

Національний технічний університет України
«Київський політехнічний інститут»
Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

Дипломна робота

Скорингові карти та прогнозування платоспроможності позичальника фінансової установи

Виконала: **Кораб Катерина, КА-41м**

Науковий керівник: к. ф.-м. н., доц. **Каніовська І.Ю.**

Дипломна робота

Предмет дослідження

Скорингові карти для прогнозування ймовірності невиклати кредиту

Об'єкт дослідження

Оцінка платоспроможності позичальника фінансової установи

Мета роботи

Застосування скорингових карт для класифікації позичальників за ймовірністю дефолту кредиту та побудова моделі прогнозу отриманих коштів за портфелем кредитних справ – грошових позик.

Практичне застосування

Побудова скорингової карти для визначення «якості» позичальника та розробка комісійної моделі фінансової установи для прогнозування зібраних коштів



Актуальність дослідження

- Неповорнення позичальником отриманих коштів залишається для банківських установ однією з основних проблем.
- Рівень простроченої заборгованості за кредитами в 20 найбільших банках України у 2015 році за результатами оцінки якості активів і стрес-тестів склав 41%.
- Використання скорингової карти для класифікації позичальників є головним способом зменшення кредитного ризику.

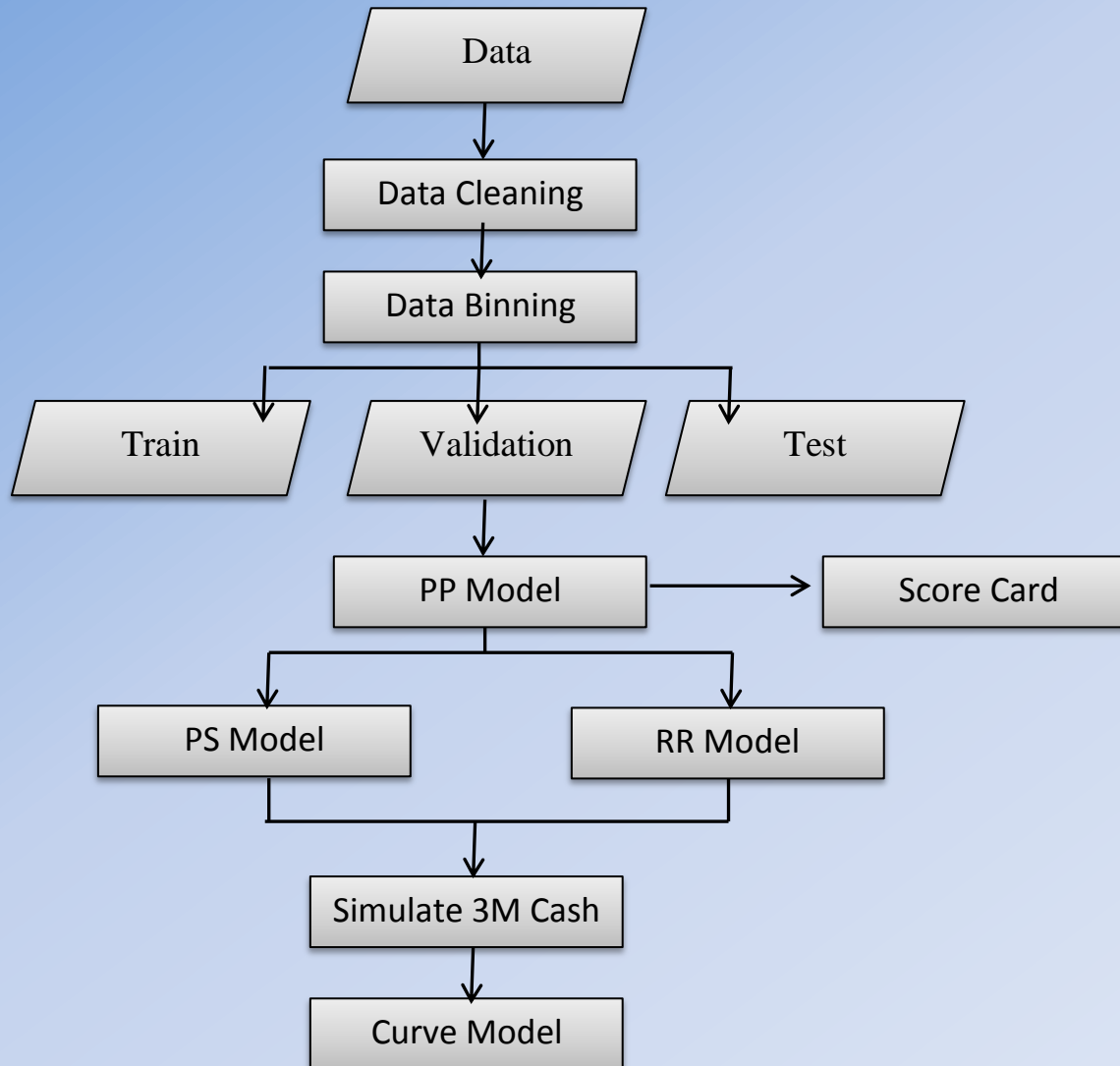
Скоринг

Скоринг - це методика оцінки кредитного ризику, яка дозволяє, оцінивши набір ознак, що характеризують позичальника, визначити, чи варто надавати йому кредит.

В основі скорингової системи лежить припущення, що люди зі схожими соціальними показниками поведуться однаково. Суть скорингу полягає в тому, що кожному параметру, що характеризує позичальника, надається реальна оцінка в балах.



Схема моделі



Логістична регресія

Логістична регресія — це вид нелінійної множинної регресії, яка аналізує функціональну залежність між кількома незалежними змінними (регресорами) і залежною змінною.

$$\ln\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1 x_i^{(1)} + b_2 x_i^{(2)} + \dots + \varepsilon_i$$

Рівняння відображає лінійну залежність ймовірності настання прострочення за кредитом (або невиплати) в залежності від значень незалежних змінних.

Константа в моделі відображає природний рівень ризику настання модельованої події за умови, що всі незалежні змінні дорівнюють нулю.

Значення коефіцієнтів при незалежних змінних, відображають ступінь їх впливу на шанс дефолту у логарифмічній шкалі, використовуються для побудови скорингової карти.

Information Value для вибору змінних моделі

Інформаційна цінність є дуже корисною концепцією для відбору змінних під час побудови моделі.

$$IV = \sum (DistributionGood_i - DistributionBad_i) \times \ln\left(\frac{DistributionGood_i}{DistributionBad_i}\right)$$

Information Value	Сила прогнозу
< 0.02	Недоцільний для прогнозу
0.02 to 0.1	Слабкий прогноз
0.1 to 0.3	Середній прогноз
0.3 to 0.5	Сильний прогноз
>0.5	Підозріло або занадто добре

ROC-крива та GINI

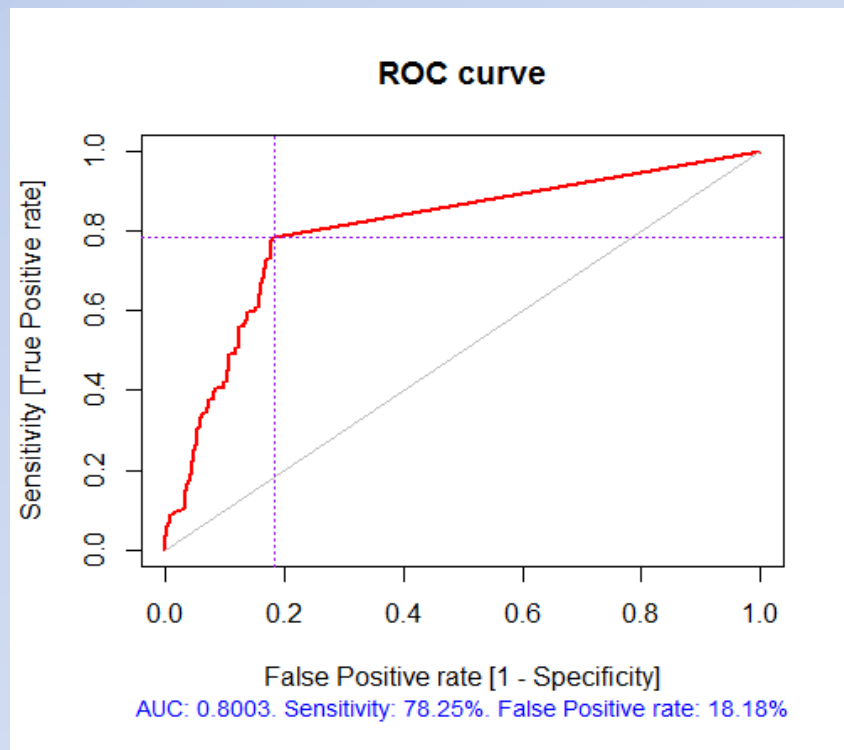
ROC-крива (Receiver Operation Characteristic) показує залежність кількості правильно класифікованих позитивних прикладів від кількості неправильно класифікованих негативних прикладів. Перші з них називаються істинно позитивними, а другі — неправильно негативними множинами.

Оскільки візуальне порівняння ROC-кривих не завжди дозволяє визначити ефективнішу модель, застосовують оцінку площі під кривими. Чисельний показник площі під кривою AUC (Area Under Curve)

Індекс GINI можна визначити через площу фігури, що знаходиться під ROC-кривою, таким чином:

$$\text{GINI} = 2 \cdot \text{AUC} - 1.$$

Інтервал AUC	Індекс GINI	Якість моделі
0,9—1,0	0,8—1,0	Відмінна
0,8—0,9	0,6—0,8	Дуже висока
0,7—0,8	0,4—0,6	Прийнятна
0,6—0,7	0,2—0,4	Середня
0,5—0,6	0—0,2	Незадовільна



Дані, що використовуються для побудови моделі

- Статистика по справах з проблемною заборгованістю за 2015-2016 роки з початковою інформацією:

Назва банку чи фінансової організації

Вид кредиту

Місце проживання

Кількість днів прострочення

Кількість мобільних телефонів

Кількість днів з останнього платежу

Вік

Сума кредиту

Стать

Сума заборгованості

Сума заборгованості за тілом кредиту до всієї заборгованості

Кількість передач в КА

```
[1] "Total DPD_bin variable IV: 0.4038"  
[1] "Total PD_TC_bin variable IV: 0.1408"  
[1] "Total PD_TC_Ratio_bin variable IV: 0.1697"  
[1] "Total IsWashed_bin variable IV: 0.1519"  
[1] "Total NumberOfWashes_bin variable IV: 0.1519"  
[1] "Total LPD_bin variable IV: 0.6515"  
[1] "Total InitMobPhones_bin variable IV: 1e-04"  
[1] "Total Age_bin variable IV: 0.0163"  
[1] "Total OtherLoans_bin variable IV: 0.0936"
```

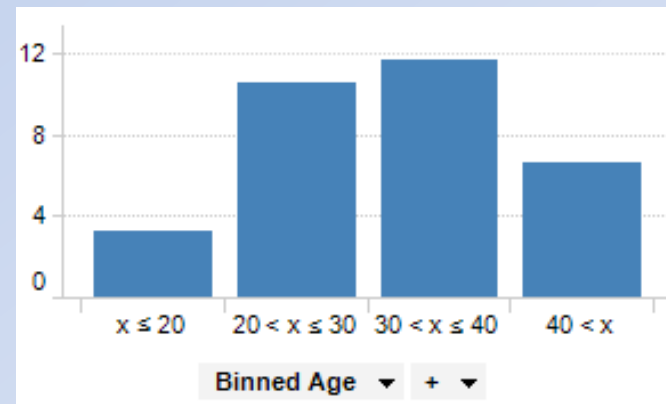
- Уся вибірка розбивається на три групи : train, validation, test

Неперервні та факторні змінні. Data binning

В скоринговій моделі можуть використовуватися в якості незалежних змінних категоріальні і кількісні предиктори.

Багато розробників скорингових систем використовують завжди метод категоризації кількісних змінних.

Категоризація кількісних змінних дозволяє досягти таких основних переваг при побудові скорингової карти: полегшити обробку викидів та екстремальних значень кількісних змінних; спростити інтерпретацію скорингової карти; відобразити складні нелінійні зв'язки.



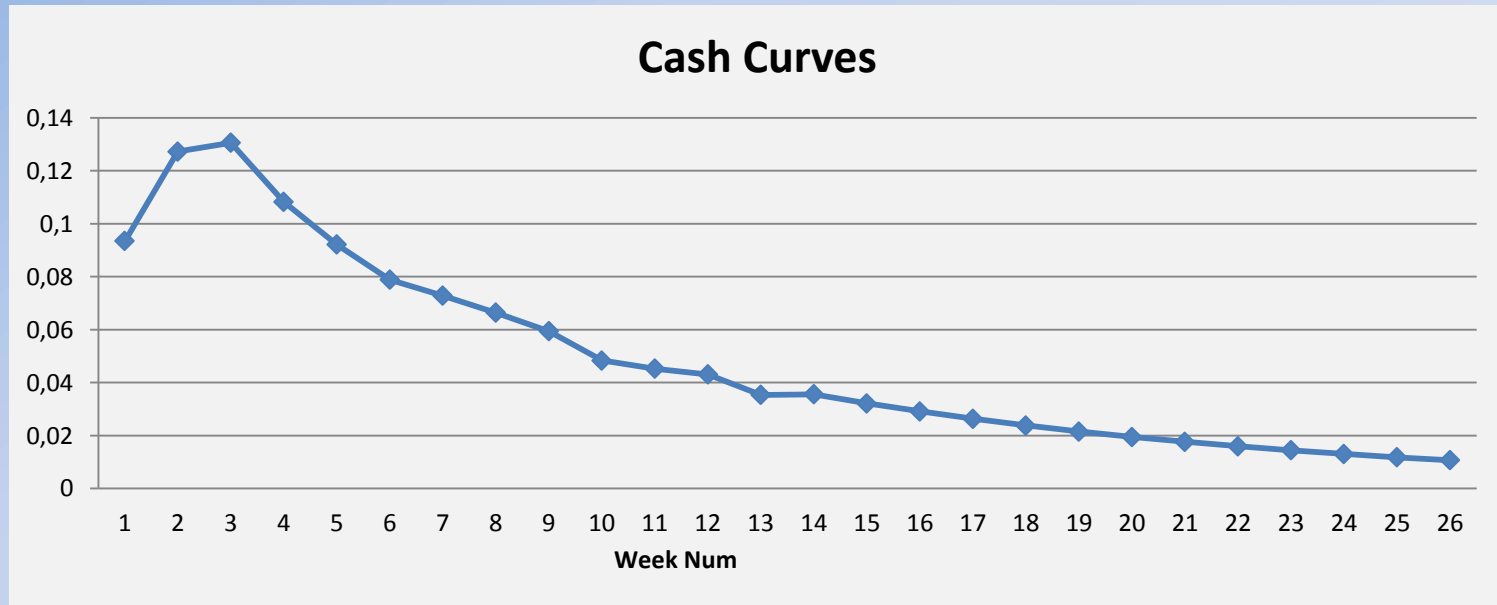
Формула моделі

PP Model → ймовірність платежу за 3 міс;

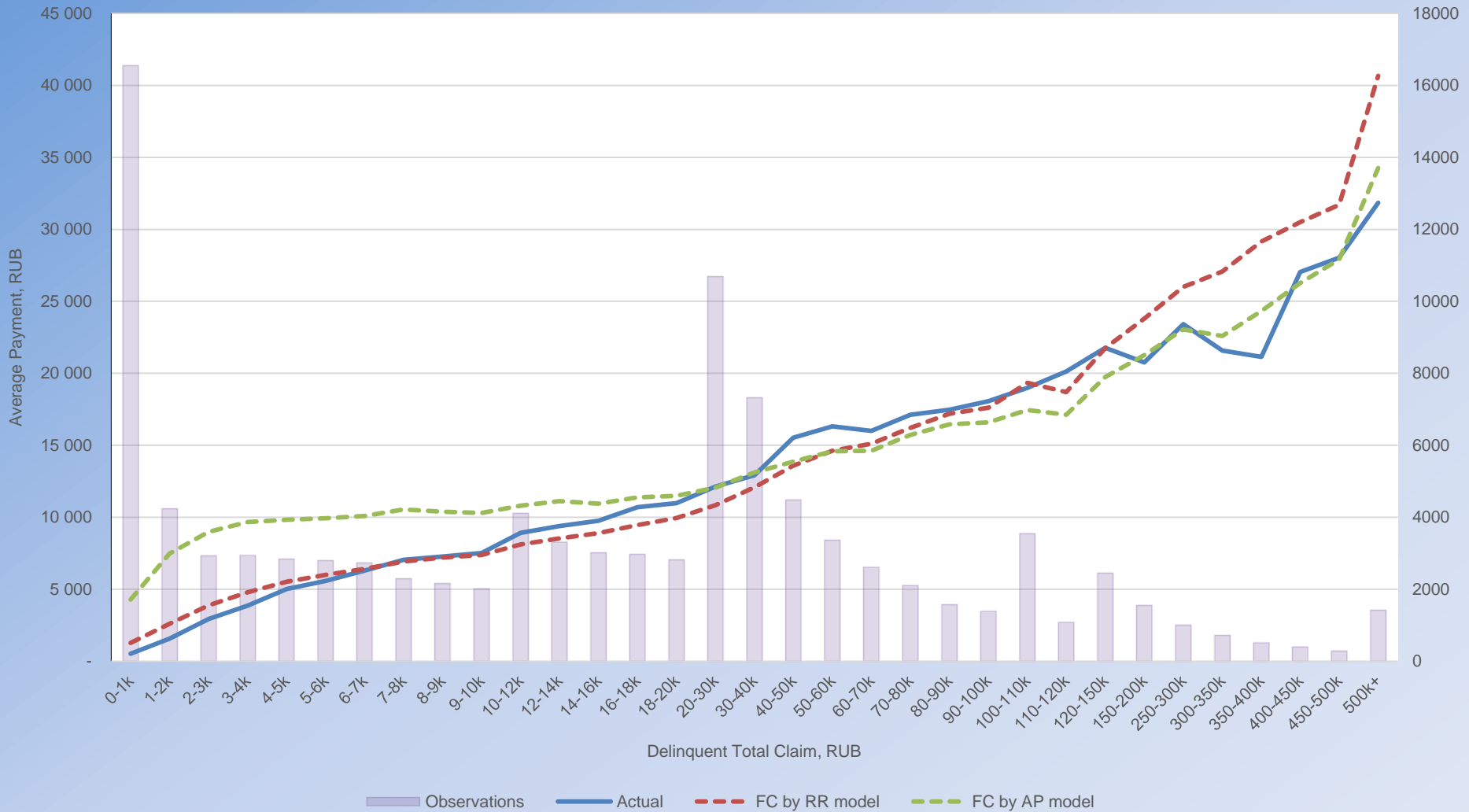
PS Model → величина платежу за 3 міс;

RR Model → відсоток від суми заборгованості, яку заплатить позичальник;

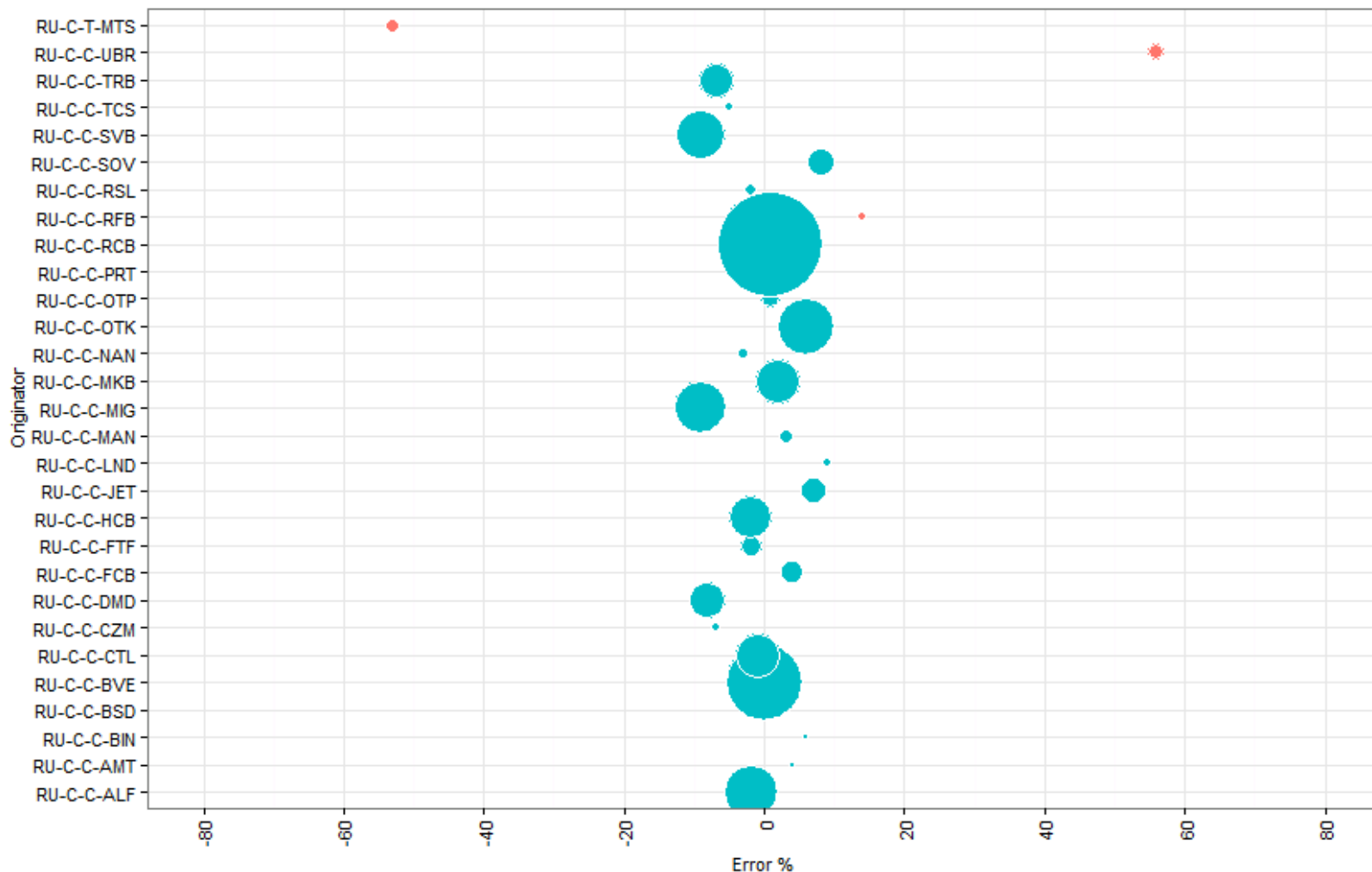
Сума, яку заплатить за 3 міс = $PP * PS$ or $PP * RR * TC$



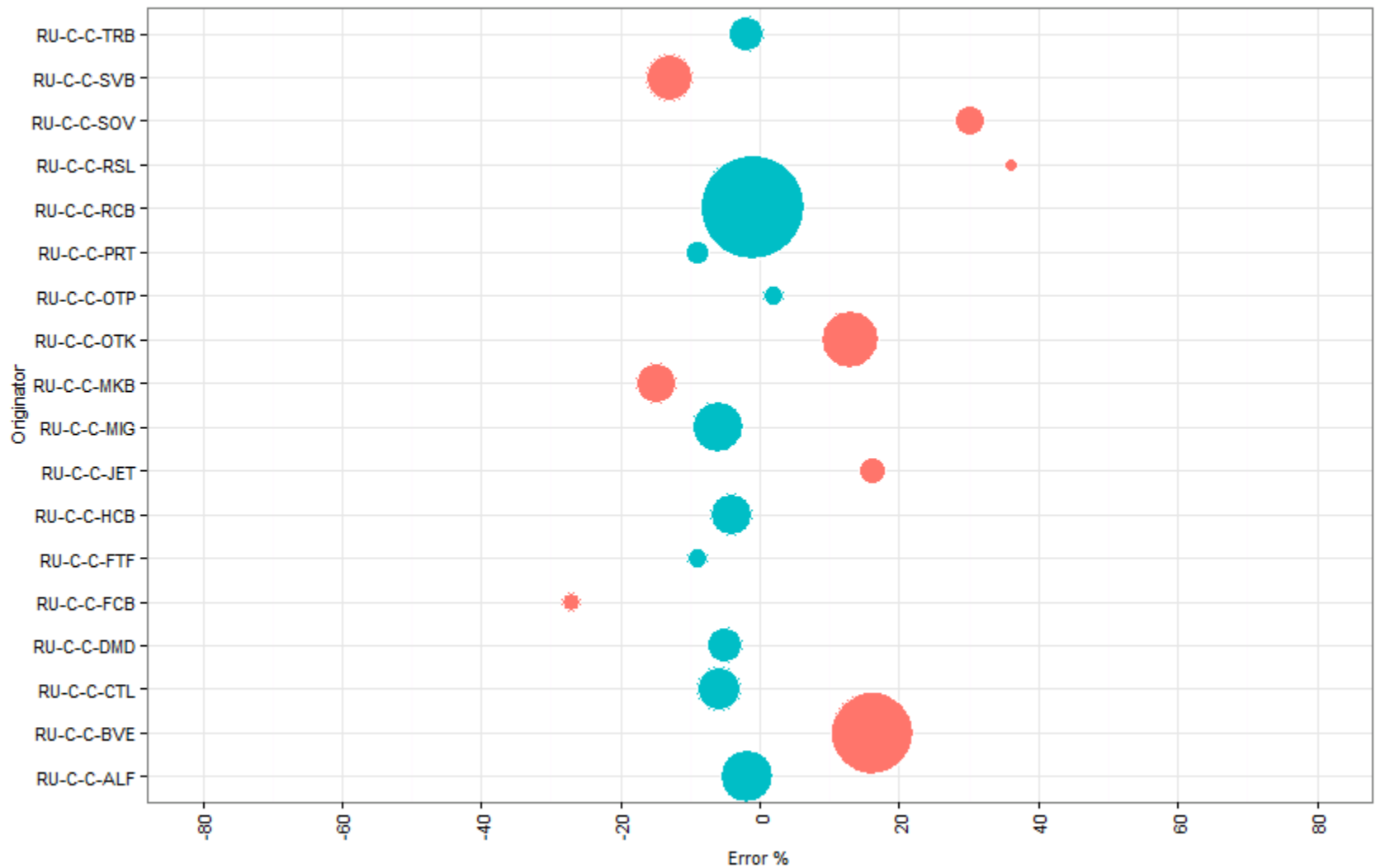
Проблема прогнозування PS Model vs RR Model



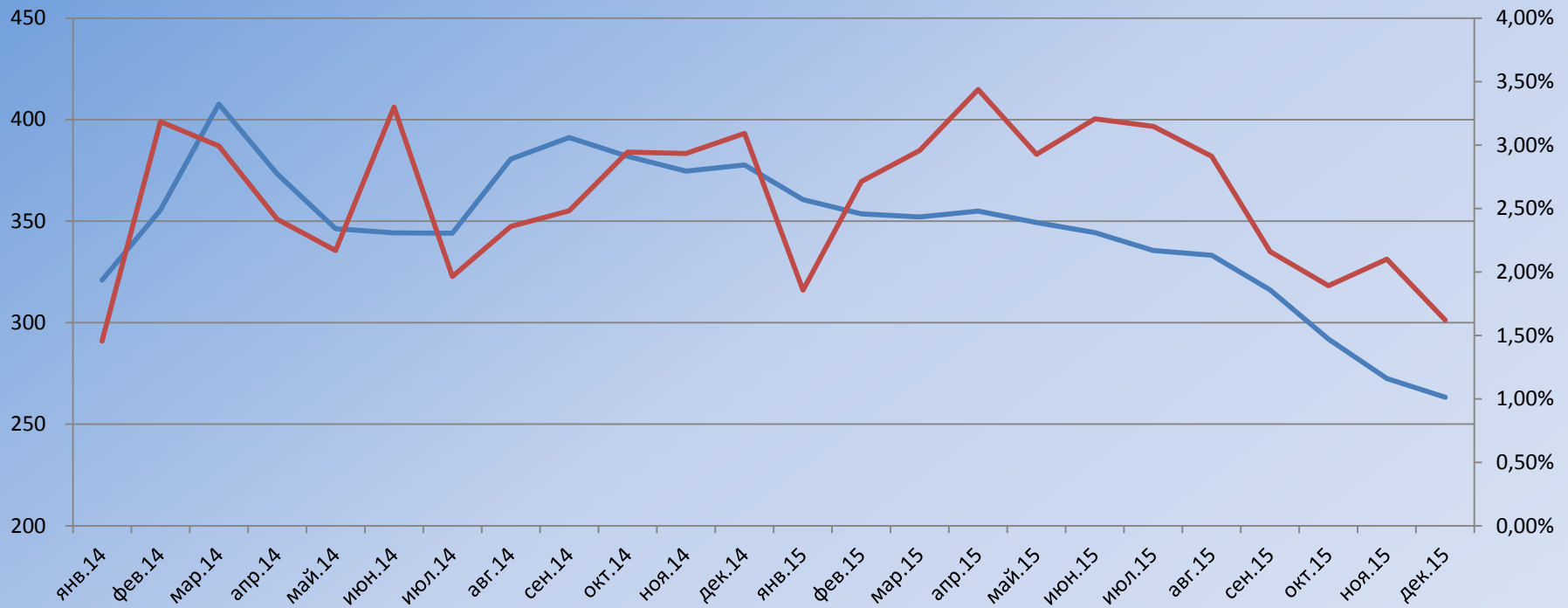
Результати моделювання для вибірки validation



Результати моделювання для вибірки test



Результати скорингової карти



Ймовірність платежу та середній скоринговий бал кредитного портфелю банку

Висновки

- Було розглянуто логістичну регресію для прогнозування ймовірності невикплати кредиту
- Для аналізу якості моделі використовується ROC-крива та коефіцієнт Джині
- Було побудовано модель для підрахунку суми грошей, яку виплатить боржник за 3 міс
- На основі моделі створена скорингова карта, яка дозволяє сортувати позичальників від «поганих» до «добрих»

ДЯКУЮ ЗА УВАГУ