

## МАШИН СУРГАЛТЫН АРГААР ХЭРЭГЛЭГЧИЙН ЗАХ ЗЭЭЛИЙГ СЕГМЕНТЧЛЭХ НЬ (БАНКНЫ КРЕДИТ КАРТ ХЭРЭГЛЭГЧДИЙН ЖИШЭЭН ДЭЭР)

Г.Намуун\*, Б.Даваасүрэн\*\*

**Хураангуй:** Энэ судалгааны ажлаар кластер шинжилгээний K-means, шаталсан кластерын арга алгоритм дээр үндэслэсэн машин сургалтын аргаар зах зээлийн сегментчлэлийг хийх боломжтой эсэхийг банкны кредит карт эзэмшигчдийн өгөгдлийг ашиглан гүйцэтгэж, гарсан үр дүнд үнэлэлт дүгнэлт хийхийг зорилоо. Кластер шинжилгээнд орох хувьсагчдыг логик, үндсэн бүрэлдэхүүний шинжилгээ, корреляцийн шинжилгээнд үндэслэн сонгосон. Судалгааны үр дүнгээр банкны кредит карт хэрэглэгчдийг хамгийн оновчтойгоор 3-4 кластерт хуваах нь зохистой болох нь тодорхойлогдов.

**Түлхүүр үгс:** Сегментчлэл, машин сургалт, кластер, кредит карт, K-means, үндсэн бүрэлдэхүүний шинжилгээ, шаталсан кластерын алгоритм

### CUSTOMER SEGMENTATION USING MACHINE LEARNING METHODS: IN CASE OF CREDIT CARD CUSTOMERS

**Abstract:** This research explores the possibility of performing market segmentation using machine learning methods on data from bank credit card customers, specifically employing the K-means cluster analysis and hierarchical clustering algorithms. Variables for cluster analysis were selected using logical, principal component analysis, and correlation analysis approaches. The study evaluated the results of the segmentation, which determined that dividing the bank's credit card users into 3-4 clusters is appropriate.

**Keywords:** Segmentation, machine learning, K-means, principal component analysis, hierarchical clustering algorithms

---

\* МУИС, Бизнесийн сургууль, (E-mail): namu.namka@gmail.com

\*\* МУИС, Бизнесийн сургууль, (E-mail): davaasurenb@num.edu.mn

## **Удиртгал**

### **Судалгааны ажлын үндэслэл**

Хэрэглэгчдийг ижил төстэй хэрэгцээ, шинж чанар болон зан төлвийн онцлогоор нь бүлэглэж зах зээлийн сегментчилэл хийж, тухайн сегмент бүрт тохирсон маркетингийн иж бүрдлийг боловсруулах шаардлага бий болж байгаатай холбогдуулан машин сургалтын аргаар хэрэглэгчдийн зах зээлийг сегмент хийсэн.

### **Судалгааны ажлын ач холбогдол, шинэлэг байдал**

Сегментчлэлийн үйл явц нь бизнес, компаниудад өөрсдийн үйлчлүүлэгчдийг зөв таних ойлгох, цаашлаад зорилтот бүлгүүдэд зориулсан үр дүнтэй, ялгаатай маркетингийн стратеги боловсруулахад тусалдаг. Хэрэглэгчийн сегментчиллийг олон хүчин зүйлийг ашиглан хийж болно. Хамгийн түгээмэл хүчин зүйл нь газарзүйн бүс нутаг, хүн ам зүй (нас, хүйс, гэр бүлийн байдал, орлого), сэтгэл зүйн (үнэт зүйл, сонирхол, амьдралын хэв маяг, бүлгийн харьяалал) болон худалдан авалтын зан төлөв (өмнөх худалдан авалт, сонголт гэх мэт) (Philip Kotler et al., 2022) юм. Тэгвэл сүүлийн жилүүдэд өгөгдөлд суурилсан бизнесийн шийдвэр гаргалтын хэрэгцээ улам бүр нэмэгдэж байна. Энэхүү судалгааны ажлын шинэлэг зүйл нь машин сургалтын аргаар хэрэглэгчдийн өгөгдөл доторх хэв шинж, холбоо хамаарлыг тогтоож, бүлэглэх юм.

### **Судалгааны ажлын зорилго зорилт**

Энэ судалгааны ажил нь банкны кредит карт хэрэглэгчдийн өгөгдөлд машин сургалтаар кластер шинжилгээ хийх аргаар маркетингийн сегмент үүсгэх зорилго тавьсан. Энэ төрлийн шинжилгээ нь машин сургалтын ангилалын удирдагчгүй сургалтад хамаарна.

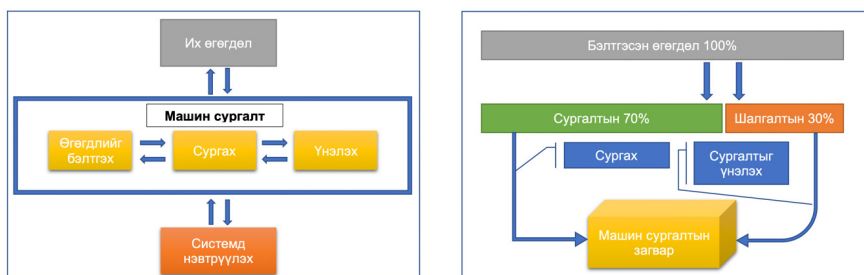
### **Онолын судалгаа**

#### **Хэрэглэгчдийн өгөгдлийг сегментчлэх машин сургалтын арга зүй**

Машин сургалт (machine learning) гэх нэр томъёог 1959 онд Америкийн компьютер тоглоом, хиймэл оюун ухааныг үндэслэгч Arthur Samuel анх томъёолсон. Машин сургалт бол өгөгдсөн өгөгдлийг ашиглан тодорхой даалгаврыг статистик, математикийн аргаар гүйцэтгэх компьютерийн шинжлэх ухааны нэгэн салбар юм. Өгөгдөл нь зураг, дуу, тоон статистик, текст гэх мэт компьютерт кодлож болох бүх зүйлс байж болно. Эдгээр өгөгдлийн нэг төрлөөс маш олныг цуглуулж тоон хэлбэрт хөрвүүлж загвар боловсруулна. Үүний дараа машин сургалтын гол даалгаврыг гүйцэтгэх суралцах процессийг тодорхой арга алгоритмын тусламжтайгаар хамгийн бага алдаатайгаар олж, загварыг тодорхойлно. Загварыг тодорхойлсны дараа түүний гүйцэтгэлийн

чанарыг шинжилж, шаардлага хангаж байвал хэрэглээнд (Зураглал 1) нэвтрүүлнэ. Өөрөөр хэлбэл, шинэ өгөгдлийн хувьд тодорхойлсон загвараа ажиллуулан танин мэдэх даалгаврыг гүйцэтгүүлнэ.

Зураг 1. Машин сургалт хийх бүдүүвч



А. Машин сургалт ерөнхий бүдүүвч    Б. Машин сургалт хийх процесс

Машин сургалтын загварыг өгөгдсөн даалгавраас хамааран дараах үндсэн 3 байдлаар (Bishop, 2006) ангилдаг. Үүнд:

- Удирдагчтай сургалт (supervised learning)
- Удирдагчгүй сургалт (unsupervised learning)
- Хүч нэмэгдүүлсэн сургалт (reinforcement learning)

Энэ судалгааны ажил нь банкны кредит карт хэрэглэгчдийн өгөгдөлд машин сургалтаар кластер шинжилгээ хийх аргаар маркетингийн сегмент үүсгэх зорилго тавьсан. Энэ төрлийн шинжилгээ нь дээрх машин сургалтын ангилалын удирдагчгүй сургалтад хамаарна. Кластер хийдэг маш олон төрлийн арга загварууд байдаг. Тухайлбал, шаталсан кластер, K-medoids, K-mean гэх мэт. Ихэнх өгөгдлийн шинжээчдийн үзэж буйгаар энгийн боловч практикт хамгийн их хэрэглэгддэг, гүйцэтгэл сайтай нь K-means арга юм.

### **K-means кластер**

Кластер бол өгөгдлийн цэгүүдийг ижил төстэйгөөр нь нэг бүлэгт, ижил төстэй бишийг нь өөр нэгэн бүлэгт ангилах техник юм. Ижил төстэй гэдэг нь ажиглалтын утгыг тоон вектор руу хөрвүүлж, стандартчлах хувиргалт хийсний дараа векторын норм ашиглан хоорондох зайг хэмжсэнээр тодорхойлогдоно. Өгөгдлийг  $(n \times m)$  хэмжээс бүхий  $[x_{it}]$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, m$  матриц хэлбэртэй гэж үзье.

K-means бол алгоритмын алхам бүрд шатлан буурдаг кластер шинжилгээний хамгийн нийтлэг арга юм. K-means алгоритмын ерөнхий ажиллах зарчим нь урьдчилан тодорхойлсон кластерийн тоо  $k$  – д тохируулан санамсаргүйгээр

хамгийн эхний кластер бүрийн төвийг тодорхойлсноор ажиглалтын утга болгоныг хамгийн ойрхон орших кластерт оруулна, үүний дараа кластер бүрд орсон ажиглалтын утгуудын дунджаар кластерийн шинэ төвийг бодож, дахин ажиглалтын утга болгоныг хамгийн ойрхон орших кластерт оруулна. Үүнийг кластерийн шинэ төвийн утга тодорхой нарийвчлалтайгаар өөрчлөгдөхгүй байх хүртэл төгсгөлөг удаа давтаснаар дараа тухайн  $k$  кластер ангилах бодлогын харьцангуй сайн кластер гарна. Аливаа нэг ажиглалтын утгын цэгийн кластерийн төвийн цэг хүртэлх зайг дараах 3 нормын аль нэгээр хэмжиж болно.

Евклид норм: 
$$d(i, j) = \sqrt{\sum_{p=1}^m (x_{i,p} - x_{j,p})^2}$$

Манхеттан норм: 
$$d(i, j) = \sum_{p=1}^m |x_{i,p} - x_{j,p}|$$

Минковскийн норм: 
$$d(i, j) = \left( \sum_{p=1}^m |x_{i,p} - x_{j,p}|^q \right)^{1/q}$$

Энд  $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$  өгөгдлийн ажиглалтын  $i$ -р утгыг тоон вектор руу хөрвүүлснийг илэрхийлнэ.

### Шаталсан кластер

Өмнөх хэсэгт авч үзсэн K-means кластер алгоритм нь урьдчилан тодорхойлсон  $k$  ширхэг кластерт ажиглалтын утгуудыг ангилдаг бол шаталсан кластерт урьдчилан тодорхойлсон кластерийн тоо байдаггүй. Шаталсан кластерт ажиглалтын утгуудыг дээрээс доошоо эсвэл доороос дээшээ гэсэн 2 чиглэлээр кластерт хуваадаг. Доороос дээшээ чиглэл буюу хамгийн эхлээд ажиглалтын утга бүр ялгаатай  $n$  ширхэг кластерт хуваагаад, дараа нь хамгийн ойрхон орших 2 кластерийг нэгтгэсээр нэг кластер хүртэл нь ангилах аргыг нэгтгэх арга (agglomerative) гэнэ. Харин эсрэгээр дээрээс доошоо буюу нийт ажиглалтын утгыг нэг кластер гэж үзэн дор хаяж 2 кластерт хувааснаар хамгийн сүүлд  $n$  ширхэг кластер 1 ажиглалтын утга үлдтэл ангилах аргыг хуваах арга (divisive) гэнэ. Үүнийг a, b, c, d, e гэсэн 5 ажиглалтын утга бүхий жишээгээр Зураг – т харуулав.

### Шаталсан кластерын нэгтгэх алгоритм

Өгөгдсөн  $n$  ажиглалтын утгын хувьд ( $n \times m$ ) хэмжээст зайн матриц (ижил төстэй байдлын) тооцоологдсон байг. Тэгвэл Johnson 1967 (Johnson, 1967) онд дэвшүүлсэн шаталсан кластерын нэгтгэх алгоритмыг авч үзье.

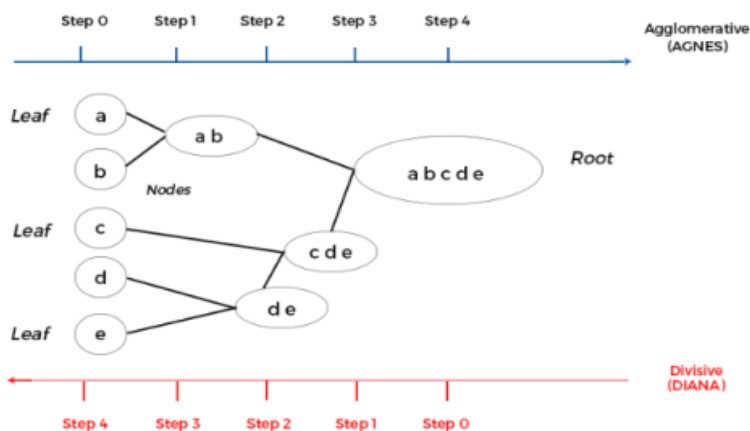
1. Ажиглалтын утга бүрийг зөвхөн өөрийг нь агуулсан  $n$  ширхэг кластерт оруулж, кластер бүрийн хоорондох зайг дээрх нормуудын аль нэгээр бодно.
2. Хамгийн ойрхон орших хоёр кластеруудыг нэгтгэж нэг кластер болгоно.

3. Шинэ кластер болон хуучин кластер бүрийн хоорондох зайг тооцоолно.
4. Алхам 2,3-г ажиглалтын бүх утга  $n$  хэмжээтэй нэг кластерт ортол нь давтана.

Дээрх алгоритмын хувьд кластеруудын нэгтгэхэд хоорондох зайг хэрхэн бодохыг хэд хэдэн янзаар олж болно. Үүнд:

1. Дан холбоос
2. Бүрэн холбоос
3. Дундаж холбоос
4. Төвийн арга
5. Вардын арга

Зураг 2. Хуваах болон нэгтгэх шаталсан кластер



Эх сурвалж: <https://www.r-bloggers.com/2017/12/how-to-perform-hierarchical-clustering-using-r/>

Шаталсан кластерт өмнө дурдсанчлан кластерийн тоо яг нэг тодорхой тоо байхгүй буюу  $n - 1$ -ийн хооронд байна. Тиймээс шаталсан кластерийг дендограмыг тайрах буюу судалгаанд тохиромжтой гэж үзсэн кластерийн тоогоор тайрч, кластерийг тодорхойлдог. Үүнийг зөв буруу гэх ямар нэгэн шинжүүр байхгүй бөгөөд судлаачийн харах өнцгөөр шийдэгдэнэ.

**Машин сургалтын аргаар кредит карт хэрэглэгчдийг сегментчлэх шинжилгээ, судалгааны үр дүн**

Машин сургалтын арга ашиглан хэрэглэгчийн сегментчлэл хийхэд шаардагдах

өгөгдөл нь тухайн байгууллагын хэрэглэгчдийн хангалттай хэмжээний өгөгдөл байх ёстой байдаг. Тиймээс Монгол улсад үйл ажиллагаа явуулж буй банкуудад хэрэглэгчдийн их өгөгдлийг нь нууцлалтайгаар ашиглах санал тавьсан боловч банкууд татгалзсан учраас олон улсын өгөгдлийн шинжээчдийн нээлттэй их өгөгдлийг судалгаанд авч ашигласан. Kaggle платформ дээрх нээлттэй ашиглах (CC0: Public Domain) <https://doi.org/10.5281/zenodo>. зөвшөөрөл бүхий кредит карт хэрэглэгчдийн 18 шинж чанарыг илэрхийлсэн хувьсагчтай, 10127 мөр өгөгдлийг судалгаанд ашиглав.

Өгөгдлийн хувьсагчийн талаарх мэдээллийг дараах хүснэгтэд харуулав.

Хүснэгт 1. Кредит карт хэрэглэгчдийн өгөгдлийн хувьсагчид

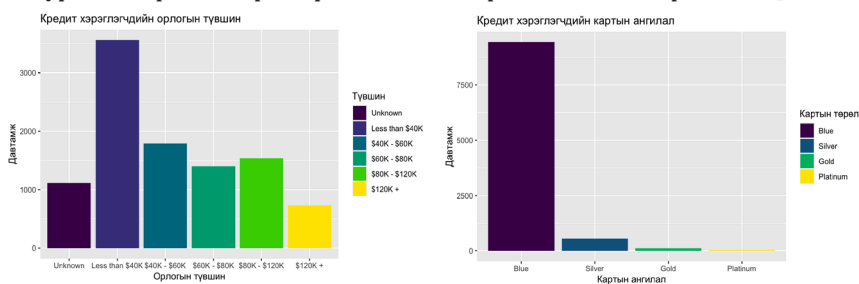
№	Хувьсагч	Тэмдэглэгээ	Төрөл	Авах утга
1	Кредит картын байнгын хэрэглэгч эсэх	Attrition_Flag	Чанарын	Attrited Customer, Existing Customer
2	Нас	Customer_Age	Тоон	26-73
3	Хүйс	Gender	Чанарын	M, F
4	Ам бүлийн тоо	Dependent_count	Тоон	0-5
5	Боловсролын түвшин	Education_Level	Чанарын	Unknown, Uneducated, High School, College Graduate, Doctorate, Post-Graduate
6	Гэр бүлийн байдал	Marital_Status	Чанарын	Divorced, Married, Single, Unknown
7	Орлогын түвшин	Income_Category	Чанарын	Unknown, Less than \$40K, \$40K - \$60K, \$60K - \$80K, \$80K - \$120K, \$120K +
8	Картын төрөл	Card_Category	Чанар	Blue, Silver, Gold, Platinum
9	Карт хэрэглэж буй хугацаа	Months_on_book	Тоон	13-103
10	Бусад данс, картын тоо	Total_Relationship_Count	Тоон	1-6
11	Сүүлийн нэг жилд идэвхгүй байсан сарын тоо	Months_Inactive_12_mon	Тоон	0-6
12	Сүүлийн нэг жилд холбоо барьсан тоо	Contacts_Count_12_mon	Тоон	0-6
13	Гүйлгээний хязгаар	Credit_Limit	Тоон	\$1000-35000
14	Зээл төлөтийн баланс	Total_Revolving_Bal	Тоон	\$0-3000
15	Дундаж худалдан авалтын харьцаа	Avg_Open_To_Buy	Тоон	0-35000
16	Гүйлгээний өөрчлөлт I-IV улирал	Total_Amt_Chng_Q4_Q1	Тоон	0-4

17	Гүйлгээний дүн	Total_Trans_Amt	Тоон	\$0-20000
18	Гүйлгээний давтамж	Total_Trans_Ct	Тоон	0-150

### Эрэн хайх өгөгдлийн шинжилгээ

Эрэн хайх өгөгдлийн шинжилгээ бол машин сургалт хийхийн өмнө зайлшгүй хийх шаардлагатай, өгөгдлийн талаар зүй тогтол, учир шалтгааны талаар нэмэлт мэдээлэл өгдөг чухал шинжилгээ юм. Өгөгдлийн хувьсагчдын хувьд орхигдсон утга байгаа эсэхийг шалгаж үзэхэд орхигдсон утга байхгүй байсан учир өгөгдлийг нөхөх шаардлагагүй гэж үзэв.

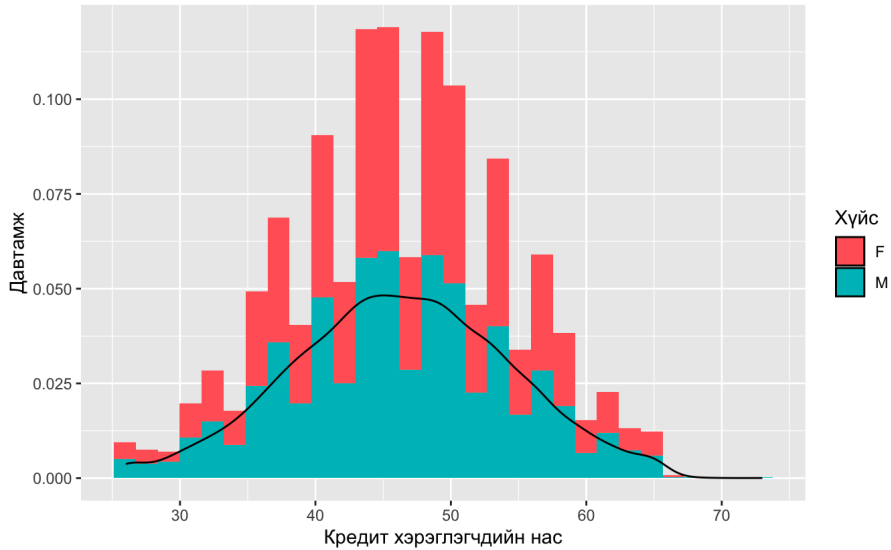
Зураг 3. Кредит карт хэрэглэгчдийн картан ангилал, орлогын түвшин



Эх сурвалж: Судлаачийн тооцоо

Дээрх зурагт харуулснаар кредит хэрэглэгчдийн 93% нь Blue буюу энгийн кредит карт, 5% нь Silver, үлдсэн хувьд Gold болон Platinum карт хэрэглэгчид байна. Кредит карт хэрэглэгчдийн орлогын түвшний ангиллаар 11% нь тодорхойгүй, 35% нь буюу хамгийн их хувь нь жилийн 40,000 доллар хүртэлх орлоготой ангилалд, 18% нь 40,000-60,000 долларын ангилалд, 14% нь 80,000-120,000 долларын ангилалд, 15% нь 60,000-80,000 долларын ангилалд, үлдсэн 7% нь 120,000 доллароос дээш ангилалд тус тус багтаж байна.

Зураг 4. Кредит карт хэрэглэгчид насны тархалт, хүйсээр



Эх сурвалж: Судлаачийн тооцоо

Зураг 3 -т кредит карт хэрэглэгчдийн насны гистограммыг хүйсийн ангиллаар харуулсан бөгөөд насны тархалт 46 дундаж бүхий 8 стандарт хазайлттай хэвийн тархалттай байна. Доорх хүснэгтэд орлого болон боловсролын түвшний контингенцийн хүснэгтийг харуулав.

Хүснэгт 2. Кредит карт хэрэглэгчдийн орлого болон боловсролын түвшний хамаарал

Контингенцийн хүснэгт		Орлогын түвшин (мянган ам.доллаароор)					
		Тодорхойгүй	\$40.0	\$40 - \$60.0	\$60.0 - \$80.0	\$80.0 - \$120.0	\$120.0 <
Боловсролын түвшин	Тодорхойгүй	145	556	269	210	219	120
	Боловсролгүй	185	522	249	195	217	119
	Ахлах сургууль	225	671	355	307	308	147
	Коллеж	108	345	183	132	175	70
	Бакалавр, магистир төгссөн	332	1139	553	422	478	204
	Докторын зэрэгтэй	70	158	70	59	57	37
	Пост доктор	47	170	111	77	81	30

Эх сурвалж: Судлаачийн тооцоо



Дээрх контингентийн хүснэгтээр орлого болон боловсролын түвшинд хамаарал байгаа эсэхийг Пирсоны Хи-квадрат тестээр шалгавал:

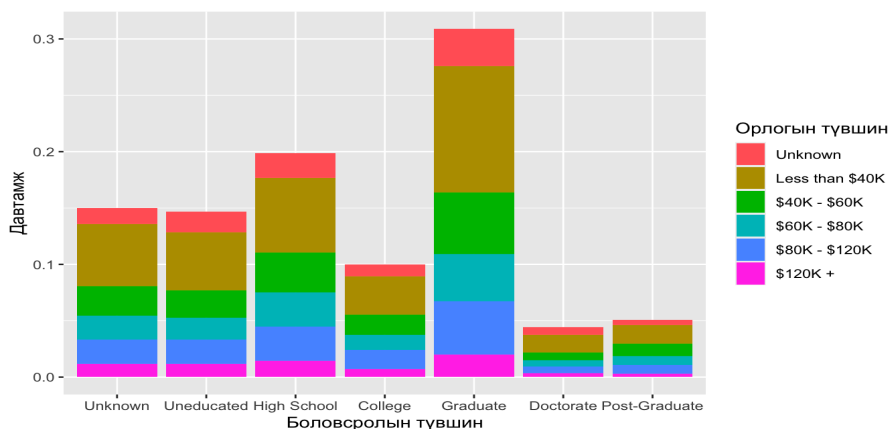
Хүснэгт 3. Пирсоны Хи-квадрат шинжилгээний үр дүн

Хи-квадрат тестийн статистик	Чөлөөний зэрэг	Магадлалын утга
45.254	30	0.03655

Эх сурвалж: Судлаачийн тооцоо

Эндээс 99%-ийн итгэх түвшинд боловсрол болон орлого хамааралгүй гэж үзэж болохоор байна. Контингентийн хүснэгтийг харьцангуй давтамжаар илэрхийлэн графикаар харуулбал Зураг 4.

Зураг 5. Кредит карт хэрэглэгчдийн боловсрол, орлогын түвшний хамаарал



Эх сурвалж: Судлаачийн тооцоо

Зураг 4 -г харуулснаар кредит карт хэрэглэгчдийн хувьд орлогын түвшинд боловсрол нөлөөлдөггүй байна.

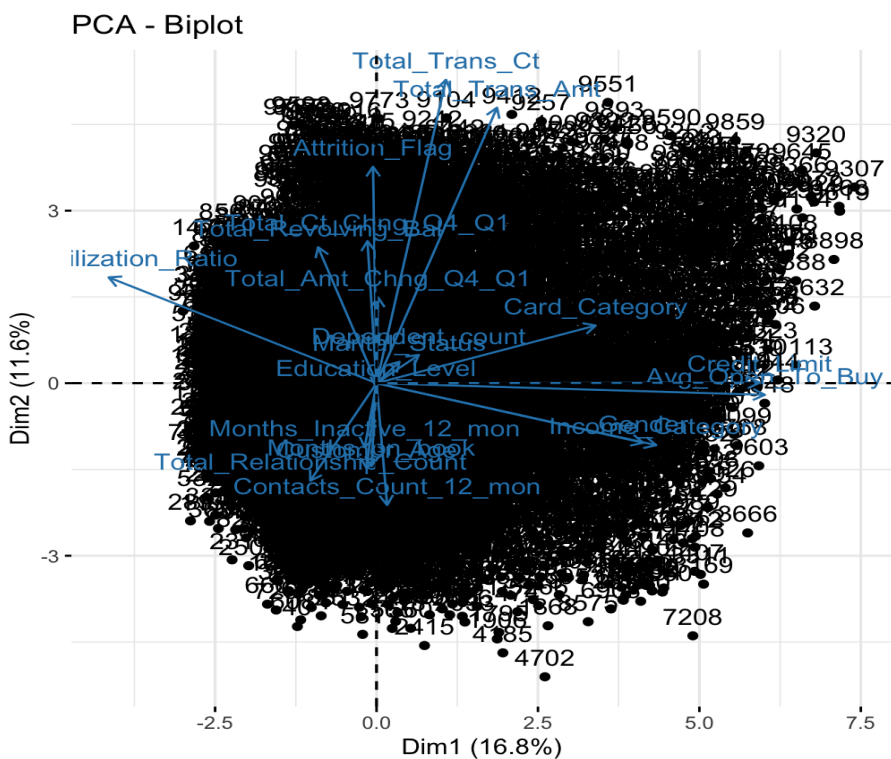
### Кредит карт хэрэглэгчдийг сегментчлэх Кластер шинжилгээ

Кредит карт хэрэглэгчдийн өгөгдөл дээр кластер шинжилгээ хийхийн зорилго нь хэрэглэгчдийн шинж чанарт тохирсон маркетингийн стратеги боловсруулахад шаардлагатай байдлаар хэрэглэгчдийг кластерт хуваах юм. Өөрөөр хэлбэл, хэрэглэгчдийг маркетингийн өөр өөр сегментэд кластерт шинжилгээгээр ангилах юм. Энэ судалгааны ажлаар зөвхөн ижил төстэй шинж чанараар кластер хувааж, кластерийн онцлогийг гаргаж ирэх бөгөөд тухайн кластерт ямар маркетингийн стратеги боловсруулах нь судалгааны дараагийн

үе шат болно.

Өмнөх хэсэгт дурьдсанчлан кредит карт хэрэглэгчдийн өгөгдөлд тэдгээрийн шинж чанарыг илэрхийлэх 18 төрлийн тоон болон чанарын хувьсагч байгаа боловч эдгээр бүх хувьсагчийг ашиглан кластер хийх нь тохиромжгүй. Учир нь зарим хувьсагчид хоорондоо хамааралтай эсвэл кластер шинжилгээнд нөлөөгүй байж болно. Тиймээс уг судалгааны ажилд дараах алхамын дагуу кластер шинжилгээнд орох хувьсагчдыг тодорхойлов. Attrition\_flag буюу кредит картын байнгын хэрэглэгч эсэх хувьсагчийг ашиглан зөвхөн байнгын хэрэглэгчийг сонгон авна. Нийт 10127 хэрэглэгчийн 8500 нь байнгын хэрэглэгч, үлдсэн нь зөвхөн урамшуулал хүртэхийн тулд кредит карт нээлгэсэн хэрэглэгчид болно.

Зураг 6. Үндсэн бүрэлдэхүүний шинжилгээ



Эх сурвалж: Судлаачийн тооцоо

Байнгын хэрэглэгчдийн ямар хувьсагчийг кластерт оруулахыг логик, үндсэн бүрэлдэхүүний шинжилгээ, корреляцийн шинжилгээний үр дүнд суурилан тодорхойлов. Хүснэгт 4 – т кластерт орох хувьсагчдыг харуулав.

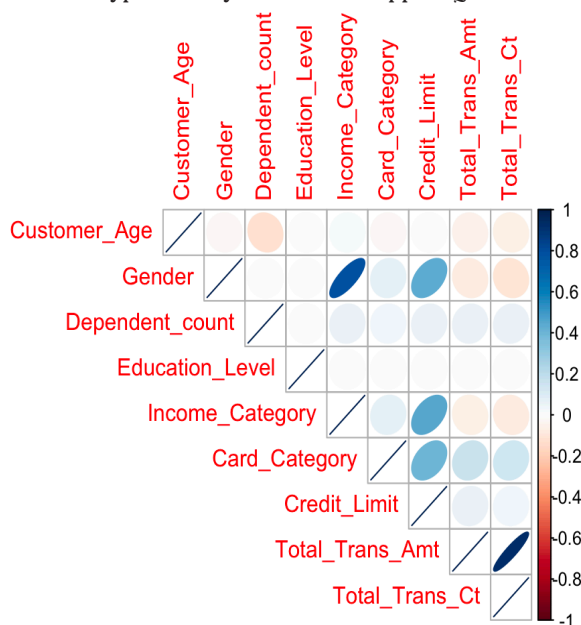
Хүснэгт 4. Кластерт бэлтгэсэн өгөгдлийн тайлбар

№	Хувьсагч	Тэмдэглэгээ	Төрөл	Авах утга
1	Нас	Customer_Age	Тоон	26-73
2	Хүйс	Gender	Чанарын	M, F
3	Ам бүлийн тоо	Dependent_count	Тоон	0-5
4	Боловсролын түвшин	Education_Level	Чанарын	Unknown, Uneducated, High School, College Graduate, Doctorate, Post-Graduate
5	Орлогын түвшин	Income_Category	Чанарын	Unknown, Less than \$40K, \$40K - \$60K, \$60K - \$80K, \$80K - \$120K, \$120K +
6	Картын төрөл	Card_Category	Чанар	Blue, Silver, Gold, Platinum
7	Гүйлгээний хязгаар	Credit_Limit	Тоон	\$1000-35000
8	Гүйлгээний дүн	Total_Trans_Amt	Тоон	\$0-20000
9	Гүйлгээний давтамж	Total_Trans_Ct	Тоон	0-150

Эх сурвалж: Судлаачийн тооцоо

Кластерт орох хувьсагчдын корреляцийн графикийг Зураг 6-т дүрслэв.

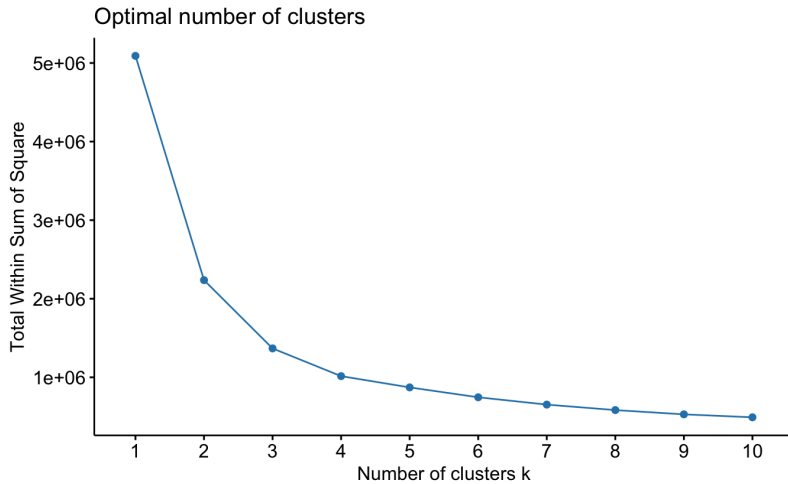
Зураг 7. Хувьсагчдын корреляци



Эх сурвалж: Судлаачийн тооцоо

K-means кластер алгоритмаар шинжилгээ хийхэд оновчтой кластерийн тоог кластер доторх квадратуудын нийлбэр тоог тодорхойлбол Зураг 8-т харуулав.

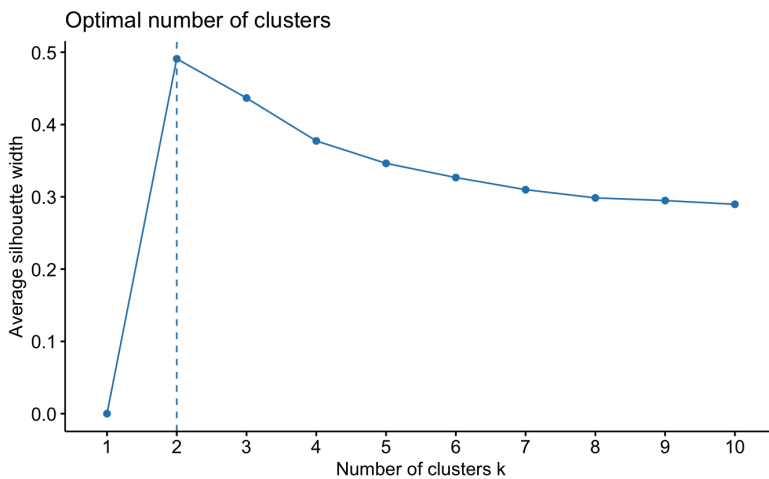
Зураг 8. Оновчтой кластерийн тоо, кластер доторх квадратуудын нийлбэрээр



Эх сурвалж: Судлаачийн тооцоо

Оновчтой кластерийн тоог мөн өөрөөр дундаж silhouette индексээр бодвол

Зураг 9. Оновчтой кластерийн тоо, Силхаутын индексээр



Эх сурвалж: Судлаачийн тооцоо

Оновчтой кластерийн тоог дээрх хоёр аргын аль нэгээр сонгох боломжтой бөгөөд хэдийгээр дундаж силхаут индексийн утгаар 2 кластер байх нь оновчтой гэж байгаа боловч маркетингийн сегмент талаас авч үзвэл хэтэрхий бага учраас кластер доторх квадратуудын нийлбэр хамгийн их буурч байгаагаар буюу 3 кластер байх нь оновчтой гэж үзлээ.

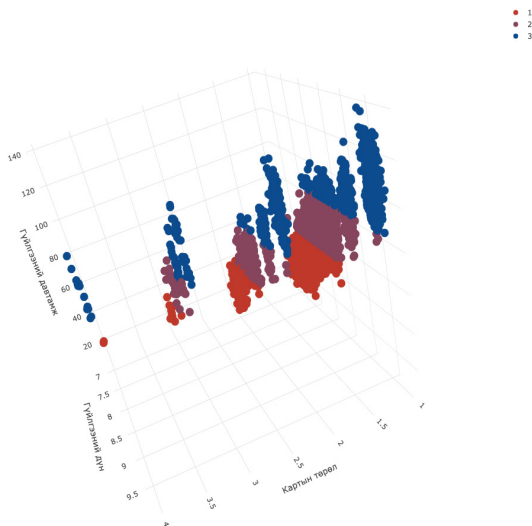
Хүснэгт 5. K-means кластерийн хэмжээ

	Кластер 1	Кластер 2	Кластер 3
Хэмжээ	2219	4750	1531

Эх сурвалж: Судлаачийн тооцоо

K-means кластерт Хүснэгт 4 дэх 9 хувьсагч орсон тул кластерийн хоёр эсвэл гурван хувьсагчдыг сонгон авч 2D эсвэл 3D байдлаар харах боломжтой. Кластерыг хэрэглэгчийн картын төрөл, гүйлгээний дүн болон давтамжаар харвал:

Зураг 10. K-means кластерийн график, гүйлгээний давтамж, дүн, картын төрлөөр

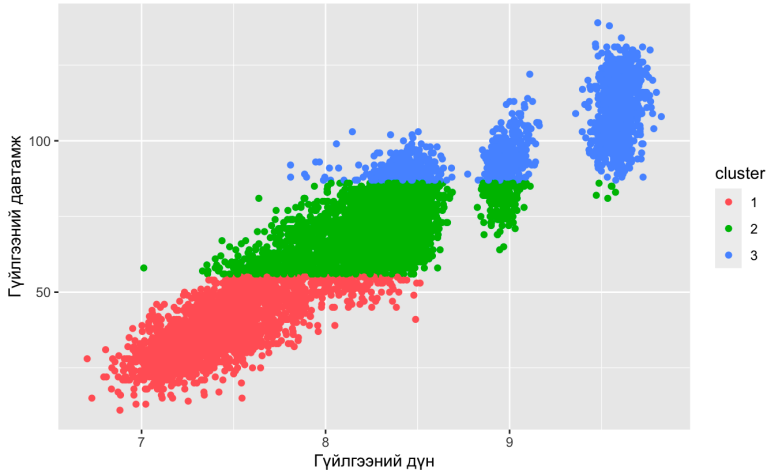


Эх сурвалж: Судлаачийн тооцоо

Зургаас харахад гүйлгээний дүн болон давтамж багатай хэрэглэгчид Кластер 1-д, Кластер 2-т гүйлгээний дүн болон давтамж дунд зэрэг бол Кластер 3-т өндөр дүнтэй олон гүйлгээ хийдэг хэрэглэгчид орсон байна. Мөн

дийлэнх хэрэглэгчид энгийн (1 гэсэн кодтой) буюу Blue картны хэрэглэгчид байгаа бөгөөд хамгийн цөөхөн буюу Platinum карттай хэрэглэгчид бараг бүгд Кластер 3-т орсон байна. Гүйлгээний давтамж болон дүнг 2D графикаар харвал:

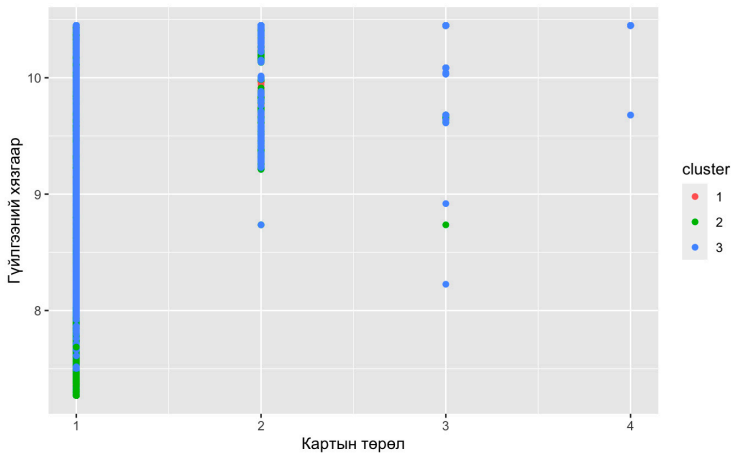
Зураг 11. K-means кластерийн график, гүйлгээний давтамж, дүн



Эх сурвалж: Судлаачийн тооцоо

Гүйлгээний хязгаар болон картны төрлийн хамаарлыг харвал:

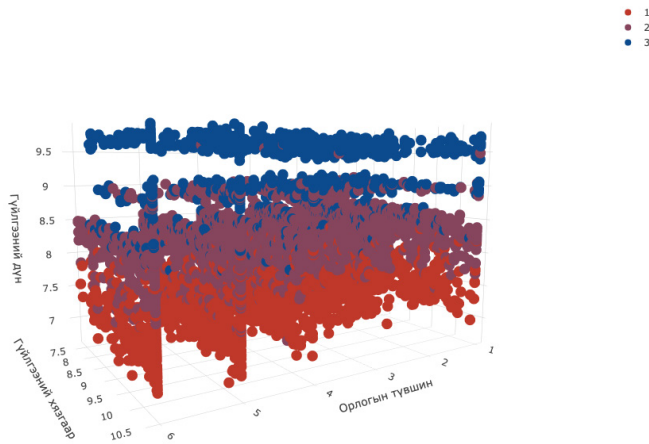
Зураг 12. K-means кластерийн график, гүйлгээний хязгаар, картны төрлөөр



Эх сурвалж: Судлаачийн тооцоо

Кластерийг орлогын түвшин, гүйлгээний хязгаар, гүйлгээний дүнгээр харвал кластеруудын хувьд орлогын түвшин болон гүйлгээний хязгаар жигд тархсан, харин гүйлгээний дүн ялгаатай байна.

Зураг 13. K-means кластерийн график, гүйлгээний дүн, хязгаар, орлогын түвшнээр



Эх сурвалж: Судлаачийн тооцоо

Өмнөх хэсэгт дурьдсанчлан шаталсан кластерийг нэгтгэх аргаар хийсэн дендограмыг Зураг 13-т дүрслэв. Зургаас үзэхэд нийт өгөгдлийг  $n = 8500$  – с 1 хүртэлх кластерт хуваасан ба эндээс судлаач тохиромжтой гэж үзсэн кластерын тоо  $k$  – р кластерыг сонгон авна. Өмнө K-means аргаас үзэхэд судалгааны өгөгдөл дээр олон кластер байх нь зохимжгүй тул дендограмын графикт үзүүлснээр  $k = 4$  байхаар тайрч авав. Шаталсан кластерын  $k = 4$  – өөр тайрч авсанд нийт ажиглалтын утгууд хэрхэн хуваагдсныг доорх хүснэгтэд харуулав.

Хүснэгт 6. Шаталсан кластерийн хэмжээ

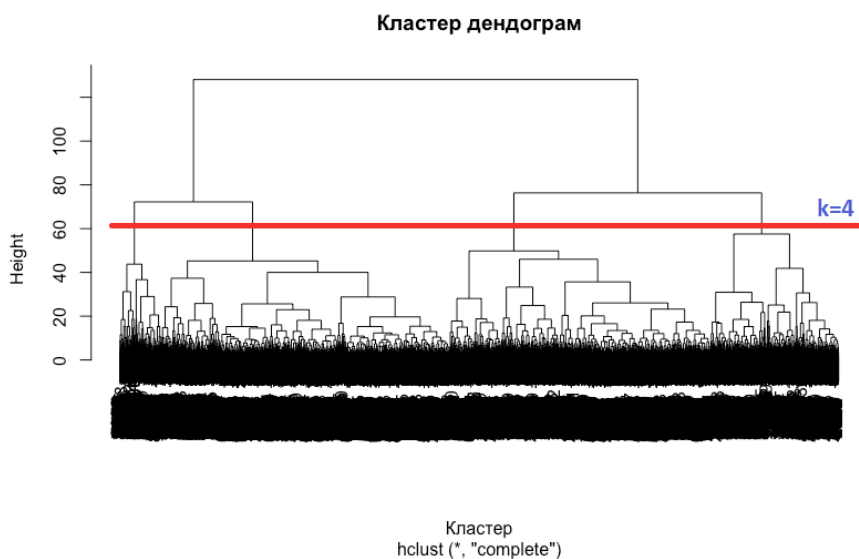
	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Хэмжээ	1513	3085	3415	487

Эх сурвалж: Судлаачийн тооцоо

Зураг 14 – д шаталсан кластерын  $k = 4$  – өөр тайрч авсан ангиллын

хувьд гүйлгээний давтамж, гүйлгээний дүн, боловсролын давтамжийн графикийг харуулав. Кластер 1 – д хамгийн бага давтамж, гүйлгээний дүн багатай бөгөөд боловсролын хувьд бүх түвшний ангиллаас жигд орсон бол Кластер 2, 3 – т гүйлгээний дүн болон давтамж дунд зэрэг, мөн боловсролын түвшин жигд байна. Харин Кластер 4 – т харьцангуй бага буюу 487 хэрэглэгч байгаа бөгөөд эдгээр хэрэглэгчид өндөр гүйлгээний дүн болон давтамжтай, боловсролын хувьд жигд байна. Эндээс шаталсан кластерийн аль ч ангилалд боловсролын түвшин жигд хуваагдаж орсон байгаа тул зөвхөн гүйлгээний дүн, давтамжаар хэрхэн кластерт хуваагдасныг Зураг 15 – т дүрслэв. Эндээс үзвэл өмнөх Зураг 10 – тай төстэй бөгөөд K-means кластерийн Кластер 2,3-ын дундаж нэг кластер нэмэгдсэн байдалтай харагдаж байна.

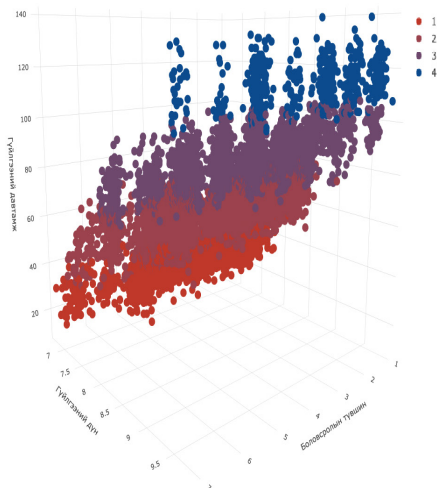
Зураг 14. Шаталсан кластерын дендограм



Эх сурвалж: Судлаачийн тооцоо

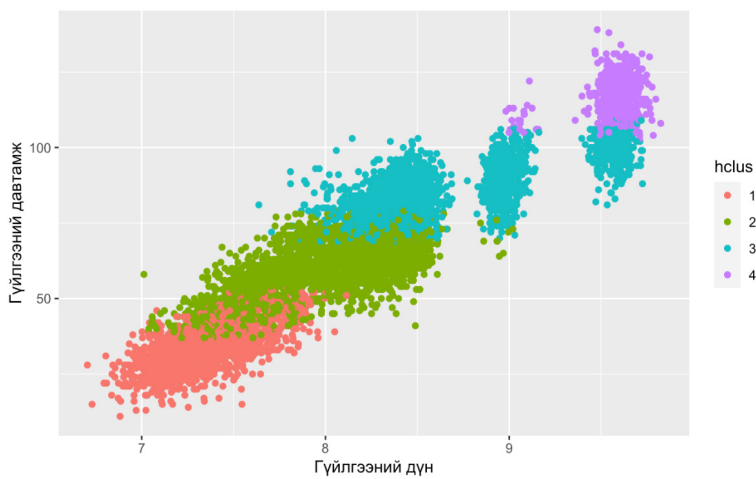


Зураг 15. Шаталсан кластерын график, гүйлгээний дүн, давтамж, боловсролын түвшнээр



Эх сурвалж: Судлаачийн тооцоо

Зураг 16. Шаталсан кластерын график, гүйлгээний дүн, давтамжаар



Эх сурвалж: Судлаачийн тооцоо

### Дүгнэлт

Судалгааны ажлын хүрээнд машин сургалтын удирдагчгүй сургалт буюу кластер шинжилгээг хэрэглэгчдийн өгөгдөл дээр сургаж, маркетингийн ялгаатай стратеги боловсруулах боломж бүхий сегментүүдийг гарган авахыг зорилоо. Судалгааны ажлыг Монголын бизнесийн байгууллагын бодит өгөгдөл дээр хийхээр төлөвлөж байсан боловч өгөгдлийн олдоц, бизнесийн нууцлал зэрэг шалтгааны улмаас боломжгүй болсон тул олон улсын нээлттэй судалгааны зориулалтаар ашиглах боломжтой өгөгдлийн санг ашиглан гүйцэтгэв.

Энэхүү судалгааны зорилго нь кредит картны нээлттэй 10127 хэрэглэгчийн 18 төрлийн хувьсагч агуулсан өгөгдөл дээр эрэн хайх өгөгдлийн шинжилгээ, машин сургалтын кластер шинжилгээ зэргийг хийж, хэрэглэгчийн зах зээлийн сегментчлэлийг ялган таних боломжтой загварыг боловсруулахыг явдал байв. Судалгааны тооцоолол, шинжилгээг R программыг ашиглан хийсэн. Онолын судалгааны шинжилгээ хийх явцад өгөгдлийн эрэн хайх шинжилгээгээр хувьсагчдын давтамж, зүй тогтол, тархалт хоорондын холбоо хамаарал байгаа эсэхийг шинжилсэн.

Судалгааны ажлын үр дүнд кластерт орох хувьсагчдыг логик зүй тогтол, үндсэн бүрэлдэхүүний шинжилгээ, корреляцийн шинжилгээ зэргийг ашиглан 18 төрлийн хувьсагчаас 9-г тодорхойлсон. Үүний дараа K-means алгоритмаар кластер хийхэд өгөгдөлд тохирсон оновчтой кластерын тоог кластер доторх квадратуудын нийлбэр, дундаж силхаут индекс аргаар тодорхойлж, хэрэглэгчдийг 3 кластерт ангилав. K-means кластерын дараа шаталсан кластерын нэгтгэх алгоритмаар кластер хийж,  $k = 4$  кластертай байхаар шаталсан кластерийн модыг тайрч аван график шинжилгээ хийв. Шаталсан кластерын ангилал нь K-means кластертай төстэй үр дүн гарсан. Практикт эдгээр 2 кластер нь маркетингийн ялгаатай стратеги боловсруулах боломжтой хэрэглэгчийн сегментүүдийг үүсгэж байна гэж үзэв. Тухайн кластер бүрийн онцлог шинж чанарыг графикийн аргаар тодорхойлсон. Ингэснээр дараагийн судалгаанд уг кластерийн онцлог шинж чанарыг ашиглан маркетингийн оновчтой стратегийг боловсруулах асуудлыг авч үзэх боломжтой юм.

### Санал зөвлөмж:

Цаашид энэхүү судалгааны ажилд бөөгнөрөл шаталсан кластер, силхаут дундаж өргөн, маркетингийн оновчтой стратеги боловсруулах бүрэн боломжтой.

Тухайлбал, K-means болон шаталсан кластерийн алгоритмын үр дүнд үндэслэн кредит карт хэрэглэгчдийг сегментчилэхтэй холбоотой дараах маркетингийн стратегийг санал болгож байна. Кластер 1 сегментийн хэрэглэгчдэд гүйлгээний давтамжийг ихэсгэх шимтгэлийг хөнгөлөх, k-means алгоритмын кластер 2 болон шаталсан кластерын алгоритмын кластер 2,3

сегментийн хэрэглэгчдэд гүйлгээний мөнгөн дүнг нэмэгдүүлэх урамшуулал зэргийг боловсруулах, k-means алгоритмын кластер 3 болон шаталсан кластерын алгоритмын кластер 4 сегментийн хэрэглэгчдийн гүйлгээний давтамж, мөнгөн дүн өндөр учир буцаан олголт эсвэл эрхийн бичиг (купон) өгөх зэрэг маркетингийн идэвхжүүлэлтийг хийх боломжтой.

Судалгааны ажлын хүрээнд хийгдсэн машин сургалтын загварыг цаашид Монголын бизнесийн байгууллагын бодит өгөгдөл дээр хийж, 5 “О” хүчин зүйлийн загвар хувьсагчдыг нэмж үнэлэн, улмаар системд нэвтрүүлж, маркетингийн үйл ажиллагаанд ашиглах бүрэн боломжтой гэж дүгнэж байна.

### **Ашигласан материал**

- A. Fuster, 2018. Predictably Unequal The Effects of Machine Learning on Credit Markets,
- B. Lantz, 2019. Machine Learning with R, Packt Publishing Ltd.
- Cant, M.C., Strydom, J.W., Jooste, C.J. & Du Plessis, P.J. 2006. Marketing Management. Cape Town: Juta & Co.
- D.James, D. Witten, T.Hastie & R. Tibshirani. 2017. An Introduction to Statistical Learning, Springer.
- Dibb, S. 1998. Market segmentation: Strategies for success. Marketing Intelligence & Planning, 16(7):394-406.
- Handley Wickham, Danielle Navarro, 2023Thomas Lin Pedersen, Elegant graphics for data analysis., ggplot-book.org.
- Leonard Kaufman, Peter J.Rosseeuw, 1990. Finding groups in data: An introduction to Cluster Analysis, New Jersey: John Wiley& Sons, Inc.
- M.Bishop, 2006. Pattern recognition and machine learning. Cambridge: Springer.
- Philip Kotler, Kevin Keller,&Alex Chernev, 2022. Marketing Management, 16th edition, Pearson.
- Waggoner, P. D. 2020.Unsupervised and Machine Learning for Clustering in Political and Social Research, Cambridge university Press.