

ПОБУДОВА СКОРИНГОВОЇ МОДЕЛІ ДИФЕРЕНЦІАЦІЇ ВИСОКО РИЗИКОВИХ ПОЗИЧАЛЬНИКІВ БАНКУ

У статті представлено загальні положення розробки процесу скорингу та застосовано механізм побудови скорингових моделей оцінювання клієнтів на підприємствах роздрібною торгівлі та банківського сектору. Виділено логічну послідовність процесу проведення скорингу, визначено види скорингових моделей та їхні особливості. Окремо надано інструменти верифікації скорингу на основі реальних даних.

Ключові слова: скоринг, кредитування, індекс значимості інформації (Information Value – IV), індекс відмінності інформації (Weight of Evidence – WoE), множинна кореляційно-регресійна модель, логістична регресія, дерево рішень, індекс стабільності популяції (Population Stability Index – PSI), тест Колмогорова – Смірнова.

CONSTRUCTION SCORING MODEL DIFFERENTIATION HIGH-RISK BORROWERS OF THE BANK

The aim of the research is to show the applicability of mathematical methods in the assessments of clients at retail enterprises and banking sector. Mathematical models of differentiation debtors were built in the work, including binary choice models using decision trees, regression and logit models, the most effective one that is the best able to distinguish potential debtors from reliable payers was selected. The article clearly identifies stages of scoring and its kinds. At the example of one of Ukrainian banks the scoring model of bank creditors differentiation, which enables financial institutions to identify the level of credit quality was built. Also, in the article the method of verification of the mathematical models was used and with the help of Kolmogorov – Smirnov test the selection one of the best of them was made. Thus, the application of scoring allows significantly to enhance the performance of financial institutions, to simplify the decision making process for granting credit, to diversify the risks connected with processing of a large number of similar data and increase profits (minimize probable losses) of the company.

Keywords: scoring, lending, Information Value, Weight of Evidence, regression model, logistic regression, decision tree, Population Stability Index, Kolmogorov – Smirnov test.

Вступ

Промислові підприємства, великі торгові мережі, банки та фінансові установи, часто мають справу з великою кількістю клієнтів. Прибуток, який можуть отримати такі компанії від окремої особи, залежить від цілого ряду її індивідуальних характеристик. Частину інформації про своїх клієнтів кожна компанія може зібрати самостійно з використанням різноманітних аплікаційних форм та опитувань, застосувавши після чого кількісний та якісний аналіз наявних клієнтів. Це дозволить в майбутньому мінімізувати ризики при роботі з клієнтами та покращити якість їхнього обслуговування. Одним з можливих шляхів вирішення такого завдання виступає проведення скорингу.

Проблеми проведення скорингу досліджено у працях вітчизняних вчених: В. Вітлінського, О. Кириченко, А. Камінського, Л. Патерікіна, О. Криклія, В. Єлейка, Р. Іванова, Б. Кишакевича, В. Серед закордонних вчених варто виділити роботи Р. Андерсона, Н. Сіддікі, Т. Крука, Т. Ліна, Дж. Белотті, Л. Томаса, С. Міллера та інших.

Метою цього дослідження є аналіз процесу проведення скорингу, дослідження сфер його застосування, виділення етапів побудови скорингової моделі та проведення її формалізації, наведення прикладу застосування скорингу банківською установою.

Виклад основного матеріалу дослідження

Перші згадки про надання позик під відсотки сягають древнього Вавилону понад 5000 років тому. Водночас кредитний скоринг, у його сучасному розумінні, зародився лише близько 60 років тому. Все почалося з того, що Рональд Фішер у 1936 почав застосовувати статистичні методи для вирішення проблеми ідентифікації груп у популяції. Він намагався розрізнити два види ірису на основі заміряних фізичних величин. У 1941 році цю ідею підхопив Д. Дюран і вирішив застосувати цю ж методику для визначення хороших та поганих позичальників. В часи Другої світової війни більшість банків почали зазнавати труднощів з видачами кредитів, адже багато аналітиків були призвані на військову службу, що призвело до нестачі кваліфікованих кадрів. Тому фінансові установи почали запроваджувати введення системи емпіричних правил, які використовували аналітики в процесі прийняття рішення про надання кредиту. Таке впровадження було одним з перших прикладів експертних систем, які допомогли залучати неспеціалістів до проведення фінансових процедур, що значно спростило саму процедуру кредитування. В результаті виявилось, що розроблені скорингові карти, в яких набиралися бали за певними характеристиками потенційного позичальника, дають набагато точніші прогнози, аніж індивідуальні висновки кредитних спеціалістів. З часом керівники фінансових установ зрозуміли, що без точного та автоматизованого

управління ризиками, вони, як кредитори, не зможуть розширити свої кредитні портфелі. З цього моменту і почався бум скорингових систем.

На даний час, скоринг застосовується не лише у банківських установах, а й на великих підприємствах різних видів діяльності. Для максимізації свого прибутку підприємствам доцільно розробляти індивідуальний підхід до роботи з кожним споживачем своїх послуг. Звичайно для великих корпорацій, які працюють щоденно з об'ємною базою клієнтів, такий варіант є неприйнятним. Аналіз клієнтської бази дає компанії можливість диференціювати клієнтів з різними характеристиками та визначати ключові з них, дослідити вплив кожного показника на деякий цільовий показник. Такий підхід можна розширити через присвоєння кожному клієнту балу відповідності цільовому показнику, на основі якого обирають споживачів товарів чи послуг підприємства. На даний час визначення такого балу є однією з стадій проведення скорингу.

У зарубіжній літературі, як правило, оперують поняттям кредитний скоринг. Оскільки у вітчизняних наукових джерелах термін "скоринг" також вживають в розумінні кредитного скорингу, то було запроваджено власну інтерпретацію цього терміну. Отже, скоринг – це процес оцінювання об'єкта за допомогою скорингової моделі, на основі наявної інформації про об'єкт оцінювання, цільового показника та факторів, які впливають на цей показник. Об'єктами, що підлягають оцінюванню, можуть виступати платоспроможність боржника, величина очікуваного доходу, оцінка відгуку на рекламу, імовірність шахрайства тощо.

Скорингова модель – це математична модель оцінювання об'єктів, що підлягають скорингу, яка слугує для сегментації (класифікації) цих об'єктів за категоріями.

Сегментація та сортування клієнтської бази дає можливість компаніям впливати на різні групи своїх клієнтів з різною інтенсивністю або діяти на них різними інструментами. Також перевагою такого підходу є те, що на основі правил присвоєння скорингового балу можна не лише проставити бал теперішнім клієнтам за певною характеристикою, а й віднести майбутніх клієнтів за апріорними характеристиками до певної групи (категорії) з метою визначення їхнього потенціалу у майбутньому. Це значить, що деякому об'єкту автоматично присвоюється бал, який вказує на ступінь довіри і цінності клієнта.

Сфери застосування скорингу різноманітні, зокрема його можуть застосовувати великі супермаркети, які продають різного роду товари (будівельні матеріали, харчову та промислову продукцію, електроніку); рекламні агенції та компанії, які працюють у сфері прямого або контент маркетингу (поштові, sms та e-mail розсилки, інтернет реклама у соціальних мережах, сайтах та блогах); фінансові компанії та банки, які надають позики та кредитні картки; колекторські компанії та організації по стягненню заборгованості; страхові компанії, медичні заклади, що працюють безпосередньо з пацієнтами, особливо у галузі трансплантології; правоохоронні органи з метою визначення потенційно небезпечних злочинців тощо.

На нашу думку, можна виділити такі основні етапи проведення скорингу (рис. 1).

Проведення скорингу можливе лише за наявності достатньої статистичної інформації щодо самих клієнтів, цільового показника та кореляційних залежностей між вхідними величинами і цільовим показником.

Розглянемо декілька прикладів застосування такого сортування (класифікації) клієнтів. Одним із застосувань скорингу є аналіз купівельних характеристик покупців в закладах роздрібною торгівлі. Метою побудови скорингової моделі є сегментація покупців закладу та розроблення ефективних інструментів стимулювання продажів, якими можуть виступати такі, як відправлення листів з рекламними акціями та розповсюдження рекламних матеріалів, так і формування акційних пропозицій та знижок. Під сегментацією будемо розуміти виділення тих груп покупців, які володіють деякими схожими характеристиками. Звичайно розсилати рекламні матеріали всім покупцям немає сенсу, так як і надавати усім знижку на окремі види товарів, оскільки така стратегія не дасть додаткового прибутку. Тому потрібно вплинути на ті сегменти покупців, які зможуть дати найбільшу віддачу.

Для виділення найбільш прибуткових сегментів і проведення аналізу характеристик покупців необхідно зібрати найбільш важливу інформацію про клієнтів, наприклад, вік людини, її місце праці, посада, стать, сімейний стан, місце проживання, трудовий стаж на одному місці праці, кількість дітей тощо. Таку інформацію об'єкти роздрібною торгівлі, як правило, беруть з різноманітних анкет, соціологічних опитувань або інформації під час видачі дисконтних карт. Зауважимо, що перевагою надання дисконтних карт є можливість отримання інформації про частоту покупок кожним клієнтом та суми, на які він здійснює покупки.

Після збору статистичної інформації важливо побудувати якісну скорингову модель, яка допомогла б виділити групи найбільш цінних покупців (як теперішніх клієнтів, так і потенційних). Основними вихідними показниками аналізу статистичних даних є кількість витрачених грошей за певний період, частота звернень у магазин, частка чоловіків та жінок у кожній з вищезазначених груп, їхній вік, рівень доходу тощо.

Провівши скоринг, можна визначити ступінь впливу кожного фактору на формування купівельного кошика споживача, оцінити середній розмір покупки покупця з наперед відомими характеристиками, оцінити імовірність здійснення покупки клієнтом. Така сегментація дозволить компанії визначити найбільш важливі сегменти, на які варто впливати, з метою максимізації прибутку або мінімізації затрат.

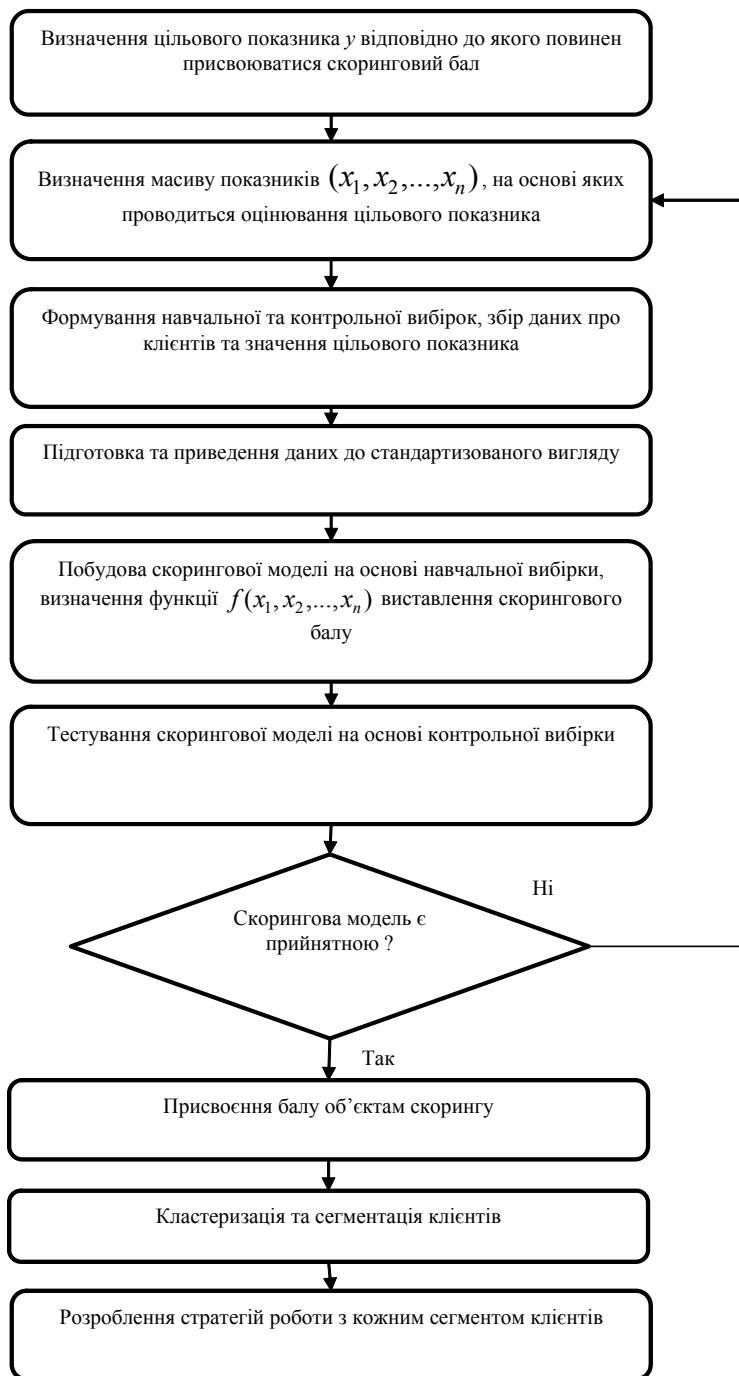


Рис. 1. Етапи проведення скорингу

Тепер розглянемо застосування кредитного скорингу на підприємствах фінансового сектору з метою визначення платоспроможності клієнтів.

На думку Томаса Ліна [2, с.1], кредитний скоринг – це набір моделей та методів прийняття рішень, які допомагають кредитору під час надання кредиту. Ці методи дозволяють визначити особи (групи осіб), які отримають кредит, суму кредиту, а також стратегії підвищення рентабельності роботи позичальника з кредитором. Проте в умовах збільшення кількості неплатників за кредитами, на нашу думку, особливу увагу варто приділити механізму обслуговування кредиту. За нестабільної фінансової ситуації застосування математичних методів роботи з боржниками, які протермінували кредитний платіж, дозволить фінансовим установам працювати більш ефективніше.

Сама методика оцінювання кредитного ризику, як правило, зводиться до врахування набору соціально-економічних показників позичальника, на основі яких роблять висновок про можливість надання йому кредиту. Також фінансова установа може здійснювати ще суміжні послуги, такі як надання інформації про стан кредитоспроможності та калькуляцію скорингового балу, приведення скорингового балу до стандартизованого виду (наприклад, за американською метрикою FICO score), надання відповідної

підтвердженої інформації про скоринговий бал у місця вимоги (банки, лізингові компанії, магазини що продають товари у кредит, тощо). Основним завданням формування скорингового рейтингу є виділення найбільш важливих характеристик клієнта, які визначають можливість повернення кредиту без врахування причин неповернення клієнтом кредитних коштів. Важливою особливістю скорингових систем в сучасних умовах є те, що завдяки сучасним інформаційним технологіям калькуляція скорингового балу проводиться практично миттєво, а рішення кредитного спеціаліста є неупередженим та об'єктивним, що ставить у рівні умови позичальників з схожими характеристиками.

Варто зазначити, що різні фінансові установи можуть визначати одного і того ж позичальника як надійного або неплатоспроможного на основі різних скорингових моделей та сукупності вхідних характеристик клієнта.

В теперішніх умовах відбувається подальший розвиток кредитного скорингу (рис. 2). Тому окрім визначення платоспроможних клієнтів (так званий аплікаційний скоринг), підприємства також:

1) виділяють потенційних споживачів своїх послуг (скоринг відгуку та скоринг збереження (втрати) клієнта), що дозволяє проводити цільові рекламні кампанії;

2) визначають можливу поведінку теперішніх клієнтів (поведінковий скоринг) за обраними параметрами (кількість протермінувань клієнтом кредитних платежів, на скільки днів це було зроблено, як змінювалася доля тіла кредиту в структурі загального боргу), що допомагає прогнозувати поведінку клієнта та зменшити імовірність перетворення клієнта у проблемного кредитора;

3) вирішують питання управління проблемними кредитами (колекторський скоринг), що дозволить обрати оптимальні інструменти роботи з кредитним портфелем проблемної заборгованості.



Рис. 2. Види кредитного скорингу

Таким чином, використовуючи кредитний скоринг, банки та інші фінансові структури можуть проводити повний цикл роботи з кредитом – від залучення та оцінювання клієнтів до повернення проблемної заборгованості, що в цілому дозволяє сформувати дохідний кредитний портфель з прийнятним рівнем ризику.

Розглянемо застосування кредитного скорингу в діяльності банківської установи. Метою побудови скорингової моделі диференціації позичальників банку за платоспроможністю є визначення імовірності протермінування кредиту терміном більше 90 днів.

Нехай кожен клієнт фінансової установи характеризується 10 параметрами:

x_1 – відношення загального залишку по кредитних картках і особистих кредитних лініях (за виключенням нерухомості), не кредитних коштів та автокредитування до суми кредитних лімітів;

x_2 – вік позичальника, роки (якщо вік від 18 до 43 років то бінарна змінна x_{21} приймає значення 1, якщо від 44 до 58, то $x_{22} = 1$, від 59 до 73, то $x_{23} = 1$, якщо від 74 до 88, то $x_{24} = 1$);

x_3 – кількість протермінувань виплати кредиту позичальником терміном 30 – 59 днів (якщо відбулось одне протермінування на 30 – 59 днів, то бінарна змінна x_{31} приймає значення 1, якщо два, то $x_{32} = 1$, якщо три, то $x_{33} = 1$, якщо чотири то $x_{34} = 1$, у всіх інших випадках ці змінні приймають значення 0);

x_4 – кількість протермінувань виплати кредиту терміном 60 – 89 днів (якщо відбулось одне протермінування на 60 – 89 днів, то бінарна змінна x_{41} приймає значення 1, якщо два, то $x_{42} = 1$, у всіх

інших випадках ці змінні приймають значення 0);

x_5 – кількість протермінувань виплати кредиту терміном більше 90 днів (якщо відбулось одне протермінування більше 90 днів, то бінарна змінна x_{51} приймає значення 1, якщо два, то $x_{52} = 1$, якщо три, то $x_{53} = 1$, у всіх інших випадках ці змінні приймають значення 0);

x_6 – місячний дохід, гривень;

x_7 – відношення щомісячних виплат боргів, аліментів, витрат на проживання до місячного валового доходу (якщо значення $\leq 50\%$ то бінарна змінна x_{71} приймає значення 1, якщо $50\% < x_7 \leq 75\%$, то $x_{72} = 1$, якщо $75\% < x_7$, то $x_{73} = 1$);

x_8 – кількість відкритих кредитів (в розстрочку, видані як позики на купівлю автомобіля чи іпотеку) і кредитних ліній (наприклад, кредитні карти);

x_9 – кількість іпотечних кредитів та позик на нерухомість;

x_{10} – кількість утриманців у сім'ї, за виключенням самого клієнта (члени подружжя, діти, тощо).

Через y позначимо цільовий показник, тобто бінарну змінну, яка показує, чи особа протермінувала виплату кредиту терміном більше 90 днів.

Статистичні дані характеристик клієнтів візьмемо з бази даних одного з великих банків України. Спочатку сформуємо навчальну вибірку (на основі 100 тисяч записів кредитних історій) та контрольну вибірку (на основі 40 тисяч записів). Перед побудовою скорингової моделі приведемо дані до стандартизованого вигляду, тобто проведемо очистку таблиці з даними від пустих значень, замінивши їх на нулі, приведемо значення дат до формату зрозумілого статистичному пакету, виправимо дані, що введені невірно (наприклад від'ємне значення віку, тощо).

Наступним етапом будемо скорингову модель диференціації позичальників банку. Побудова скорингової моделі складатиметься з 4 етапів:

1. Формування груп за кожним з вхідних показників.
2. Введення нових змінних.
3. Вибір та побудова математичних моделей скорингу.
4. Аналіз побудованих скорингових моделей.

Першим кроком створюємо групи по кожному з показників та розраховуємо для них індекси інформаційності, зокрема індекс значимості інформації – Information Value (IV), який показує вплив кожної групи показника на кінцевий результат, та індекс відмінності інформації – Weight of Evidence (WoE), який показує приведену до логарифмічної шкали відмінність між частками боржників по кожній з груп. З допомогою цих показників, можна підібрати мінімальну кількість груп, за якими частки неплатоспроможних (“поганих”) та платоспроможних (“хороших”) клієнтів значно відрізняються.

Графік, який ілюструє один з можливих поділів віку боржника, зображено на рис. 3.

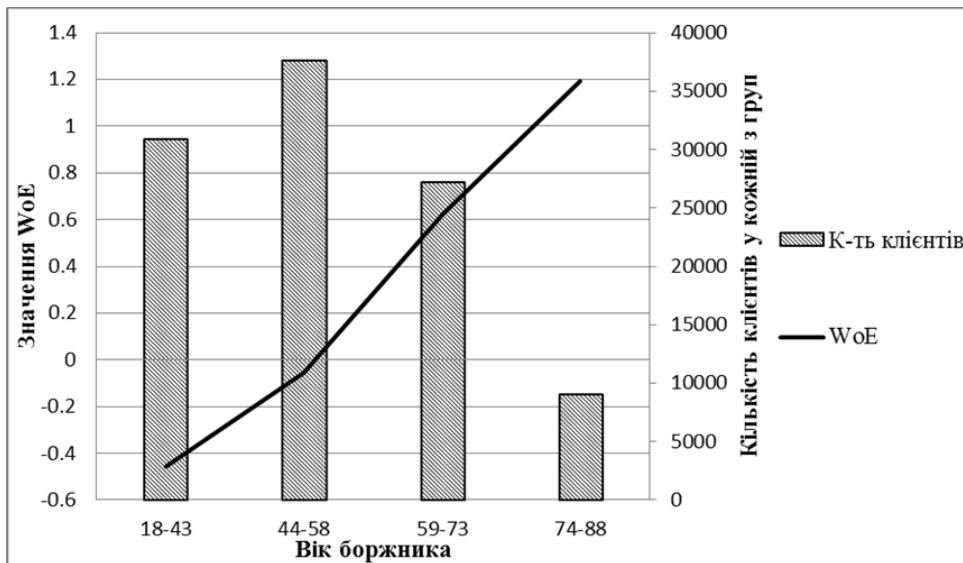


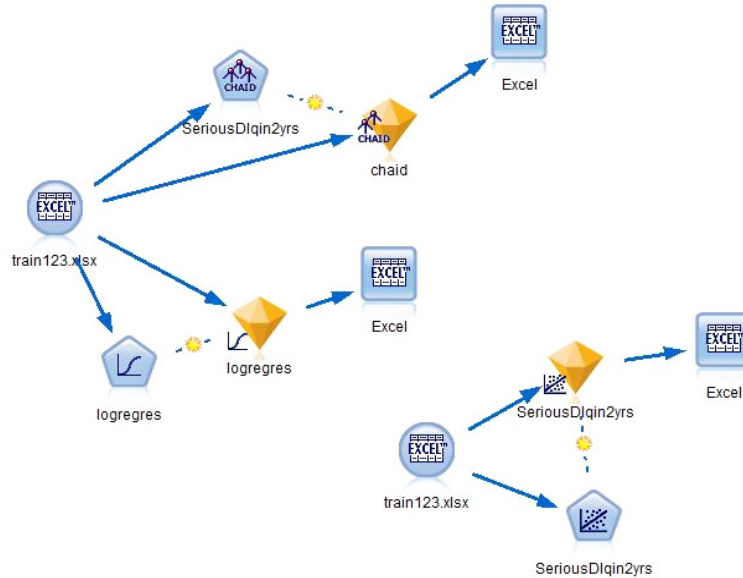
Рис. 3. Поділ віку боржника на 4 групи, та відповідність індексу WoE кожній з цих груп

Зазначимо, що чим більше відрізняється значення індексу WoE від нуля, тим чіткіше видно відмінності між “хорошими” та “поганими” клієнтами по кожній з груп.

Наступним кроком буде введення нових змінних в залежності від типів отриманих даних та їх ваги.

Це робиться з метою визначення показників, які прямо не подані серед сукупності початкових вхідних даних. Це можуть бути бінарні змінні, наприклад, наявність у боржника більше двох членів сім'ї; логарифмічні, наприклад логарифм натурального доходу; відношення показників, наприклад частка тіла кредиту до суми всього кредиту, тощо.

На третьому кроці здійснюють побудову математичної моделі залежності результуючої змінної від вхідних. На нашому прикладі проведено побудову скорингових математичних моделей у середовищі IBM SPSS Modeler, за допомогою алгоритмів побудови рівнянь регресії, CHAID (одна з модифікацій дерева рішень, де використовується автоматичний детектор взаємодії χ^2 (хі – квадрат)) та логістичної регресії [1, с. 308]. Загальний вигляд побудованих моделей у SPSS Modeler зображено на рис. 4



Пояснення позначень:





- | | | | |
|---|--------------------------------------|---|----------------------------------|
|  | - модель дерева рішень |  | - структура дерева рішень |
|  | - модель множинної лінійної регресії |  | - коефіцієнти множинної регресії |
|  | - модель логістичної регресії |  | - коефіцієнти logit - моделі |
|  | - вхідні дані |  | - вихідні дані |

Рис. 4. Загальний вигляд процесу побудови математичних моделей у IBM SPSS Modeler (побудовано автором)

Після цього проведемо аналіз побудованих моделей. Спершу визначаємо найбільш вагомий показник з допомогою індексу Information Value, який дозволяє визначити вхідний фактор, що найбільш чітко диференціює навчальну вибірку на “хороших” та “поганих” позичальників. Наприклад, для регресії таким показником є відношення залишку кредитних коштів до суми кредитних лімітів (рис. 5).

Якщо якийсь один показник Information Value має значення близьке до 1, це означає що, він сильно корелює з цільовим показником у.

Наступним кроком є застосування математичної моделі до контрольної вибірки, що дасть нам прогнозовані значення результуючої змінної у вигляді скорингового балу.

Для перевіряння адекватності поділу контрольної вибірки на групи, в залежності від скорингового балу в розрізі вхідних показників обчислюють показник Population Stability Index (PSI). Він допоможе визначити ступінь відмінності контрольної вибірки від навчальної, та показники, які заслуговують особливої уваги під час застосування побудованої математичної моделі. Наприклад, під час застосування скорингової моделі протягом достатньо довгого періоду часу в кредитному портфелі можуть відбутися значні зміни. При цьому контрольна вибірка може значно відрізнятись від тестової, що жодним чином не відображено у початкових даних. А якщо зміни справді відбулися, то такі відхилення потрібно враховувати. Показник PSI обчислюють за допомогою співвідношення

$$PSI = \sum_{i=1}^n \left(\frac{x_{ip}}{\sum_{i=1}^n x_{ip}} - \frac{x_{if}}{\sum_{i=1}^n x_{if}} \right) \cdot \ln \left(\frac{\frac{x_{ip}}{\sum_{i=1}^n x_{ip}}}{\frac{x_{if}}{\sum_{i=1}^n x_{if}}} \right), \quad (1)$$

де $N = \overline{1, n}$ – кількість інтервалів скорингових балів, x_{ip} – кількість об'єктів у i -му інтервалі в навчальній вибірці, x_{if} – кількість об'єктів у i -му інтервалі в контрольній вибірці.

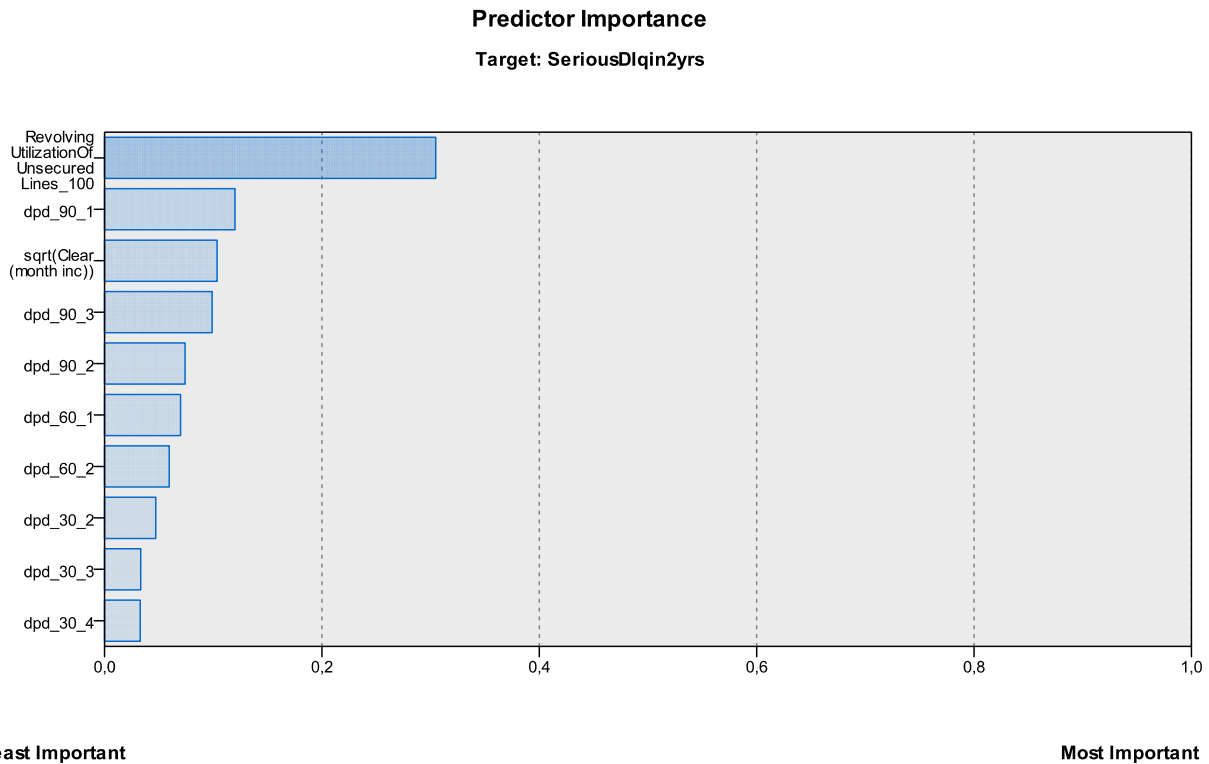


Рис. 5. Вага вхідних показників у моделі регресії (обчислено автором)

Зауважимо, що показник PSI розраховують для однакових інтервалів скорингових балів кожного з вхідних показників.

При цьому, якщо значення $PSI > 0,25$ то значення цього показника у контрольній вибірці значно відрізняються від значень тестової вибірки. У такому випадку, цей показник потрібно виключити з аналізу та знову побудувати математичну модель.

Для перевіряння якості побудованих скорингових моделей, проведено тест Колмогорова – Смірнова для кожної з моделей (для порівняння прогнозних результатів з фактичними). Приклад графіку, де зображено якість диференціації хороших клієнтів від поганих з допомогою лінійної регресії зображено на рис. 6.

Виявивши найбільш точну модель визначення неплатоспроможних клієнтів, її можна застосовувати у майбутньому для прийняття рішення про видачу кредиту новому клієнту.

Дослідивши дані одного з українських банків, виявлено, що найкраще описує імовірність протермінування кредиту множинна лінійна кореляційно – регресійна модель, яка має вигляд:

$$y = 0.00005629 x_2 + 0.000003851 x_7 + 0.001609 x_8 + 0.002667 x_9 - 0.0000007049 \ln(x_1) - 0.02523 \ln(x_2) + 0.003413 \ln(x_7) + 0.00001575 x_6 + 0.09713 \ln(x_6) - 0.009692 \sqrt{x_1} - 0.0009316 \sqrt{x_7} - 0.005445 \sqrt{x_6} + 0.002149 x_{10} - 0.004841 x_{22} - 0.007683 x_{23} + 0.000839 x_{24} - 0.009641 x_{71} - 0.0154 x_{72} - 0.006905 x_{73} + 0.1208 x_{41} + 0.1832 x_{42} + 0.0514 x_{31} + 0.117 x_{32} + 0.1668 x_{33} + 0.1564 x_{34} + 0.1791 x_{51} + 0.3002 x_{52} + 0.3503 x_{53} + 0.002017 x_1 - 0.4119.$$

Залежно від сфери застосування, скоринг може служити для розв'язання різних задач, але його

завдання будуть схожими – визначити тих клієнтів, співпраця з якими принесе компанії найбільший прибуток або дозволить мінімізувати рівень ризиків.

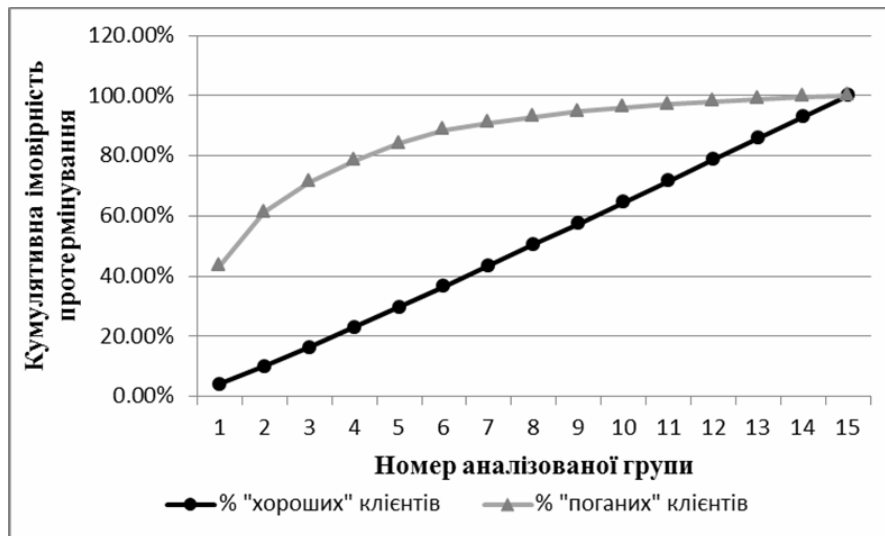


Рис. 6. Ілюстрація якості диференціації клієнтів прогнозною моделлю

Висновки

Проведення скорингу дозволяє підприємствам диверсифікувати ризики, пов'язані з обробленням великої кількості однотипних даних з інформацією про клієнтів, і, як наслідок, збільшити прибуток або мінімізувати можливі втрати підприємства. Одним з можливих варіантів застосування скорингу є використання моделі бінарного вибору, яка дозволяє передбачити імовірність настання певної події, використовуючи ряд вхідних характеристик. Методика що наведена у даній праці дозволяє спрогнозувати вихід позичальника банку у розстрочку на термін більше 90 днів, беручи до уваги його персональні дані. Застосування математичного інструментарію для аналізу бази даних клієнтів, споживачів чи пацієнтів, в залежності від сфери роботи установи, відкриває нові можливості щодо формування стратегій роботи з великими вибірками людей, що в майбутньому покращить ефективність роботи підприємств та підвищить якість надаваних ними послуг.

Література

1. Здрок В. В. Основи економетричних досліджень : навч. посіб. / В. В. Здрок. – Львів : ЛНУ імені Івана Франка, 2013. – 358 с.
2. Lyn C. Thomas, David B. Edelman, Jonathan N. Crook. Credit Scoring and its Applications (SIAM monographs on mathematical modeling and computation), 2002. 248 p.

References

1. Zdrok V. V. Osnovy ekonometrychnykh doslidzhen: navch. posib. / V. V. Zdrok – Lviv: Lviv National University of Ivan Franko, 2013. – 358 p. [in Ukrainian]
2. Thomas, L. C. Credit Scoring and its Applications / Lyn C. Thomas, David B. Edelman, Jonathan N. Crook. p. cm. - (SIAM monographs on mathematical modeling and computation), 2002 – 248 p.

Надійшла 08.11.2014; статтю представляє к. т. н. Здрок В. В.