C5.0 алгоритмд суурилсан сэтгэл ханамжийн ангиллын дүрэмт загварын судалгаа

Б.Одгэрэл[1], Ц.Жавзмаа[2], Т.Ахыт[2], Ч.Баасандорж[2], У.Нямдавваа[2], Б.Оюунцэцэг[3], Л.Ажнай[2], Л.Пүрэвдолгор[2]

[1]АШУҮИС, Сувилахуйн сургууль, Ерөнхий эрдмийн тэнхим
[2]АШУҮИС, Био-Анагаахын сургууль, Физик мэдээлэл зүйн тэнхим
[3]АШУҮИС, Био-Анагаахын сургууль, Харилцааны ур чадварын тэнхим
Цахим шуудан: odgerel.b@mnums.edu.mn, Утас: 99879002

Түлхүүр үг:
Cross-validation
Дүрэмд суурилсан загвар
Шийдвэрийн мод
Likert төрлийн өгөгдөл
Сэтгэл ханамжийн ангилал
Хиймэл оюун

Товч утга:
Үндэслэл: Likert төрлийн асуулгад суурилсан өгөгдлийг сэтгэл ханамжийг үнэлэхэд өргөн ашигладаг боловч бодит хэрэглээнд зөвхөн ангиллын өндөр нарийвчлал бус, загварын тайлбарлах чадвар чухал шаардлагатай болдог. Дүрэмд суурилсан машин сургалтын аргууд нь эдгээр хоёр шаардлагыг зэрэг хангах боломж олгодог. Сүүлийн жилүүдэд машин сургалтын аргуудыг ашиглан ангилал, таамаглал хийх судалгаа эрчимтэй хөгжиж байгаа ч сэтгэл ханамжийн Likert төрлийн асуулгад C5.0 ангиллын алгоритм ашиглсан судалгаа хомс байгаа тул уг судалгааг хийх үндэслэл болсон. **Зорилго:** Сэтгэл ханамжийн хоёр ангиллыг (High/Medium) таамаглахад C5.0 ангиллын алгоритмыг 10-fold cross-validation аргаар хэрэгжүүлж, загварын гүйцэтгэл болон тайлбарлах чадварыг үнэлэх. **Арга, аргачлал:** Судалгааг АШУҮИС-ийн Био-Анагаахын сургуулийн Физик мэдээлэл зүйн тэнхимийг түшиглэн Эрүүл мэндийн мэдээллийн технологич мэргэжлээр төгссөн төгсөгчдийн сэтгэл ханамжийн судалгааны өгөгдлийг ашиглан хийсэн. Судалгаанд Likert төрлийн дөрвөн хувьсагч (V1–V4) бүхий өгөгдлийг ашигласан. C5.0 загварыг параметрийн торон хайлтаар сургаж, 10-fold cross-validation-ын хүрээнд accuracy болон confusion matrix ашиглан үнэлсэн. Мөн эцсийн дүрэмд суурилсан загвараас IF–THEN дүрмүүдийг гарган авч, шийдвэрийн логикийг flowchart хэлбэрээр дүрсэлсэн. **Үр дүн:** Судалгааны үр дүнд дүрэмд суурилсан C5.0 загвар хамгийн оновчтой гэж сонгогдож, cross-validation-ын хүрээнд тогтвортой гүйцэтгэл үзүүлсэн. Confusion matrix-ийн үр дүн нь High ангиллыг өндөр нарийвчлалтай илрүүлж байгааг харуулсан. Гарсан дүрмүүдээс V4 болон V2 хувьсагчид сэтгэл ханамжийг тодорхойлох голлох нөлөөтэй болох нь илэрхий болсон. **Дүгнэлт:** C5.0 алгоритмыг cross-validation-ын хүрээнд ашиглах нь Likert өгөгдөлд суурилсан сэтгэл ханамжийн ангиллыг үр дүнтэй, ойлгомжтой байдлаар гүйцэтгэх боломжтой ба дүрэмд суурилсан бүтэц нь загварыг тайлбарлах, шийдвэр дэмжих хэрэглээнд ашиглахад тохиромжтой давуу тал болж байгаа нь тус тус батлагдлаа.

Үндэслэл: Сүүлийн жилүүдэд асуулгад суурилсан өгөгдөл болон Likert төрлийн хэмжилтийн шкалуудыг ашиглан хэрэглэгчийн болон үйлчилгээний сэтгэл ханамжийг ангилах, таамаглах судалгаа эрчимтэй хөгжиж байна.[1,2] Ийм төрлийн өгөгдөл нь субъектив үнэлгээнд суурилдаг тул уламжлалт статистик аргуудаас гадна машин сургалтын ангиллын алгоритмуудыг ашиглах нь илүү уян хатан, өгөгдлийн далд хамаарлыг илрүүлэх боломж олгодог гэж үздэг.[3] Гэсэн хэдий ч бодит судалгаанд зөвхөн ангиллын өндөр нарийвчлал төдийгүй шийдвэрийн логикийг ойлгомжтой тайлбарлах чадвар чухал шаардлага болдог.

Шийдвэрийн мод болон дүрэмд суурилсан ангилагчид нь өгөгдлийн нөхцөлт бүтэц, ангиллын шалгуурыг IF–THEN хэлбэрээр ил тод илэрхийлэх давуу талтай. Эдгээр аргуудын нэг болох C5.0 алгоритм нь Quinlan-ийн боловсруулсан C4.5 алгоритмын сайжруулсан хувилбар бөгөөд мэдээллийн ашиг (information gain), boosting, дүрэм гаргалт зэрэг

боломжуудыг агуулдаг.[4] C5.0 нь ангиллын өндөр гүйцэтгэлтэй байхаас гадна, тухайн ангилалд хүргэж буй логикийг тайлбарлах боломж олгодгоороо "Хар хайрцаг" шинжтэй загваруудаас ялгардаг.

Машин сургалтын загварын гүйцэтгэлийг үнэлэхэд cross-validation арга түгээмэл хэрэглэгддэг. Ялангуяа өгөгдлийн хэмжээ хязгаарлагдмал үед нэг л training/test хуваалт дээр тулгуурлах нь загварын бодит гүйцэтгэлийг хэт үнэлэх эрсдэлтэй байдаг. Харин k-fold cross-validation нь өгөгдлийг олон дэд багцад хуваан ээлжлэн шалгах замаар загварын ерөнхийшүүлэх чадварыг илүү найдвартай үнэлэх боломжийг олгодог.[5] Kohavi-ийн судалгаанд cross-validation нь ангиллын загварын бодит гүйцэтгэлийг үнэлэхэд үр дүнтэй арга болохыг онцолсон байдаг.

Сүүлийн жилүүдэд Explainable AI (XAI) чиг хандлага эрчимтэй хөгжиж, нийгэм болон эрүүл мэндийн салбарын судалгаанд тайлбарлах боломжгүй загваруудыг ашиглахад болгоомжтой хандах шаардлага тулгарч байна. Rudin-ийн тэмдэглэснээр өндөр ач холбогдолтой шийдвэрүүдэд ашиглагдах загварууд нь заавал ойлгомжтой, ил тод байх ёстой бөгөөд тайлбарлах боломжгүй "Хар хайрцаг" загваруудыг дараа нь тайлбарлах оролдлого хийхээс илүүтэйгээр анхнаасаа тайлбарлагдахуйц загвар сонгох нь зүйтэй.[6] Энэ хүрээнд C5.0-ийн дүрэмд суурилсан ангилагч нь тайлбарлах чадвар болон гүйцэтгэлийн хоорондын зохистой тэнцвэрийг хангах боломжтой арга хэдийч манай улсад энэ чиглэлээр хийгдсэн судалгааны ажил хомс байгаа нь энэхүү судалгааны ажлын үндэслэл боллоо.

Зорилго: Likert төрлийн хэмжилтэд суурилсан өгөгдөл дээр C5.0 ангиллын алгоритмыг 10-fold cross-validation аргаар хэрэгжүүлж, сэтгэл ханамжийн хоёр ангиллыг (High/Medium) таамаглах загварын гүйцэтгэл болон тайлбарлах чадварыг үнэлэх.

Арга, аргачлал: Судалгааг АШУҮИС-ийн Био-Анагаахын сургуулийн Физик мэдээлэл зүйн тэнхимийг түшиглэн Эрүүл мэндийн мэдээллийн технологич мэргэжлээр төгссөн төгсөгчдийн сэтгэл ханамжийн судалгааны өгөгдлийг ашиглан хийсэн.
Сэтгэл ханамжийг илэрхийлэх Likert төрлийн 1-5 онооны хэмжилтийн хувьсагчид (V1–V4) болон ангиллын хувьсагч SatisfactionBinary (High/Medium) бүхий өгөгдлийг ашигласан.[7] Бүх тооцоолол, загвар сургалт, cross-validation болон дүрслэлийг R орчинд C50, caret, ggplot2, DiagrammeR зэрэг багцуудыг ашиглан хэрэгжүүлсэн.

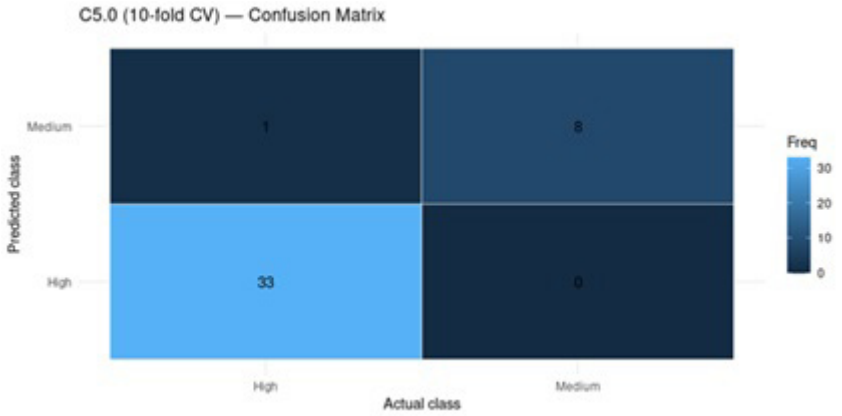
Ангиллын загварыг C5.0 алгоритмаар байгуулсан. Загварын гүйцэтгэлийг үнэлэхдээ 10-fold cross-validation аргыг ашигласан.[5] Сонгогдсон хамгийн сайн загварыг Accuracy хэмжүүрээр оновчилж, out-of-fold таамаглал дээр үндэслэн confusion matrix-ийг тооцсон. Confusion matrix нь High/Medium ангиллын зөв ба буруу ангилалтын тоог нэг дор харуулж, ангиллын алдааны бүтэц (аль ангилал руу андуурах хандлагатай гэх мэт)-ийг тайлбарлахад ашиглагддаг.[8] Нэмж, fold бүрийн accuracy-г тооцож баганан графикаар дүрслэн cross-validation-ын тогтвортой байдлыг харуулсан. C5.0-ийн тайлбарлах чадварыг нэмэгдүүлэх зорилгоор summary(finalModel) ашиглан IF–THEN дүрмүүдийг гаргаж, хамгийн өндөр хамрах хүрээ/итгэлцэлтэй дүрмүүдийн логикийг flowchart/decision logic хэлбэрээр дүрсэлсэн.[9]

Үр дүн:
Cross-validation-ын оновчтой загварын сонголт: C5.0 ангиллын загварыг 10-н удаагийн алдааны матриц (cross-validation) аргаар сургаж, нарийвчлал (accuracy)-ийн хэмжүүрийг ашиглан хамгийн оновчтой тохиргоог сонгосон. Параметрийн торон хайлтын үр дүнд дүрэмд суурилсан (rules) загвар нь бусад хувилбаруудаас илүү гүйцэтгэлтэй гарсан. Энэхүү үр дүн нь судалгаанд ашигласан өгөгдлийн хувьд энгийн, дүрэмд суурилсан C5.0 загвар хамгийн тохиромжтой болохыг харуулж байна (Хүснэгт 1).

Хүснэгт 1. C5.0 загварын оновчтой параметрүүд

Параметр	Утга
Загварын төрөл (model)	Rules
Boosting давталт (trials)	1
Winnow (feature selection)	FALSE
Оновчлолын хэмжүүр	Accuracy
Validation арга	10-fold cross-validation

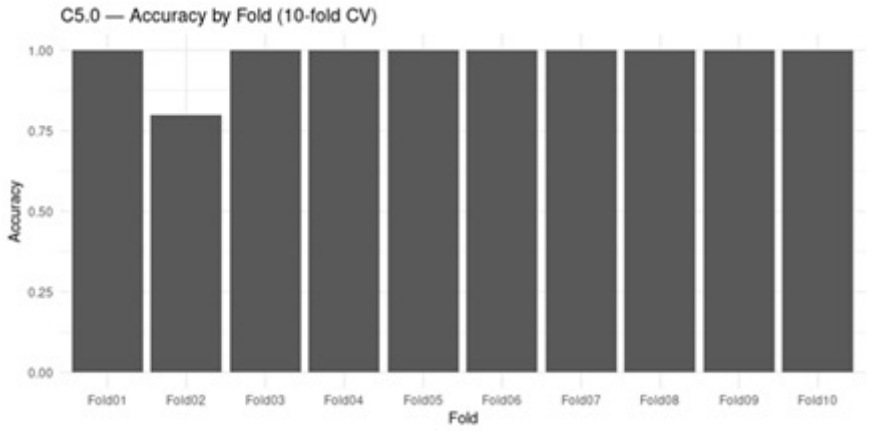
Ангиллын гүйцэтгэл ба confusion matrix: Оновчтой C5.0 загварын out-of-fold таамаглал дээр үндэслэн ангиллын гүйцэтгэлийг үнэлэхэд, High ба Medium ангиллуудын хоорондын ялгалт тогтвортой, High ангиллыг алдаагүй, харин Medium ангиллын нэг ажиглалт High руу андуурагдсан байна. Энэ нь C5.0 загвар High сэтгэл ханамжийг илрүүлэхдээ илүү мэдрэг, харин Medium ангилалд бага зэргийн алдаа гарч байгааг илтгэнэ (Зураг 1).



Зураг 1. Алдааны матрицын дүрслэл (Confusion matrix-ийг heatmap).

C5.0 загварын out-of-fold таамаглалд үндэслэсэн confusion matrix-ийг heatmap хэлбэрээр дүрсэлж, баганаар бодит (actual), мөрөөр таамагласан (predicted) ангиллыг илэрхийлсэн.

Cross-validation-ын тогтвортой байдал: Хөндлөн баталгаажуулалт (10-fold cross-validation)-ын хүрээнд нугалаа (fold) бүрийн нарийвчлал (accuracy)-ийг тооцоход, гүйцэтгэлийн хэлбэлзэл харьцангуй бага, загварын ерөнхийшүүлэх чадвар тогтвортой байгааг харуулсан (Зураг 2).



Зураг 2. C5.0 загварын нугалаа бүрийн хөндлөн баталгаажуулалтын навийчлал.

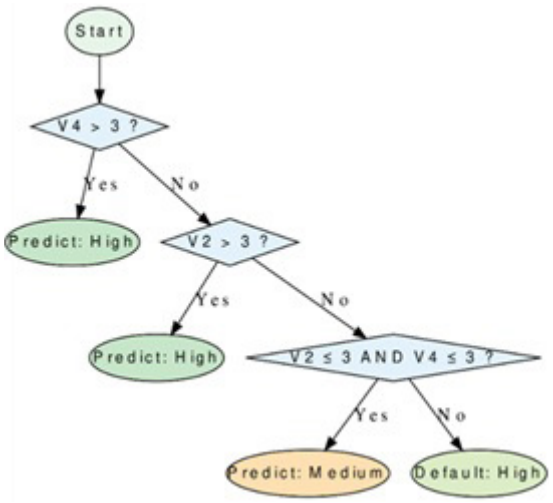
Энэ нь загвар нэг л хуваалт дээр санамсаргүй өндөр үр дүн үзүүлээгүй, харин өгөгдлийн дэд олонлог бүр дээр ойролцоо гүйцэтгэл үзүүлж байгааг илтгэнэ.

C5.0 дүрмүүд ба тайлбарлах үр дүн: C5.0-ийн rule-based загвараас IF–THEN хэлбэрийн ойлгомжтой дүрмүүд гарсан нь загварын тайлбарлах чадварыг нэмэгдүүлсэн (Хүснэгт 2).

Хүснэгт 2. C5.0 загвараас гарсан гол дүрмүүд

Дүрэм №	Нөхцөл	Таамагласан ангилал	Итгэл (Confidence)
Rule 1	V4 > 3	High	0.968
Rule 2	V2 > 3	High	0.939
Rule 3	V2 ≤ 3 ∧ V4 ≤ 3	Medium	0.900
Default	Бусад тохиолдол	High	–

Эдгээр дүрмээс харахад V4 болон V2 хувьсагчид сэтгэл ханамжийн түвшинг тодорхойлоход голлох нөлөө үзүүлж байна. Ялангуяа V4>3 болон V2>3 нөхцөлүүд нь High ангилалд хүчтэй хамааралтай бол, харин V2 ба V4 хоёулаа 3-аас бага буюу тэнцүү үед Medium ангилал тодорхойлогдож байна (Зураг 3).



Зураг 3. C5.0 загвараас гарсан сэтгэл ханамжийн ангиллын шийдвэрийн мод.

C5.0 дүрэмд суурилсан ангилагчийн шийдвэрийн логикийг харуулсан шийдвэрийн мод нь V4-ийн утга 3-аас их үед сэтгэл ханамжийг High гэж ангилна. Хэрэв V4≤3 бол дараагийн шалгалтаар V2>3 эсэхийг үнэлж, уг нөхцөл биелсэн тохиолдолд High ангилалд оруулна. Харин V2 ба V4 хоёулаа 3-аас бага буюу тэнцүү үед сэтгэл ханамжийг Medium гэж ангилна. Эдгээр нөхцөлд хамаарахгүй үлдсэн тохиолдлуудыг анхдагч байдлаар High ангилалд оруулна.

Дээрх үр дүнгүүдээс дараах дүгнэлтийг хийж болно. Үүнд:

- Нэгдүгээрт, C5.0-ийн дүрэмд суурилсан загвар нь 10-fold cross-validation-ын хүрээнд тогтвортой гүйцэтгэл үзүүлсэн.
- Хоёрдугаарт, загварын гаргасан IF–THEN дүрмүүд нь сэтгэл ханамжийн ангиллыг ойлгомжтой, ил тод байдлаар тайлбарлах боломж олгож байна.
- Гуравдугаарт, V2 болон V4 хувьсагчид сэтгэл ханамжийн ангилалд голлох нөлөөтэй хүчин зүйлс болох нь илэрхий болсон.

Хэлцэмж: Энэхүү судалгаанд сэтгэл ханамжийн өгөгдөлд C5.0 ангиллын алгоритмыг 10-fold cross-validation аргаар хэрэгжүүлж, High болон Medium гэсэн хоёр ангиллыг таамаглахад загварын гүйцэтгэл болон тайлбарлах чадварыг үнэлэв. Үр дүнгээс харахад C5.0-ийн дүрэмд суурилсан (rule-based) хувилбар нь параметрийн хайлтын хүрээнд хамгийн оновчтой загвар болж сонгогдсон нь өгөгдлийн бүтэц харьцангуй энгийн, хувьсагчдын ялгах чадвар тодорхой байгаатай холбоотой гэж үзэж байна. Энэ нь эрүүл мэндийн мэдээлэл зүйн боловсролын судалгаанд илэрсэн сэтгэл ханамжийн хүчин зүйлүүд тодорхой, бүтэцтэй байгааг харуулсан фактор шинжилгээний үр дүнтэй нийцэж байна.[10,11]

Quinlan-ийн тэмдэглэснээр C5.0 алгоритм нь decision tree загвараас илүү цөөн, ойлгомжтой дүрмүүд үүсгэх хандлагатай бөгөөд энэ нь загварын тайлбарлах чадварыг нэмэгдүүлдэг. Энэхүү судалгаанд гарсан IF–THEN дүрмүүд нь сэтгэл ханамжийн ангиллыг тодорхойлох гол хувьсагчдыг ил тод харуулсан нь боловсролын чанарын үнэлгээнд практик ач холбогдолтой юм.[4]

Confusion matrix болон cross-validation-ийн үр дүнгээс харахад загвар High ангиллыг өндөр нарийвчлалтай илрүүлж, Medium ангилалд харьцангуй бага хэмжээний андуурал үзүүлсэн. Энэ нь сэтгэл ханамжийн өгөгдөлд өндөр үнэлгээ давамгайлах хандлага ажиглагддагтай холбоотой бөгөөд өмнөх судалгаанд төгсөгчдийн сэтгэл ханамж дунджаас дээш үнэлэгдсэнтэй нийцэж байна.[10,12] Гэсэн хэдий ч fold бүрийн гүйцэтгэл тогтвортой байсан нь загвар хэт тохиролд (overfitting) ороогүй, ерөнхийшүүлэх чадвар боломжийн байгааг илтгэнэ. Kohavi-ийн судалгаанд cross-validation нь жижиг өгөгдлийн багцад загварын гүйцэтгэлийг найдвартай үнэлэх арга болохыг онцолсон байдаг.[5]

Судалгааны чухал үр дүнгийн нэг нь C5.0 загвараас гарсан дүрэмд суурилсан тайлбар юм. Тухайлбал, V4 болон V2 хувьсагчид сэтгэл ханамжийн ангилалд голлох нөлөө үзүүлж байгааг илрүүлсэн. V4>3 болон V2>3 нөхцөлүүд нь High ангилалтай хүчтэй хамааралтай байсан бол, эдгээр хувьсагчид 3-аас бага буюу тэнцүү үед Medium ангилал тодорхойлогдсон. Энэ нь фактор шинжилгээгээр илэрсэн “Сургалтын чанар”, “Практик ур чадвар” зэрэг хүчин зүйлүүд сэтгэл ханамжид нөлөөлдөгтэй уялдаж байна.[10,13]

Explainable AI (XAI)-ийн өнөөгийн чиг хандлагын хүрээнд дүрэмд суурилсан загварууд нь нийгэм, боловсрол, эрүүл мэндийн судалгаанд өндөр ач холбогдолтойд тооцогдож байна. Rudin-ийн үзэж байгаагаар тайлбарлах боломжгүй “Хар хайрцаг” загваруудыг дараа нь тайлбарлах оролдлого хийхээс илүүтэйгээр анхнаасаа ойлгомжтой загварыг ашиглах нь илүү найдвартай шийдэл юм. Энэ утгаараа C5.0-ийн дүрэмд суурилсан загвар нь зөвхөн өндөр гүйцэтгэлтэй байхаас гадна шийдвэр гаргалтын логикийг ил тод харуулж байгаагаараа давуу талтай.[6]

Нөгөө талаас, төгсөгчдийн ажил эрхлэлтийн өндөр хувь (≈85%) болон ажил олгогчдын өндөр үнэлгээ нь боловсролын үр дүнг баталгаажуулж байгаа хэдий ч сэтгэл ханамжийн зарим хүчин зүйлүүд дунд түвшинд үнэлэгдсэн нь загварын ангиллын үр дүнтэй уялдаж байна.[10,14] Иймээс өгөгдөлд илэрсэн хэв маяг нь бодит боловсролын системийн сул болон давуу талуудыг тусгасан гэж үзэж болно.

Гэсэн хэдий ч энэхүү судалгаанд ашигласан өгөгдлийн хэмжээ харьцангуй бага (n=42) байгаа нь загварын үр дүнг ерөнхийшүүлэхэд тодорхой хязгаарлалт үүсгэж байна. Цаашид илүү олон өгөгдөл цуглуулах, олон ангиллын загвар турших, мөн энэхүү дүрэмд суурилсан загварыг боловсролын чанарын үнэлгээ болон шийдвэр дэмжих системд нэвтрүүлэх шаардлагатай байна. Мөн өгөгдөл олборлолт, кластер шинжилгээ зэрэг аргуудыг хавсарган ашиглах нь үр дүнг илүү гүнзгийрүүлэх боломжтой.[15]

Дүгнэлт: C5.0 алгоритмыг cross-validation-ын хүрээнд ашиглах нь Likert өгөгдөлд суурилсан сэтгэл ханамжийн ангиллыг үр дүнтэй, ойлгомжтой байдлаар гүйцэтгэх боломжтой ба дүрэмд суурилсан бүтэц нь загварыг тайлбарлах, шийдвэр дэмжих хэрэглээнд ашиглахад тохиромжтой давуу тал болж байгаа нь тус

тус батлагдлаа. Цаашид илүү олон ажиглалт бүхий өгөгдөл дээр турших, ангиллын түвшинг нарийсгах, эсвэл дүрэмд суурилсан загварыг шийдвэр дэмжих системд нэвтрүүлэх судалгааг өргөжүүлэх боломжтой.

Ном зүй:

1. *Zeithaml VA, Bitner, M. J., &amp, Gremler DD. Services Marketing (7th ed.). 2018;*
2. *Oliver RL. Staisfaction: A behoviorol prespective on the consumer. McGraw-Hill. 1997;*
3. *Han J, Kamber, M., & Pei, J. Data Mining: Concepts and Techniques. 2012.*
4. *Quinlan JR. C4.5: Programs for Machine Learning. Morgan Kaufmann; 1993.*
5. *Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Proceedings of IJCAI; 1995.*
6. *Rudin C. Stop explaining black box machine learning models for high stakes decisions. Nature Machine Intelligence. 2019;*
7. *He H, &amp, Garcia EA. Learning from imbalanced data IEEE Transactions on Knowledge and Data Engineering 21(9). 2009:1263–1284.*
8. *Fawcett T. An Introduction to ROC Analysis. Pattern Recognition Letters. 2006;*
9. *Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software. 2008;*
10. *Luvsantseren Pea. Reform of the Health Informatics Curriculum in Mongolia: Needs and Development Perspectives. IEEE. 2025;*
11. *Han JP, J. and Tong, H. Data mining: Concepts and techniques, Morgan Kaufmann, . 2011;*
12. *Solongo M. Health information curriculum satisfaction survey. Ulaanbaatar2015.*
13. *Galbadrakh C. Health informatics curriculum satisfaction survey. Ulaanbaatar2022.*
14. *Mongolia Go. Development Policy (2015–2025) Monitoring Report. 2015;*
15. *Kaufman LR, P. J. Finding groups in data: An introduction to cluster analysis, John Wiley & Sons. 2009;*

Research on a satisfaction classification model based on the C5.0 algorithm

Odgerel B [1], Javzmaa Ts[2], Akhit T[2], Baasandorj Ch[2], Nyamdavaa U[2], Oyuntsetseg S[3], Ajnai L[2], Purevdolgor L[2]

[1]Department of Basic Sciences, School of Nursing, MNUMS
[2]Department of Physics & Informatics, School of Biomedicine, MNUMS
[3]Department of Communication Skills. School of Biomedicine, MNUMS
E-mail: odgerel.b@mnums.edu.mn, Tel: +976-99879002

Abstract:

Background: Likert-scale questionnaire data is widely used to evaluate satisfaction; however, in practical applications, both high classification accuracy and model interpretability are essential requirements. Rule-based machine learning methods provide the opportunity to satisfy both criteria simultaneously. Despite the rapid growth of research using machine learning for classification and prediction in recent years, there is a scarcity of studies employing the C5.0 classification algorithm for Likert-scale satisfaction data, which serves as the rationale for this study.

Aim: To implement the C5.0 classification algorithm using a 10-fold cross-validation method to predict two levels of satisfaction (High/Medium) and to evaluate the model's performance and interpretability.

Materials and Methods: he study was conducted at the Department of Physics and Informatics, School of Bio-Medicine, MNUMS, utilizing satisfaction survey data from graduates majoring in Health Information Technology. Data consisting of four Likert-scale variables (V1–V4) were used. The C5.0 model was trained using a parametric grid search and evaluated within a 10-fold cross-validation framework using accuracy and a confusion matrix. Furthermore, IF–THEN rules were extracted from the final rule-based model, and the decision logic was visualized in the form of a flowchart.

Results: The results indicated that the rule-based C5.0 model was selected as the optimal model, demonstrating stable performance throughout the cross-validation process. The confusion matrix results showed high precision in identifying the "High" satisfaction class. Among the generated rules, it became evident that variables V4 and V2 have the most significant influence on determining satisfaction levels.

Conclusion: Utilizing the C5.0 algorithm within a cross-validation framework enables effective and intelligible classification of satisfaction based on Likert data. It was confirmed that the rule-based structure offers distinct advantages for model interpretation and decision-support applications.

Keywords: Cross-validation, Rule-based model, Decision tree, Likert-scale data, Satisfaction classification, Artificial intelligence.