{\sum_{i=1}^n (\mu_{A_k}(a_i))^m}. \quad (3.4)$$

Биз (3.4) формуладан фойдаланиб, ҳар бир кластер учун ковариация матрицаларини аниқлаймиз. Учинчи кластер учун ковариация матрицаси, яъни қизил муҳимлилик тоифаси учун қуйидаги шаклга эга:

$$B^3 = \begin{bmatrix} 0,4924 & 0,4542 & 0,4808 & 0,4493 & 0,4152 & 0,4393 & 0,4544 & 0,4743 \\ 0,4542 & 0,4891 & 0,5079 & 0,4551 & 0,3935 & 0,4161 & 0,5092 & 0,5344 \\ 0,4808 & 0,5079 & 0,5972 & 0,4911 & 0,4311 & 0,4549 & 0,5449 & 0,5695 \\ 0,4493 & 0,4551 & 0,4911 & 0,5340 & 0,4769 & 0,5210 & 0,5874 & 0,6106 \\ 0,4152 & 0,3935 & 0,4340 & 0,4769 & 0,4927 & 0,6020 & 0,5404 & 0,5280 \\ 0,4393 & 0,4161 & 0,4311 & 0,5210 & 0,5266 & 0,5266 & 0,6190 & 0,6123 \\ 0,4544 & 0,5092 & 0,5549 & 0,5874 & 0,5404 & 0,6190 & 0,8236 & 0,7867 \\ 0,4743 & 0,5344 & 0,5695 & 0,6106 & 0,5280 & 0,6123 & 0,7867 & 0,8273 \end{bmatrix}$$

4-қadam. X тўплам ва кластер марказларидан объектлар орасидаги масофани ҳисоблаш

$$D_{B_j^k} = (x_j^i - v_j^k) \cdot [(\det(B_j^k)) \cdot (B_j^k)^{-1}] \cdot (x_j^i - v_j^k)^T.$$

X тўплам ва кластер марказлари бўлган объектлар орасидаги масофани 2.5-формуладан фойдаланиб ҳисоблаймиз. Учинчи кластер учун матрица қуйидаги шаклга эга

$$D_{B^3} = \begin{bmatrix} 1,4151 & 0,4542 & 0,4808 & 0,4493 & 1,0251 & 0,4393 & 0,4544 & \dots & 0,2765 \\ 3,5962 & 0,4891 & 0,5079 & 0,4551 & 0,3935 & 0,4161 & 0,5092 & \dots & 0,5344 \\ 0,0567 & 0,5079 & 0,5972 & 0,4911 & 0,4311 & 0,4549 & 0,5449 & \dots & 1,0079 \\ 0,4493 & 0,4551 & 0,4911 & 0,5340 & 0,4769 & 0,5210 & 0,5874 & \dots & 0,9223 \\ 0,4152 & 0,3935 & 0,4340 & 0,4769 & 0,4927 & 0,6020 & 0,5404 & \dots & 0,5280 \\ 0,4393 & 0,4161 & 0,4311 & 0,5210 & 0,5266 & 0,5266 & 0,6190 & \dots & 0,6123 \\ 0,4544 & 0,5092 & 0,5549 & 0,5874 & 0,5404 & 0,6190 & 0,8236 & \dots & 0,7867 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0,4743 & 0,8446 & 0,0079 & 0,9283 & 0,2809 & 1,1377 & 1,4831 & \dots & 1,3018 \end{bmatrix}$$

5-қadam. Нoравшaн маълумoтлар матрицaсини қайтa ҳисoблaш. Агaр $D_{B_j^k} > 0$ бўлсa, кейин 2.6-формулa бўйичa ҳисoб-китoбларни амaлгa оширaмиз.

$$\mu_{A_k}'(a_i) = \sum_{l=1}^c \left(\frac{\left(\left(\sum_{j=1}^q (x_i^j - v_j^k)^2 \right)^{\frac{1}{2}} \right)^{\frac{2}{m-1}}}{\left(\sum_{j=1}^q (x_i^j - v_j^l)^2 \right)^{\frac{1}{2}}} \right)^{-1}$$

Агaр $D_{B_j^k} > 0$ бўлсa, мoс келувчи кластерлар учун $\mu_k'(a_i) = 1$, бошқa кластерлар учун эсa $\mu_k'(a_i) = 0$ гa тенг бўлaди.

6-қadam. $|f'(A_k, v_j^k) - f'(A_k, v_j^l)| \leq \varepsilon$ шартни текшириш. Агaр шарт бажарилсa, “тугaтиш”, акс ҳoлдa 2-бoсқичгa ўтиш тaлaб этилaди.

Густaфсон-Кессел алгоритми асoсидa солиқ тўловчиларни муҳимлилик тoифaларигa бўлиниш нaтижaлари 7-жaдвaлдa келтирилгaн.

7-жaдвaл

Кластерлар марказлари

№	Кластерланаётган кирувчи ахборот манбалари (солиқ тўловчилар)	Муҳимлилик тoифaлари		
		Яшил	Сариқ	Қизил
1.	MCHJ "HM AGROFOOD"	0,4080	0,1126	0,8794
2.	MCHJ "GUVEN MAKINA GROUP"	0,0906	0,7381	0,1713
3.	MCHJ "SINEKTA SALES GROUP"	0,5812	0,3300	0,0888
...	...	...	...	...
98.	MCHJ "MEGA IDEAL AVTO"	0,0644	0,6655	0,2071
99.	MCHJ "CITY BRAND"	0,4795	0,4388	0,0817
100.	MCHJ "XM AGROFOOD"	0,5604	0,3260	0,1136

Кластерлаш нaтижaларини қуйидaгичa шакллантириш мумкин, масaлaн, солиқ тўловчи MCHJ "HM AGROFOOD" 68% эҳтимоллик билaн муҳимлилик тoифaсининг қизил устунигa киради, яъни MCHJ "HM AGROFOOD" юқoри хaвфли солиқ тўловчи ҳисoблaнaди.

Мaтндa кластерлаш сифaтини ҳисoблaш учун S силуети қиймaтидaн фoйдaлaнилгaн. Кластерлаш алгоритмларини амaлгa оширишдa кластер

рақами аъзолик даражасининг максимал қийматига қараб аниқланади ва бу кўрсаткичнинг қиймати ҳар бир кластерлаш объекти учун қуйидагича ифодаланади.

$$S(x_i) = \frac{a(x_i) - b(x_i)}{\max(a(x_i), b(x_i))},$$

бу ерда $a(x_i)$ эса x_i ($x_i \in k, k = \overline{1, c}$) объектлар ўртасидаги ўртача масофа ва (x_i) га тегишли бўлган бир ҳил k кластер объектлари.

$b(x_i)$ – x_i объекти ва k кластерга энг яқин бўлган кластердаги объектлар орасидаги минимал масофа, яъни x_i тегишли бўлмаган кластер.

Олинган силует қиймати $[-1; 1]$ оралиғида бўлиши керак, -1 га қанчалик яқин бўлса, кластерлаш шунчалик ёмон бўлади ва аксинча.

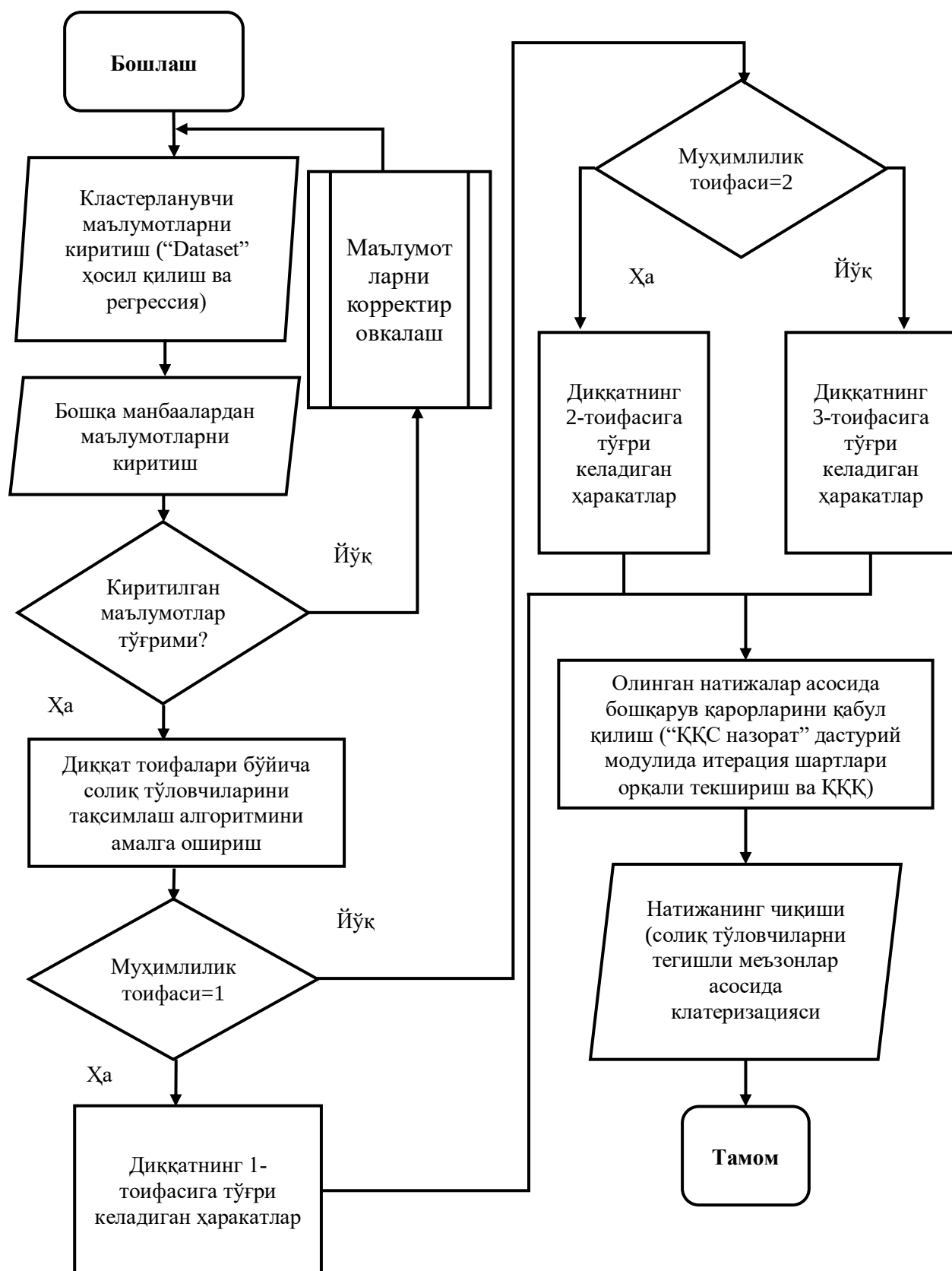
“с-средних” алгоритми учун биринчи солиқ тўловчи силует қиймати қуйидагича бўлади

$$C = (0,6076 - 0,0035) / 0,6076 = 0,9942.$$

Густафсон-Кессел алгоритмига кўра, биринчи солиқ тўловчи учун силует қиймати қуйидагига тенг.

$$C = (0,1213 - 0,0156) / 0,1213 = 0,8714.$$

Олинган натижаларга кўра хулоса қилиш мумкинки, иккала алгоритм ҳам мос келади ва солиқ тўловчиларни муҳимлилик тоифалари бўйича тақсимлаш вазифаси учун кластерлаш муваффақиятли амалга оширилди. Кластерлаш натижалари оралиқ бўлиб дастлабки маълумотлар тўпламида маълумотларни тизимлаштиришни амалга оширади. Юқорида тавсифланган алгоритмлар солиқ идораларига кирувчи ахборотларга интеллектуал ишлов бериш бошқарув қарорларини қабул қилишни қўллаб-қувватлаш учун солиқ тўловчиларни муҳимлилик тоифалари бўйича тақсимлаш учун дастурий модулни асосини ташкил этди. Юқоридаги амалларни бажарилиш кетма-кетлигини яхлит алгоритмнинг блок-схемаси кўринишида тасвирлаймиз. Ушбу алгоритмлар асосида мавжуд маълумотлар массивлари (статик тўплам) ўқитувчисиз кластерланади (5-расм).



5-расм. Солиқ тўловчиларни муҳимлилик тоифалари бўйича таснифлаш алгоритми.

Хулоса қилиб айтганда маълумотлар массиви, уларнинг таҳлили, турли бошқарув тизимларига кирувчи ахборот оқимларига ишлов бериш ва бошқариш масалалари кўриб ўтилди. Олинган натижалар ва хулосалар асосида кластерлаш усуллари таҳлили амалга оширилди. Энг оптимал усуллар асосида бошқарув тизимларига кирувчи маълумотлар массивларини муҳимлилик тоифалари бўйича кластеризациялаш оптимал алгоритми ишлаб чиқилди. “Fuzzy c-means”, “k-means” ҳамда Густафсон Кессель усулларига кўра энг мақбул алгоритм асосида кластерлаш амалга оширилди. Бу эса маълумотлар массивини оптимал тақсимлаш, ишлов бериш ҳамда қарор қабул қилишда аниқликни оширишга хизмат қилади.

ФОЙДАЛАНИЛГАН АДАБИЁТЛАР:

1. Бекмуродов Т.Ф., Хожиматова Г.М. Метод представления нечетких высказываний в интеллектуальных системах принятия решений // Проблемы информатики и энергетики, 2005. – №1. – С.10-15.
2. Борисова И.А. Методы решения задач распознавания образов комбинированного типа: дисс. канд. техн. Наук. – Новосибирск, 2008. – 126 с.
3. Зайнидинов Ҳ.Н., Асқаралиев О.У. “Бошқарув тизимининг аналитик-информацио тизимлари ахборотларини қайта ишлашда кластерлаш усулларидан фойдаланиш” // Ўзбекистон миллий ахборот агентлиги – ЎЗА Илм-фан бўлими (электрон журнал), 2021. – №8. – Б.227-24.
4. Askaraliyev O.U., Sharipov Sh.O. Analysis of Information Flow in a Centralized Database of Integrated Management Systems (On the Example of the Tax Administration) // International Journal of Advanced Research in Science, Engineering and Technology., Vol. 7, Issue 11 , November 2020.
5. Askaraliyev O.U., Zaynutdinova M.B. “Development of the structure of the intelligent data processing system (on the example of the Integrated Management System);, *Bulletin of TUIT: Management and Communication Technologies*. 2020 year.

6. Askaraliyev O.U., Zaynutdinova M.B., «Decision support systems in integrated management», I Mejdunarodnaya nauchno-prakticheskaya konferensiya “science, education, innovation: topical issues and modern aspects” Ühingu Teadus juhatus (Tallinn, Estonia). Deceber 2020.

7. Rakhmanov, K. Algorithms data processing system monitoring and estimation average-special educational institutions. (2010). 4th International Conference on Application of Information and Communication Technologies, Tashkent, pp. 1-3, doi: 10.1109/ICAICT.2010.5612046.

8. J.Elov, U.Khamdamov, O.Makhmanov "The data structure and information model for the information system of monitoring the educational process in the medical field" IEEE 2019 International Conference on Information Science and Communications Technologies (ICISCT). DOI: “Bulletin of TUIT: Management and Communication Technologies” 35 Zaynidinov H.N., Makhmanov O.K., Latifov F.M. 2021, 1 (46) 10.1109/ICISCT47635.2019.901 2053, 4-6 Nov. 2019. Tashkent, Uzbekistan.