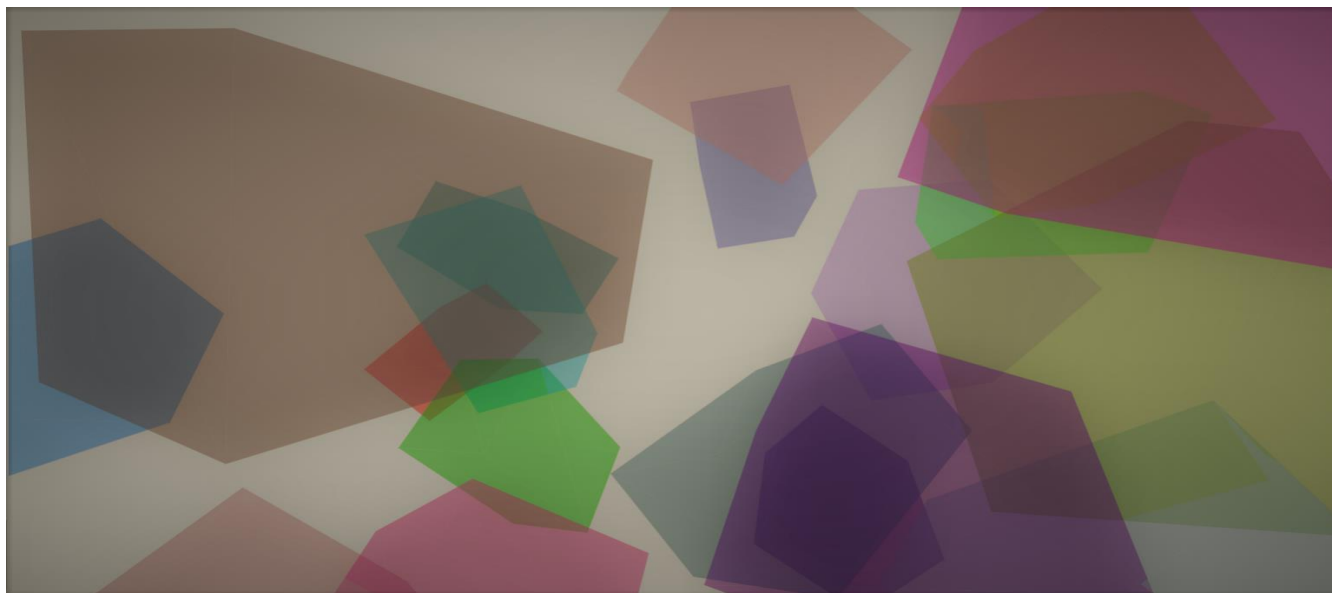




ВАЛИДИРАНЕ НА МОДЕЛИ ЗА СКОРИНГ И РЕЙТИНГ ЧРЕЗ НЕВРОННА МРЕЖА

д-р Анатолий Антонов, д-р Венцислав Николов
28 декември 2016

Еврориск Системи ООД
Ул. Генерал Киселов 31
9002 Варна, България
Тел. +359 52 612 367
Факс +359 52 612 371
info@eurorisksystems.com
www.eurorisksystems.com

Резюме: В статията се представя валидиране и калибриране на модели за определяне на кредитен скоринг и рейтинг със статистически методи чрез съпоставка на резултатите от модела и алтернативен модел основан на невронна мрежа и изчисление на различни статистически параметри. Показан е прототип на софтуерна система за анализ и оценки с изчисление на грешка, стандартно отклонение, корелация, *cumulative accuracy profile*, както и натрупване и разглеждане на исторически статистики от загуби от неизпълнения.

Ключови думи: кредитен рейтинг, скоринг, валидиране, анализ, калибриране, невронни мрежи.

Валидирането на моделите, които включват подбрани обективни количествени фактори, субективни качествени фактори, а както и калибрирането на тегловни коефициенти, степени на насищане, задължителни условия (К.О. проверки) и в крайна сметка причисляване към скоринг/рейтинг-нива е итеративен процес и е на отговорност на банковата институция.

Производителят и внедрителят ще участват при подготовката и провеждането на валидирането пред надзорните органи чрез анализи, подготовка на примери, участие в дискусиите по валидирането и калибриране с помощта на статистически методи и регресионни процедури, напр. чрез невронни мрежи и изчисление на CAP (*Cumulative Accuracy Profile*). В литературата съществуват редица анализи и студии по валидиране на моделите (виж Google: Basel III rating system validation), които могат да се използват за удовлетворяване на изисквания на надзорния орган.

В този материал се демонстрират някои принципи и се показват примери за валидиране чрез невронни мрежи. Еврориск Системи ООД разполага с богат опит по използване на невронни мрежи в различни финансови области като клъстеризация и предсказване на фактори, разпознаване на разпределения, имплицитен рейтинг и многофакторни модели, които са внедрени в различни модули на финансови институции.

Виж:

http://www.eurorisksystems.com/documents/Mathematical_Modelling_Methods_For_Time_Series.pdf

1. Задачата за валидиране

Дадено:

Множество от кредитоискатели с данни за тях (възраст, спестявания, приходи, имущество и т.н.), като за всеки кредитоискател е даден кредитен скоринг/рейтинг, определени в миналото чрез обследваната рейтингова система.

Цел:

Построяване на алтернативен валидиращ математически модел чрез известните данни за кредитоискатели и техните скоринги/рейтинги, така че:

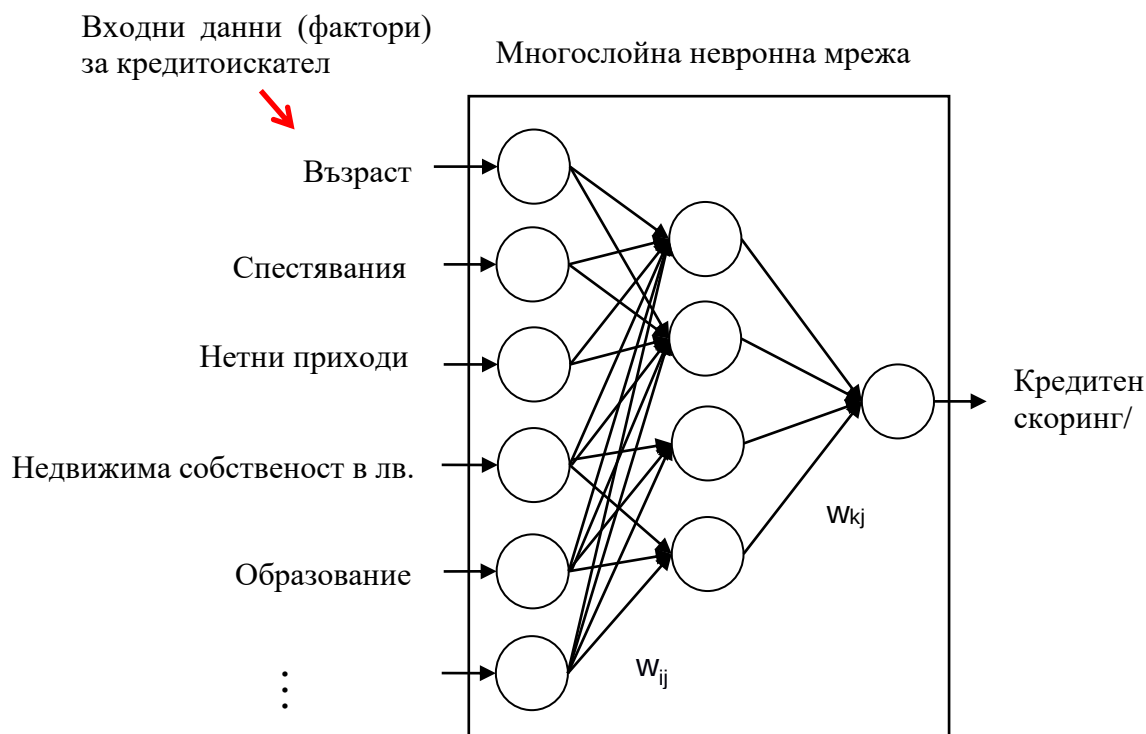
- при нови входни данни за кредитоискател или при нов кредитоискател с неизвестен скоринг/рейтинг, моделът да може да определи стойността на скоринг/рейтинга и да го сравни с резултата от рейтинговата система.
- след калибрация (обучение) на математическия модел да е възможна оценката на скоринга/рейтинга на скорирания от рейтинговата система кредитоискатели срещу математическия модел с цел валидация и определяне на статистики като допустима грешка, стандартно отклонение, корелация, CAP и др.

2. Невронна мрежа

Реализираният математически модел е базиран на многослойна невронна мрежа, която работи с преобразувани по подходящ начин и предварително обработени символни и числови данни (фактори) на кредитоискателите. Слоевете са както следва:

- Входен слой: броят на невроните = броя на входните фактори, напр. 24
- Скрит слой: броят на невроните се определя вътрешно по правила, напр. 10
- Изходен слой: броят на невроните се свързва с броя на резултатите, напр. 2 неврона за скоринг и за рейтинг.

Структурата на мрежата е показана на фиг. 1. Невроните са свързани с тегловни връзки W_{ij} и W_{kj} , които се настройват автоматично в процеса на обучение на мрежата. По този начин мрежата реализира йерархичен скоринг модел на две нива. Невроните се моделират чрез полиномна функция от изходите на невроните-предшественици и вътрешна нелинейна предавателна функция към изхода на неврона.

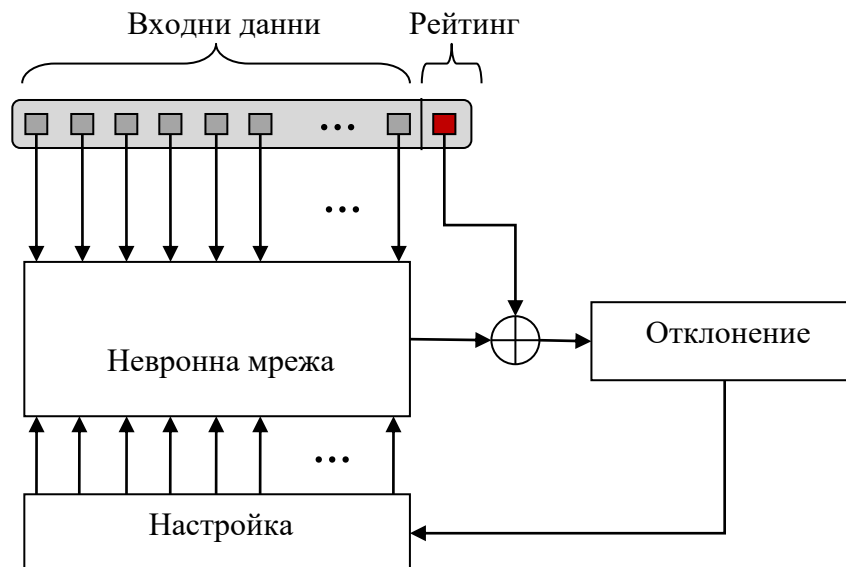


Фиг. 1 Структура на невронната мрежа

Работата на мрежата се основава на два етапа:

1) Първи етап: обучение на мрежата

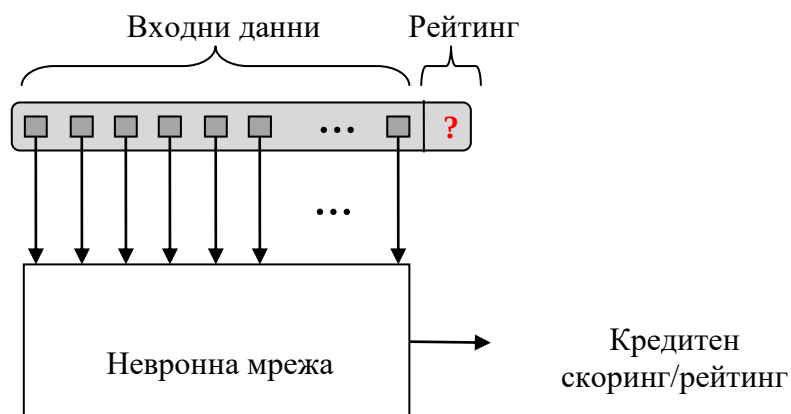
Множеството от входни данни (фактори) за всеки кредитоискател се асоциират към предварително определените скоринги/рейтинги и чрез тях се настройва вътрешната структура на невронната мрежа, така че тя да "научи" зависимостите между входните фактори и изходните резултати чрез настройка на тегловните връзки. Този етап е показан на фиг. 2., обучението се извършва автоматично чрез множество епохи (последователни сесии на обучение).



Фиг. 2 Обучение на невронната мрежа

2) Втори етап: използване на мрежата

Подават се нови входни данни на предварително обучената невронна мрежа, която генерира стойността на неизвестния кредитен скоринг/рейтинг.



Фиг. 3 Определяне на кредитен скоринг/рейтинг чрез обучената невронна мрежа

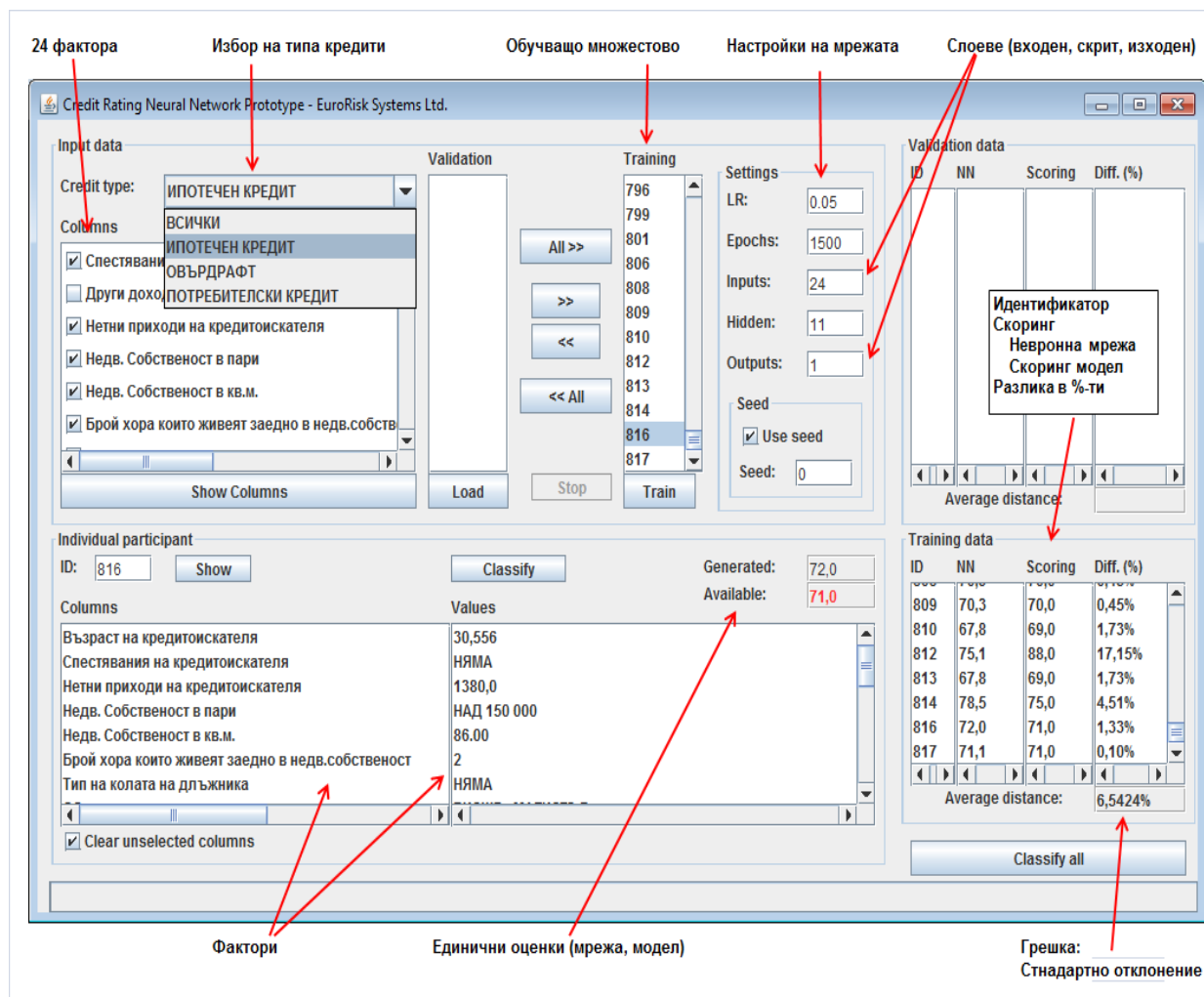
3. Прототипен модул

Работата на невронната мрежа ще бъде онагледена чрез прототип, който работи с реални исторически данни и реализира задачата за валидиране от т. 1 по-горе.

Множество от 24 входни данни (фактори) за 821 кредитоискатели, групирани в 3 групи по типа кредити (ипотечен, потребителски или овърдрафт) се прочитат от невронната мрежа

(фиг. 4.). Избира се типът на кредитите и след настройка мрежата се пуска за обучение с всички кредитоискатели в групата. След това е възможно извършването на единични оценки или на оценки за всички кредитоискатели.

Принципът за валидиране тук се базира на съпоставка на всеки единичен скоринг от рейтинговия модел към всеки единичен скоринг от обучената невронна мрежа, чиито акумулирано знание съответства на опита от всички скорирания по време на обучението.



24 фактора

Избор на типа кредити

Обучаващо множество

Настройки на мрежата

Слоеве (входен, скрит, изходен)

Фактори

Единични оценки (мрежа, модел)

Грешка: Стандартно отклонение

Фиг. 4 Прототип на невронната мрежа: обучение и валидиране за ипотечни кредити

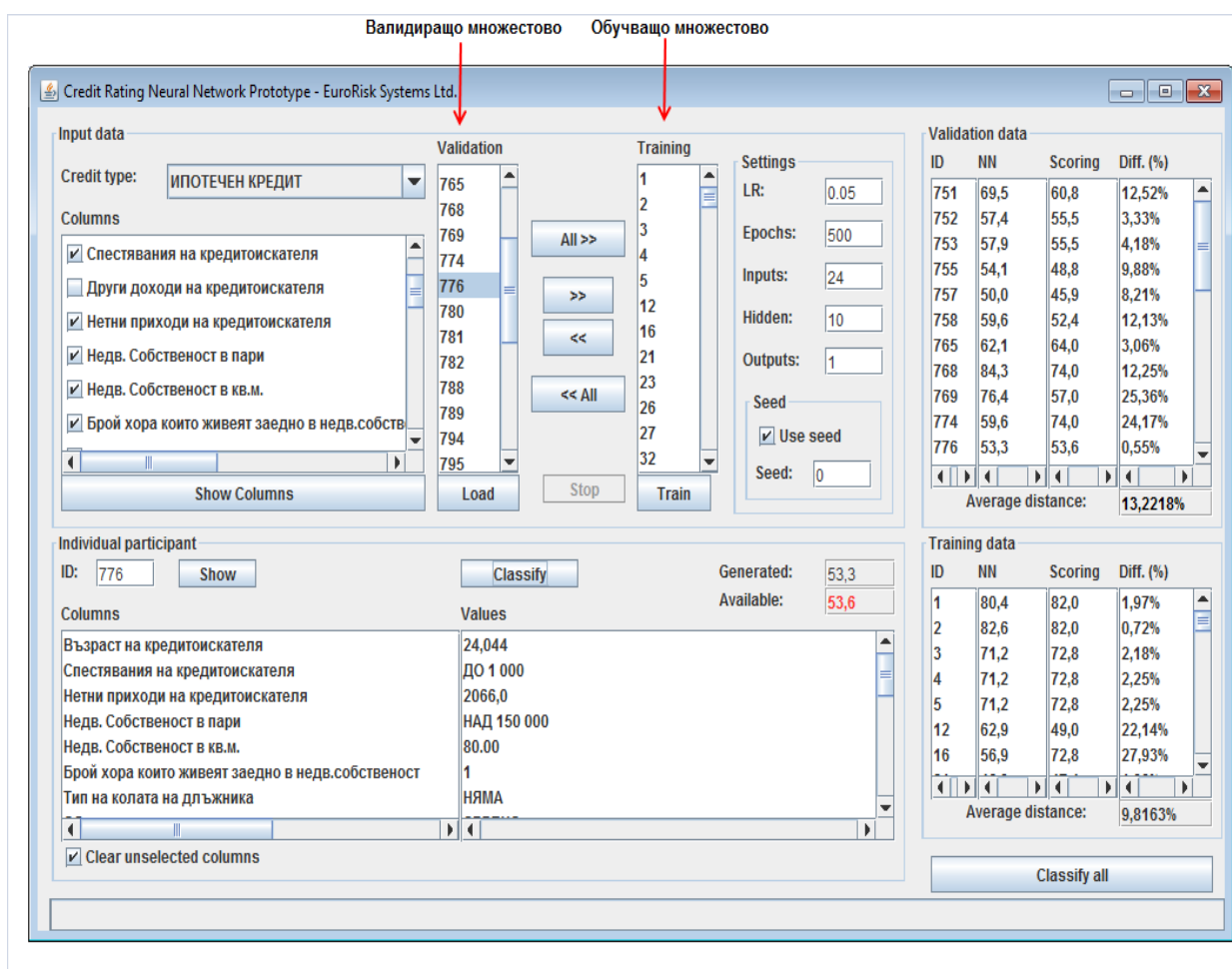
Резултатът от съпоставката се обработва статистически и може да се използва за оценки и валидация на скоринговия/рейтинговия модел, напр. грешка, стандартно отклонение, корелация и CAP. Статистиките показват доколко правилата за скорирание в скоринговия модел са консистентни относно обобщаване и съответстват на обща стратегия или се базират на много изключения и частни решения.

Номер	Фактор	Пример: кредит	Пример: ипотечен	Пример: овъдрафт
1	Възраст		51,62	47,38
2	Спестявания		1 000 - 10 000	НЯМА
3	Нетни приходи		1434	0
4	Недв. Собственост в пари		НАД 150 000	НАД 150 000
5	Недв. Собственост в кв.м.		72	200
6	Брой хора които живеят заедно в недв. собственост		3	1
7	Тип на колата		НАД 3 ГОДИНИ	НАД 3 ГОДИНИ
8	Образование		СРЕДНО	ВИСШЕ - МАГИСТЪР
9	Семеен статус		ЖЕНЕН/ОМЪЖЕНА	ЖЕНЕН/ОМЪЖЕНА
10	В какъв тип предприятие работи		АД	ЕАД
11	Сектор в който работи		ДРУГИ	ДРУГИ
12	Служител в:		МЕСТНА КОМПАНИЯ	МЕСТНА КОМПАНИЯ
13	Източник на приходите		ПОСТОЯНЕН ТД	ПОСТОЯНЕН ТД
14	Общ стаж		25	23
15	Тип на служителя		МЕСТНО ЛИЦЕ	МЕСТНО ЛИЦЕ
16	Област на кредитополучателя		СОФИЯ	ПЛОВДИВ
17	Присвояване към целеви групи на банката		НЕ	ЦЕЛЕВА ГРУПА I
18	Тип на кредита - ипотечен, потребителски, овъдрафт		ИПОТЕЧЕН КРЕДИТ	ОВЪРДРАФТ
19	Размер на кредита		61360	1500
20	Лихва		8,15	10,95
21	Валута на кредита		EUR	BGN
22	Срок на погасяване на кредита в мес.		216	24
23	Амортизационна схема		АНЮИТЕТЕН ПЛАН	
24	Метод за доказване на доходите		ДОКАЗАН ДОХОД	
	Скоринг		72,80	58,40

Фиг. 5 Фактори за скоринг/рейтинг и два примера:

Таблицата на фиг. 5 показва списъка от основни фактори и 2 примера за 2-ма кредитоискателя. Параметрите на кредитоискателите се обработват числено преди да се подадат на невронната мрежа. На последния ред се вижда скорингът получен от скоринговия модел. Скорингът заема стойности между 0% и 100%. Данните и скорингите за всички кредитоискатели от избраната група служат за обучение на мрежата.

Прототипът от фиг. 6 реализира обучение само с част (обучаващо множество) от данните и скорингите на кредитоискателите, останалата част (валидиращо множество) се явява непозната за невронната мрежа. Скорирането на валидиращото множество при този вариант показва доколко мрежата може да се справя с непознати данни и резултати скорирани със скоринговия модел, т.е. каква е мощността на дискриминантната функция на скоринговия модел.



Валидиращо множество Обучаващо множество

Credit Rating Neural Network Prototype - EuroRisk Systems Ltd.

Input data: Credit type: ИПОТЕЧЕН КРЕДИТ

Columns: ☒ Спестявания на кредитоискателя, ☐ Други доходи на кредитоискателя, ☒ Нетни приходи на кредитоискателя, ☒ Недв. Собственост в пари, ☒ Недв. Собственост в кв.м., ☒ Брой хора които живеят заедно в недв.собств.

Validation: 765, 768, 769, 774, 776, 780, 781, 782, 788, 789, 794, 795

Training: 1, 2, 3, 4, 5, 12, 16, 21, 23, 26, 27, 32

Settings: LR: 0.05, Epochs: 500, Inputs: 24, Hidden: 10, Outputs: 1, Seed: 0, Use seed: ☒

Individual participant: ID: 776, Show, Classify, Generated: 53,3, Available: 53,6

Columns: Възраст на кредитоискателя (24,044), Спестявания на кредитоискателя (ДО 1 000), Нетни приходи на кредитоискателя (2066,0), Недв. Собственост в пари (НАД 150 000), Недв. Собственост в кв.м. (80.00), Брой хора които живеят заедно в недв.собственост (1), Тип на колата на длъжника (НЯМА)

Validation data:

ID	NN	Scoring	Diff. (%)
751	69,5	60,8	12,52%
752	57,4	55,5	3,33%
753	57,9	55,5	4,18%
755	54,1	48,8	9,88%
757	50,0	45,9	8,21%
758	59,6	52,4	12,13%
765	62,1	64,0	3,06%
768	84,3	74,0	12,25%
769	76,4	57,0	25,36%
774	59,6	74,0	24,17%
776	53,3	53,6	0,55%

Average distance: 13,2218%

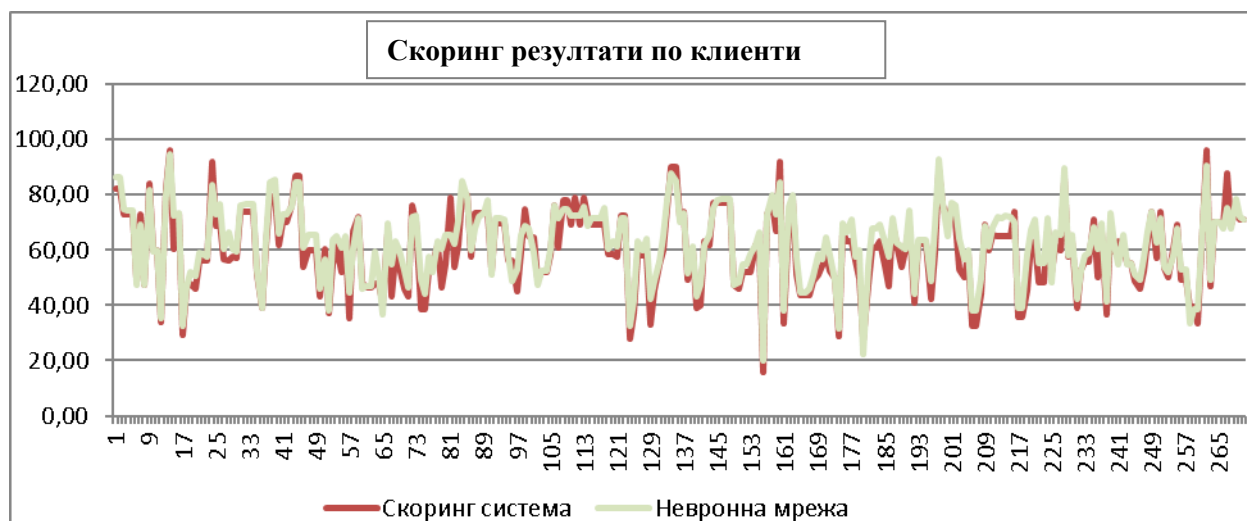
Training data:

ID	NN	Scoring	Diff. (%)
1	80,4	82,0	1,97%
2	82,6	82,0	0,72%
3	71,2	72,8	2,18%
4	71,2	72,8	2,25%
5	71,2	72,8	2,25%
12	62,9	49,0	22,14%
16	56,9	72,8	27,93%

Average distance: 9,8163%

Classify all

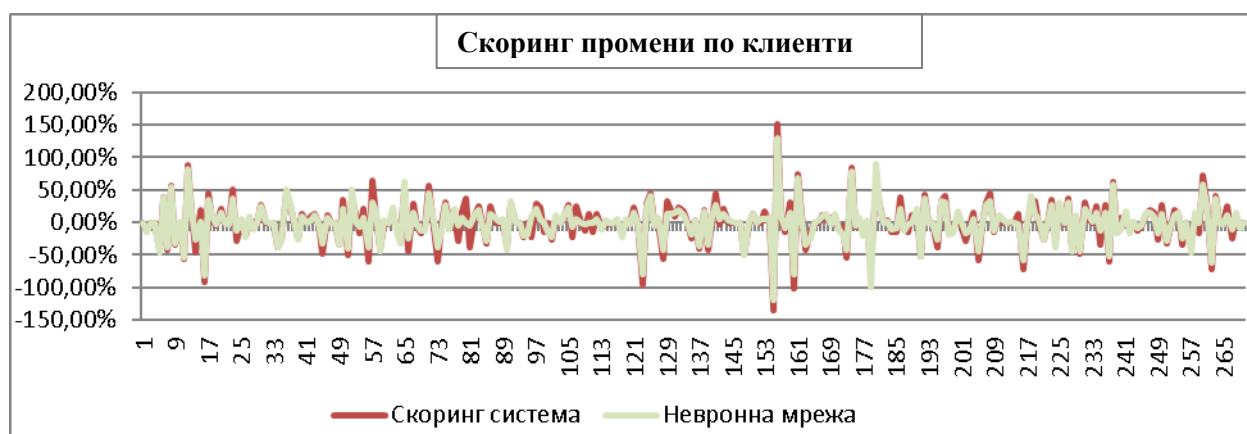
Фиг. 6 Прототип на невронната мрежа: обучение и скорирание на нови ипотечни кредити



Фиг. 7 Скоринг система срещу невронна мрежа

Съпоставката на скорингите от скоринг модела срещу скорингите от обучената невронна мрежа е представена графично на фиг. 7 за групата на ипотечните кредити. Наблюдава се определена "уравновесеност и балансираност" на мрежата, която не реагира на всички дребни промени от скоринг модела. Това поведение е очаквано, тъй като скорингите на мрежата се основават на интегрирани "знания" от обучителната фаза.

Относителните промени са представени на фиг. 8, където се наблюдава подобно поведение.



Фиг. 8 Скоринг система срещу невронна мрежа

Съпоставката на скорингите за първите 29 кредитоискатели за ипотечни кредити (общо 271 кредитоискателя) е представена в таблицата на фиг. 9. Статистиката показва висока положителна корелация = 0,9431 между поведението на скоринговия модел и невронната мрежа. Стандартното отклонение на разликата в двата скоринга е 5,27%, което означава

добро съответствие и стабилност на скоринговия модел. Тези статистически резултати могат да доведат до заключението, че скоринговият модел е стабилен и в крайна сметка до неговото валидиране.

Забележки:

Валидирането на скоринговите и рейтингови модели може да се извършва и чрез историческа статистика от загуби от неизпълнения, което изисква съхранение, синхронизация и натрупване на такива данни от по дълга предистория. Тогава рейтинговата система може да се валидира и калибрира по историческите данни.

В тук описаната методика валидирането е авторегресивно, т.е. множеството от скоринги на самия скорингов модел определя скоринг "знанията".

Определянето на значимите фактори може да се извърши чрез изключване на отделните фактори и оценка на промяната на CAP (Cumulative Accuracy Profile), което е съпоставимо с определянето на значимостта на изключения фактор.

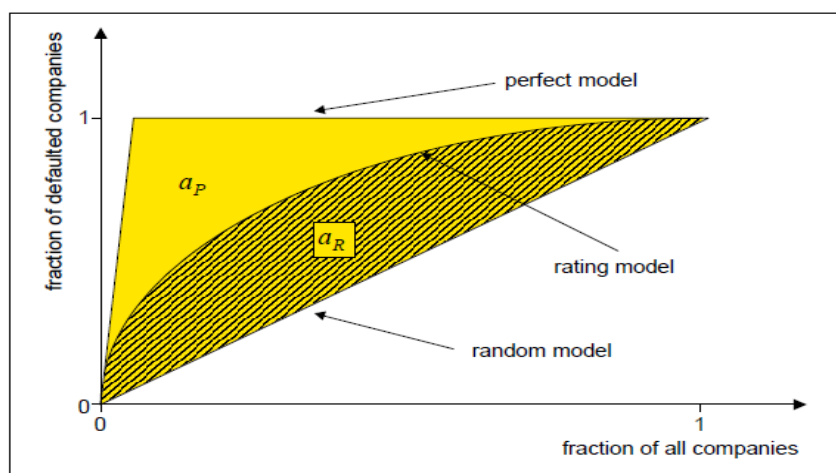
Обучение		Станд. Откл.			5,27%	
Средна стойност		62,67	60,26	Корелация		0,9431
Скоринг резултати				Логаритм. промяна		
Номер на клиент	Невронна мрежа	Скоринг система	Разлика в %-ти	Невронна мрежа	Скоринг система	
1	86,40	82,00	5,13%	-0,23%	0,00%	
2	86,20	82,00	4,85%	-14,32%	-11,90%	
3	74,70	72,80	2,52%	-0,27%	0,00%	
4	74,50	72,80	2,26%	0,00%	0,00%	
5	74,50	72,80	2,26%	-45,43%	-39,59%	
12	47,30	49,00	3,64%	38,19%	39,59%	
16	69,30	72,80	5,04%	-37,56%	-42,91%	
21	47,60	47,40	0,40%	54,14%	57,22%	
23	81,80	84,00	2,68%	-32,00%	-33,65%	
26	59,40	60,00	1,01%	1,17%	0,00%	
27	60,10	60,00	0,25%	-53,21%	-56,80%	
32	35,30	34,00	3,61%	82,56%	88,04%	

33	80,60	82,00	1,74%	16,02%	15,76%
39	94,60	96,00	1,52%	-26,75%	-46,34%
41	72,40	60,40	16,57%	1,51%	18,95%
46	73,50	73,00	0,73%	-81,60%	-90,95%
48	32,50	29,40	9,64%	35,39%	45,63%
50	46,30	46,40	0,26%	11,61%	2,97%
53	52,00	47,80	8,15%	-2,73%	-4,27%
54	50,60	45,80	9,40%	15,19%	20,11%
57	58,90	56,00	4,86%	-0,34%	0,00%
58	58,70	56,00	4,61%	-2,24%	0,00%
59	57,40	56,00	2,42%	37,36%	49,64%
62	83,40	92,00	10,31%	-13,46%	-29,06%
64	72,90	68,80	5,56%	4,82%	1,73%
68	76,50	70,00	8,54%	-23,30%	-20,72%
73	60,60	56,90	6,14%	9,29%	-1,24%
76	66,50	56,20	15,53%	-9,79%	2,11%
99	60,30	57,40	4,81%	-1,50%	-0,70%
-----	-----	-----	-----	-----	-----
271	ипотечни	кредити	-----	-----	-----

Фиг. 9 Валидиране на ипотечни кредити: скоринг система срещу невронна мрежа

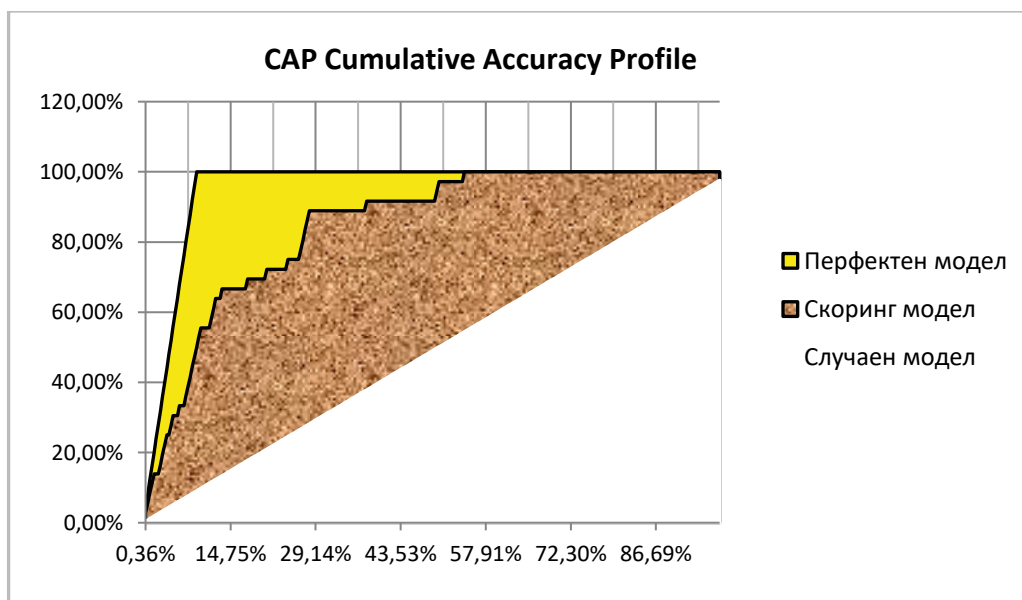
Следващата оценка е базирана на построение за CAP, фиг. 10., където се извършва съпоставка на скоринговия модел към идеалния (перфектен) модел. Площта под кривата на скоринговия модел се съотнася към площта под идеалния модел, т.е. $CAP = a_p / (a_p + a_r)$. CAP се движи между 0% и 100%, качествените скоринг модели имат висок CAP.

Нивото от скоринг = 48% е параметър за скорингите от скоринговия модел и от невронната мрежа и служи за асоцииране на кредитоискателите към категорията "неизпълнение", ако скорингът е по-малък от 48%. Така 36 кредитоискатели са в категорията "неизпълнение" при невронната мрежа, 54 са в "неизпълнение" при скоринговия модел, а 25 са в "неизпълнение" едновременно при скоринговия модел и при невронната мрежа.



Фиг. 10 CAP (Cumulative Accuracy Profile)

Построението на CAP с данните от скориянето чрез скоринговия модел и чрез невронната мрежа е представено на фиг. 11. От графичното представяне може да се отчете CAP = 80%. При това скорингът от невронната мрежа се приема да съответства на очакванията (вероятности) за бъдещи неизпълнения или изпълнения.



Фиг. 11 Построение на CAP след скориянето