

# Өгөгдөл Олборлолтын Аргыг Ашиглан Жижиглэн Худалдааны Борлуулалтын Орлогыг Нэмэгдүүлэх нь

Намсрайдоржийн Мөнхцэцэг, Ивановын Бямбасүрэн\*,  
Сэрээтэрийн Батбаяр, Бөхсүрэнгийн Энхтуул

*Монгол Улсын Их Сургууль, Хэрэглээний Шинжлэх Ухаан, Инженерчлэлийн Сургууль,  
Улаанбаатар 14200, Монгол Улс*

\*Холбоо барих зохиогч: [byambasuren@seas.num.edu.mn](mailto:byambasuren@seas.num.edu.mn); ORCID:0000-0001-5236-7857

Өгүүллийн мэдээлэл: Хүлээн авсан: 2022.09.17; Зөвшөөрөгдсөн: 2022.10.01;  
Нийтлэгдсэн: 2022.12.26

**Хураангуй:** Энэхүү өгүүлэлд бид жижиглэн худалдаа хийдэг дэлгүүрийн хувьд зарагдсан барааны түүхэн өгөгдөлд тулгуурлан бараа, татан нийлүүлэлтийн таамаглалыг өгөгдөл олборлолтын argiоi болон association rule-г ашиглан орлогыг нэмэгдүүлэхийн тулд худалдан авагчийн нэг удаагийн худалдан авалтын өгөгдөлд дүн шинжилгээ хийж, өгөгдөл олборлолтын өнгөрсөнд болсон зүйлс яагаад тохиолдсон болохыг ойлгох (descriptive analytics) арга алгоритмуудын талаар болон тэдгээрийг хэрхэн оновчтойгоор ашиглах талаар өгүүлнэ. Одоо үед аж ахуй нэгж, бизнесийн байгууллага бүр түүнчлэн хувь хүн хүртэл үйл ажиллагааны түүхэн мэдээллээс өгөгдөл олборлолтын арга техникүүдийг ашиглан шаардлагатай мэдээллийг мэдлэг болгон авч цаашдын ажил, амьдралын үйл ажиллагаандаа ухаалгаар ашиглаж байна. Өгөгдөл олборлолтын хувьд бусад хэрэгслүүдээр илрүүлэх боломжгүй эдийн засгийн үр өгөөж өгөх боломжтой нууцлаг зүй тогтол, холбооснуудыг ч илрүүлж чаддагаараа илүү давуу талтай.

**Түлхүүр үгс:** Машин сургалтын алгоритмууд, сагсны шинжилгээ, хэрэглэгчийн зан төлөв, Argiоi, Хамаарлын шинжилгээ

## 1. Оршил

Өнөөдрийн өрсөлдөөнт нийгэмд хувиараа багахан хэмжээний борлуулалт хийдэг дэлгүүрүүдийн хувьд булан бүрд байрлах өөртэй нь ижил жижиг дэлгүүрүүд, түргэн үйлчилгээний цэгүүд түүнчлэн том хэмжээний супермаркет мөн бөөний төвүүдтэй өрсөлдөн хэрхэн орлоготой байх вэ гэдэгт дан ганц бизнесийн холч ухаан, хашир туршлага нөлөөлөхөөс гадна өдөр тутмын бараа борлуулалтын мэдээллийг кассын машины программ хангамж ашиглан бүртгэж байгаа тохиолдолд уг өгөгдөл дээр тулгуурлан хэрхэн зөв, ухаалгаар борлуулалтын пирамидыг оновчтой зохион байгуулж орлогыг нэмэгдүүлж, зарлагыг бууруулахад хариу өгөхөд өгөгдөл олборлолтын техникүүдийг ашиглах бүрэн боломжтой болсон байна. Жижиг дэлгүүрийн эзэн бараа борлуулалтын түүхэн өгөгдөлд тулгуурлан sales forecasting буюу борлуулалт таамаглалын анхааруулга өгдөг программ хангамж ашигласнаар ямар барааг сарын аль өдрүүдэд, хэдий хэмжээний тоо ширхгээр татах, барааг лангуунууд дээр байрлуулахад аль барааг, аль барааны ойролцоо байрлуулбал худалдан авагчийн сэтгэлд нийцтэй байж сагсны агуулга ихсэх вэ?, ямар хүнд ямар барааг санал болгох, хэнд ямар урамшуулал тооцох вэ? гэх мэтчилэн борлуулалтын өгөгдлөөс орлогыг нэмэгдүүлэх нуугдмал өгөгдлийг гарган авах боломжтой. Борлуулалтын таамаглалын зорилго нь бизнесийн ухаалаг шийдвэр гаргахад ашиглаж болох

мэдээллээр хангах явдал юм [1]. Жишээлбэл, хэрэв таамаглал бүтээгдэхүүн, үйлчилгээний борлуулалтыг 30%-иар нэмэгдүүлэхийг харуулж байгаа бол илүү том бизнес, эрэлтийг хангахын тулд нэмэлт ажилтан авах шаардлагатай болно. Нөгөөтэйгүүр борлуулалтын хомсдолын таамаглал нь зардлыг бууруулах эсвэл маркетингийн бодлогоо өөрчлөх зэрэг урьдчилан арга хэмжээ авах замаар үр нөлөөг бууруулах боломжийг олгоно.

## 2. Өгөгдөл олборлолтын алгоритмууд

Өгөгдөл олборлолт (data mining) нь их хэмжээний өгөгдлөөс тухайн өгөгдлийн талаарх ойлголт авах болон ирээдүйд болох зүйлсийг таамаглах боломжийг олгодог зүй тогтол, холбоо хамаарлуудыг илрүүлэх зорилгоор өгөгдлийг шинжлэн судлах процесс юм. Тодорхой мэдээлэл, мэдлэгийг өгөгдөл олборлолтын үр дүнд илрүүлэхдээ машин сургалтын төрөл бүрийн алгоритмуудыг ашигладаг.

Түгээмэл хэрэглэгддэг алгоритмуудыг төрлөөр нь нэгтгэвэл:

- Association (Хамаарал) [2] - Нэг тодорхой үзэгдэлд хамааралтай тохиолдлуудыг илрүүлнэ. Хамгийн түгээмэл хэрэглээ нь худалдан авалтын сагсны шинжилгээнд ашиглагдах хамаарлын зүй тогтлын илрүүлэлт юм. Жишээлбэл, хүнсний дэлгүүрийн борлуулалтын өгөгдөлд хийсэн шинжилгээгээр чипс худалдан авсан үйлчлүүлэгчдийн 65 хувь нь кола худалдан авсан болох нь тогтоогдсон гэж үзье. Энэ мэдээлэл нь менежерүүдэд тухайн бүтээгдэхүүнүүдийг хамтад нь урамшуулал зарлах, тавиурын байрлалыг нь ойртуулах гэх мэт шийдвэрүүдийг гаргах боломж олгоно.
- Sequential (Дараалсан) [2] - Үйл явдлууд хугацаагаар холбогддог. Жишээлбэл, ямар нэг байшин зарагдсанаас хойш хоёр долоо хоногийн дотор шинэ хөргөгч зарагдах магадлал 70 хувь, мөн нэг сарын хугацаанд шарах шүүгээ зарагдах магадлал 45 хувь байдаг байж болно.
- Classification (Ангилалт) [3] - Өгөгдлийг тодорхой мэдэгдэж буй шинж, дүрмүүдэд үндэслэн ялгаатай ангиудад хуваахад ашиглагддаг. Түгээмэл хэрэглэгддэг алгоритмууд нь Naïve Bayes, decision tree, regression, neural network зэрэг алгоритмууд ордог. (supervised) Жишээ нь: Үйлчлүүлэгчдийн дундаас өрсөлдөгч компани руу шилжих магадлалтай хэсгийг ангилах (churn modeling).
- Clustering (Бүлэглэлт) [4] - Төсөөтэй шинж чанар бүхий объектуудыг бүлэглэх алгоритмууд. Ангилалтын алгоритмуудтай төсөөтэй боловч ямар бүлгүүдэд хуваах нь урьдчилан тодорхойгүй байдаг. (unsupervised) Жишээ нь: Ижил худалдан авалтын хэвшилтэй хэрэглэгчдийг илрүүлэх.
- Prediction (Таамаглалт) [4] - Мэдэгдэхгүй байгаа утгыг өгөгдлийн бусад талбаруудаас хамааруулан таамаглана. Хэрвээ таамаглах хувьсагчийн утга, тасралтгүй байвал регрессийн алгоритмууд өргөн хэрэглэгддэг. Жишээ нь: Борлуулалт ирэх саруудад ямар байхыг таамаглах.

Цахим өгөгдлийн хэмжээ жилээс жилд огцом нэмэгдэж байгаа нь уг өгөгдлийг олборлох замаар оновчтой шийдвэр гаргах, бизнесийн үйл ажиллагааг сайжруулах боломжийг олгодог. Өгөгдөл олборлолт нь машин сургалтын алгоритмуудыг бодит өгөгдөлд нэвтрүүлж үр дүн гарган авах үйл ажиллагаа бөгөөд өгөгдөл олборлолтын алгоритмуудын чанар сайжирсаар байгаа боловч тодорхой шинжилгээний асуудалд хамгийн сайн тохирох өгөгдөл олборлолтын алгоритмыг сонгох нь хүндрэлтэй асуудлуудын нэг байсаар байна.

### **3. Жижиглэн худалдааны салбарт ашиглагдах өгөгдөл олборлолтын хэрэглээ**

Өгөгдөл олборлолт нь бизнесийн зорилгоос хамааран олон салбарт хэрэгжих боломжтой. Үүнд: өнөөгийн бизнесийн ертөнцөд үйлдвэрлэл, банк санхүү, эрүүл мэнд, даатгал, харилцаа холбоо, анагаах ухаан, жижиглэн худалдаа гэх мэт олон салбар мэдээлэл олборлох технологийг ашиглаж байна. Үүнээс жижиглэн худалдааны салбарт дараах байдлаар ашиглах боломжтой [5]. Үүнд:

#### **Хэрэглэгчийг олж авах, хадгалах**

Шинэ хэрэглэгчдийг татах нь одоо байгаа хэрэглэгчээс илүү үнэтэй байдаг. Тиймээс одоо байгаа худалдан авагчдын худалдан авалтын зан төлөвийг мэдсэнээр, тодорхой бүтээгдэхүүнийг худалдан авах үйлчлүүлэгчдийн хэрэгцээ, сонирхлыг урьдчилан таамаглаж чадна. Энэ төрлийн шинжилгээг хийснээр жижиглэн худалдаачид хямдрал зарлах, бараа санал болгох замаар одоо байгаа үйлчлүүлэгчдээ хадгалж, шинэ үйлчлүүлэгчдийг татах, худалдан авагчдыг олж авах боломжтой.

#### **Зах зээлийн сагсны шинжилгээ**

Зах зээлийн сагсны шинжилгээ нь ассоциаци буюу холбоо, хамаарлын дүрмийн дагуу ямар барааг, ямар бараатай хамтад нь худалдаж авах боломжтойг ойлгох арга техник юм. Энэ нь янз бүрийн бүтээгдэхүүнүүдийн хоорондын холбоог харуулах замаар үйлчлүүлэгчид, худалдан авалтын хэв маягийн талаар тодорхой зөвлөмж, санал өгдөг. Энэ төрлийн техникийн үр дүнгээр бараа, бүтээгдэхүүний хоорондын холбоог ашиглан лангууны өрөлт хийх, эд зүйлсийн байршил, сурталчилгааг хослуулах замаар борлуулалтын орлогыг шийдвэрлэхэд тустай. Ингэснээр хэрэглэгчид бараагаа хялбархан олох боломжтой бөгөөд мөн худалдан авагчид бараануудаа зэрэг зарах боломжийг олгодог.

#### **Хэрэглэгчийн сегментчилэл ба зорилтот маркетинг**

Сегмент гэдэг нь зах зээлийг тодорхой тэмдэгтээр хэд хэдэн хэсэгт хуваах явдал юм. Өгөгдлийн олборлолтыг зан төлөвт тулгуурлан хэрэглэгчдийг бүлэглэх эсвэл бүлэг үүсгэн ашиглаж болно. Энэ төрлийн мэдээлэл нь кластер дахь ижил төстэй үйлчлүүлэгчдийг тодорхойлох, үнэ цэнтэй үйлчлүүлэгчдийг хадгалах, зорилтот маркетингийн хариу үйлдэл үзүүлэх магадлалтай хүмүүсийг тодорхойлох боломжтой.

### **4. Хэрэглэгчийн худалдан авалтын мэдээлэлд хийсэн хамаарлын шинжилгээ**

Зах зээл дээр худалдан авагчийн сагсанд дүн шинжилгээ хийх аргыг жижиглэнгийн худалдааны орлогыг нэмэгдүүлэхэд ихэвчлэн ашигладаг бөгөөд үйлчлүүлэгчид аль бүтээгдэхүүнийг хамтад нь хамгийн их худалдан авдаг болохыг тодорхойлох арга юм. Үүнд борлуулалтын цэгийн (POS) гүйлгээний өгөгдөлд дүн шинжилгээ хийж, хамт зарагдсан бараануудыг мэдээллээс хамааран өөр, өөр бараануудын хоорондын хамаарлыг тодорхойлдог. Энэхүү шинжилгээг хийснээр худалдан авагчдын худалдан авалтын хэв маягийг тодорхойлох, ялангуяа худалдан авагчид ямар, ямар бүтээгдэхүүнийг хамтад нь худалдан авах боломжтойг борлуулалтын таамаглалаас мэдэх боломжтой [5]. Худалдан авалтын сагсанд дүн шинжилгээ хийж байгаа бидэнд мэдэгддэг үлгэр жишээ жишээнүүд бол idemu-гийн хичээлүүдийг худалдаж авахад заавал хамааралтай, эсвэл төстэй хичээлийг санал болгоод 2 болон түүнээс дээш хичээл худалдан авбал бага үнээр багцыг санал болгодог. Мөн Amazon вэб сайт дээрээс бүтээгдэхүүнийг сонгосны дараа сонгосон бүтээгдэхүүний талаарх дэлгэрэнгүй мэдээллийг харуулахаас гадна “Хамтдаа байнга худалдаж авдаг” бүтээгдэхүүнүүдийг харуулдаг.

Жижиглэнгийн худалдаан дээр сагсны шинжилгээний үр дүнд үндэслэн аль бүтээгдэхүүнийг бие биетэйгээ ойрхон байрлуулах лангууны өрөлтийг зөв хийснээр борлуулалтын орлогыг нэмэгдүүлэх боломжтой. Жишээлбэл, хэрэв үйлчлүүлэгчид үхрийн амталсан мах эсвэл тахианы мөч худалдан авч байгаа бол аль төрлийн амтлагч

хачир, шингэн уух бүтээгдэхүүнийг авахыг санал болгохыг энэ шинжилгээгээр мэдэж боломжтой. Энэхүү мэдээлэл нь бүтээгдэхүүн байршуулах, борлуулалтыг нэмэгдүүлэх, борлуулалтыг тооцоолох талаар шийдвэр гаргахад тусална. Мөн борлуулалтын урамшуулал зарлах, үнэнч хэрэглэгчийн тодорхойлох, багцлах зэрэг үнийн стратегид нөлөөлдөг [6].

Зах зээлийн сагсны шинжилгээ хийх хэд хэдэн арга байдаг ч холбоо, хамаарлыг олох нь хамгийн өргөн хэрэглэгддэг аргуудын нэг хэвээр байсаар байна. Энэ нь худалдан авалтын гүйлгээний өгөгдөлд хамтдаа зарагдсан бүтээгдэхүүнүүдийн давтамжийг тоолж, цаашид хамтад нь зарагдах магадлалыг тодорхойлдог арга юм. Энэхүү шинжилгээг хийхдээ Python дээр априори алгоритм болон холбоо хамаарлын дүрмийг ашиглан жижиг дэлгүүрийн түүхэн өгөгдөлд тулгуурлан сагсанд хамаарлын шинжилгээг хэрхэн хэрэгжүүлэх талаар туршилт хийлээ.

#### 4.1. Хамаарлын шинжилгээнд ашигласан өгөгдөл

Энэхүү илтгэлд ашиглаж байгаа өгөгдөл нь 2021 оны 9-р сараас 2022 оны 10-р сар хүртэлх 13 сарын өдөр тутмын бараа борлуулалтын 50000 гаруй гүйлгээний мэдээллийг ашиглалаа. Энэхүү дэлгүүр нь хүнс болон барааны гэсэн 2 тасагтай, Хүнсний тасагт өдөр тутам хүнс, барааны тасагт гоо сайхны болон, ахуйн хэрэглээний өдөр тутмын бараа зарагддаг. Өдөрт 120 гаруй хүнд үйлчилдэг. Хэрэглэгчдийн хувьд тухайн дэлгүүрийн ойролцоо оршин суудаг бөгөөд ихэвчлэн өдөр тутмын хэрэглээний бараа, бүтээгдэхүүний худалдан авалт сонхилдог. Өгөгдөл олборлолтын арга техникүүдийг ашиглан цуглуулсан өгөгдлийн онцлогоос хамаарч таамаглах, ангилах, бүлэглэх, хамаарал илрүүлэх гэх мэтчилэн төрөл бүрийн алгоритмыг ашиглан шинжилгээ хийх боломжтой [7]. Дэлгүүрийн зай талбай багатай учраас олон нэр төрлийн бараа худалдан авсан тохиолдолд байрлуулах лангуу хомсдолтой ба ихэнх үйлчлүүлэгчдийн хувьд байнгын авдаг зүйлээ ахин авах магадлалтай тул тодорхой тооны, борлуулалт сайтай бүтээгдэхүүнүүдийг хэрэглэгчдэд хамгийн хялбар байдлаар санал болгох зорилготойгоор гүйлгээний түүхэн өгөгдөлд хамаарлын шинжилгээ буюу association analysis хийж гүйцэтгэе.

Ихэнх жижиг дэлгүүрүүдийн POS машины өгөгдөл нь json эсвэл csv өргөтгөлтэй файлаар тодорхой бүтэцтэй байна. Бидний сонгосон дэлгүүрийн барааны өгөгдлийн формат нь дараах байдалтай байлаа. Үүнд: items-нэгж борлуулалт, products-борлуулагдсан бараа, customer-худалдан авагчийн талаарх мэдээлэл, payments-төлбөрийн мэдээлэл. Зураг 1 дээр харууллаа.

```

{
  "items": [
    {
      "invoice_id": "102622090109211549408",
      "pay_status": "PAID",
      "payment_type_name": "CARD",
      "total_amount": "10500",
      "total_fee": "105",
      "created_date": "2022-09-01 09:21:11",
      "customer_type": "Хүнс хүй",
      "customer_reg_no": null,
      "customer_name": null,
      "products": [
        {
          "product_id": "99220418191334322066",
          "measure_type_id": "41",
          "vat": "1",
          "barcode": "99220418191322388283",
          "measure_type_name": "эмрхэр",
          "tax": "0",
          "quantity": "1",
          "amount": "10500.00",
          "item_org_id": "20000850",
          "is_tax": "N",
          "name": "laperla модон сан",
          "subtotal": "10500.00",
          "is_vat": "Y",
          "customer_name": null,
          "org_id": "20000850",
          "org_name": "сөргөйст монголиа",
          "user_id": "100000546",
          "user_name": "Yunaraa",
          "terminal_id": "30000018",
          "serial_no": "pb11214v22023",
          "paid_amount": "10500",
          "unpaid_amount": "0",
          "created_date": "2022-09-01 09:21:15",
          "payments": [
            {
              "payment_type": "CARD",
              "name": "Kapt",
              "amount": "10500"
            }
          ]
        }
      ]
    }
  ]
}

```

Зураг 1: JSON формат бүхий POS машины борлуулалтын өгөгдөл

Энэхүү гүйлгээний мэдээлэл бүхий файлыг дүн шинжилгээ хийхийн тулд боловсруулалт хийхдээ excel хүснэгт рүү оруулахдаа node.js дээрх дараах кодыг бичлээ.

```

1 var fs = require("fs");
2 var http = require("http");
3 var data = fs.readFileSync("data.json");
4 var writeStream = fs.createWriteStream("file.csv");
5 console.log("Program Start");
6 http.createServer(function (request, response) {
7   response.writeHead(200, { 'Content-Type': 'application/json; charset=utf-8' });
8   var obj = JSON.parse(data);
9   var header = "";
10  var row = "";
11  var itemsArray = Object.keys(obj.result.items);
12  for (var j = 0; j < 1; j++) {
13    var valueItem = itemsArray[j];
14    //console.log(obj.result.items[valueItem]);
15    var itemsItemArray = Object.keys(obj.result.items[valueItem]);
16    for (var i = 0; i < itemsItemArray.length; i++) {
17      console.log("Items items items " + itemsItemArray[i]);
18      obj.result.items[valueItem][itemsItemArray[i]];
19      header = header + "\t" + itemsItemArray[i];
20    }
21    console.log(j);
22    header = header + "\n";
23  }
24  writeStream.write(header, "utf8");
25  var itemsArray = Object.keys(obj.result.items);
26  for (var j = 0; j < itemsArray.length; j++) {
27    var valueItem = itemsArray[j];
28    //console.log(obj.result.items[valueItem]);
29    var itemsItemArray = Object.keys(obj.result.items[valueItem]);
30    row = "";
31    for (var i = 0; i < itemsItemArray.length; i++) {
32      console.log("Items items items " + itemsItemArray[i]);
33      obj.result.items[valueItem][itemsItemArray[i]];
34    }
35    row = row + "\n";
36    writeStream.write(row, "utf8");
37    console.log(j);
38  }
39  writeStream.close();
40  response.end(data);
41  //listen(8081);
42  console.log("Server running at http://127.0.0.1:8081/");
}

```

Зураг 2: JSON формат бүхий файлыг excel рүү экспортолох скрипт

3-р мөрөнд Data.json файлыг сервер рүү ачаалан JSON/parse функцээр задлан уншиж 8-р мөрөнд obj хувьсагчид онооно. 4-р мөрөнд export хийх file.csv файлыг үүсгэж байна. 11-р мөрөнд for давталтаар item бүрээр гүйн нийт борлуулалтыг гаргаж байна. Борлуулалт бүрийн хувьд 25-р мөрөнд itemsArray объектын массив руу зарагдсан бараа бүрийг авч 31-р мөрөнд for давталтаар бараа бүрийг уншиж авч мөр мөрөөр нь уншин авч 38-р мөрөнд Stream буюу урсгал руу бичихэд энэ нь csv файл руу бичигдэж байна. Энэхүү гүйлгээний мэдээлэлд хийсэн хамаарлын шинжилгээний үр дүнд тулгуурлан ямар бараанууд хамгийн их зарагддаг, ямар бараа, аль бараатай хамт зарагддаг гэдгийг таамаглан цаашид барааны татан авалтыг бүрдүүлэх, барааны лангуунууд дээр ямар бараануудыг зэрэгцүүлэн байрлуулахад ашиглах, мөн худалдан авагчдад шаардлагатай барааг санал болгох гэх мэт оновчтой шийдвэрүүдийг гаргах боломжтой юм.

Хүснэгт 1: Өдрийн дундаж өгөөжийн голлох статистик үзүүлэлтүүд.

invoice_id	product_id	vat	barcode	n	tax	quantity	amount	item_org_id	is	tax name	subtotal	is_vat
102622090109211	992204181913343	1	99220418191322300000	ц	0	1	10500	20000850	N	laperla модрн сам	10500 Y	
102622083121013	992111290955476	1	1000010	ц	0	1	10000	20000850	N	Балхиц Хөргөгчийн наалт	10000 Y	
102622083121013	992207072050084	1	99000000000000000000	ц	0	1	6000	20000850	N	Малла жижиг үсний боолт	6000 Y	
102622083121013	992204181656526	1	99220418165548000000	ц	0	1	6000	20000850	N	Малла автомат хавчаар том	6000 Y	
102622083120403	992205131855073	1	99220513185333600000	ц	0	2	10300	20000850	N	Агула эдийн саван кг-аар	20600 Y	
102622083120403	992205101508186	1	99220510150741300000	ц	0	1	9750	20000850	N	prisa шүдний оо	9750 Y	
102622083120403	992206101247174	1	6970872050103	ц	0	1	23000	20000850	N	tea Хушгатай улаан чагаа 268гр	23000 Y	
102622083120403	992206101245115	1	6970872054415	ц	0	1	23000	20000850	N	tea Хушгатай инжир 250гр	23000 Y	
102622083120403	992208181828307	1	99000000000000000000	ц	0	1	25000	20000850	N	Малла Хоолны ягаан цүхх	25000 Y	
102622083120140	992208291728072	1	4640007322611	ц	0	1	7000	20000850	N	Смолка Гацuur 10ш	7000 Y	

Хүснэгт 1-т Жижиг ахуйн барааны өдөр тутмын гүйлгээний мэдээллийн хүснэгтэн дээр гүйлгээний дугаар, барааны дугаар, баркод, нэр, хугацаа, барааны үнэ гэх мэтчилэн нийт 50 мянга гаруй гүйлгээний мэдээлэл байна.

## 4.2. Apriori алгоритм

Энэхүү шинжилгээг хийхдээ Python хэлний num.py, panda сангуудын априори алгоритмаар холбоо хамаарлын дүрмийг ашиглан жижиг дэлгүүрийн түүхэн өгөгдөлд тулгуурлан сагсны шинжилгээг хэрхэн хэрэгжүүлэх талаар туршилт хийлээ. Априори нь 1-рт өгөгдлийн багц дахь байнга тохиолддог бүх шинж чанаруудыг олоод 2-рт хамгийн чухал харилцааг тодорхойлохын тулд support - хамаарал хүчинтэй байхын тулд давтагдах шаардлагатай тоо, confidence - хамт байх магадлал гэсэн 2 хэмжигдэхүүнд суурилсан холбоодын дүрмийг ашиглан ажилладаг [8].

**Minimum support** нь тухайн хамаарал хүчинтэй байхын тулд давтагдах шаардлагатай тоог хэлнэ. Жишээлбэл, хэрвээ А барааг нэг л хэрэглэгч худалдан аваад, мөн Б барааг хамт авсан гэж үзье. Хэрвээ minimum support-н тоог тодорхойлоогүй бол бид “А барааг худалдан авсан хэрэглэгч бүр Б барааг авдаг.” гэсэн дүгнэлт хийхэд хүрнэ. Ийм учраас бид үнэн зөв байдлыг хангахын тулд “minimum support” - г тохируулж өгөх шаардлагатай. Энэ утга нь байгаа өгөгдлийн хэмжээ болон тархалтаас бүрэн хамаарна. Бидний шинжилгээний хувьд хамгийн багадаа 5 байхаар тогтооё. Энэ нь тухайн бараанууд хамтдаа хамгийн нь багадаа 5 удаа зэрэг зарагдсан байх шалгуурыг тогтоож өгнө.

**Minimum rule probability** нь өгөгдсөн support-н хүрээнд дүрэм зөв байх магадлал. Жишээлбэл, 100 хэрэглэгч А барааг худалдан авснаас, 10 нь Б барааг худалдан авсан бол захиалсан бол энэхүү 10% магадлалд үндэслэн хамаарлыг тооцох нь хэтэрхий бага байх талтай. Бидний хувьд хамгийн багадаа 40% байхаар авч үзье.

Apriori алгоритмын зорилго нь давтагдсан тоогоороо тогтоосон Minimum support-с байгаа зүйлүүдийн олонлогийг олоод, зүйлүүдийн олонлогт байгаа элементүүдээр боломжит бүх дүрмүүдийг гаргаж, түүний хамт байх магадлал буюу confidence нь бидний тогтоосон Minimum rule probability-с багагүй байвал эдгээр зүйлүүд нь хоорондоо хамгийн их хамааралтай гэсэн дүрмийг гаргадаг.

### Алгоритм Apriori

```

 $F_1 = \{\text{Давтамж } 1 - \text{бараа}\};$ 
for ( $k = 2; F_{k-1} \neq \emptyset; k++$ ) do begin
     $C_k = \text{apriori-gen}(F_{k-1});$ 
    foreach transaction  $t \in \mathcal{D}$  do begin
         $C_t = \text{subset}(C_k, t);$ 
        foreach candidate  $c \in C_t$  do
             $c.\text{count}++;$ 
    end
     $F_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\};$ 
end

Answer =  $\cup_k F_k;$ 

```

(4.1)

$F$ -Minimum rule probability буюу minsup-с их байх хамт байх магадлалын хувь,  $C_k$ -нэг худалдан авалтанд дах хамт зарагдах бараанууд,  $t$ -гүйлгээний дугаар,  $k$ -бараа,  $\mathcal{D}$ -нийт гүйлгээний олонлог.

## 5. Дүн шинжилгээ

### 5.1. Борлуулалтын орлогод хийсэн хамаарлын шинжилгээ

Цуглуулсан өгөгдөлд тулгуурласан хамаарлын шинжилгээг хийхдээ дараах алхмуудыг гүйцэтгэсэн. Үүнд:

1. Шаардлагатай Pandas болон num.py, matplotlib(өгөгдлийг графикаар харах) сангуудыг import хийж оруулж ирсэн.
2. Product.xlsx хүснэгтээс өгөгдлийг оруулсан/Зураг 3/.
3. Өгөгдлийг цэвэрлэж, сагс бүр дээрх барааны олонлогийг df.groupby функцээр байгуулсан.
4. Өгөгдлийн *apriori* алгоритмыг хэрэгжүүлэхэд боломжтойгоор бэлтгэсэн. Үүнд сагс бүрд байгаа барааны тоо ширхгийг *encode\_values* функц ашиглан байгаа барааны тоог 1, байхгүй барааны тоог 0-р сольсон. Нэг гүйлгээнд хамгийн багадаа 2 бараатай гүйлгээнүүдийг сонгосон/Зураг 4/.
5. Байнга хамт зарагддаг бараануудын багцыг үүсгэх. Багцыг үүсгэхдээ *apriori* алгоритмыг ашигласан ба *min\_sup*-г 3% байхаар зааж өгсөн. Энэ нь хамтдаа зарагддаг бараанууд дор хаяж 3%-с багагүй давтамжтай зарагддаг байх багцуудыг үүсгэнэ.
6. Цаашид хамт зарагдах бараануудын хамаарлыг таамаглах/Зураг 5/
7. Үр дүнг хэвлэх, зөвлөмж боловсруулах.

```
#Import all relevant libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
#Load the file into pandas
df = pd.read_excel("Product.xlsx")
```

Зураг 3: Шаардлагатай сангуудыг импортлон, өгөгдлийг оруулах

```
# Саяг бэлтгэх: Гүйлгээ тус бүр дээр ямар ямар бараа борлуулагдсныг олох өгөгдөл бэлтгэх
basket = df.groupby(['invoice_id', 'name'])["quantity"].count().unstack().fillna(0)
print(basket.head(100))
# Тухайн сагсанд байгаа өгөгдлүүдийг тоо хэмжээг 1 байхгүйг 0 болгох
def encode_values(x):
    if x <= 0:
        return 0
    if x >= 1:
        return 1
#Тухайн өгөгдлийн олонлогийн нэг өгөгдөлд бүрт encode_values функцийг хэрэгжүүлэх
basket_encoded = basket.applymap(encode_values)
```

Зураг 4: Гүйлгээний дугаарт харгалзах бараануудаар гүйлгээний олонлогийг үүсгэж байгаа хэсэг

```
#Байнга хамт зарагдсан бараануудын багцыг үүсгэх
frequent_itemsets = apriori(basket_filtered, min_support=0.03, use_colnames=True).sort_values("support",ascending=False)
# Чаашид хамт зарагдах бараануудын хамаарлыг тогтоох
assoc_rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1).sort_values("lift",ascending=False).reset_index(drop=True)
print(assoc_rules)
```

Зураг 5: Apriori болон Assosotive analysis хийх

support	itemsets		
19 0.185404	(бор цаасан уут )	12 0.069034	(ecolife хулсан сам том хүн)
6 0.143984	(ecolife пад (ширхэг))	16 0.067061	(laperla хатуу шампунь ангижруулагчтай)
11 0.112426	(ecolife хулсан нүүрний браш)	9 0.063116	(ecolife тоглоомтой саван)
10 0.108481	(ecolife хавтгай луффа спонж)	15 0.055227	(laperla хатуу шампунь)
7 0.106509	(ecolife сойз том хүн хулсан)	0 0.053254	(Ecolife гонзгой луффа спонж)
13 0.078895	(ecolife хулсан сам хүүхэд)	14 0.053254	(ecolife ширүүн угаагч)
5 0.078895	(ecolife ган соруул)	2 0.043393	(aguna гал тогооны саван)
18 0.076923	(Хулсан халбага савх ком)	20 0.039448	(aguna эдийн саван кг-аар, aguna угаалга нунтаг )
3 0.072978	(aguna угаалга нунтаг )	17 0.033531	(Биеийн массажны сойз)
1 0.072978	(Нова талх)	23 0.033531	(ecolife пад (ширхэг), ecolife хавтгай луффа с...
4 0.069034	(aguna эдийн саван кг-аар)	25 0.033531	(ecolife сойз хүүхэд хулсан, ecolife сойз том ...
8 0.069034	(ecolife сойз хүүхэд хулсан)	21 0.031558	(бор цаасан уут , aguna угаалга нунтаг )
		22 0.031558	(бор цаасан уут , aguna эдийн саван кг-аар)
		24 0.031558	(ecolife хулсан нүүрний браш, ecolife пад (шир...

Зураг 6: Apriori алгоритмын үр дүн

Энд эхний 26 ширхэг барааг харууллаа. Үр дүнгээс харахад “Бор цаасан уут” нь 0.185404-ийн support буюу давтагдан зарагдсан тоотой байгаа нь нийт худалдан авалтын багцад бор цаасан уутыг бусад бараануудтай 18%-д мөн ecolife пад-г бусад зарагдсан бараануудтай 14%, 2-р баганын дээр байгаа 12-р бараа буюу ecolife хулсан том хүнд зориулсан нь 0.069034% буюу 6%-тай байгаа нь бусад барааг зарагдахад дээрх хувиудын хэмжээтэйгээр хамт борлуулсан гэдгийг харуулж байна. Энэ шинжилгээнээс бид ямар бараа бусад бараатай хамт зарагдах магадлалтай олж харсан бол худалдан авагчийн нэг удаагийн худалдан авалтад аль барааг сонгосны дараа аль барааг сонгох магадлалыг мэдсэний үндсэн дээр барааны лангууны өрөлт болон санал болголт, багцын оновчтой худалдааг үр ашигтайгаар хийхэд хамаарлын судалгааг хийдэг.

Энэ үр дүн дээр тулгуурлан цаашид хамт зарагдах бараануудын хамаарлыг таамаглах-даа *association\_rules* дүрмийг ашиглан үр дүнг Зураг 7-д харууллаа.

	antecedents		consequents	
0	(aruna угаалга нунтаг )	(aruna эдийн саван кг-аар)	(aruna эдийн саван кг-аар)	(aruna угаалга нунтаг )
1	(aruna эдийн саван кг-аар)	(aruna угаалга нунтаг )	(aruna угаалга нунтаг )	(aruna эдийн саван кг-аар)
2	(ecolife сойз хүүхэд хулсан)	(ecolife сойз том хүн хулсан)	(ecolife сойз том хүн хулсан)	(ecolife сойз хүүхэд хулсан)
3	(ecolife сойз том хүн хулсан)	(ecolife сойз хүүхэд хулсан)	(ecolife сойз хүүхэд хулсан)	(ecolife сойз том хүн хулсан)
4	(aruna эдийн саван кг-аар)	(бор цаасан уут )	(бор цаасан уут )	(aruna эдийн саван кг-аар)
5	(бор цаасан уут )	(aruna эдийн саван кг-аар)	(aruna эдийн саван кг-аар)	(бор цаасан уут )
6	(aruna угаалга нунтаг )	(бор цаасан уут )	(бор цаасан уут )	(aruna угаалга нунтаг )
7	(бор цаасан уут )	(aruna угаалга нунтаг )	(aruna угаалга нунтаг )	(бор цаасан уут )
8	(ecolife пад (ширхэг))	(ecolife хавтгай луффа спонж)	(ecolife хавтгай луффа спонж)	(ecolife пад (ширхэг))
9	(ecolife хавтгай луффа спонж)	(ecolife пад (ширхэг))	(ecolife пад (ширхэг))	(ecolife хавтгай луффа спонж)
10	(ecolife пад (ширхэг))	(ecolife хулсан нүүрний браш)	(ecolife хулсан нүүрний браш)	(ecolife пад (ширхэг))
11	(ecolife хулсан нүүрний браш)	(ecolife пад (ширхэг))	(ecolife пад (ширхэг))	(ecolife хулсан нүүрний браш)

  

	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	0.072978	0.069034	0.039448	0.540541	7.830116	0.034410	2.026221
1	0.069034	0.072978	0.039448	0.571429	7.830116	0.034410	2.163051
2	0.069034	0.106509	0.033531	0.485714	4.560317	0.026178	1.737344
3	0.106509	0.069034	0.033531	0.314815	4.560317	0.026178	1.358708
4	0.069034	0.185404	0.031558	0.457143	2.465653	0.018759	1.500571
5	0.185404	0.069034	0.031558	0.170213	2.465653	0.018759	1.121934
6	0.072978	0.185404	0.031558	0.432432	2.332375	0.018028	1.435240
7	0.185404	0.072978	0.031558	0.170213	2.332375	0.018028	1.117180
8	0.143984	0.108481	0.033531	0.232877	2.146700	0.017911	1.162158
9	0.108481	0.143984	0.033531	0.309091	2.146700	0.017911	1.238970
10	0.143984	0.112426	0.031558	0.219178	1.949531	0.015371	1.136718
11	0.112426	0.143984	0.031558	0.280702	1.949531	0.015371	1.190071

Зураг 7: Хамаарлын дүрэм ашигласан үр дүн

*Association rule* функц нь холбоо хамаарлын таамаглалыг гаргахдаа *support*, *confidence*, *lift*, *leverage* гэсэн параметрүүдийг тооцон харуулдаг [9]. Үүнд: *Antecedents*-Үндсэн сонгогдсон бараа буюу  $X$ , *Consequents*-Үндсэн сонголтын дараа сонгох бараа буюу дагалдан сонголтын бараа  $Y$ ,  $Support-X \Rightarrow Y$   $X$  барааг сонгосон хүн  $Y$ -г барааг сонгох магадлал,  $Confidence-Y \Rightarrow X$   $Y$  барааг сонгосон хүн  $X$  барааг сонгох магадлал.

$$confidence(X \Rightarrow Y) = \frac{support(X \cup Y)}{support(X)} \quad (5.1)$$

$$lift(X \Rightarrow Y) = \frac{support(X \cup Y)}{support(X) * support(Y)} \quad (5.2)$$

*Lift* буюу нь үндсэн болон дагалдан сонголтууд хир олон удаа хамт сонгогдож байгаад  $X$ ,  $Y$ -г сонгох магадлал хир нөлөөлж байгаа харьцааг. Хэрэв  $X$  ба  $Y$ -н сонголт нь бие биеэсээ бүрэн хамааралгүй бол *lift* 1 байна.

$$leverage(X \Rightarrow Y) = support(X \cup Y) - support(X) * support(Y) \quad (5.3)$$

*Leverage* (*Lift*-тэй бараг төстэй) нь үндсэн болон дагалдан сонголтууд хир олон удаа хамт сонгогдож байгаад  $X$ ,  $Y$ -г сонгох магадлал хир нөлөөлөхийг харуулахдаа харьцаагаар биш ялгааг ашигладаг [10].

Хамаарлын дүрмийг хэрэгжүүлсний дараах үр дүнгийн тайлбарыг 0 дугаарт байгаа *aruna* угаалгын нунтаг манай  $X$  бараа, *aruna* эдийн саван манай  $Y$  барааны хувьд дараах байдлаар хийнэ. Үүнд:

*Antecedent support*-Нийт гүйлгээний 7%-тай  $X$  бараа хамт зарагдана.

*Consequent support*-Нийт гүйлгээний 6%-тай нь бараа хамт зарагдана.

*Support*-Нийт гүйлгээний 0.039448 хувиар илэрхийлбэл нийт гүйлгээний 3.9%-д нь энэ 2 бараа зэрэг зарагдана.



*Confidence-Y*-г сонгосон хүн *X*-г сонгох магадлал 5% байна.

*Lift-X* болон *Y* бараануудыг хамт авахад *X*-г сонгох магадлал \**Y*-г сонгох магадлалын харьцаа нь 70% байна.

*Leverage-X* болон *Y* бараануудыг хамт авахад *X*-г сонгох магадлал \**Y*-г сонгох магадлалын ялгавар нь 3% байна.

## 6. Дүгнэлт

Энэхүү өгүүлэлд өгөгдөл олборлолтыг жижиглэн худалдааны борлуулалтын орлогыг сайжруулахын тулд түүхэн өгөгдөлд хамаарлын шинжилгээ хийх боломжтой аргийг алгоритмыг хэрэгжүүлсэн туршлагын талаар орууллаа. Одоо үед янз бүрийн салбаруудад бизнесийн оновчтой шийдвэр гаргахын тулд өгөгдөл олборлолтын аргуудыг өргөнөөр ашиглаж байна. Өгөгдөл олборлолтын олон алгоритмууд байдаг ба жижиглэн худалдааны дэлгүүрт борлуулалтын орлогоос хамаарч аль бараа ер нь бүх зарагдаж байгаа бараануудтай хамт зарагддаг, аль бараа аль бараатай зохицон зарагддаг гэдгийг ажиглах нь борлуулалтын орлогод их өөрчлөлт үзүүлэх боломжтой учраас бид аргийг алгоритмыг ашиглан хамаарлын шинжилгээг хийлээ. Энэхүү үр дүнг ашиглан ямар маркетингийн стратеги хэрэгжүүлэх нь тухайн дэлгүүрийн эзний шийдвэрээс хамаарах бөгөөд жижиглэн худалдааны чиглэлээр төрөл бүрийн борлуулалтын стратегиудыг хэрэгжүүлэх үнэн зөв бодит шинжилгээний өгөгдлөөр хангагдах боломжтой. Жишээ нь: анхны сонголтын бүтээгдэхүүний дараа, сонголт өндөртэй 2 дахь барааг сонгох зөвлөмжийг санал болгох, хамт зарагддаг бараануудыг лангууны өрөлтийн хувьд ойрхон байрлуулах, багцын үнэ тогтоох, ямар бараа хир олон зарагддагаас хамаарч барааны татан нийлүүлэлтийн зөв бодлого барих гэх мэтчилэн маш олон стратегиудыг хэрэгжүүлэх боломжтой.

## Ном зүй

- [1] B. M. Ramageri, and B. L. Desai, "Role of Data Mining in Retail Sector," *International Journal on Computer Science and Engineering*, Vol. 5, pp. 47-50, 2013.
- [2] N. R. Mabroukeh, and C. I. Ezeife, "A taxonomy of sequential pattern mining algorithms," *ACM Computing Surveys*, Vol. 43, pp. 1-41, 2010, doi: <https://doi.org/10.1145/1824795.1824798>.
- [3] W. Banzhaf, P. Nordin, R.E. Keller, F.D. Francone, "Genetic Programming An Introduction," *Morgan Kaufmann Publishers*, 1998.
- [4] E. Forgey, "Cluster analysis of multivariate data: Efficiency vs. interpretability of classification," *Biometrics*, 21, pp. 768, 1965, doi: <https://doi.org/10.2307/2528096>.
- [5] Dhanabhakyaam, and M. Punithavalli, "A Survey on Data mining Algorithm for Market Basket Analysis," *Global Journal of Computer Science and Technology*, Vol. 11, Issue-1, ver. 1.0, ISSN: 0975-4350, July., 2011.
- [6] M. Gera, and S. Goel, "Data Mining-Techniques," *Methods and Algorithms: A Review on Tools and their Validity*, *International Journal of Computer Applications*, Vol. 113, Issue-18, pp. 22-29, 2015, doi: <https://doi.org/10.5120/19926-2042>.
- [7] A. Rakesh, and S. Ramakrishnan, "Fast Algorithms for Mining Association Rules," 1994.
- [8] Н. Мөнхцэцэг, Н. Батгүшиг, "Ном захиалгын мэдээлэлд хийсэн хамаарлын шинжилгээ," *ШУА-ийн Физик, технологийн хүрээлэн*, №. 43, х. 40-46, 2016.
- [9] M. Z. Amin, and A. Ali, "Machine Learning & Deep Learning Researcher Co-Founder of Wavy AI Research Foundation An Intuitive Guide of Apriori Algorithm with Practical Implementation," 2019.
- [10] J. Ranjan, and Ghaziabad, "A Review of Data Mining Tools in Customer Relationship Management," *Journal of Knowledge Management Practice*, Vol. 9, no. 1, March., 2008.

# Increasing Retail Sales Revenue Using Data Mining Techniques

Munkhtsetseg Namsraidorj, Byambasuren Ivanov\*, Batbayar Sereeter, Enkhtuul Bukhsuren

*School of Engineering and Applied Sciences, National University of Mongolia, Ulaanbaatar 14200, Mongolia*

*\*Corresponding author: byambasuren@seas.num.edu.mn; ORCID:0000-0001-5236-7857*

*Article Info: Received: 2022.09.17; Accepted: 2022.10.01; Published: 2022.12.26*

---

**Abstract:** In this study, we analyze a customer's single purchase data to increase revenue by using data mining apriori and association rules to make product and withdrawal predictions based on historical sales data for a retail store and understand why things happened in the past through data mining. (descriptive analytics) tells about algorithms and how to use them optimally. Nowadays, every enterprise and business organization and individuals use data mining techniques from the historical information of operations to take the required information as knowledge and use it wisely in their future work and life activities. Data mining has the advantage of uncovering hidden patterns and links that can provide economic benefits that other tools cannot detect.

**Key words:** Machine Learning Algorithms, Basket Analysis, User Behavior, Apriori, Associative Analysis

---