

MACHINE LEARNING APPLICATIONS IN ACTUARIAL RISK ASSESSMENT AND PRICING

Bazarragchaa Barsbold^I, Dorjkhoo Enkhtaivan^{II}

Abstract: This study explores the potential of applying machine learning models in actuarial premium calculations by classifying policyholders based on their risk levels. Among the various insurance products available in the market, health insurance known for its high loss ratio was selected as the focus of this research. Based on population morbidity data, the most significant variables were identified and used to develop and evaluate machine learning models that classify risk levels according to each category of the International Classification of Diseases (ICD). Furthermore, for each ICD category, the study examines the distribution patterns and descriptive statistics of the corresponding treatment cost data. Based on this analysis and the machine learning outputs, risk assessment recommendations are proposed using the z-score methodology.

Keywords: general insurance, machine learning, risk assessment, actuarial analysis

ДААТГАЛЫН АКТУАР ТООЦООЛОД МАШИН СУРГАЛТ АШИГЛАХ БОЛОМЖ

Хураангуй: Энэхүү судалгааны ажлаар машин сургалтын загваруудыг ашиглан даатгуулагчдыг эрсдэлийн хэмжээгээр ангилж, хураамжийн актуар тооцоололд хэрэглэх боломжийн талаар авч үзнэ. Даатгалын салбарт борлуулагдаж байгаа олон бүтээгдэхүүнээс хохирлын харьцаа өндөр эрүүл мэндийн даатгалыг сонгосон. Хүн амын өвчлөлийн өгөгдөлд үндэслэн хамгийн ач холбогдол өндөртэй хувьсагчуудыг сонгон авч, олон улсын өвчний ерөнхий ангилал тус бүрээр эрсдэлийг ангилах машин сургалтын загварыг байгуулж үнэлэлт дүгнэлт өгнө. Мөн өвчний олон улсын ерөнхий ангилал тус бүрт харгалзах эмчилгээний зардлын дүнгийн тархалтын хэлбэр, тодорхойлогч статистик ба машин сургалтын загварт үндэслэн z-score ашиглах замаар эрсдэлийн үнэлгээ хийх санал зөвлөмжийг тусгалаа.

Түлхүүр үгс: ердийн даатгал, машин сургалт, эрсдэлийн үнэлгээ, актуар тооцоолол

^I National University of Mongolia, Department of Applied Mathematics (E-mail): Barsbold@num.edu.mn

^{II} Actuary Analytics LLC (E-mail): Do.enkhtaivan@gmail.com

Оршил

Английн нэрт математикч Алан Матисон Тюринг 1950 онд анх машин хангалттай оюунлаг эсэхийг тодорхойлох Тюринг тест судалгааны ажлаараа хиймэл оюун ухааныг (artificial intelligence) үндэслэсэн бол машин сургалт (machine learning) гэх нэр томъёог 1959 онд Америкийн компьютер тоглоом, хиймэл оюун ухааныг үндэслэгчдийн нэг Артур Самуэль анх томъёолсон. Хиймэл оюун ухаан түүхэн хөгжлийн явцад 1973 онд анхны өвөлтэй^{III} нүүр тулж судалгааны ажлууд, хөрөнгө оруулалтууд зогсонги байдалд орсон бол 1988 оны үеэс 2 дахь том давалгааны өвөл болжээ. 2000-аад оны үеэс компьютерын хүчин чадал эрс сайжирч ирсэнтэй холбоотойгоор хиймэл оюун ухаан, машин сургалт, гүн сургалт (deep learning), нейроны сүлжээ (neural network) маш эрчимтэй, өндөр төвшинд хөгжсөөр байна. Одоогоор гадны улс орнууд даатгалын салбартаа дараах 6 ерөнхий чиглэлд хиймэл оюун ухааныг ашиглаж байна. Үүнд:

1. Эрсдэлийн үнэлгээ ба хураамж тооцоолох;
2. Нөхөн төлбөрийн процесс хялбаршуулах;
3. Хууран мэхлэлт илрүүлэх;
4. Хэрэглэгчийн үйлчилгээнд виртуал туслахууд ашиглах;
5. Хэрэглэгчийг сегментчилэх;
6. Их өгөгдөлд анализ хийх ба урьдчилан таамаглах.

Харин манай орны хувьд дээрх зургаан чиглэлээс 2 ба 4 дэх асуудлын хүрээнд зөвхөн гадны хөгжүүлсэн бэлэн загварыг авч ашиглаж байна. Энэ нь өртгийн хувьд үнэтэй бөгөөд гадаадын даатгалын салбараас ялгагдах өгөгдөл, бизнесийн онцлог, хууль эрх зүйн орчин зэргийг тусгаж олон хувилбараар тасралтгүй хөгжүүлэх боломж хязгаарлагдмал юм.

Мөн Монгол улсын арилжааны даатгалын салбар сүүлийн жилүүдэд эрчимтэй хөгжиж оролцогч талууд болон даатгалын компаниуд нэмэгдэж, Санхүүгийн зохицуулах хорооноос холбогдох багц дүрэм, журмыг шинэчлэн баталж байна. Даатгагчийн төлбөрийн чадварын шалгуур үзүүлэлтийг эрсдэлд суурилсан хөрөнгийн аргачлалаар тооцож, даатгалын компани эрсдэлээ илүү нарийвчлалтай ахисан төвшинд үнэлэх, андеррайтингийн эрсдэл буюу даатгалын гэрээний төвшинд үнэлгээ хийх шаардлагатай болсон.

Судалгааны арга зүй

Энэ хэсэгт бид даатгуулагчийн эрүүл мэндийн эрсдэлийг үнэлэхэд ашигласан машин сургалтын 9 загварууд, тэдгээрийн онцлог, давуу ба сул тал зэргийг товч авч үзнэ.

А. Машин сургалт

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 (Зураг 1) нэвтрүүлнэ. Өөрөөр хэлбэл, шинэ өгөгдлийн хувьд тодорхойлсон загвараа ажиллуулан танин мэдэх даалгаврыг гүйцэтгүүлнэ.

^{III} First AI winter

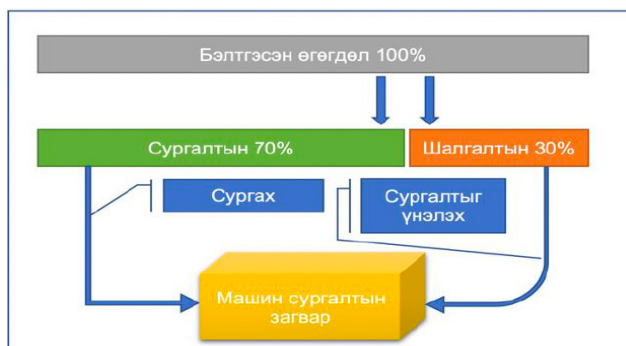
Зураг 1 Машин сургалт ерөнхий бүдүүвч



Машин сургалтын загварыг өгөгдсөн даалгавраас хамааран дараах үндсэн 3 байдлаар ангилдаг. Үүнд:

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

Зураг 2 Машин сургалт хийх процесс



Энэ судалгааны ажил нь даатгуулагчдын эрүүл мэндийн эрсдэлийг ангилах даалгавар учраас удирдагчтай сургалтын загвар хэрэглэнэ. Удирдагчтай сургалтын нэг төрлийн даалгавар нь ангилах^{IV} бодлого юм. Ангилах бодлогыг олон төрлийн машин сургалтын загвараар хийж болох бөгөөд судалгаанд дараах загваруудыг авч үзнэ. Үүнд:

- Дэмжих вектор машин (Support Vector Machine – SVM)
- k-хамгийн ойр хөрш (k-Nearest Neighbour - KNN)
- Шийдвэрийн мод (Decision Tree – DT)
- Санамсаргүй ой (Random Forest - RF)
- Ада идэвхижүүлэгч (Ada Boost - ADA)
- Градиент идэвхижүүлэгч (Gradient boost - GB)
- Стохастик градиент идэвхижүүлэгч (Stochastic Gradient Boosting - SGB)
- Нэмэлт мод (Extra Trees - ET)
- Саналд үндэслэсэн ангилагч (Voting Classifier - VC)

^{IV} Classification problem

Машин сургалт хийхэд өгөгдсөн өгөгдлийг сургалтын болон шалгалтын гэсэн хоёр хэсэгт хуваан (Зураг2) дээрх загваруудыг оновчлолын^V арга алгоритмуудыг ашиглан хамгийн бага алдаатайгаар сургана.

Судалгаанд ашигласан машин сургалтын загварууд

В. Дэмжих вектор машин. (Support Vector Machine)

Гол зорилго нь ангиудыг хамгийн сайн ялгаж чадах гипер хавтгайг олох бөгөөд ангиудын хоорондох зайг хамгийн өргөн байлгахыг зорьдог. Давуу тал:

- Өндөр нарийвчлалтай.
- Олон хэмжээст өгөгдөлтэй ажиллах чадвар өндөр.
- Олон төрлийн кернел ашиглах боломжтой.
- Зөвхөн support vector өгөгдөлд анхаарал хандуулдаг тул санах ойн хувьд үр ашигтай.
- Хэт үнэлэлтээс^{VI} сэргийлдэг.

Сул тал:

- Өгөгдөл ихтэй үед удаан.
- Бүх төрлийн кернел тохиромжтой байдаггүй.
- Шугаман бус өгөгдөлд тохируулж ажиллахад төвөгтэй.
- Кернел параметруудийг оновчтой сонгоход нэмэлт цаг, туршилт шаардана.

С. *k*-хамгийн ойр хөрш. (*k*-Nearest Neighbour)

Шинэ өгөгдлийн цэгийн анги (эсвэл утга)-ийг тодорхойлохдоо тухайн цэгээс хамгийн ойр байгаа *K* хөрш өгөгдлийг ашиглах бөгөөд өгөгдөл хоорондын зайг тооцоолохдоо Евклидийн зай (Euclidean Distance), Манхэттен зай (Manhattan Distance) зэрэг зайн хэмжүүрийг ашиглана. Давуу тал:

- Онолын хувьд ойлгоход хялбар, кодчилж ашиглахад энгийн.
- Өгөгдөл сайтар хуваагдсан үед оновчтой үр дүн гарна.
- Сургалтын үе шат байхгүй, өгөгдөл нэмэхэд шууд ажиллах боломжтой.
- Тоон өгөгдөл дээр үр дүнтэй.

Сул тал:

- Шинэ өгөгдлийг ангилах үед бүх өгөгдөлтэй харьцуулж зай тооцоолдог тул том өгөгдлийн хувьд удаан.
- Бүх өгөгдлийг санах ойд хадгалдаг тул санах ойн хэрэглээ ихэсдэг.
- Ойрын хөршид шуугиант^{VII} өгөгдөл орвол гүйцэтгэлд сөрөг нөлөө үзүүлнэ.
- Өгөгдлийг жигдлэхгүй^{VIII} бол алдаатай үр дүн гарах магадлалтай.

Д. Шийдвэрийн мод. (Decision Tree)

Энэ алгоритм нь өгөгдлийг бүлэглэх эсвэл урьдчилан таамаглах үйл явцыг мод бүтэцтэйгээр дүрсэлдэг. Модны зангилаануудад өгөгдлийн шинж чанаруудыг ашиглан өгөгдлийг хуваана. Хуваалтыг gini impurity эсвэл entropy хэмжүүрүүдийн үндсэн дээр гүйцэтгэдэг. Давуу тал:

- Онол болон кодчилал нь энгийн, ойлгомжтой.
- Үр дүнг модны бүтэц дээр суурилан амархан тайлбарлаж болдог.
- Тоон болон чанарын өгөгдөл дээр үр дүнтэй ажиллана.

^V Optimization

^{VI} Overfitting

^{VII} Noise

^{VIII} Scale ашиглах

- Чухал шинж чанаруудыг автоматаар тодорхойлох боломжтой.
- Сургалтын болон таамаглалын үед хурдан ажилладаг.

Сул тал:

- Хэт үнэлэлтэнд өртөмтгий.
- Шуугианд мэдрэмтгий.
- Бага өгөгдөлтэй үед тохиромжгүй.
- Хэмжээ томсох тусам алдаа нэмэгддэг.
- Олон ангилалтай өгөгдөлд асуудалтай.

Е. Санамсаргүй ой. (*Random Forest*)

Удирдагчтай машин сургалтын арга бөгөөд олон шийдвэрийн мод алгоритмоос бүрдэх, түүнийг сайжруулсан ансамбль^{IX} загвар юм. Давуу тал:

- Хэт үнэлэлтийг багасгадаг.
- Шуугианы эсрэг тэсвэртэй.
- Өгөгдлийн олон төрлийн хэлбэрийг дэмждэг.
- Шинж чанаруудын ач холбогдлыг үнэлж, оновчтой шинж чанар сонгоход тусалдаг.
- Шийдвэрийн мод алгоритмтай харьцуулахад илүү нарийвчлалтай үр дүн гаргадаг.
- Гиперпараметруудыг^X тохируулан оновчлол хийх боломжтой.
- Их хэмжээний өгөгдөл дээр сайн ажилладаг.

Сул тал:

- Шийдвэрийн мод олон байх тусам сургалтын болон таамаглал хийх хугацаа уртасдаг.
- Өндөр хамааралтай хувьсагчуудыг зөв сонгохгүй бол гүйцэтгэл буурах магадлалтай.
- Өндөр хамааралтай шинж чанаруудыг зөв сонгохгүй бол гүйцэтгэл буурах магадлалтай.

Ғ. Ада идэвхижүүлэгч. (*Ada Boost*)

Машин сургалтын ансамбль сургалтын арга бөгөөд сул загваруудыг (weak learners) ашиглан хүчирхэг загвар үүсгэдэг. Ихэвчлэн шийдвэрийн мод зэрэг энгийн алгоритмыг ашиглаж, давтсан сургалтаар алдааг сайжруулдаг. Давуу тал:

- Олон энгийн загварыг нэгтгэснээр гүйцэтгэлийг сайжруулдаг.
- Алгоритмын үр дүнг харьцангуй хялбар ойлгож, хувьсагчийн ач холбогдлыг үнэлж болно.
- Зарим шуугиант өгөгдөлд зохицон ажилладаг.
- Чухал хувьсагчуудыг автоматаар тодорхойлох чадвартай.

Сул тал:

- Хэт олон шуугиант өгөгдөлтэй үед сайн гүйцэтгэл гаргахгүй.
- Давталтын тоо ихсэх тусам сургалтын хугацаа уртасна.
- Бага өгөгдлийн хувьд хэт үнэлсэн загвар гаргаж магадгүй.
- Гадна утгууд гүйцэтгэлд сөргөөр нөлөөлж болзошгүй.

Г. Градиент идэвхижүүлэгч. (*Gradient boost*)

Ансамбль сургалтын аргын нэг бөгөөд сул загваруудыг дараалсан байдлаар

^{IX} Ensemble

^X Hyperparameter

сайжруулж, хүчирхэг загвар үүсгэдэг. Давуу тал:

- Олон төрлийн өгөгдөл дээр маш сайн гүйцэтгэлтэй.
- Хоёр болон олон ангилалд үр дүнтэй.
- Төрөл бүрийн алдааны функц дээр суурилан ажиллах боломжтой.
- Чухал хувьсагчуудыг автоматаар тодорхойлох чадвартай.
- Learning rate, max depth, number of iterations зэрэг параметруудийг тохируулснаар хэт үнэлэлтийг хянах боломжтой.

Сул тал:

- Давталтын тоо их, градиент тооцоолол хийдэг тул сургалт удаан явагддаг.
- Шуугианд мэдрэмтгий.
- Гиперпараметруудыг тохируулахад төвөгтэй.
- Том өгөгдөл дээр санах ой болон процессын нөөц ихээр зарцуулдаг.

Н. *Стохастик градиент идэвхижүүлэгч. (Stochastic Gradient Boosting)*

Энэхүү загвар нь градиент идэвхижүүлэгчийн нэг хувилбар бөгөөд алгоритмын гүйцэтгэл, хурдыг сайжруулах зорилгоор өгөгдөлд санамсаргүй түүвэр ашигладаг.

Давуу тал:

- Хэт үнэлэлтийг бууруулдаг.
- Бүх өгөгдөл дээр биш, түүвэр дээр сургалт явуулдаг тул сургалтын хугацааг багасгана.
- Хувьсагчуудын ач холбогдлыг үнэлэх боломжтой.
- Ангилалч, регресс зэрэг олон төрлийн асуудлыг шийдэхэд тохиромжтой.

Сул тал:

- Гиперпараметруудыг зөв тохируулахгүй бол гүйцэтгэл буурч магадгүй.
- Шуугианд мэдрэмтгий.
- Их хэмжээний өгөгдөлд санах ой шаарддаг.
- Олон загвараас бүрдэх тул үр дүнг тайлбарлахад хэцүү.

И. *Нэмэлт мод. (Extra Trees)*

Тус алгоритм нь шийдвэрийн мод загварт үндэслэсэн, санамсаргүй ой загвартай төстэй машин сургалтын ансамбль сургалтын арга юм. Моднуудыг илүү санамсаргүй байдлаар сургаж, дисперсийг бууруулахыг зорьдог. Давуу тал:

- Хуваалтын цэгийг оновчлох шаардлагагүй тул шийдвэрийн мод болон санамсаргүй ойгоос хурдан.
- Санамсаргүй байдал ихтэй тул дисперсийг бууруулж, загварын тогтвортой байдлыг сайжруулдаг.
- Олон төрлийн өгөгдөл дээр сайн гүйцэтгэлтэй.
- Санамсаргүй сонголт нь хэт үнэлэлтийг бууруулахад тустай.
- Олон мод зэрэгцээ ажиллах боломжтой тул тооцооллын хурдыг нэмэгдүүлнэ.
- Шийдвэрийн модтой адил хувьсагчуудын ач холбогдлыг үнэлэх боломжтой.

Сул тал:

- Шуугианд мэдрэмтгий.
- Их хэмжээний өгөгдөл дээр тооцоолол удаан байж болно.
- Гиперпараметруудыг тохируулах шаардлагатай:

Ж. *Саналд үндэслэсэн ангилагч. (Voting Classifier)*

Ансамбль сургалтын нэг арга бөгөөд хэд хэдэн үндсэн загваруудын гаралтын үндсэн дээр эцсийн шийдвэр гаргадаг. Үндсэн загваруудын нийлбэрээр таамаглалыг сайжруулж, гүйцэтгэлийг нэмэгдүүлэхэд чиглэнэ. Давуу тал:

- Олон загварыг хослуулж ашигласнаар ганц загварын гүйцэтгэлээс илүү тогтвортой, сайн үр дүн гаргана.
- Өөр өөр зарчимтай загваруудыг нэгтгэн сул талыг багасгаж, өгөгдлийн онцлогт илүү сайн нийцүүлж чадна.
- Нэг загварын өгөгдөл хэт мэдрэг байдал нь төдийлөн нөлөөлөхгүй.
- Нэг загвараас хамаарал багатай тул хэт үнэлэлт хийх магадлал багасна.
- Ангилалч болон регрессийн асуудалд хоёуланд нь ашиглах боломжтой.

Сул тал:

- Ганцхан загвар дээр үр дүнгүй.
- Олон үндсэн загвар ашиглах нь сургалт болон тооцооллын хугацааг нэмэгдүүлдэг.
- Загваруудын жин, параметруудийг тохируулах шаардлагатай үед төвөгтэй болж магадгүй.

Судалгааны хэсэг

Судалгааны зорилго: Энэ судалгааны ажлаар даатгуулагчийн ерөнхий 7 мэдээлэл ба тухайн хүнд илэрч болох шинж тэмдгийн 34 асуултыг үндэслэн өвчний олон улсын ангиллаас^{XI} даатгалын гэрээнд түгээмэл тусгагддаг 13 ерөнхий ангилал тус бүрээр өвдөх эрсдэлийг тооцоолох машин сургалтын загварууд байгуулсан. Өвчний олон улсын ангилал тус бүрийн машин сургалтын загвараас гарсан үр дүнд эрсдэлийн үнэлгээний оноог харгалзуулж, түүний нийлбэр оноогоор тухайн даатгуулагчийн эрсдэлийн төвшин тодорхойлогдоно.

Нийт эрсдэлийн оноог сонгон авсан интервалуудад хувааж, тус бүрт харгалзах даатгалын хураамжийн актуар тооцооллыг хийж гүйцэтгэснээр практикт шууд ашиглах боломжтой гэж бид үзэж байна. Өөрөөр хэлбэл, тухайн даатгуулагчаас асуумж авсны дараа уг хүн өвчний олон улсын ерөнхий ангиллын аль бүлгээр өвдөх өндөр эрсдэлтэйг, түүний нийт эрсдэлийн төвшин ямар хэмжээнд байгааг, тухайн даатгуулагчаас ямар хэмжээний хураамж авах хамгийн тохиромжтой бөгөөд зорилтот хохирлын харьцаанд байхыг шууд тооцоолох систем юм.

К. *Судалгаанд ашигласан өгөгдөл*

Энэхүү судалгааны ажилд Монгол улсын бүх хүн амын өвчлөлийн 2020-2022 он хамрах 3 жилийн бодит өгөгдлийг ашигласан. Тухайн мэдээллийн эх сурвалж нь “Эрүүл мэндийн хөгжлийн төв” бөгөөд зөвхөн судалгааны ажлын зорилгоор ашигласан тул хувь хүний нууцын тухай хуультай холбоотойгоор овог, нэр, регистрын дугаар зэрэг мэдээллийг агуулаагүй. Уг өгөгдөл нь нас, хүйс, боловсрол, албан тушаал, мэргэжил, үндсэн онош, эмнэлгийн төрөл зэрэг 22 талбарыг агуулсан 1.2 сая мөр мэдээлэл юм.

Л. *Өгөгдлийг машин сургалтанд бэлтгэх*

Өгөгдлийг машин сургалтанд дараах байдлаар бэлтгэсэн. Үүнд:

- Өгөгдлийн бүх талбарын хүрээнд орхигдсон, хоосон утга нийт өгөгдлийн 1%-иас хэтрэхгүй байсан тул орхигдсон утгыг хассан.
 - Тоон хувьсагч бүрийн тодорхойлогч статистикийг тооцоолж, 3 сигма гадна утга буюу эрс ялгаатай өгөгдлийг хассан.
 - Ердийн даатгалын эрүүл мэндийн даатгалын бүтээгдэхүүний хувьд эрсдэлийг үнэлэх систем тул даатгалын гэрээнд тусгагддаггүй олон улсын өвчний ангиллыг

^{XI} International Classification of Diseases (ICD)

хассан. Жишээлбэл, (F00-F99) - Сэтгэцийн эмгэгүүд, (S00-T98) - Гэмтэл хордлого болон гадны шалтгаант өвчлөл зэрэг. Үүний дараа олон улсын өвчний 13 ерөнхий ангилал үлдсэн бөгөөд энэхүү ангилал нь бидний машин сургалтын загварын хамаарах хувьсагч^{XII} юм.

- Даатгалын гэрээнд тусгагддаггүй насны хязгаар буюу 18-аас доош ба 65-аас дээш насны бүлгийг хассан.
- Feature extraction: Өгөгдлийг задлан шинжилж олон улсын өвчний ерөнхий ангилал тус бүрийн шинж тэмдгийг тодруулах анхан шатны 34 асуулгыг нэмсэн.
- Оршин суугаа хаяг, ажлын газар, албан тушаал, мэргэжил гэсэн 4 хувьсагчийг^{XIII} машин сургалтын үр дүнг сайжруулах зорилгоор дамми хувьсагч болгосон. Олон зэрэг хамаарлын алдаа^{XIV} гарахаас сэргийлж оршин суугаа хаягийн хувьд “Орхон”, ажлын газрын хувьд “бусад”, албан тушаалын хувьд “бусад”, мэргэжлийн хувьд “туслах ажилтан” гэсэн талбаруудыг тус тус хассан.
- “Хүйс” чанарын хувьсагчийн хувьд эрэгтэй бол 1, эмэгтэй бол 0 байхаар тоон хувьсагч руу хөрвүүлсэн.
- “Боловсрол” чанарын хувьсагчийн хувьд боловсролгүй бол 0, бага боловсролтой бол 1, бүрэн дунд боловсролтой бол 2, мэргэжлийн болон техникийн боловсролтой бол 3, дээд боловсролтой бол 4 байхаар тоон хувьсагч руу хөрвүүлсэн.
- Хамаарах хувьсагч буюу “үндсэн онош”-ийн хувьд А өвчний ангиллаар өвдсөн бол 1, үгүй бол 0, В өвчний ангиллаар өвдсөн бол 1, үгүй бол 0 гэх мэтчилэн тоон хамаарах хувьсагч руу хөрвүүлсэн.
- Олон улсын өвчний ерөнхий ангиллын шинж тэмдгийн 34 асуулга тус бүр нь тухайн хүнд уг шинж тэмдэг илэрсэн бол тийм буюу 1, илрээгүй бол үгүй буюу 0 утга авахаар тоон хувьсагч руу хөрвүүлсэн. Ангилал тус бүр дээр харгалзан 2-3 асуулга байсан бөгөөд эдгээр нь тухайн ангилал тус бүр дээр шууд шугаман хамааралтай байсан тул харгалзах асуултуудыг нэмж шинэ хувьсагч үүсгэж өөрчилсөн. Дээрх байдлаар машин сургалтанд өгөгдлийг бэлтгэсний дараа хамаарах хувьсагч нь олон улсын өвчний ерөнхий ангилал буюу 1, үл хамаарах хувьсагч нь 74 болсон. Тайлбарлагч хувьсагчдын корреляцийг Хавсралт 1-ээс үзнэ үү.
- Олон улсын өвчний ерөнхий ангилал тус бүрийн авах утга буюу хамаарах хувьсагчийн авах утга нь тэнцвэргүй өгөгдөл^{XV} байсан тул харгалзах санг ашиглан тохиргоо хийв.
- Машин сургалтын үр дүнг сайжруулах зорилгоор бүх үл хамаарах хувьсагчдыг хамгийн сүүлд харгалзах санг ашиглан стандартчилав.

М. *Машин сургалт*

Машин сургалтыг нийтлэг хэрэглэгддэг ихэнх програмчлалын хэлний орчинд хийж болно. Сүүлийн жилүүдэд Python хэлийг машин сургалтад хэрэглэх хандлага давамгайлахын хирээр уг хэл дээр бичигдсэн машин сургалтын сангийн хөгжүүлэлт эрчимтэй явагдаж байгаа зэрэг олон давуу талтай. Иймээс Python 3.10 дээрх pandas, numpy, sklearn, matplotlib, joblib зэрэг санг ашиглан Macbook Pro (3.1GHz Intel core i5, 8GB RAM, Intel Iris Plus Graphics 650) компьютер дээр сургалтыг хийв.

Энэхүү судалгааны ажлын хүрээнд машин сургалтын дэмжих вектор машин, k-хамгийн ойр хөрш, шийдвэрийн мод, санамсаргүй ой, ада идэвхижүүлэгч, градиент

^{XII} Label

^{XIII} Feature

^{XIV} Multicollinearity

^{XV} Imbalanced data

идэвхижүүлэгч, стохастик градиент идэвхижүүлэгч, нэмэлт мод, саналд үндэслэсэн ангилагч гэсэн 9 төрлийн загварыг олон улсын өвчний 13 ерөнхий ангилал тус бүр дээр сургаж, холбогдох үнэлгээний аргачлалаар шинжлэв.

Машин сургахад бэлтгэсэн өгөгдөл хангалттай их хэмжээнд байсан учраас 80:20 харьцаатайгаар санамсаргүй түүврийн аргачлалаар сургалтын болон шалгалтын өгөгдөлд хуваан сургасан.

N. *Машин сургалтын үр дүн*

Сургасан загваруудад тестийн өгөгдлийг оруулж accuracy, confusion matrix, precision, recall, F1-score үнэлгээнүүдийг үндэслэн өвчний ерөнхий А ангиллын хувьд дэмжих вектор машин, В ангиллын хувьд ада идэвхижүүлэгч, С ангиллын хувьд дэмжих вектор машин, D ангиллын хувьд саналд үндэслэсэн ангилагч, Е ангиллын хувьд ада идэвхижүүлэгч, G ангиллын хувьд ада идэвхижүүлэгч, Н ангиллын хувьд дэмжих вектор машин тус тус хамгийн гүйцэтгэл сайн байна. Харин загваруудыг хөгжүүлэх цаг хугацааны хязгаарлалтаас шалтгаалан одоогоор I,J,K,L,M,N оношийн ангиллуудад зөвхөн градиент идэвхижүүлэгч загварыг сонгосон.

Судалгаанд ашигласан санамсаргүй ой, ада идэвхижүүлэгч, градиент идэвхижүүлэгч, стохастик градиент идэвхижүүлэгч, нэмэлт мод, саналд үндэслэсэн ангилагч нь сүүлийн жилүүдэд хамгийн их хөгжүүлэлт хийгдэж байгаа ансамбль хүчирхэг загварууд юм.

Хүснэгт 1 Машин сургалт загваруудын оновчтой таамаглах хувь, олон улсын өвчний ерөнхий ангилал ба загвар тус бүрээр

Загварууд	A	B	C	D	E	G	H
Ada boost	74%	77%	73%	69%	75%	71%	67%
Voting Classifier	74%	77%	73%	69%	75%	71%	67%
SVM	75%	76%	74%	69%	75%	70%	68%
SGB	73%	77%	73%	69%	75%	70%	67%
Extra Trees	74%	77%	74%	68%	74%	70%	66%
Gradient Boost	74%	77%	73%	69%	74%	70%	67%
Random Forest	73%	76%	73%	69%	74%	70%	67%
Dicision Tree	72%	75%	72%	68%	74%	70%	65%
KNN	70%	71%	69%	62%	70%	66%	62%
Загвар	I	J	K	L	M	N	-
Gradient Boost	75%	66%	66%	71%	68%	70%	-

O. *Эрсдэлийн үнэлгээний оноо*

Бид сургасан загваруудаа ашиглан даатгуулагч бүр тухайн олон улсын өвчний ерөнхий ангилал тус бүрээр өвчлөх эсэхийг таамагласны дараа ашиглах доор дурьдсан үнэлгээний оноог боловсруулсан.

Олон улсын өвчний ерөнхий ангилал тус бүрт харгалзах эмчилгээний зардалын дүн буюу эрүүл мэндийн даатгалын сангаас олгогдсон төлбөрүүдийн тодорхойлогч статистик, медиан, дээд квартиль, хамгийн их утгуудыг олсон. Уг олсон 13 утга бүрийн хувьд z score-г тооцоолж, тэдгээрийн нийлбэр дээр сөрөг утга авахаас зайлсхийж тогтмол тоо нэмсэн. Өөрөөр хэлбэл, А өвчний ангиллын хувьд эрүүл мэндийн даатгалын сангаас гарсан зардлын медиан 42.2 мянган төгрөг, В өвчний ангиллын хувьд эрүүл

мэндийн даатгалын сангаас гарсан зардлын медиан 35.1 мянган төгрөг гэх мэт байсан бөгөөд тэдгээрийн хувьд z-score тооцож олсон. Дээрх зарчмын дагуу дээд квартиль ба хамгийн их утгуудын хувьд мөн z-score олж, гурван z-score нийлбэр дүнгээр эрсдэлийн үнэлгээний оноог боловсрууллаа.

Эрсдэлийн үнэлгээний оноог зөвхөн медианы хувьд z-score ашиглах нь учир дутагдалтай юм. Жишээлбэл, C ангилалд харгалзах хавдрын өвчлөлийн бүлгийн эрүүл мэндийн даатгалын сангаас олгосон зардлын медиан 189.9 мянган төгрөг байхад хамгийн их утга нь 41.4 сая төгрөг байна.

Хүснэгт 2 Эрсдэлийн үнэлгээний онооны тооцоолол

Өвчний ангилал	медиан	дээд квартиль	хамгийн их утга	z score медиан	z score дээд квартиль	z score хамгийн их утга	Эрсдэлийн үнэлгээний оноо
A	42,198	70,608	14,512,040	-0.45	-0.53	-0.68	1.33
B	35,147	62,400	6,027,630	-0.63	-0.62	-0.86	0.88
C	189,900	413,246	41,409,860	3.31	3.24	-0.10	9.44
D	73,998	176,929	26,277,320	0.35	0.63	-0.43	3.56
E	60,160	108,344	29,926,870	0.00	-0.11	-0.35	2.53
G	44,677	89,900	44,690,040	-0.39	-0.32	-0.03	2.24
H	33,687	62,474	26,343,390	-0.67	-0.62	-0.42	1.27
I	47,679	86,300	46,004,700	-0.31	-0.36	-0.01	2.31
J	48,665	83,931	34,513,700	-0.28	-0.38	-0.25	2.06
K	64,750	118,398	198,595,100	0.12	0.00	3.23	6.34
L	36,660	54,913	28,202,980	-0.59	-0.70	-0.38	1.30
M	53,427	131,752	28,901,390	-0.16	0.13	-0.37	2.59
N	48,867	89,500	79,331,440	-0.28	-0.32	0.69	3.08

Дүгнэлт

Энэхүү судалгааны ажлын зорилгын хүрээнд ердийн даатгалын компанийн эрүүл мэндийн даатгалын бүтээгдэхүүн худалдан авах гэж буй даатгуулагчийн эрсдэлийг үнэлэх системийг боловсруулж, хөгжүүлэлтийг Odoо 17 программ дээр гүйцэтгэсэн. Уг систем нь дараах алхмын дагуу ашиглагдана. Үүнд:

Алхам 1 (Хавсралт 2a)

Системд нэвтэрнэ.

Алхам 2 (Хавсралт 2b)

Даатгуулагчийн ерөнхий мэдээлэл ба шинж тэмдгийн асуулгуудыг үнэн зөв, бүрэн бөглөж эрсдэл үнэлэх хэсгийг дарна.

Алхам 3 (Хавсралт 2c)

- Алхам 2-т заасны дагуу бөглөсөн оролтын утгууд олон улсын өвчний оношийн ерөнхий ангилал тус бүрээр таамаглах машин сургалтын загварууд (Хүснэгт 1-т харуулсан тайлбарлах чадвартай) руу орж тухайн даатгуулагчийг таамаглана. Таамагласан гаралтын утга нь даатгуулагч тухайн ерөнхий оношоор өвдөх өндөр магадлалтай бол тийм буюу 1, эсрэг тохиолдолд үгүй буюу 0 гэсэн утга авна.
- Машин сургалтын гаралтын утга тийм буюу 1 бол Хүснэгт 2-т харуулсан

оношийн ерөнхий ангилалд харгалзах эрсдэлийн үнэлгээний оноог авна. Харин машин сургалтын гаралтын утга үгүй буюу 0 бол эрсдэлийн үнэлгээний оноонд 0 утга харгалзана.

- Оношийн ерөнхий ангилал тус бүрт харгалзах эрсдэлийн үнэлгээний оноонуудын нийлбэрээр тухайн даатгуулагчийн эрсдэлийн төвшин тодорхойлогдоно.

- Даатгуулагчийн авч болох нийт эрсдэлийн онооны боломжит бүх утгыг тухайн даатгалын компани бизнес төлөвлөгөө, эрсдэлийн удирдлагын бодлого, борлуулалтын төлөвлөгөө, даатгалын гэрээ зэрэгтэй уялдуулан хэд хэдэн интервалд хувааж ашиглана. Тухайн интервалд харгалзах даатгалын хураамжийн актуар тооцооллыг олон улсын оношийн ерөнхий ангилал тус бүрээр нарийвчлалтай хийвэл энэхүү системтэй илүү нийцтэй, хоршиж ажиллах бөгөөд хялбар байна. Даатгалын цэвэр хураамжийг тооцоолохдоо түгээмэл хэрэглэдэг аль ч аргыг сонгон хэрэглэж болно. Жишээлбэл, хохирлын хэмжээ, хохирлын давтамжийг олон улсын оношийн ерөнхий ангилал тус бүрээр тооцоолж, хязгаарын гол теорем ашиглах замаар харьцангуй ачааллын коэффициент, цэвэр хураамжийг олж болно. Бидний судалгааны ажлын цар хүрээнд ердийн даатгалын компани тус бүрийн эрүүл мэндийн даатгалын гэрээ өөр бөгөөд даатгалын үнэлгээ, хамгаалах эрсдэл, өөрийн хүлээх хариуцлага, хэвтэн эмчлүүлэх зардлын дэд хязгаар, хэвтэн эмчлүүлэх эмнэлэг зэрэг олон үзүүлэлтүүд ялгаатай тул даатгалын хураамжийн актуар тооцоолол хийгдээгүй.

Цаашид судалгааны ажлын төвшинд дараах зүйлсийг хийх шаардлагатай гэж бидний зүгээс үзэж байна.

- Монгол улсын хүн амын өвчлөлийн өгөгдлийг ашиглан эрсдэлийн үнэлгээнд ач холбогдолтой хамаарах хувьсагчийн судалгаа хийх. Жишээлбэл, эмнэлэгт хэвтсэн хоног, биеийн жин, артерын даралт гэх мэт

- Хувьсагч сонгоход (Feature selection) илүү нарийвчилсан судалгаа хийх.

- Үндсэн бүрэлдэхүүний шинжилгээг (Principal component analysis) дэлгэрэнгүй хийх.

- Эрсдэлийн үнэлгээний оноог илүү нарийвчлан тогтоох, практик ач холбогдлыг судлах хэрэгтэй.

- Машин сургалтын загварууд болон эрсдэлийн үнэлгээний оноотой уялдуулан даатгалын хураамжийн актуар тооцооллыг шинээр гүйцэтгэх.

- Hyperparameter tuning хийх шаардлагатай.

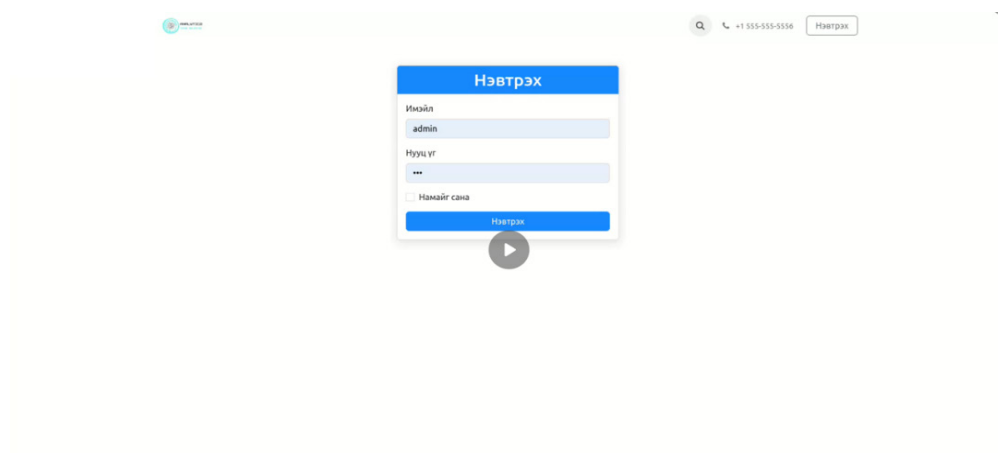
Ашигласан материал

- Alex J.Smola and Bernhard Scholkopf. (2004). A tutorial on support vector regression. *Statistics and Computing* 14, 199-222.
- Bellotti, T., Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert System* 36 (2), 3302-3308
- Chanawee Chavaltada, Kitsuchart Pasupa, and David R. Hardoon. (2017). A Comparative Study of Machine Learning Techniques for Automatic Product Categorisation. Springer.
- Cuicui Luo, Desheng Wu, Dexiang Wu. (2017). A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*.
- Flavio Barboza, Herbert Kimura, Edward ALtman. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications* 83, Elsevier, 405-417.
- Fuster, A. (2018). Predictably Unequal The Effects of Machine Learning on Credit Markets. *The Effects of Machine Learning on Credit Markets*.
- G. James, D. Witten, T.Hastie and R. Tibshirani. (2017). An Introduction to Statistical Learning. Springer.
- M.Bishop, C. (2006). *Pattern recognition and machine learning*. Cambridge: Springer.
- Nicolas Chapados, Yoshua Bengio. (2002). Estimating car insurance premia: a case study in High-dimensional data inference. *Advances in Neural Network*.
- Peter Addo, Dominique Guegan, Bertrand Hassani. (2018). Credit Risk Analysis using Machine and Deep Learning models. *Maison des Sciences Economiques*, 106-112.
- Sebastian Raschka and Vahid Mirjalili. (2017). *Python Machine Learning*. Birmingham: Packt Publishing.

The heatmap displays the correlation matrix for 100 variables. The diagonal is white, indicating a correlation of 1. The color scale ranges from dark purple (low correlation) to light yellow (high correlation). The matrix shows a block-like structure with high correlations within groups and lower correlations between groups.

Эх сурвалж: Судлаачдын тооцоо

Хавсралт 2а Системд нэвтрэх хэсэг



Хавсралт 2б Мэдээлэл оруулах хэсэг

Эрсдлийн үнэлгээний маягт

Ерөнхий асуултууд

Нас	Хүйс
23	Эмэгтэй
Хөл	Боловсрол
Улаанбаатар	Дээд боловсрол
Ажлын газар	Албан тушаал
Бусад	Албан хаагч
Мэргэжил	
Боловсролын салбарын ажилтан	

Танд ойрын үед дараах шинж тэмдэгүүд илэрч байна уу?

Халуурлах	Суулгах	Арьсны тууралт	Тунгалагийн булчирхайн томрол
Үгүй	Үгүй	Үгүй	Үгүй
Үгүй			
Тийм	Хэвийн бус цус алдах	Жин хасагдах	Идэвг
Үгүй	Үгүй	Үгүй	Үгүй

SEARCH +1 555-555-5556 Administrator

Амьсгаадах Үгүй	Арьс цайвар болох Үгүй	Хоолны дуршил өөрчлөгдөх Үгүй	Бөөний жиж өөрчлөгдөх Үгүй
Хэт ядрах Үгүй	Талстай өвдөх Үгүй	Таталт өгөх Үгүй	Булчлагчийн сулрал Үгүй
Хараа муудах Үгүй	Нүд улайх Үгүй	Нүлэмс гоожиж Үгүй	Сонсгол буурах Үгүй
Чихээс шингэн уунал гарах Үгүй	Чих өвдөх Үгүй	Цээжээр өвдөх Үгүй	Хал хавагнах Үгүй
Ханиалгах Үгүй	Хэвлийн өвдөлт Үгүй	Веелжих Үгүй	Арьс сагатнах, улайх Үгүй
Үе мөчний өвдөлт Үгүй	Хавагнах Үгүй	Хөдөлгөөн хязгаарлагдах Үгүй	Шээсний өөрчлөлт Үгүй
Өвдөлттэй шээх Үгүй	Хэвлийн доод хэсэг өвдөх Үгүй		

Эрсдэл үнэлэх Цэвэрлэх

Хавсралт 2с Даатгуулагчийн эрсдэлийг үнэлэх, тооцоолол хийх хэсэг

Үр дүн

Нийт оношийн бүлэг	Нийт оноо
A оношийн бүлэг (Халдварт өвчин): Үгүй	A оноо: 0.0
B оношийн бүлэг (Халдварт өвчин): Үгүй	B оноо: 0.0
C оношийн бүлэг (Халдвар болон хорт хавдрын өвчин): Үгүй	C оноо: 0.0
D оношийн бүлэг (Халдвар болон цус, дархлааны өвчин): Үгүй	D оноо: 0.0
E оношийн бүлэг (Дрогид шүүрлийн болон бодисын солилцооны өвчин): Үгүй	E оноо: 0.0
G оношийн бүлэг (Мэдрэлийн тогтолцооны өвчин): Үгүй	G оноо: 0.0
H оношийн бүлэг (Нүдний болон чихний өвчин): Үгүй	H оноо: 0.0
I оношийн бүлэг (Зүрх судасны өвчин): Үгүй	I оноо: 0.0
J оношийн бүлэг (Амьсгалын замын өвчин): Тийм	J оноо: 2.068003
K оношийн бүлэг (Хоол боловсруулах тогтолцооны өвчин): Үгүй	K оноо: 0.0
L оношийн бүлэг (Арьс болон арьсан доорх эдийн өвчин): Тийм	L оноо: 1.307663
M оношийн бүлэг (Яс бүлчлэгийн тогтолцооны өвчин): Үгүй	M оноо: 0.0
N оношийн бүлэг (Шээс болон нөхөн үржигүйн тогтолцооны өвчин): Үгүй	N оноо: 0.0

Эрсдэлийн дүн

Нийт эрсдэлийн үнэлгээний оноо: 3.375666

Нийт хураамж

900000

Form руу буцах