

Комплексний скоринговий підхід до аналізу характеристик позичальників кредитів

СТУДЕНТ: ПУДЛО ІЛЛЯ ВІТАЛІЙОВИЧ

ГРУПА: КА-41М

НАУКОВИЙ КЕРІВНИК:

Д.Т.Н., ПРОФ. БІДЮК ПЕТРО ІВАНОВИЧ

Об'єкт, предмет і мета дослідження



Об'єкт: статистичні дані стосовно клієнтів банку – позичальників кредиту.

Предмет: моделі і методи аналізу даних позичальників кредитів, критеріальна база для аналізу якості класифікації клієнтів.

Мета: підвищення якості класифікації клієнтів за допомогою нових розроблених моделей.

Актуальність роботи



Кредитування є найбільш прибутковим і одночасно ризикованим видом банківської діяльності.

Висока частка проблемного кредитного портфелю може спричинити банкрутство банку, а через його положення в економіці, до цілого ряду банкрутств, пов'язаних з ним суб'єктів господарювання.

Тому управління кредитним ризиком є необхідною частиною стратегії і тактики розвитку будь-якого банку.

ПОСТАНОВКА ЗАДАЧІ

- Проаналізувати проблему виникнення кредитного ризику та супутні ризики, які виникають в процесі кредитування
- Виконати огляд сучасних методів і підходів до оцінювання характеристик позичальників кредиту.
- Зібрати статистичні дані, необхідні для побудови математичних моделей і виконання обчислювальних експериментів.
- Розробити математичні моделі для оцінювання кредитоспроможності клієнтів (аплікаційну та поведінкову).
- Виконати оцінювання якості побудованих моделей.
- Виконати порівняльний аналіз отриманих результатів моделювання і оцінювання кредитоспроможності клієнтів.

ОСНОВНІ ПІДХОДИ ДО РОЗВ'ЯЗАННЯ ЗАДАЧІ

- ▶ Логістична регресія
- ▶ Класифікаційні дерева
- ▶ Random forests
- ▶ Нейронні мережі
- ▶ Ланцюги Маркова
- ▶ Байєсівські мережі

Чому регресія?

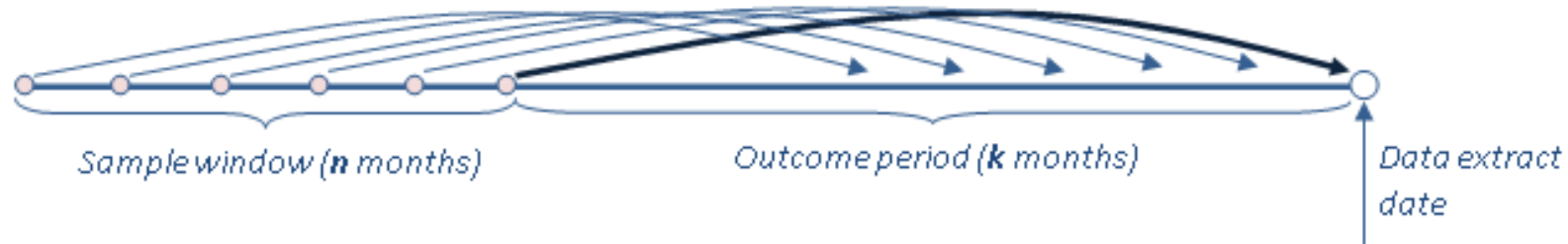
Основний статистичний метод, який в останні 20-30 років асоціюється в банках з кредитним скорингом – це логістична регресія. Поряд з іншими більш витонченими та складнішими моделями вона має такі суттєві переваги:

1. Простота інтерпретації: Вагові коефіцієнти отримані в результаті розрахунків легко переводяться в скорингові бали, які в свою чергу інтерпретуються просто: «краща група клієнтів – більший бал».
2. Поряд з іншими статистичними методами, які можуть давати несуттєвий приріст до предикативної сили моделі, логістична регресія є однією з найбільш стабільних, тобто такою що зберігає свою силу протягом довгого періоду часу і не потребує частої перебудови.

Формування вибірки та вибір цільової змінної

Аплікаційна карта розрахована на прогноз дефолту по кредиту протягом певного фіксованого часу після його видачі.

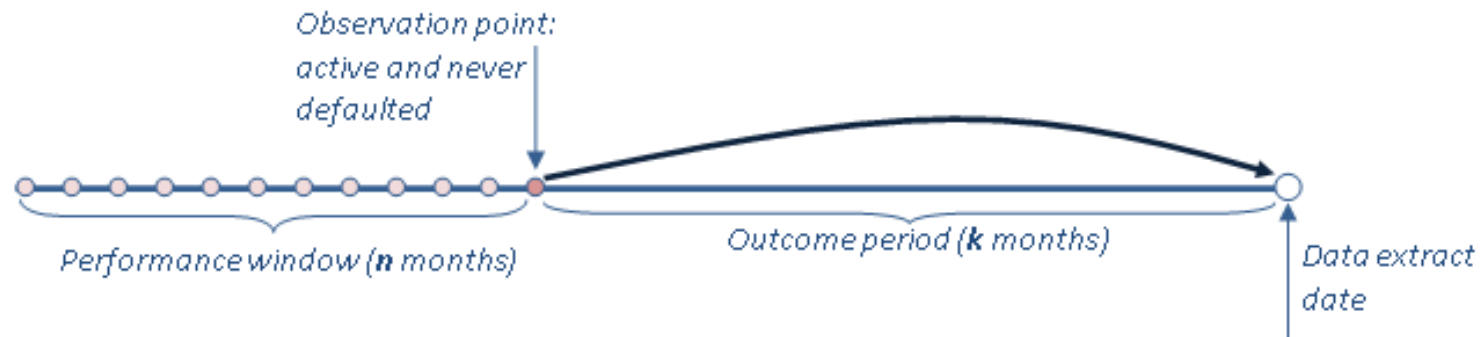
Дана модель оцінює популяцію, що заходить в банк використовуючи лише соціо-демографічні характеристики клієнтів.



Формування вибірки та вибір цільової змінної

Класична **поведінкова карта** призначена для прогнозування дефолту активного недефолтного клієнта.

Дана модель проводить оцінку позичальника з використанням його наявної історії.



Предикативна сила

Предикативна сила характеристики визначається з допомогою показника IV (informative value):

$$IV = \sum_{i=1}^k (\%good_i - \%bad_i) \cdot \ln \left(\frac{\%good_i}{\%bad_i} \right)$$

де k – кількість атрибутів відповідної характеристики, $\%good_i$ та $\%bad_i$ – відповідно відсотки добрих та поганих клієнтів, які попали в атрибут « i ».

Сегментація характеристик

Сегментація проводиться як для неперервних характеристик при розбитті їх на категоріальні так і для категоріальних при наявності атрибутів з малою часткою в вибірці та атрибутів подібних між собою.

Показник WOE (weight of evidence) для кожного атрибуту характеристики розраховується за формулою

$$WOE_i = \ln \left(\frac{\%good_i}{\%bad_i} \right),$$

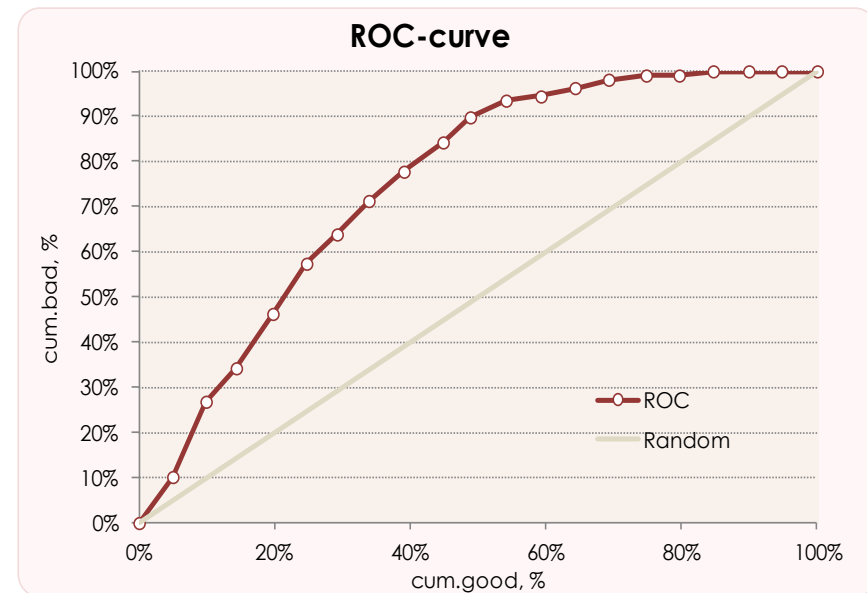
де $\%good_i$ та $\%bad_i$ – відповідно відсотки добрих та поганих клієнтів, які попали в атрибут « i ».

Оцінка якості моделі

ROC-крива

Для кожного скорингового бала (від найменшого по найбільший) визначається пара точок (x, y) : x – частка добрих клієнтів (від всіх добрих) які набрали бал менший або рівний заданого; y – частка поганих клієнтів (від всіх поганих) які набрали бал менший або рівний заданого.

- ▶ Точка (x, y) на кривій інтерпретується так: модель вгадує $(1 - x)\%$ добрих клієнтів;
- ▶ модель вгадує $y\%$ поганих клієнтів;
- ▶ модель не вгадує $x\%$ добрих клієнтів (**помилка I-го роду**);
- ▶ модель не вгадує $(1 - y)\%$ поганих клієнтів (**помилка II-го роду**);



Оцінка якості моделі

Показник GINI визначається як відношення площі між ROC-кривою та $y = x$ (випадковою моделлю) до площі ідеальної моделі (верхній трикутник на графіку).

$$GINI = 2 \int_0^1 (ROC(x) - x) dx = 2AUC - 1$$

K-Статистика: Для кожного значення *Score* рахується значення кумулятивного розподілу «поганих» – $CB(Score)$ та «добрих» – $CG(Score)$. Статистика *KS* розраховується наступним чином:

$$KS = \max_{Score} (CB(Score) - CG(Score))$$

Стабільність моделі

Індекс стабільності популяції *PSI* (population stability index) розраховується за наступною формулою:

$$PSI = \sum_{i=1}^n (\%New_i - \%Old_i) * \ln \left(\frac{\%New_i}{\%Old_i} \right),$$

де $\%New_i$ – відсоток нової популяції в скорбенді i , $\%Old_i$ – відсоток старої популяції в скорбенді i .

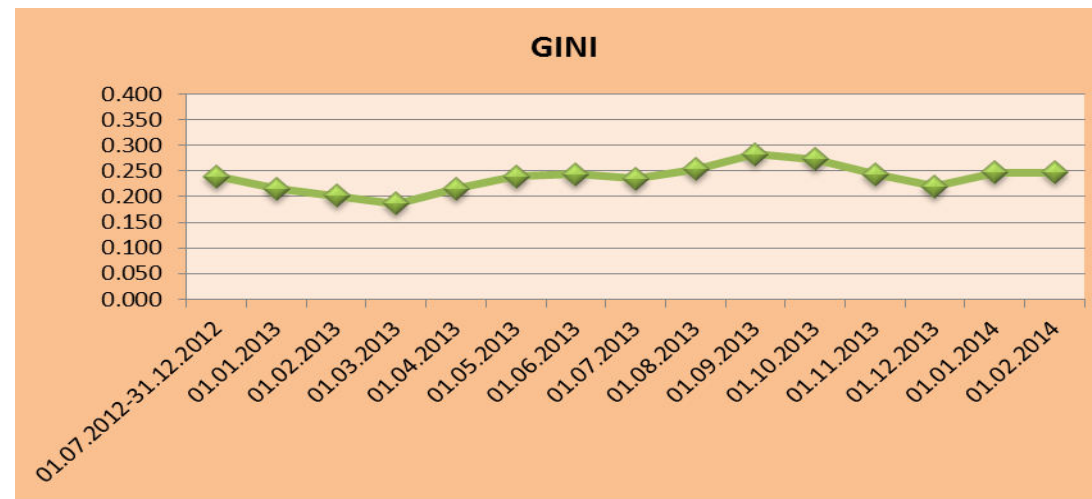
Якщо показник *PSI* приймає значення:

- $\leq 10\%$ – популяція стабільна;
- $10\% - 30\%$ – спостерігається зсув в популяції;
- $> 30\%$ – популяція нестабільна.

Аплікаційна модель для револьверних карток з грейсовим періодом

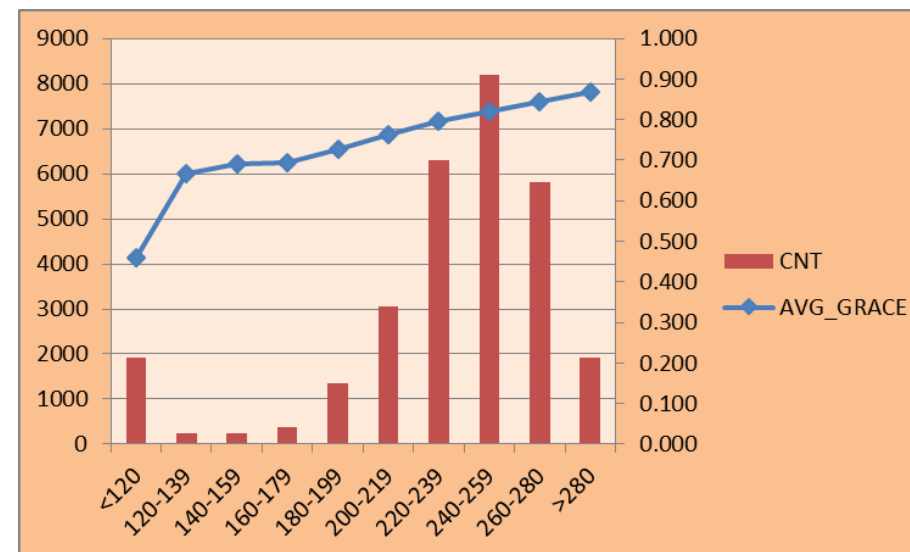
- ▶ *APPBIRTHDATE* - вік клієнта у заявці
- ▶ *APPFAMILYSTATUS* - сімейний стан
- ▶ *APPPOSTCATEG* - займана позиція
- ▶ *APPSEX_APPBIRTHDATE* - схрещений параметр статі та віку
- ▶ *APPSEX_APPEDUCATION* - схрещений параметр статі та освіти
- ▶ *DEP_NUMACT_DEP_SUM* - схрещений параметр кількості активних депозитів та їх суми
- ▶ *APPWPERIODG* - робочий стаж
- ▶ *CHAR10* - кількість місяців до планованої дати закриття
- ▶ *COMPRO* - форма влаштування
- ▶ *CHR4* - кількість активних депозитів за 2 роки
- ▶ *FLAT* - наявність житла

У результаті подальшого присвоєння скорингово балу для кожного значення параметру та подальшої оцінки її якості, було визначено що побудована математична модель з індексом GINI 24%.



Аплікаційна модель для револьверних карток з грейсовим періодом

Згідно зі створеною моделлю, граничні значення скорингового балу можуть коливатися від -92 до 205.



Стабільність моделі

16

Stab.Ind	1/1/2013	2/1/2013	3/1/2013	4/1/2013	5/1/2013	6/1/2013	7/1/2013	8/1/2013	9/1/2013	10/1/2013	11/1/2013	12/1/2013	1/1/2014	2/1/2014
APPBIRTHDATE	4.06%	4.53%	4.99%	2.01%	1.62%	1.64%	1.86%	1.79%	2.78%	3.58%	3.42%	1.99%	4.75%	3.80%
COMPPO	2.93%	2.58%	1.65%	0.84%	0.66%	0.67%	0.87%	0.78%	0.58%	0.74%	1.14%	0.28%	1.39%	1.20%
APPPOSTCATEG	0.02%	0.03%	0.01%	0.01%	0.02%	0.01%	0.00%	0.00%	0.01%	0.00%	0.01%	0.00%	0.01%	0.02%
APPWPERIODG	1.57%	1.59%	1.94%	0.35%	0.62%	0.61%	0.97%	0.72%	1.23%	1.57%	1.46%	0.74%	3.23%	2.34%
APPFAMILYSTATUS	0.18%	0.42%	0.22%	0.13%	0.27%	0.35%	0.64%	0.56%	1.13%	1.16%	1.73%	0.46%	0.55%	0.18%
FLAT	0.02%	0.06%	0.34%	0.00%	0.00%	0.03%	0.01%	0.21%	0.66%	0.72%	0.85%	1.60%	1.78%	0.91%
CHAR10	5.50%	5.95%	8.94%	12.42%	9.87%	6.57%	6.60%	5.62%	4.35%	5.58%	2.74%	8.91%	13.54%	12.57%
CHR4	1.01%	1.14%	1.34%	0.18%	0.01%	0.06%	0.02%	0.15%	0.59%	0.54%	0.30%	0.86%	3.70%	1.94%
APPSEX_APPBIRTHDATE	3.30%	3.15%	2.12%	1.06%	0.68%	0.97%	0.99%	0.91%	1.24%	1.62%	1.81%	1.20%	4.76%	3.61%
APPSEX_APPEDUCATION	1.05%	0.62%	0.73%	0.12%	0.18%	0.12%	0.11%	0.18%	0.28%	0.13%	0.51%	1.04%	1.51%	0.91%
DEP_NUMACT_DEP_SUM	1.02%	0.64%	0.60%	0.11%	0.74%	0.72%	0.55%	1.73%	3.38%	3.29%	6.44%	16.48%	37.89%	34.01%

Shift	1/1/2013	2/1/2013	3/1/2013	4/1/2013	5/1/2013	6/1/2013	7/1/2013	8/1/2013	9/1/2013	10/1/2013	11/1/2013	12/1/2013	1/1/2014	2/1/2014
APPBIRTHDATE	-0.20	-0.19	-0.16	-0.06	-0.12	-0.08	-0.13	-0.12	-0.08	-0.08	-0.09	0.06	0.23	0.24
COMPPO	0.50	-0.01	-0.09	-0.12	-0.03	-0.02	0.01	-0.05	-0.06	0.05	0.02	-0.13	-0.11	0.19
APPPOSTCATEG	-0.04	-0.05	-0.02	-0.02	-0.04	-0.04	-0.02	-0.02	0.03	0.01	0.03	0.00	0.01	0.01
APPWPERIODG	-0.25	-0.25	-0.23	-0.09	-0.18	-0.11	-0.19	-0.15	-0.14	-0.14	-0.13	0.09	0.32	0.32
APPFAMILYSTATUS	-0.03	-0.08	-0.08	-0.06	-0.09	-0.10	-0.14	-0.13	-0.18	-0.18	-0.23	-0.11	-0.06	-0.05
FLAT	0.07	0.12	0.29	-0.01	0.01	-0.09	-0.05	-0.22	-0.39	-0.41	-0.44	-0.60	-0.63	-0.46
CHAR10	1.31	1.20	0.76	0.99	0.25	-0.86	-0.71	-0.14	-0.24	-0.28	-0.81	1.07	0.77	0.95
CHR4	0.65	0.70	0.75	0.29	-0.06	0.10	0.10	-0.03	-0.58	-0.57	-0.08	-0.54	-1.40	-0.98
APPSEX_APPBIRTHDATE	-1.06	-0.85	-0.42	-0.18	-0.35	-0.27	-0.24	-0.46	-0.41	-0.24	-0.42	0.02	0.51	0.52
APPSEX_APPEDUCATION	-0.21	-0.06	0.65	0.31	0.15	0.29	0.36	-0.41	-0.59	-0.20	-0.73	-1.02	-1.13	-0.83
DEP_NUMACT_DEP_SUM	2.88	2.36	2.29	-0.39	-2.24	-2.67	-2.19	-4.29	-6.29	-6.24	-7.16	-11.31	-20.30	-18.31

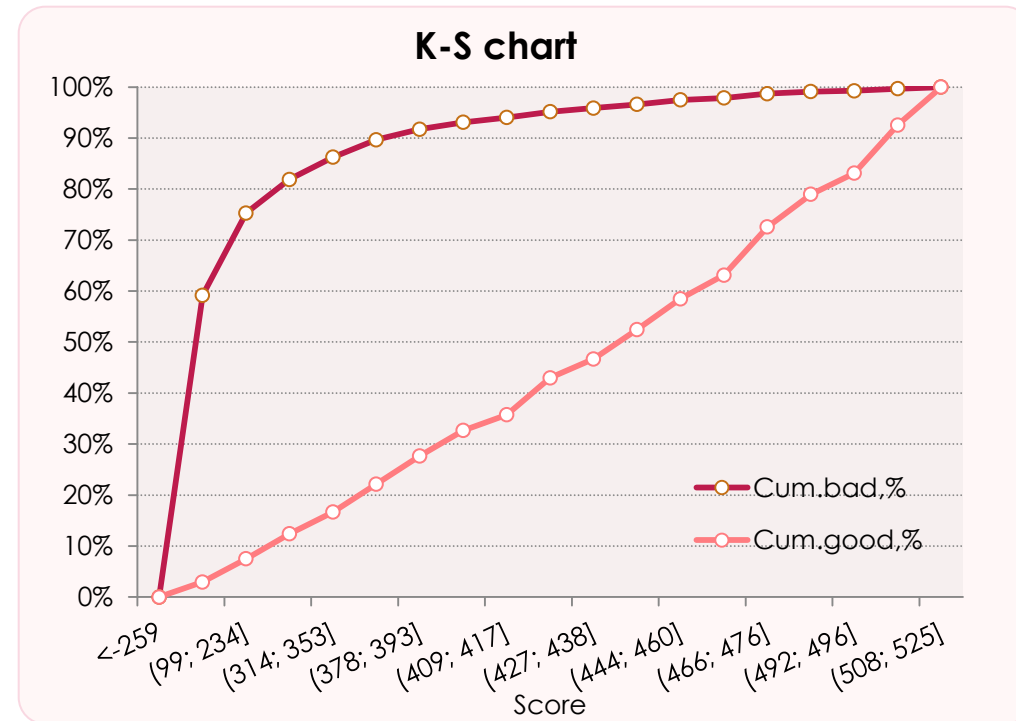
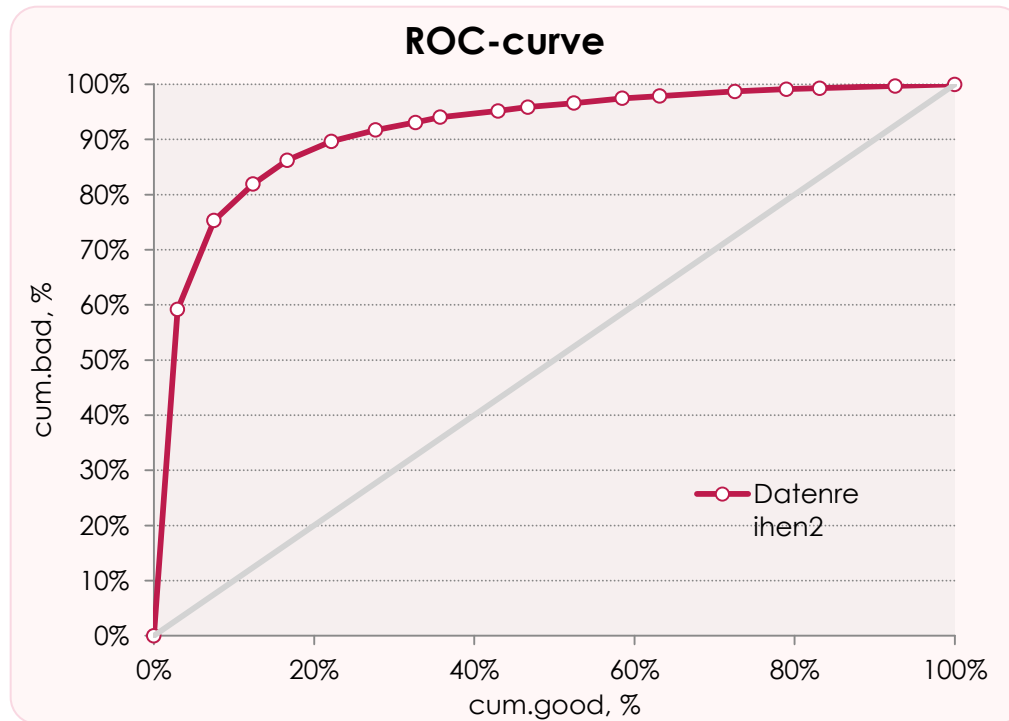
Поведінкова модель для паперових кредитів

- ▶ GRP_CHAR11 - кількість активних кредитів з балансом > 50 грн;
- ▶ GRP_CHAR14 - кількість місяців від відкриття договору до першого бакета 1;
- ▶ GRP_CHAR19_ALL - кількість місяців після першої просрочки >7;
- ▶ GRP_CHAR26_ALL - максимальна кількість днів просрочки за останні 3 місяці;
- ▶ GRP_CHAR65_ALL - відношення залишку на дату зрізу до ліміту на дату зрізу;
- ▶ GRP_CHAR98_ALL - максимум з відношень прострочених заборгованостей по тілу до ліміту в кожному з місяців за останні 3 місяці;
- ▶ GRP_CHAR125 - кількість місяців з платежами за останні 3 місяці;
- ▶ GRP_CHAR136 - максимальна кількість послідовних місяців без змін в сумах платежів за місяць за останні 6 місяців;

Показник **GINI** для побудованої моделі складає **83.1%**, що відповідає моделі із високою предикативною здатністю.

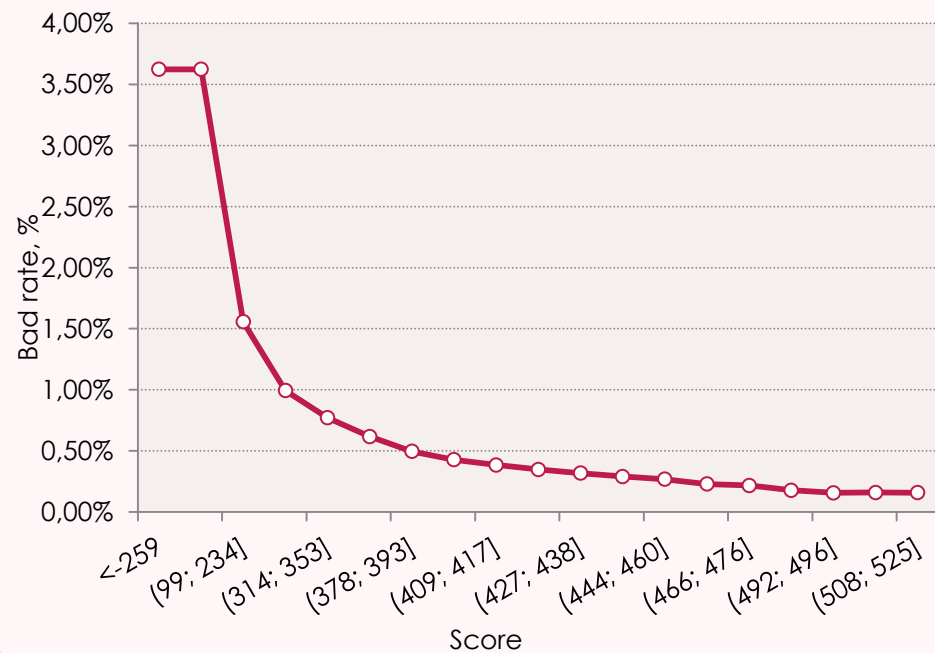
Значення **K-S статистики** складає **69.6%**, що ще раз підтверджує якість побудованої моделі.

ROC-curve & K-S

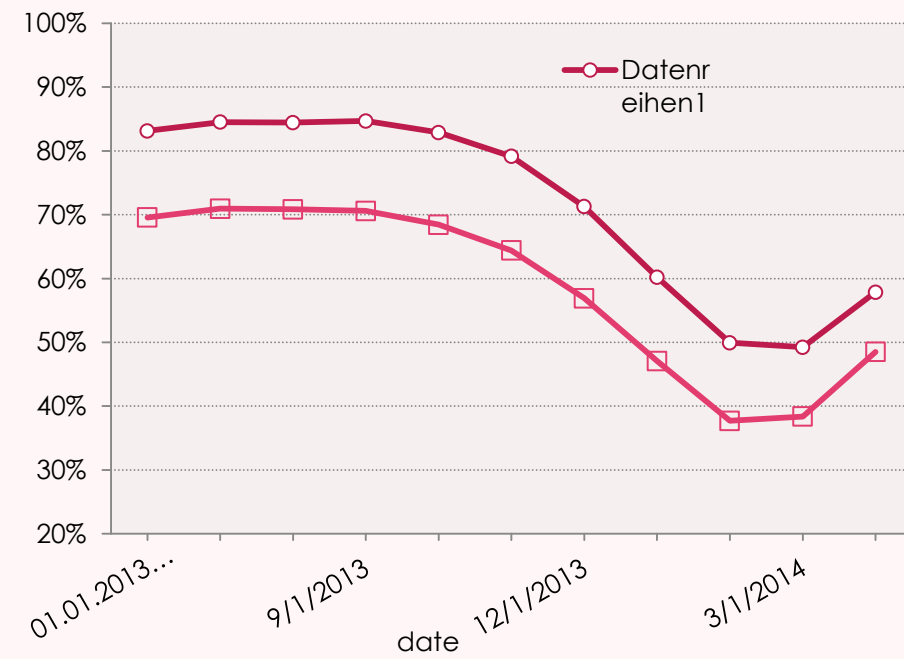


Bad Rate & GINI dynamics

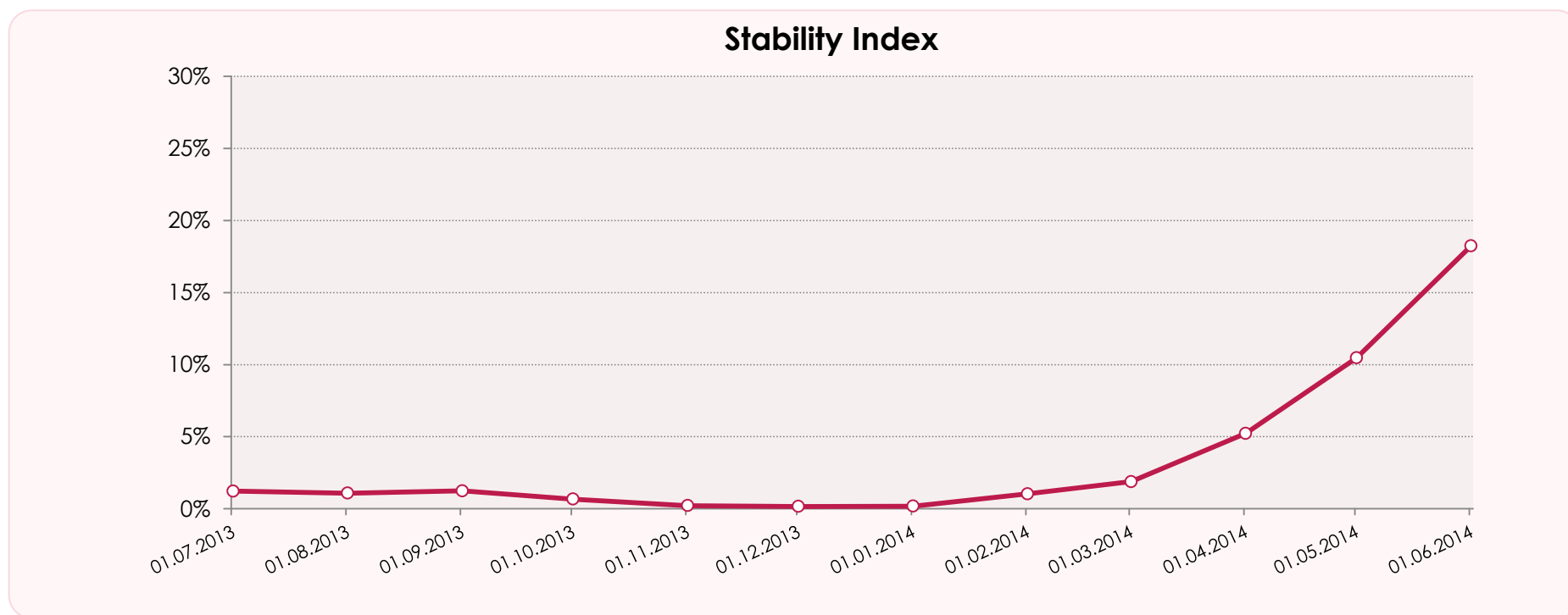
LIFT curve



GINI, KS



Стабільність



Стабільність моделі почала різко погіршуватися починаючи із січня – лютого 2014 року, що пов'язано із початком політичної та економічної кризи в Україні

Зміщення

Максимальна к-сть послідовних місяців без змін в сумах платежів за місяць за останні 6 місяців

char136	Score	2013/07	2013/08	2013/09	...	2014/04	2014/05	2014/06	2013.06 (Orig)	Stab.Ind	Shift
<1	24	44.7%	42.9%	37.4%	...	40.4%	43.1%	50.3%	39.9%	2.4%	2.5
[1,2)	49	13.1%	15.6%	20.5%	...	13.8%	10.6%	9.1%	16.3%	4.3%	-3.6
[2,5)	64	19.9%	19.6%	20.9%	...	26.7%	25.7%	17.6%	25.8%	3.1%	-5.2
>=5	81	22.3%	21.9%	21.2%	...	19.1%	20.6%	23.0%	18.0%	1.2%	4.0
		100.0%	100.0%	100.0%	...	100.0%	100.0%	100.0%	100.0%	11.0%	-2.2

Змінна є зміщеною. Зміщення по скоринговому балу складає -2.2

Зміщені параметри

- ▶ Кількість місяців від відкриття договору (активації карти) до першого бакета 1
- ▶ Максимум з відношень прострочених заборгованостей по тілу до ліміту в кожному з місяців за останні 3 місяці. По всіх кредитах клієнта.
- ▶ Максимальна к-сть послідовних місяців без змін в сумах платежів за місяць за останні 6 місяців

Наукова новизна

- ❖ Побудовано дві нові математичні моделі для оцінювання кредитоспроможності та дохідності
- ❖ На відміну від класичної моделі прогнозування дефолту клієнта, побудована модель з неперервною прогнозованою змінною для оцінки частки повернених відсотків по кредиту.
- ❖ Розширення області застосування скорингового підходу.

Перспективи подальших досліджень

- ▶ Розробка нових типів скорингових моделей на основі інтегрованого використання методів інтелектуального аналізу даних, зокрема, моделей ймовірнісного типу: байєсівська регресія, мережі, ланцюги Маркова, метод опорних векторів (SVM) і т. ін., з метою розширення повноти аналізу характеристик позичальників кредитів.
- ▶ Автоматизація розробки скорингових моделей з використанням системи R , а саме реалізація автоматичного підбору характеристик та їх сегментація.

ВИСНОВКИ

В результаті роботи була проаналізована проблема кредитного ризику, розглянуті супутні проблеми, що виникають в процесі кредитування.

Побудовано дві моделі для прогнозування:

- ▶ дефолту клієнта (поведінкова модель для паперових кредитів)
- ▶ недоотримання відсотків по кредиту (аплікаційна модель для револьверних карток з грейсовим періодом)

Отримано стабільні моделі з високою предикативною здатністю.

Визначено напрями подальшого розвитку системи для аналізу характеристик позичальників кредитів.

Публікації

- ▶ System analysis and information technology: 18-th International conference SAIT 2016, Kyiv, Ukraine, May 30 – June 2, 2016. Proceedings. – ESC “IASA” NTUU “KPI”, 2016
- ▶ Системні науки та кібернетика, 2016, №1, “Прогнозування дохідності банківських продуктів з використанням скорингового підходу”

Дякую за увагу!