

ANALYZING THE ECONOMIC ENVIRONMENT USING MACHINE LEARNING

Dulguuntuya Bataa^I, Bayanjargal Darkhijav^{II}, Davaasuren Batsukh^{III}

Abstract: Studying and evaluating a country's trade and business environment ensures economic stability and enhances organizational competitiveness. Business environment assessments serve as a foundation for identifying entrepreneurs' challenges and determining future development directions. This research aimed to apply machine learning techniques to classify the economic environment data that received the lowest evaluation in the 2020 joint study conducted by the Mongolian National Chamber of Commerce and Industry and the Business School. The study employed machine learning methods, including K-Means clustering, Principal Component Analysis (PCA), and Decision Tree algorithms.

Keywords: K-Means clustering, Principal Component Analysis (PCA), feature importance

МАШИН СУРГАЛТ АШИГЛАН ЭДИЙН ЗАСГИЙН ОРЧНЫГ ШИНЖЛЭХ НЬ

Хураангуй: Улсын худалдаа, бизнесийн орчныг судалж, үнэлгээ хийх нь эдийн засгийн тогтвортой байдлыг хангаж, байгууллагуудын өрсөлдөх чадварыг нэмэгдүүлэхэд чухал ач холбогдолтой. Бизнесийн орчны үнэлгээ нь бизнес эрхлэгчдийн тулгамдсан асуудлыг тодорхойлох, цаашдын хөгжлийн чиглэлийг тогтооход суурь мэдээлэл болдог. Энэхүү судалгааны зорилго нь Монголын Үндэсний Худалдаа Аж Үйлдвэрийн Танхим болон Бизнесийн сургуулийн хамтран 2020 онд хэрэгжүүлсэн “Монголын бизнесийн орчны үнэлгээний судалгаа”-ны өгөгдөлд тулгуурлан хамгийн доогуур үнэлгээтэй эдийн засгийн орчныг машин сургалтын аргаар ангилж, кластерчлах явдал юм. Судалгаанд K-Means кластерчлэл, Үндсэн бүрэлдэхүүн хэсгийн шинжилгээ (PCA), шийдвэрийн мод (Decision Tree) зэрэг машин сургалтын аргуудыг ашигласан.

Түлхүүр үгс: K-Means кластерчлэл, үндсэн бүрэлдэхүүн хэсгийн шинжилгээ (PCA), шинж чанарын ач холбогдол

^I National University of Mongolia, School of Information Technology and Electronics, Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences, (E-mail): dulguuntuya_b@mas.ac.mn, ORCID ID: <https://orcid.org/0000-0001-7465-1592>

^{II} National University of Mongolia, School of Information Technology and Electronics. (E-mail): bayanjargal@num.edu.mn

^{III} National University of Mongolia, Business school, (E-mail): davaasurenb@num.edu.mn

1. Оршил

Аж ахуйн нэгжүүдийн өгсөн үнэлгээнд тулгуурлан эдийн засгийн орчныг кластерчлах нь зах зээлийн нөхцөлийн ялгааг илрүүлж, илүү нарийвчилсан, зорилтот бодлогын шийдвэр гаргахад чухал ач холбогдолтой. Монгол Улсад 2017 оноос эхлэн гурван жил тутамд бизнесийн орчны үндэсний хэмжээний судалгаа хийгдэж, олон үзүүлэлтийн дундаж утгад тулгуурлан ерөнхий чиг хандлагыг тогтоож ирсэн (Танхим, 2020). Өмнөх судалгаанд тоон судалгааны аргачлал, статистик дүн шинжилгээ, корреляцийн шинжилгээ зэрэг уламжлалт аргуудыг ашиглан үзүүлэлтүүдийн уялдаа хамаарлыг тодорхойлж, нийтлэг чиг хандлагыг нэгтгэн дүгнэж ирсэн байдаг (Б.Даваасүрэн, Д.Энх-Отгон, М.Должин, Н.Отгонсайхан, , 2018). Гэсэн хэдий ч дундаж үнэлгээ нь байгууллагуудын дотоод бүтэц, өвөрмөц нөхцөл байдал, тулгамдсан сорилтыг ихэвчлэн бүрэн тусгаж чаддаггүй сул талтай.

Сүүлийн жилүүдэд машин сургалтын аргуудыг эдийн засаг, нийгэм, санхүү, эрүүл мэнд, байгаль орчин, уур амьсгалын өөрчлөлт зэрэг олон чиглэлд өргөнөөр ашиглах болсон нь судалгааны шинэ хандлага болж байна. Машин сургалтын аргууд нь өгөгдлийн дотоод бүтэц, нуугдмал хамаарлыг илрүүлэх чадвартай тул уламжлалт статистикийн аргаас илүү нарийвчилсан дүн шинжилгээ хийх боломжийг бүрдүүлдэг. Иймээс байгууллагын үнэлгээнд тулгуурлан эдийн засгийн орчныг машин сургалтын аргаар шинжлэх нь бодлого боловсруулахад шаардлагатай илүү нарийн, мэдээлэлд суурилсан дүгнэлт гаргах нөхцөлийг бүрдүүлнэ. Жишээлбэл, K-Means кластерчлалын алгоритм нь өндөр хэмжээст өгөгдлийг дотоод бүтэц дээр нь үндэслэн үр дүнтэйгээр бүлэглэдэг тул байгууллагын үнэлгээнд тулгуурласан бизнесийн орчныг кластерчлахад тохиромжтой. Мөн Random Forest алгоритмыг ашиглан кластер бүрийг бүрэлдүүлэхэд хамгийн их нөлөө үзүүлж буй хувьсагчдыг тодорхойлж болно.

Энэ судалгаанд 2020 оны эдийн засгийн орчны судалгааны өгөгдлийг ашигласан. 2020 оны байдлаар Монгол улсад 76,300 гаруй аж ахуйн нэгж үйл ажиллагаа явуулж байсан бөгөөд судалгаанд төр болон хувийн хэвшлийн төлөөллөөс бүрдсэн 4,070 байгууллага оролцсон. Тэдгээрийн ихэнх нь хувийн хэвшлийн байгууллага байсан. Өнөөгийн байдлаар аж ахуйн нэгжийн тоо 110 мянгаас давсан нь бизнесийн орчныг үнэлэхэд шинэ сорилт, боломжуудыг бий болгож байна. Иймээс энэхүү судалгааны зорилго нь эдийн засгийн орчны гурван тулгуур бүрэлдэхүүн болох татварын орчин, санхүүгийн зах зээл, хүний нөөц ба хөдөлмөрийн зах зээл гэсэн чухал үзүүлэлтүүдэд тулгуурлан байгууллагуудыг кластерчлах, мөн эдийн засгийн орчны үнэлгээнд хамгийн нөлөө бүхий хүчин зүйлсийг машин сургалтын аргаар тодорхойлоход оршино.

2. Арга зүй

2.1 Өгөгдөл

Энэхүү судалгаанд 2020 онд зохион байгуулагдсан “Монголын бизнесийн орчны үнэлгээний судалгаа”-ний өгөгдлийг ашиглав. Судалгааг бизнесийн төлөөлөгчдийн санаачилгаар, төр болон хувийн хэвшлийн хамтын оролцоотойгоор улсын хэмжээнд хэрэгжүүлж, нийт 4,070 аж ахуйн нэгжийг хамруулсан байна (Танхим, 2020).

2020 оны байдлаар Монгол Улсад 76,300 гаруй аж ахуйн нэгж үйл ажиллагаа явуулж байсан бөгөөд энэхүү судалгаанд тэдгээрийн төлөөллийг аймаг, нийслэлийн хэмжээнд санамсаргүй түүврийн аргаар сонгосон. Судалгаанд нийт 4,070 байгууллага оролцсон бөгөөд оролцогчдын ойролцоогоор 30 хувь нь Улаанбаатар хотод, үлдсэн 70 хувь нь 21 аймгийн бизнесийн байгууллагуудаас бүрдсэн байна.

Бизнесийн чиглэлийн хувьд дотоод худалдаа, жижиглэнгийн үйлчилгээ хамгийн өндөр буюу 46 хувийг эзэлсэн бол үлдсэн хэсгийг нийтийн хоол үйлчилгээ, бөөний болон гадаад худалдаа, экспортын чиглэлүүд голчлон бүрдүүлжээ. Борлуулалтын түвшний хувьд оролцогч байгууллагуудын 47 орчим хувь нь 50 сая төгрөгөөс бага борлуулалттай байсан бол үлдсэн байгууллагуудын ихэнх нь 51–300 сая төгрөгийн борлуулалтын орлоготой байв. Ажиллагчдын тооны хувьд, 66.5 хувь нь 10 хүртэл ажилтантай, 23 хувь нь 10–50 ажилтантай, 6 хувь нь 50–200 ажилтантай, харин 3 орчим хувь нь 200 ба түүнээс дээш ажилтантай аж ахуйн нэгжүүд байв. Энэ нь судалгаанд жижиг, дунд байгууллагуудын оролцоо давамгайл байсныг харуулж байна.

Судалгаанд бизнесийн орчныг тав үндсэн чиглэлд ангилсан бөгөөд үүнд:

1. Эдийн засгийн орчин,
2. Засаглал ба хууль эрх зүйн орчин,
3. Нийгэм соёлын орчин,
4. Дэд бүтэц ба технологийн орчин,
5. Байгууллагын дотоод орчин багтана.

Эдгээрээс энэхүү судалгаанд эдийн засгийн орчны үзүүлэлтүүдийг сонгон авч шинжилгээ хийв. Эдийн засгийн орчин дараах гурван дэд бүрэлдэхүүн хэсгээс бүрдэнэ:

1. Хүний нөөц ба хөдөлмөрийн зах зээл — ажилтны хандлага, мэдлэг ур чадвар, тогтвор суурьшил, хүний нөөцийн олдоц зэрэг үзүүлэлтүүд.

2. Татварын орчин — татвар төлөх хугацаа, татварын урамшуулал ба хөнгөлөлт, орлогын татвар болон нэмэгдсэн өртгийн татварын хувь хэмжээ зэрэг үзүүлэлтүүд.
3. Санхүүгийн зах зээл — банк, санхүүгийн үйлчилгээний төрөл, хөрөнгө оруулалтын бодлого, зээлийн хүүний түвшин зэрэг үзүүлэлтүүд.

Эдгээр гурван чиглэлийн хүрээнд нийт 33 үзүүлэлтийн нарийвчилсан үнэлгээ цуглуулсан бөгөөд оролцогч байгууллагууд үзүүлэлт бүрийг 0–5 хүртэлх онооны үнэлгээний системээр үнэлсэн. Үүнд 0 нь “маш муу”, 5 нь “маш сайн” гэсэн утгыг илэрхийлнэ. Энэхүү үнэлгээний өгөгдөл нь судалгааны шинжилгээний үндсэн санг бүрдүүлсэн.

2.2 Өгөгдөл цэвэрлэгээ ба нөхөлт

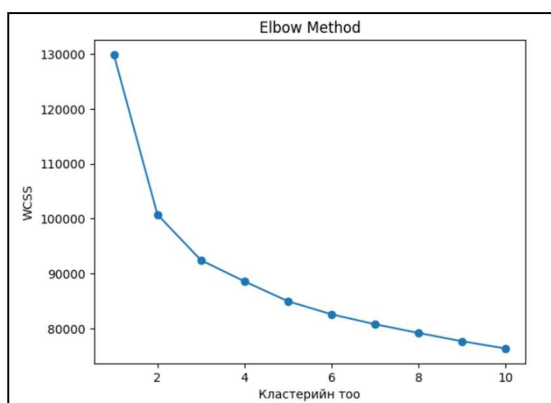
Машин сургалтын аргуудыг хэрэглэхийн өмнө өгөгдлийн чанарыг үнэлэх, дутуу болон зөрчилтэй мэдээллийг илрүүлэх нэн чухал алхам юм. Анхан шатны өгөгдлийн шалгалтаар зарим хувьсагчид дутуу утгатай байгааг илрүүлэв. Дутуу өгөгдөл нь кластерчлал болон статистик дүн шинжилгээний үр дүнг гажуудуулах эрсдэлтэй тул нөхөх алхам хийгдсэн. Нийт 4,070 байгууллагын өгөгдлөөс 3.3% орчимд дутуу утга илэрсэн бөгөөд эдгээр нь гол төлөв санхүүгийн зах зээл болон татварын орчны үзүүлэлтүүдэд төвлөрч байв. Эхний шатанд хувьсагч тус бүрийн дутуу мэдээллийн түвшинг шалгаж, дутуу утга нь нийт мэдээллийн 50%-иас давсан байгууллагуудыг шинжилгээнээс хассан. Үлдсэн өгөгдлийн хүрээнд дутуу утгуудыг нөхөхөд хамгийн ойрын хөршүүдийн арга (K-Nearest Neighbors, KNN)-г ашигласан. KNN нөхөлт нь өгөгдлийн дотоод бүтэц, хамаарлыг аль болох хадгалж, тухайн нэгжтэй төстэй хэв шинжтэй байгууллагуудын мэдээллээр дутуу утгыг бодитойгоор ойролцоолуулдаг давуу талтай. KNN нь тухайн байгууллагатай төсөөтэй бусад байгууллагуудыг (хөршүүд) олон хэмжээст орон зайд олж, тэдний утгуудын дунджийг авч дутуу мэдээллийг тооцоолох статистик арга юм (Breiman, 2001). Энэхүү судалгаанд $k = 5$ гэж тогтоосон бөгөөд энэ нь тухайн нэгжийн дутуу утгыг хамгийн ойрын таван хөршийн утгын дундажаар нөхсөнийг илэрхийлнэ. Хөршүүдийг тодорхойлоход 0–5 хүртэл нэг ижил хэмжээсээр үнэлэгдсэн 33 үзүүлэлтийн хоорондын зайг ашигласнаар үзүүлэлт хоорондын жингийн зөрүү гарах эрсдэл буурсан. Нөхөлтийн дараа бүх үзүүлэлтийн утга 0–5 хүрээний гадна эсэхийг шалгаж, шаардлагатай тохиолдолд дахин хяналт хийв. Эцэст нь өгөгдлийн бүрэн бүтэн байдлыг хангаж, боловсруулсан 3,936 байгууллагын өгөгдөл цаашдын кластерчлал, PCA шинжилгээ болон Random Forest загварчлалд ашиглагдсан.

2.3 Өгөгдлийн боловсруулалт ба кластерчлал

Өгөгдлийн боловсруулалт ба кластерчлалын шат нь өгөгдлийн чанарыг сайжруулах, дотоод бүтцийг илүү тодорхой илрүүлэх, улмаар машин сургалтын алгоритмуудын гүйцэтгэлд нөлөөлөх орчны хүчин зүйлсийг багасгах зорилготой. Дутуу утгуудыг нөхсөний дараа шинжилгээний нарийвчлал болон харьцуулах боломжийг сайжруулахын тулд нэмэлт урьдчилсан боловсруулалтуудыг хийв. Энэ шатанд өгөгдлийн тархалтын хэлбэр, хэмжээсний зөрүү болон хэт утгуудын (outlier) нөлөөг шалгаж, стандартчилах алхмыг хэрэгжүүлсэн. Нэгдүгээрт, үзүүлэлтүүдийн тархалтын хэлбэр болон хэмжээсний зөрүүгээс үүдэх нөлөөг бууруулах зорилгоор бүх 33 үзүүлэлтийг z-score хувиргалтаар стандартчилсан. Өөрөөр хэлбэл, хувьсагч бүрийг өөрийн дундаж ба стандарт хазайлтаар хэмжиж (дундаж = 0, стандарт хазайлт = 1) төвшинд оруулснаар дараагийн шинжилгээнд тэнцүү жинтэйгээр оролцох нөхцөл бүрдүүлэв.

Үүний дараа байгууллагуудын эдийн засгийн орчны үнэлгээнд тулгуурлан K-Means кластерчлал хийв. K-Means-ний зорилго нь өгөгдлийг ижил төстэй шинж чанартай хэсгүүдэд ангилдаг хяналтгүй сургалтын алгоритм бөгөөд өгөгдлийн цэгүүдийг урьдчилан тодорхойлсон k тооны кластеруудад хуваах замаар гүйцэтгэгддэг (Firdaus, G. A. A., & Wulandhari, L. A. , 2024).

Кластерын оновчтой тоог тодорхойлохын тулд эхлээд тохой (Elbow) арга ашиглаж, кластерын тоо өсөхөд дотоод вариаци хэрхэн буурч байгааг графикаар үнэлэв. Энэ арга нь муруйн “тахир” цэгийг тодорхойлж, хамгийн тохиромжтой k -г сонгоход ашиглагддаг. Нэмэлт баталгаажуулалтаар силует индекс болон Calinski–Harabasz оноо зэрэг хэмжүүрийг ашиглаж харьцуулсны үндсэн дээр нийтлэг дүнгээр $k = 3$ -ыг сонгов. Сонголтын дараа санамсаргүй эхлүүлэлтэд мэдрэг эсэхийг Adjusted Rand Index (ARI)-ээр шалгахад $ARI \approx 0.994 (\pm 0.002)$ гарч, сонгосон кластерууд өндөр тогтвортой болох нь батлагдав.



Зураг 1. Кластерын тоог сонгох “тохой” аргын график

2.4 Хувьсагчийн нөлөөллийн хэмжүүр

Кластер бүрийн онцлог шинж, ялгаатай үзүүлэлтүүдийг тогтоохын тулд 33 үзүүлэлтийг ашиглан байгууллагын кластерын ($k = 3$) ангиллыг таамаглах гол хувьсагчдын ач холбогдлын шинжилгээг хийв. Энэхүү зорилгоор Random Forest (RF) ангилагчийг хэрэглэж, тухайн хувьсагч кластерийг ялган танихад ямар хэмжээгээр хувь нэмэр оруулж буйг үнэлсэн.

Random Forest нь олон тооны шийдвэрийн модны дундаж дээр суурилан ангилал хийдэг хяналттай сургалтын алгоритм бөгөөд хувьсагч бүрийн ангиллын гүйцэтгэлд үзүүлж буй нөлөөллийг (feature importance) статистикийн аргаар тооцох боломжийг олгодог. Түүнчлэн энэхүү арга нь overfitting буюу загварын хэт тохирлын эрсдэл бага, үр дүнг тайлбарлахад хялбар байдаг нь давуу талд тооцогдоно (Sruthi, 2023).

Загварыг сургалтын өгөгдөлд суурилан бүтээж, байгууллагын кластерын шошгыг (label) ашиглан таамаглал хийсэн. Үүний үр дүнд ангилалд хамгийн их нөлөө үзүүлж буй хувьсагчдыг ач холбогдлын оноо (Mirka Saarela, Susanne Jauhiainen., 2021) (importance score)-ны дагуу эрэмбэлж, хамгийн өндөр нөлөө үзүүлсэн хувьсагчид тодорхойлогдож, эдийн засгийн орчныг тодорхойлох гол ялгаатай үзүүлэлтүүдийг гаргаж авсан.

3. Үр дүн

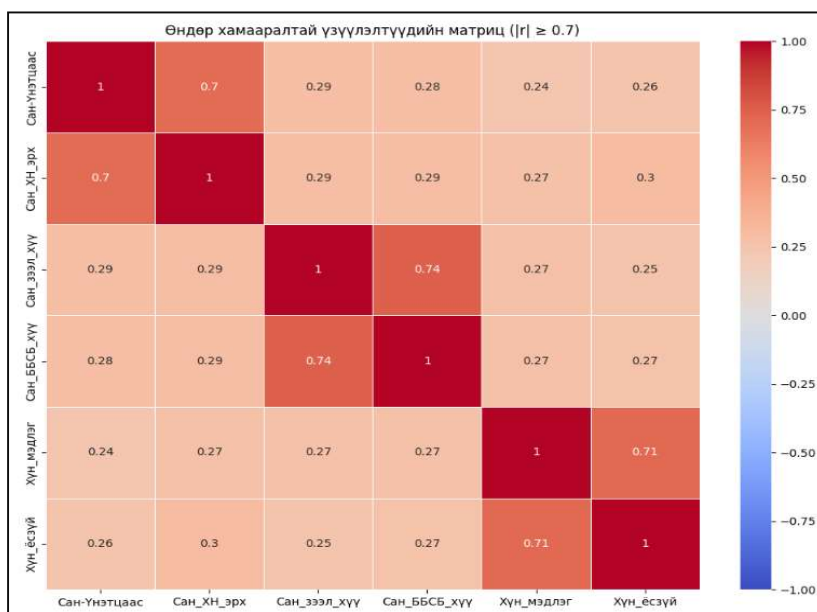
3.1 Корреляцийн шинжилгээ

Кластерчлалын өмнөх шатанд үзүүлэлтүүдийн хоорондын уялдаа хамаарлыг тодорхойлох зорилгоор корреляцийн шинжилгээ хийсэн бөгөөд Зураг 2-т өндөр

хамааралтай үзүүлэлтүүдийг ($|r| \geq 0.7$) харуулсан бөгөөд өнгөний градиент нь хамаарлын чиглэл болон хүчийг илтгэнэ. Үүнд улаан өнгө нь эерэг, цэнхэр нь сөрөг хамаарлыг илэрхийлж байна. Үр дүнгээс харахад:

- Санхүүгийн үйлчилгээ авах боломж ба санхүүгийн эрхийн хүртээмж ($r = 0.7$) хооронд өндөр эерэг хамааралтай байна. Энэ нь санхүүгийн хүртээмж сайжрах тусам үйлчилгээ авах боломж дагаад нэмэгддэгийг харуулж байна.
- Зээл авах нөхцөл ба ББСБ-ын үйл ажиллагаа ($r = 0.74$) нь мөн хүчтэй эерэг хамааралтай гарсан нь санхүүгийн зах зээл дэх хоёр үзүүлэлт хоорондоо нягт уялдаатай байгааг илтгэнэ.
- Хүний нөөцийн хүрээнд мэдлэгийн түвшин ба ёс зүй ($r = 0.71$) үзүүлэлтүүдийн хооронд өндөр хамаарал ажиглагдсан нь ажиллах хүчний чанарыг тодорхойлох үзүүлэлтүүд хоорондоо уялдаа холбоотойг харуулж байна.

Ерөнхийдөө, нийт 33 үзүүлэлтээс зөвхөн цөөн хосуудын хооронд ($r \geq 0.7$) өндөр хамаарал илэрсэн нь мэдээллийн давхардал хязгаарлагдмал, хувьсагчид харьцангуй бие даасан шинж чанарыг илэрхийлж буйг баталж байна. Иймээс дараагийн кластерчлалын шинжилгээнд бүх үзүүлэлтүүдийг хадгалан ашиглах нь өгөгдлийн бүтцийг алдагдуулахгүй, олон хэмжээст орон зайд гажуудал үүсгэх эрсдэл бага гэж үзэв.



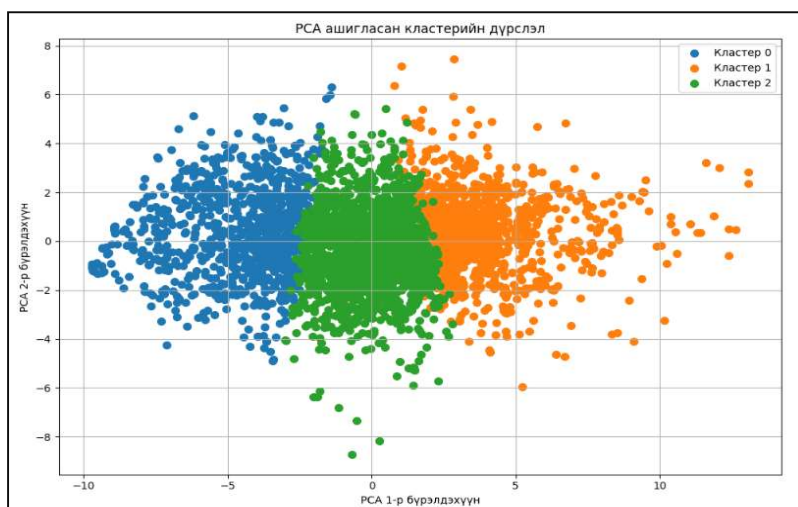
Зураг 2. Корреляцийн шинжилгээний үр дүн

3.2 Кластерчлалын үр дүн

РСА нь өндөр хэмжээст (33 үзүүлэлт) мэдээллийг хамгийн их вариаци бүхий хувьсагчдыг ашиглан хэмжээсийг бууруулдаг. (Зураг 3). РСА аргын тусламжтайгаар байгууллагуудыг 3 кластерт хуваасан. Үүнд :

- Кластер 0 (цэнхэр) — зүүн талд төвлөрч, ерөнхий үнэлгээ харьцангуй бага аж ахуйн нэгжүүд давамгайлж байна.
- Кластер 1 (ногоон) — төвийн бүсэд байрлаж, хоёр талын кластеруудтай давхцах шинжтэй; хил орчимд тархалт харьцангуй өндөр байгаа нь дундын буюу шилжилтийн хэв шинжтэй байгууллагуудыг илтгэнэ.
- Кластер 2 (улбар шар) — баруун талд төвлөрч, ерөнхий үнэлгээ өндөр байгууллагууд зонхилжээ.

РСА-гийн эхний бүрэлдэхүүн (PC1) дагуу кластерууд шугаман дараалал үүсгэн сунасан нь өгөгдлийн гол өөрчлөлт “ерөнхий үнэлгээний түвшин”-тэй шууд холбоотойг харуулна. Харин хоёр дахь бүрэлдэхүүн (PC2) нь татвар, санхүү, хүний нөөц зэрэг дэд чиглэлийн ялгааг илүү нарийвчлан тусгаж, босоо чиглэлд бага боловч ач холбогдолтой тархалтыг харуулж байна. Кластеруудын огтлолцож буй бүс, ялангуяа Кластер 1-ийн хил орчимд байгаа байгууллагууд нь их болон бага үнэлгээтэй бүлгийн аль алиных нь шинжийг зэрэг агуулсан, шилжилтийн байр суурьтай гэж үзэж болно. Энэхүү зураг нь дотоод баталгаажуулалтын үзүүлэлтүүдтэй (ж. Silhouette, CH, DB) нийцэж, гурван кластер тодорхой хэмжээнд салангид боловч бүрэн тусгаарлагдаагүй байгааг харууллаа.

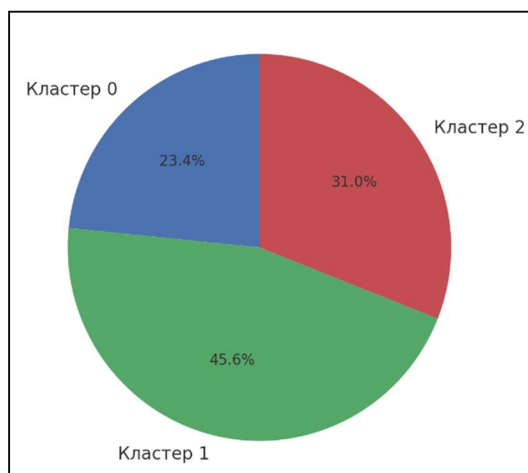


Зураг 3. Үндсэн бүрэлдэхүүн хэсгийн шинжилгээний үр дүн (РСА)

3.3 Кластеруудын бүтэц, онцлог

Өмнөх хэсэгт тодорхойлсон 3 кластерийг байгууллагын хувьд авч үзвэл нийт 3,936 байгууллагаас 33 үзүүлэлтээр гурав кластерт ангилагдсан. Кластеруудын тархалт нь дараах байдалтай байна (Зураг 4):

- Кластер 0 – нийт байгууллагын 23.4% (921 байгууллага),
- Кластер 1 – нийт байгууллагын 45.6% (1,793 байгууллага),
- Кластер 2 – нийт байгууллагын 31.0% (1,222 байгууллага).



Зураг 4. Кластерийн эзлэх хувь

Ийнхүү Кластер 1 хамгийн том, Кластер 2 дундаж, харин Кластер 0 хамгийн жижиг сегмент болж байна. Энэ нь бодлогын хувьд эхний ээлжид Кластер 1-д чиглэсэн арга хэмжээг авч үзэх боломжтойг (хэмжээгээрээ хамгийн олон байгууллагыг хамарч байгаа тул), харин Кластер 0 нь дэмжлэг шаардлагатай, харьцангуй эрсдэлтэй бүлэг болохыг харуулж байна.

Хүснэгт нэгт кластер бүрийн гол статистик үзүүлэлтийг нэгтгэн харуулсан бөгөөд дундаж оноо нь эдийн засгийн орчны үнэлгээний 0–5 онооны системд суурилсан.

Хүснэгт 1. Кластеруудын гол шинж чанарууд

Кластер	Байгууллагын тоо	Дундаж оноо	Салбар	Үйл ажиллагааны хугацаа	Жилийн борлуулалт	Ажилтны тоо
Кластер 0	921	1,83	Жижиглэн худалдаа, хоол үйлчилгээ	6–10 жил, 11+ жил	50 сая хүртэл	10 хүртэл, 10–50 хүн
Кластер 1	1793	2,55	Жижиглэн худалдаа, хоол үйлчилгээ	11+ жил	50 сая хүртэл	10 хүртэл,
Кластер 2	1222	3,19	Жижиглэн худалдаа, хоол үйлчилгээ, бөөний худалдаа	1–5 жил, 11+ жил	50–300 сая	10 хүртэл,

Кластер 0 нь нийт 921 байгууллагыг (23.4%) хамарч, дундаж үнэлгээ нь 1.83 буюу хамгийн бага түвшинд байна. Энэ бүлгийн байгууллагуудын ихэнх нь жижиглэн худалдаа болон хоол үйлчилгээний салбарт хамаарах бөгөөд үйл ажиллагааны хугацаа нь 6–10 жил болон 11-ээс дээш жилээр давамгайлдаг. Жилийн борлуулалтын хэмжээ ихэвчлэн 50 сая хүртэл, ажиллагсдын тоо микро болон жижиг (≤ 10 болон 10–50 хүн) ангилалд ордог.

Кластер 1 нь 1,793 байгууллагыг (45.6%) багтааж, дундаж үнэлгээ нь 2.55 буюу дунд түвшинд байна. Салбарын хувьд жижиглэн худалдаа, хоол үйлчилгээ давамгайлж, 11-ээс дээш жил ажилласан байгууллагууд зонхилно. Тэдний жилийн борлуулалтын хэмжээ ихэвчлэн 50 сая төгрөг хүртэл бөгөөд ажиллагсдын тоо голдуу 10-аас бага байна.

Кластер 2 нь нийт 1,222 байгууллагыг (31.0%) хамарч, дундаж үнэлгээ 3.19 буюу хамгийн өндөр үзүүлэлттэй байна. Салбарын хувьд жижиглэн худалдаа, хоол үйлчилгээ давамгайлсан хэв шинжтэй ч бөөний худалдааны оролцоо илүү тодорхой харагдана. Үйл ажиллагааны хугацаа нь 1–5 жил болон 11-ээс дээш жилээр

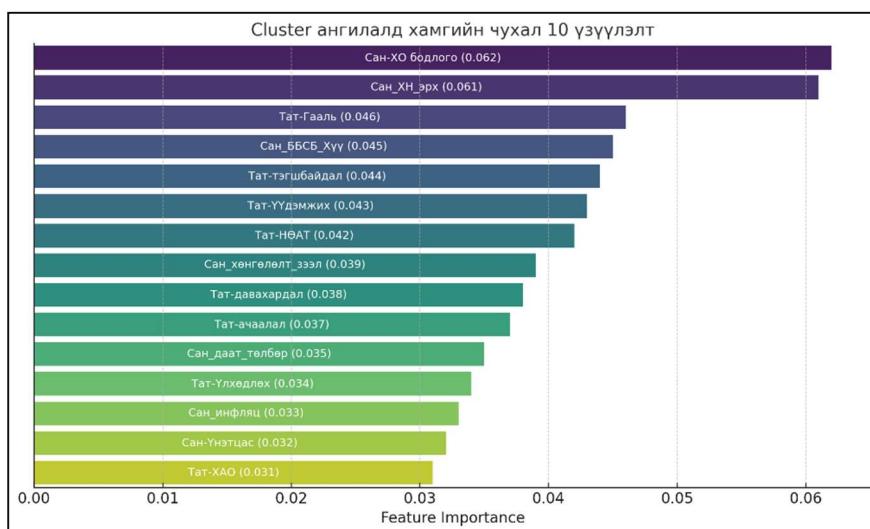
төвлөрөх хандлагатай. Жилийн борлуулалт 50–300 саяын хооронд хэлбэлзэж, ажиллагсдын тоо ихэвчлэн 10-аас доош байна.

3.4 Кластерт нөлөөлөх хүчин зүйлс

Кластеруудын ялгаа нь зөвхөн дундаж оноонд бус, тодорхой үзүүлэлтүүдийн нөлөөллөөс хамаарч байгааг илүү гүнзгийрүүлэн судлахын тулд хувьсагчийн ач холбогдлын шинжилгээ хийв. Энэ нь кластер бүрийг ялгах гол хүчин зүйлсийг илрүүлэх, цаашдын бодлогын чиглэлийг тодорхойлох зорилготой юм.

Зураг 5 нь стандарчилсан 33 үзүүлэлтээр сургасан Random Forest ангилагчийн пермутацид суурилсан ач холбогдол-оор эрэмбэлсэн чухал-10 хувьсагчийг үзүүлж байна.

Үр дүнгээс харахад “Хөрөнгө оруулалтын урамшууллын бодлого” болон “Хувьцаа эзэмшигчдийн эрхийн хамгаалалт” нь кластеруудыг ялгахад хамгийн их нөлөөлсөн хүчин зүйлсээр тодорч байна. Харин “Хувь хүний орлогын албан татвар”, “Үнэт цаасны зах зээлд хандах боломж” зэрэг үзүүлэлтүүд нь харьцангуй бага ач холбогдолтой гарсан байна.



Зураг 5. Кластер ялгах хувьсагчдын ач холбогдлын шинжилгээ

Кластер 0 ба 1-д сайжруулах шаардлагатай үзүүлэлтүүд

Кластерын дотоод ялгааг илүү тодруулах, бизнесийн орчныг сайжруулахад чиглэсэн

бодлогын зөвлөмж боловсруулахын тулд кластер хоорондын үзүүлэлтүүдийн дундаж зөрүүг харьцуулан шинжиллээ. Энэхүү дүн шинжилгээ нь харьцангуй доогуур үнэлгээтэй бүлгүүд (Кластер 0 ба Кластер 1)-ийн сайжруулах шаардлагатай чиглэлүүдийг тодорхойлох зорилготой юм.

Доорх хүснэгтүүдэд эдийн засгийн орчны сайжруулахад хамгийн их шаардлагатай чухал-10 үзүүлэлтүүдийн үр дүнг харууллаа. Үзүүлэлтүүдийг кластер 2 (хамгийн өндөр үнэлгээтэй бүлэг)-ын дундажтай харьцуулсан зөрүүний хэмжээгээр эрэмбэлсэн. Дээрээс доош эрэмбэлэгдсэн утгууд нь хамгийн том зөрүүтэй, өөрөөр хэлбэл бодлогын түвшинд илүү их анхаарал, дэмжлэг шаардах чиглэлүүдийг илэрхийлнэ.

Хүснэгт 2. Кластер 0 үзүүлэлтүүд

Үзүүлэлтүүд	З өрүү
Үйлдвэрийг дэмжсэн татварын орчин	1. 639
Нэмэгдсэн өртгийн албан татвар	1. 582
Татварын шударга байдал, тэгш байдал	1. 576
Хөрөнгө оруулалт, санхүүгийн эх үүсвэрийн хүртээмж	1. 531
Хувь хүний орлогын албан татвар	1. 526
Санхүүгийн эх үүсвэр олох боломж	1. 520
Нийгмийн даатгалын шимтгэлийн хувь хэмжээ	1. 503
Татварын ачаалал	1. 467
Онцгой албан татварын хувь хэмжээ	1. 452
Үл хөдлөх хөрөнгийн албан татварын хувь хэмжээ	1. 450

Хүснэгт 3. Кластер 1 үзүүлэлтүүд

Валютын ханшийн нолоо	0.8
	962
ББСБ-ын зээлийн хүү	0.8
	663
Хөрөнгө оруулалтыг дэмжих бодлого	0.8
	505
Засгийн газрын болон олон улсын хөнгөлөлттэй зээл	0.8
	328
Гадаад валютын олдоц	0.8
	148
Банкны зээлийн хүү	0.7
	790
Үйлдвэрийг дэмжсэн татварын орчин	0.7
	435
Банкны зээл	0.7
	366
Банкны зээлийн хүү	0.7
	340
Ажилтнуудын тогтвор суурьшил	0.7
	320

Эдгээр үр дүнгээс дүгнэхэд, Кластер 0-д татварын орчны сайжруулалт, Кластер 1-д санхүүгийн бодлого онцгой ач холбогдолтой болохыг харуулж байна. Иймд бодлого боловсруулах явцад тус кластеруудын онцлог ялгааг харгалзан, чиглэсэн арга хэмжээг хэрэгжүүлэх нь илүү үр дүнтэй байх боломжтой.

4. Дүгнэлт

Энэхүү судалгаанд Монголын бизнесийн орчныг байгууллагуудын үнэлгээнд тулгуурлан машин сургалтын аргаар шинжилж, эдийн засгийн орчны онцлогийг илэрхийлсэн гурван үндсэн кластерийг тодорхойлов. Кластерууд нь эдийн засгийн орчны үнэлгээ, байгууллагын бүтэц, орлогын түвшин болон салбарын чиглэлээрээ ялгаатай байгааг харуулсан нь бизнесийн орчны ялгаатай нөхцөлд нийцсэн, зорилтот бодлогын шийдвэрийг нийтлэг дундаж бус, бүлэг тус бүрийн өвөрмөц онцлогт тулгуурлан боловсруулах шаардлагатайг илтгэж байна.

Random Forest шинжилгээгээр хөрөнгө оруулалтын орчин, санхүүгийн хүртээмж, хувьцаа эзэмшигчдийн эрхийн хамгаалалт зэрэг үзүүлэлтүүд хамгийн чухал нөлөөтэй болох нь тогтоогдов. Энэ нь бизнесийн эдийн засгийн тогтвортой хөгжлийг хангахын тулд татварын дарамтыг бууруулах, санхүүгийн хүртээмжийг нэмэгдүүлэх, хүний нөөцийн тогтвортой байдлыг дэмжих зэрэг бодлогын чиглэлд анхаарлаа төвлөрүүлэх шаардлагатайг харуулж байна. Ийнхүү өгөгдөлд суурилсан ангиллын дүн шинжилгээ нь Монголын бизнесийн орчныг илүү нарийвчилсан байдлаар ойлгож, төр болон хувийн хэвшлийн зорилтот эдийн засгийн бодлого боловсруулахад чухал суурь мэдээлэлийг ач холбогдолтой гэж үзэж байна.

Ашигласан материал

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.
- Firdaus, G. A. A., & Wulandhari, L. A. . (2024). Company clustering based on financial report data using k-means. *Computer Science and Information Technologies*, 5(3), 243–253.
- Mirka Saarela, Susanne Jauhiainen. (2021). Comparison of feature importance measures as explanations for classification models. . Springer.
- Sruthi, E. R. (2023). Understand random forest algorithms with.
- Б.Даваасүрэн, Д.Энх-Отгон, М.Должин, Н.Отгонсайхан. . (2018). Монголын бизнесийн орчны өнөөгийн нөхцөл байдал, сайжруулах боломж. *Journal of Business and Innovation*, 4(1), 4–19.
- Танхим, М. Ү. (2020). Монголын бизнесийн орчны үнэлгээний судалгаа 2020. Улаанбаатар .