

**ЗЭЭЛДЭГЧИЙН ЗЭЭЛЖИХ ЧАДВАРЫГ ОРЧИН
ҮЕИЙН СТАТИСИК АРГУУДААР ҮНЭЛЭХ НЬ**

Судлаач, Б. Шинэцэцэг

АГУУЛГА

- ✓ ИЛТГЭЛИЙН ХУРААНГУЙ
- ✓ УДИРТГАЛ
- ✓ ИЛТГЭЛИЙН ҮНДСЭГ ХЭСЭГ
- ✓ ДҮГНЭЛТ
- ✓ СУДАЛГААНД АШИГЛАСАН НОМ ЗҮЙ

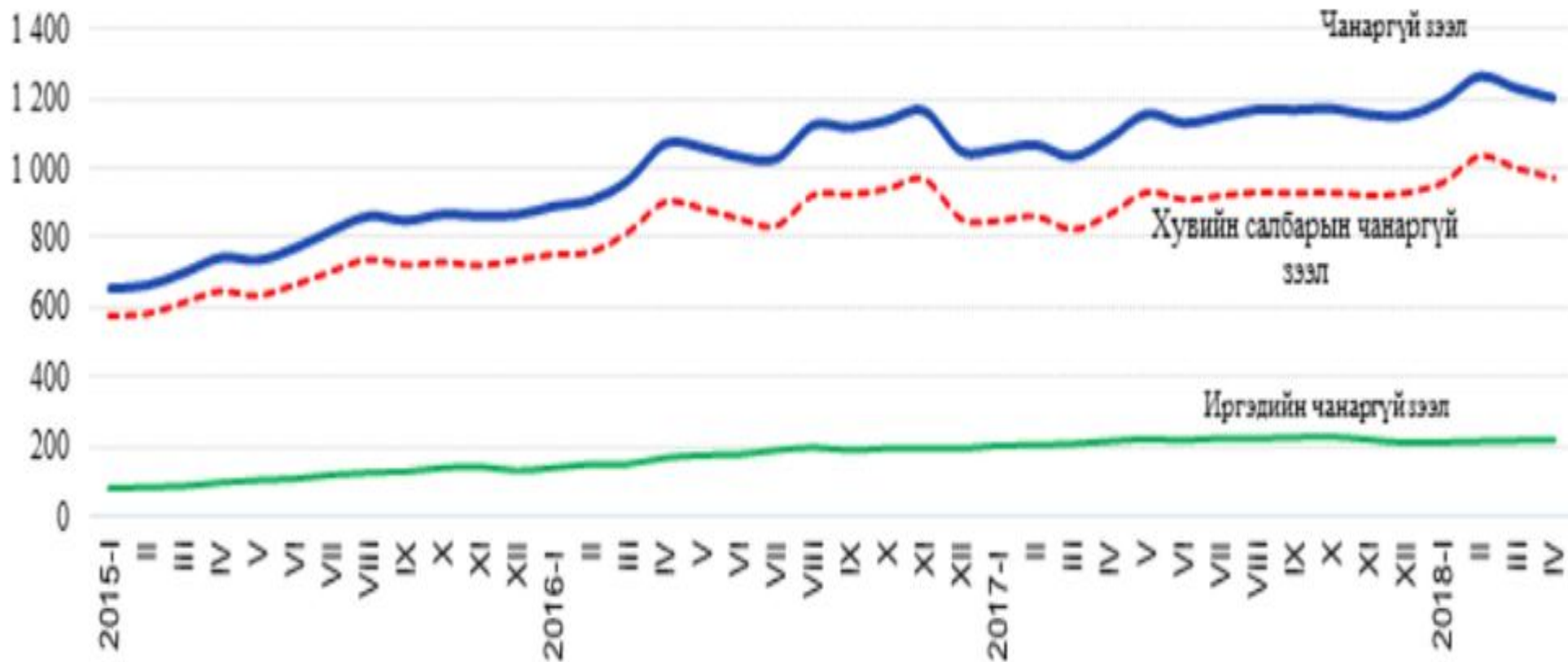
ИЛТГЭЛИЙН ХУРААНГУЙ

Судалгааны зорилго нь орчин үеийн статистик аргуудаар Монгол улсын банк секторын зээлдэгчдийг зээлжих чадварыг үнэлэхэд оршино.

Түлхүүр үг: Өгөгдөл боловсруулалт, ложит загвар, дискриминант шинжилгээ

Тайлбар: Манай улсын хувьд энэхүү чиглэлээр өмнө нь судалгаа хийгдэж байгаагүй учир судалгаа хийх үндэслэл болсон. Энэхүү сонгосон аргыг ашигласнаар зээлдэгчийн үнэлэмжийг зөв тодорхойлж ангилал хийснээрээ банкны зээлийн эрсдэлийг бууруулах практик ач холбогдолтой.

УДИРТГАЛ



Тайлбар: Хугацаа хэтэрсэн зээлийн өрийн үлдэгдэл 2018 оны 4 дүгээр сарын эцэст 979.2 тэрбум төгрөг болж, өмнөх сараас 43.3 (4.6%) тэрбум төгрөгөөр өсөж, өмнөх оны мөн үеэс 30.1 (3.0%) тэрбум төгрөгөөр буурлаа. Хугацаа хэтэрсэн зээлийн өрийн үлдэгдэл нийт зээлийн өрийн үлдэгдлийн 6.8 хувийг эзэлж, өмнөх сараас 0.2 пунктээр өсөж, өмнөх оны мөн үеэс 1.1 пунктээр буурсан байна.

ЗЭЭЛЖИХ ЧАДВАРЫГ ТААМАГЛАХ ЗАГВАРУУДЫН ТҮҮХЭН ХӨГЖИЛ БА СТАТИСТИК АРГУУД

Параметр (зээлийн зэрэглэлийн загварууд)

- Шугаман регрессийн загварууд (Linear regression model)
- Пробит болон ложит загвар (Probit models and Logit models)
- Дискриминант шинжилгээ (Discriminant analysis)

Параметр бус (зээлийн зэрэглэлийн загварууд)

- Шийдвэрийн мод (Decision tree буюу DT)
- Хиймэл оюун ухааны сүлжээ (Artificial Neural networks буюу ANN)
- Олон хувьсагчтай регресс (Multivariate adaptive regression splines)
- Математик програмчлалын загвар (Mathematical programming)
- Шинжилгээний шатлалын үйл явц (Analytical hierarchy process)
- К- хамгийн ойрхон хөршүүд (K-nearest neighbours буюу KNN)
- Эксперт систем (Expert system)
- Бэйсийн сүлжээ (Bayesian network буюу BN)
- Тулгуур вектор машин (Support vector machine буюу SVM)
- Амьдрах чадварын шинжилгээ (Survival analysis буюу SA)
- Генетик програмчлалын загварууд (Genetic programming models)
- Опцион үнийн загвар (Option pricing model буюу OPM)
- Ангилалын дүрмүүд (Rule based classification буюу RBC)

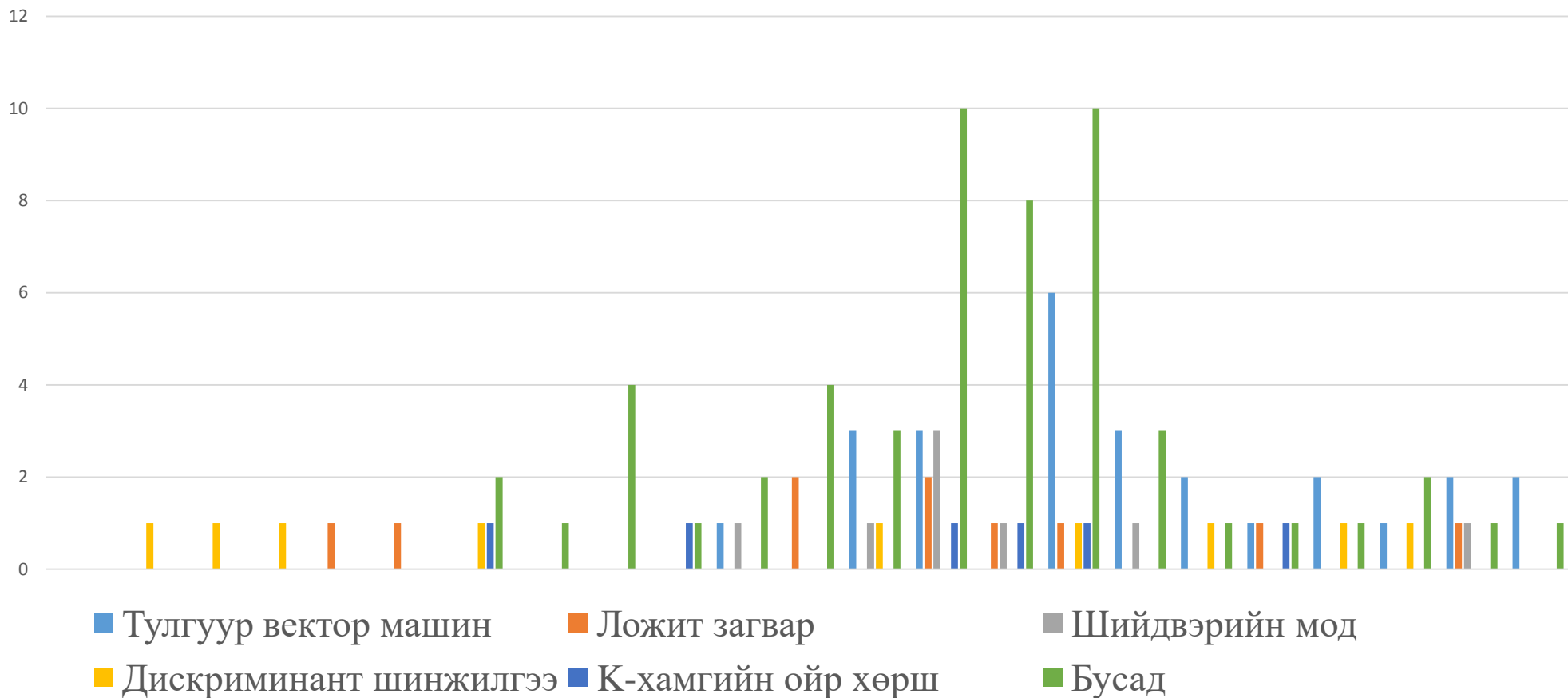
Тайлбар: 2011 онд А. Керамати болон Н. Юусефи (Keramatı & Yousefi, 2011) нар нь өгөгдөл боловсруулах арга (Data mining technique) дээр тулгуурлаж 2000 – 2010 оны хооронд зээлийн зэрэглэл тогтооход ашиглаж буй орчин үеийн статистик аргуудаас таамаглах чадвар, өндөр параметрийн болон параметрийн бус 16 аргыг тогтоосон.

Зээлжих чадварын загваруудад харьцуулалт хийсэн судалгааны ажил

Судлаачийн нэр	Он	Судалгааны ажлын тоо	Зээлжих загварын тоо	Харьцуулалтын хамрах хугацаа
Аббас Керамати, Нилүүфер Юусефи	2011	81	10	1968 – 2010
Шин-Чен Хуанг, Мин-Юу Дэй	2013	24	12	1944 – 2008
Хассен Сэбзевари, Мехти Солеймэни, болон Еаман Нүүрбахш	2008	13	7	1984 – 2005
Миний судалгаа	2017	106	16	1930 - 2017

Тайлбар: Зээлжих чадварыг таамаглах загварын түүхэн хөгжил нь тухайн үеийн бодит хэрэгцээ болон загвар боловсруулахад ашиглагдаж буй статистик, эконометрик аргуудын хөгжилтэй салшгүй холбоотой. Миний судалж мэдсэнээр 1944 – 2016 оны хооронд хийгдсэн 106 гаруй судалгаа байгаа бөгөөд эдгээрт хамгийн түгээмэл 39 статистик болон эконометрик аргуудыг ашигласан байна.

Зээлжих чадварыг боловсруулахад түгээмэл ашиглаж буй аргууд



Тайлбар: Эдгээр загварууд нь гол төлөв 2000, 2004-2010 онуудад илүү их судлагдсан бөгөөд ложит, шийдвэрийн мод, дискриминант, тулгуур вектор, K-тэй хамгийн ойр хөршүүд зэрэг загварууд түлхүү ашиглагдсан байна.

Судалгаанд ашигласан түүвэр ба арга зүй

Дискриминант шинжилгээ

Ложит загвар

Шийдвэрийн мод

К- хамгийн ойр хөрш

Дискриминант шинжилгээ (Discriminant analysis)

Гол ойлголт, нэр томъёо: дискриминант коэффициент, группийн төв, өөрийн утга, тухайн параметрийн багц тус бүрийн язгуур, хамаарал, Вилксын ламбда, ангиллын матриц, хайрцаг-М, түүврийн хэмжээ зэрэг ойлголтуудыг дискриминант шинжилгээ нь өөртөө багтаасан байдаг.

$$\begin{aligned} Z &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \dots \dots \beta_p X_p; S(\beta) \\ &= \frac{\beta^T \mu_1 - \beta^T \mu_2}{\beta^T C \beta} \gg S(\beta) = \frac{\vec{z}_1 - \vec{z}_2}{\text{Variance of } Z \text{ within groups}}; \beta \\ &= C^{-1} (\mu_1 - \mu_2); C = \frac{1}{n_1 + n_2} (n_1 C_1 + n_2 C_2); \Delta^2 = \beta^T ((\mu_1 - \mu_2)) \end{aligned}$$

Тайлбар: Зээлжих зэрэглэлд дискриминант шинжилгээг ашигласан судлаачид 1968 Altman E. I, 1977 Eisenbeis R. A, 1981 Martell & Fitts, 1984 Taffler R. J, Abassi B, 2000 Yobas M. B, Crook J. N, Ross P, 2006 Kumar K, Bhattacharya S, 2009 Abdou H.A

Ложит загвар (Logit model)

Ложит загвар нь $Y_i = f(\beta, X_i) + u_i$ байдлаар тодорхойлогдох ба Y_i нь 0 эсвэл 1 гэсэн утга авна. Өөрөөр хэлбэл тухайн зээл эрсдэлтэй, эсвэл эрсдэлгүй гэсэн утгыг илэрхийлнэ.

$$Z = \log[p(1 - p)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Тайлбар: Зээлжих зэрэглэлд ложит загварыг ашигласан судлаачид 1989 Cox & Snell, 1999 Laitinen E. K, 2005 Tang T. C, Chi L. C, 2007 Sohn S. Y, Kim H. S, 2008 Luo J. h, Lei H. Y, 2009 Liang Y, Xin H

Шийдвэрийн мод (Decision Tree)

Таамаглалын загвар болох Шийдвэрийн мод нь алгоритмын нэг чухал хэлбэр юм. Шийдвэрийн модны арга нь дотроо хэд хэдэн төрөлтэй. Үүнд:

- Interactive Dichotomizer3 (ID3)
- Classification and Regression Trees (CART)
- QUEST algorithm (QUEST)
- Chi-squared
- Automatic Interaction Detector (CHAID)
- Multivariate Adaptive Regression Splines (MARS)

$$R(S) = \frac{1}{n} \sum_i (y_i - h(t_i))^2$$

Тайлбар: Зээлжих зэрэглэлд шийдвэрийн мод аргыг ашигласан судлаачид 1972 David Sparks, 2004 Mues C, Baesens B, Files C. M, Vanthienen J, 2006 Lee T. S, Chiu C. C, Chou Y. C, and Lu C. J, 2007 Zhao H, 2008 Bastos J. A, 2015 Quinlan J. R

К-тэй хамгийн ойр хөршүүд (K- nearest neighbor)

KNN хамгийн ойр хөршүүд регрессийн арга нь хувьсагчуудын алгоритм тооцоолоход ашигладаг. К хамгийн ойр олон хувьсагчтай хөршүүдтэйгээ урвуу зайн жигнэсэн дунджаар тооцно. Богино хугацаанд ашиглагддаг параметр бус арга юм.

$$D(x_i, y) = \sqrt{\sum_k (x_k - y_k)^2}$$

Тайлбар: Зээлжих зэрэглэлд К-тэй хамгийн ойр хөрш аргыг ашигласан судлаачид 1996 Henley & Hand, 2000 Paredes R, Vidal E, 2003 Hand D. J, Vinciotti V, 2007 Islam M. J, Wu Q. M. J, Ahmadi, M, Sid-Ahmed M. A, 2008 Marinakis Y, Marinaki M, Doumpos M, Matsatsinis N, 2009 Li F. C, 2012 Henley W. E, Hand D. J

Зээлжих чадварыг таамаглах загварын боловсруулалт

Хамааран хувьсагч болон үл хамааран хувьсагчийн сонголт



Вариацийн шинжилгээ (ANOVA)

	Wilks' Lambda	F	df1	df2	Sig.
x9	0.889	3.242	1	26	0.083
x11	0.797	6.638	1	26	0.016
x 13	0.894	3.08	1	26	0.091
x21	0.889	3.25	1	26	0.083
x22	0.889	3.25	1	26	0.083
x23	0.9	2.902	1	26	0.1
x26	0.842	4.875	1	26	0.036

Тайлбар: Зээлдэгчийн зээлжих чадварын загваруудад түгээмэл ашиглаж байгаа 40 харьцаанаас (үл хамааран хувьсагчаас) зээлжих чадварт хүчтэй нөлөөлөх магадлалтай үл хамааран хувьсагчийг илрүүлэхийн тулд вариацийн шинжилгээ (ANOVA) ашиглав. Жишээлбэл: Wilk's Lambda – ийн магадлал нь хувьд бүгд 5 хувиас бага гарсан тул эдгээр үзүүлэлт нь зээлдэгчийн зээлжих чадварт хүчтэй нөлөөлдөг байх магадлал өндөртэйг харуулж байна.

Каноник дискриминант функцийн Эйгийн утга, Вилксийн лямбда

Function	Eigen value	% of Variance	Cumulative %	Canonical Correlation	Test of Function (s)	Wilk's Lambda	Chi-square	df	Sig.
1	1.036 ^a	100.0	100.0	.713	1	.491	15.99	7	.025

Тайлбар: Дискриминантын шинжилгээний хувьд ямар ч тохиолдолд цор ганц эйгийн утга байх бөгөөд статистик ач холбогдолтой загварын хувьд эйгийн утга нь 1 – ээс их гардаг. Хүснэгт харвал каноник корреляцын коэффициент нь $r_c = 71.3$ хувь ($> 35\%$ бол ач холбогдолтой), Вилксийн лямбдагийн магадлал нь $p\text{ value} = 0.025$ (< 0.05) гарсан. Иймд тус дискриминантын функц нь статистик ач холбогдол өндөртэй, маш сайн үнэлэгдсэн ба дараах хэлбэртэй гарч байна.

Үргэлжлэл

Дискриминантын шинжилгээ хийх үед зааглалтын цэгийг (cut off point) сонгохдоо дискриминантын функцийн хувьд хязгаарт оршдог. Энэхүү аргад тулгуурлан загвар боловсруулдаг судлаачид гол төлөв групп тус бүрийн төв цэгүүдийн дунджаар зааглалтын цэгийг тооцдог. Өөрөөр хэлбэл $COP = E(\text{centroid1} + \text{centroid2}) = E(E(Z_0) + E(Z_1))$. Энд Z_0 болон Z_1 тус бүр эрсдэлтэй болон эрсдэлгүй компаниудын загварын утга байна. Тиймээс бидний боловсруулсан дискриминант функцийн төв цэгүүд нь тус бүр $E(Z_0) = -1.699$, $E(Z_1) = 0.566$ гарсан тул зааглалтын цэг нь $COP = -0.566$ болно.

Төв цэг 1	Зааглалтын цэг	Төв цэг 2
-1.699	-0.566	0.566

Эх сурвалж: Судлаачийн тооцоолол

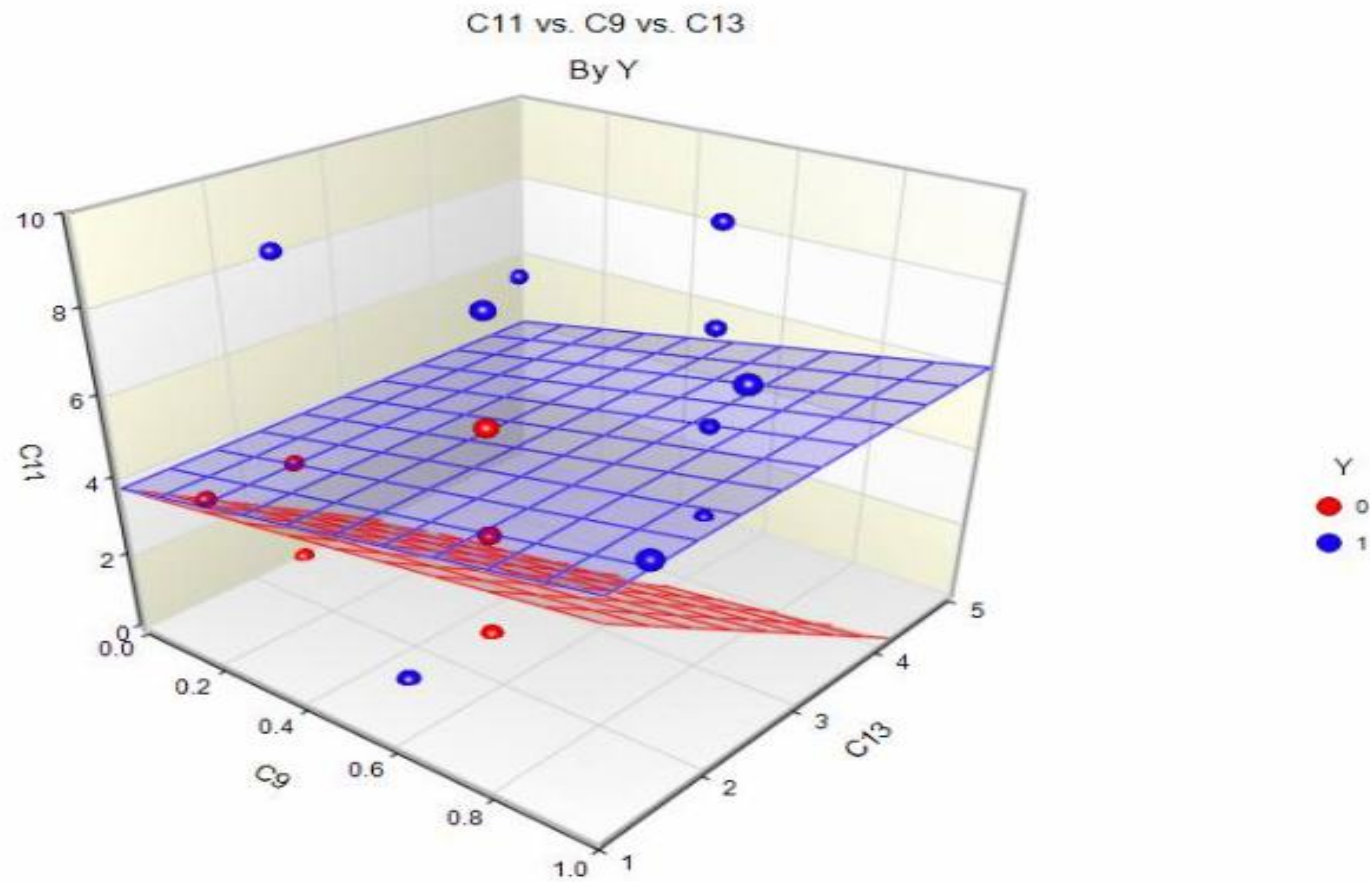
Дискриминант шинжилгээний үр дүн

Үзүүлэлтүүд				
Z		Таамагласан бүлэгийн гишүүдийн тоо		Нийт
		0	1	
Тоо	Эрсдэлгүй	6	1	7
	Эрсдэлтэй	1	20	21
Хувиар	Эрсдэлгүй	85.7	14.3	100.0
	Эрсдэлтэй	4.8	95.2	100.0

$$Z = - 6.405 + 1.359 * X9 + 0.198 * X11 + 0.397 * X13 + 0.295 * X21 + 0.148 * X22 + 0.205 * X23 + 0.134 * X26$$

Тайлбар: Ихэнх судлаачид зааглалтын цэг дээрх аргаар тооцдог хэдий ч групп тус бүрийн түүврийн хэмжээ, дундаж болон төв цэг хоорондын зай зэргээс шалтгаалан үр дүнд сөрөг нөлөө үзүүлэх эрсдэлтэй байх болно. Дискриминант функцийг ангиллын үр дүн эрсдэлгүй нь 85.7 % таамаглаж байгаа бол эрсдэлтэй харилцагчийг 95.2% ба ерөнхий таамаглал нь 92.9 байна.

Дискриминант шинжилгээний Гурав ди (3D) зааглалтын график



Эх сурвалж: Судлаачийн тооцоолол

Ложит загварын үр дүн, Зааглалтын цэг

0.5 үеийн хүлээгдэж буй таамаглал

	Estimated Equation		Total
	Dep = 0	Dep = 1	
$P(\text{Dep} = 1) \leq C$	6	1	7
$P(\text{Dep} = 1) \geq C$	1	20	21
Нийт	7	21	28
Эрсдэлгүй	85.7	14.3	100%
Эрсдэлтэй	4.8	95.2	100%

$$\ln Z / 1 - Z = 8.506 + 1.209 * X9 + 3.03 * X26 - 0.844 * X11 + 0.024 * X22$$

Тайлбар: Үүнээс үзвэл ложит загварын үнэн таамаглах ерөнхий чадвар (жигнэсэн дундаж) нь 92.9 хувь буюу таамаглах чадвар маш сайн байгааг харуулж байна. Тэр дундаа эрсдэлтэй харилцагчийг үнэн таамаглах чадвар 95.2 хувь гарлаа.

Шийдвэрийн мод загварын үр дүн

Z		Таамагласан бүлгийн гишүүдийн тоо		Нийт
		0	1	
Тоо	Эрсдэлгүй	0	7	7
	Эрсдэлтэй	0	21	21
Хувь	Эрсдэлгүй	0	100	100.0
	Эрсдэлтэй	0	75	75

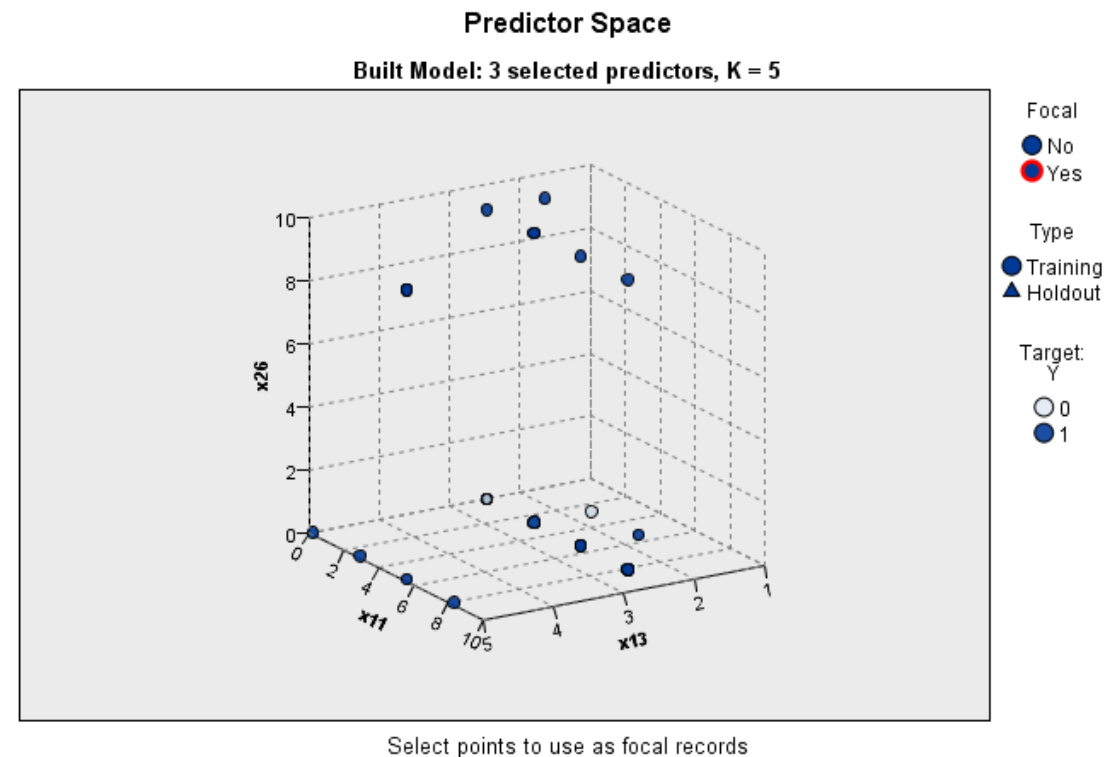
Тайлбар: Таамаглалын загвар болох Шийдвэрийн мод нь алгоритмын нэг чухал хэлбэр юм. Шийдвэрийн модны арга нь дотроо Interactive Dichotomizer3 (ID3), Classification and Regression Trees (CART), QUEST algorithm (QUEST), Chi-squared, Automatic Interaction Detector (CHAID), Multivariate Adaptive Regression Splines (MARS) зэрэг олон төрөлтэй бөгөөд дараах үнэлгээг хийсэн.

K-хамгийн ойр хөрш загварын ангилалын үр дүн

Z		Таамагласан бүлгийн гишүүдийн тоо		Нийт
		0	1	
Тоо	Эрсдэлгүй	7	0	7
	Эрсдэлтэй	2	19	21
Хувь	Эрсдэлгүй	100.0	14.3	100.0
	Эрсдэлтэй	4.8	90.5	100.0

Тайлбар: KNN хамгийн ойр хөршүүд регрессийн арга нь хувьсагчуудын алгоритм тооцоолоход ашигладаг. K хамгийн ойр олон хувьсагчтай хөршүүдтэйгээ урвуу зайн жигнэсэн дунджаар тооцно. Богино хугацаанд ашиглагддаг параметр бус арга юм. Тооцоололоос авч үзэхэд зээлжих зэрэглэлийг тайлбарлах ерөнхий чадвар нь 95.25 гарч байна.

K-хамгийн ойр хөрш Гурав ди (3D) зааглалтын график



This chart is a lower-dimensional projection of the predictor space, which contains a total of 7 predictors.

Эх сурвалж: Судлаачийн тооцоолол

Таамаглах ерөнхий чадварын харьцуулалт

Загвар	Дискриминант шинжилгээ	Ложит загвар	Шийдвэрийн мод	K-хамгийн ойр хөрш
Таамаглах ерөнхий чадвар	92.9	92.9	75	95.25

Эх сурвалж: Судлаачийн тооцоолол

Олон улсын судлаачидтай хийсэн харьцуулалт

Загвар	Бидний судалгаа	Н. Крүүк	К. Кумар	Н. Абдуо	А. Керамити	Н. Крүүк
Дискриминант шинжилгээ	92.9	85.6	90.1	80.9	91.5	83.4
Ложит загвар	92.9	74	82.8	87	85.9	90.2
Шийдвэрийн мод	75	79	71.2	85.2	80.3	73.8
К-хамгийн ойр хөрш	95.25	81.9	89.5	92.8	93.4	80.9
Матемитик дундаж	89.01	80.1	83.4	86.5	87.8	82.1
Геометр дундаж	88.6	80.0	83.0	86.4	87.6	81.9

Эх сурвалж: Судлаачийн тооцоолол

Судалгааны дүгнэлт

- Бид судалгаандаа 1968 - аас 2017 оны хооронд 106 зээлийн зэрэглэлийн судалгаа, 39 төрлийн статистик аргууд 16 төрлийн статистик аргыг хамгийн түгээмэл зээлдэгчийн зээлжих чадварыг үнэлэхэд ашигладаг байна.
- Бид судалгаандаа Дискриминант шинжилгээ, Ложит загвар, Шийдвэрийн мод, К- хамгийн ойр хөрш агуудыг ашигласан.
- Олон улсын судлаачидтай харьцуулалт хийхэд 88.6% геометр дундажтай гарсан нь бидний судалгааны ач холбогдол өндөр байгааг харуулж байна.
- К- хамгийн ойр хөрш арга нь 95.25% буюу хамгийн өндөр таамаглах чадвартай гарсан.

Санал зөвлөмж

- Судалгааг сайжруулахын тулд түүврийн хэмжээг нэмэгдүүлэх хэрэгтэй.
- МонголБанкны зээлийн эрсдэл тооцох зөвлөмжийн 2.3 – т заасны дагуу Зээлдэгчийн хувьд үйл ажиллагааны мөнгөн орлогоос зээлийг төлөх болон төлөхгүй гэсэн 2 хувилбар байдаг тул холбогдох магадлалыг нь дискрет сонголтын (discrete choice) эконометриксийн аргуудаар гарган авч болно. Үүнд пробит (probit) болон ложит (logit) загварын аргууд нь хамгийн өргөн дэлгэр ашиглагддаг.
- Судалгаанд ашиглаж буй зээлдэгчийн мэдээлэл нь нууцын зэрэглэл өндөртэй байдаг тул түүврийн хэмжээг нэмэгдүүлэхэд учир дутагдалтай.

Судалгаанд ашигласан ном зүй

- [1] Boyes, W. J., Hoffman, D. L., & Low, S. A. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics*, 40(1), 3-14.
- [2] M. B. Yobas, J. N. Crook, O. Ross (2000). Credit scoring using neural and evolutionary techniques. *IMA Journal of Mathematics Applied in Business and Industry*, 11, 111 – 125
- [3] Abbas. Keramati, Niloofar. Yousefi (2011). A proposed Classification of Data Mining Techniques in Credit Scoring. *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia*
- [4] Abbas. Keramati, Niloofar. Yousefi (2011). A proposed Classification of Data Mining Techniques in Credit Scoring. *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia*
- [5] Shin-Chen. Huang, Min-Yuh Day. (2013). A comparative study of data mining techniques for credit scoring in banking. *Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on*
- [7] Hassan Sabzevari, Mehdi Soleymani and Eaman Noorbakhsh. (2008). Credit scoring with data mining techniques.
- [7] Altman, E. I., 1968, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance*, 23(4), 589–609
- [8] Lee, G., Sung, T.K., Chang, N., 1999. Dynamics of modeling in data mining: interpretive approach to bankruptcy prediction. *J. Manage. Inform. Systems* 16, 63–85.
- [9] Martell, T.F., Fitts, R.L., 1981. A quadratic discriminant analysis of bank credit card user characteristics. *J. Econ. Bus.* 33, 153–159

Үргэлжлэл

- [10] Cox, D.R., Snell, E.J., 1989. Analysis of Binary Data, Chapman & Hall, London
- [11] Hosmer, D.W., Lemeshow, S., 1989. Applied Logistic Regression, Wiley, New York.
- [12] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees. Wadsworth.
- [13] Davis, R. H., Edelman, G. B., & Gamberman, A. J. (1992). Machine learning algorithms for credit card app
- [14] Hand, D. J., & Vinciotti, V. (2003). Choosing k for two-class nearest neighbour classifiers with unbalanced classes. Pattern Recognition Letters, 24, 1555–1562.
- [15] Henley, W. E. (1995). Statistical aspects of credit scoring. PhD dissertation, The Open University, Milton Keynes, UK.
- [16] Henley, W. E., & Hand, D. J. (1996). A k-nearest neighbour classifier for assessing consumer risk. Statistician, 44, 77–95.
- [17] Islam, M. J., Wu, Q. M. J., Ahmadi, M., & Sid-ahmed, M. A. (2007). Investigating the performance of na Bayes classifiers and k-nearest neighbor classifiers. In International conference on convergence information technology, November 21–23, 2007 (pp. 1541–1546)
- [18] Zhang, Q., Wei, X. And Xu, J., 2003. Global exponential stability of Hopfield neural networks with continuously distributed delays, Physics Letters A, 315 (6), pp