

Міністерство освіти і науки України
Сумський державний університет
Навчально-науковий інститут бізнес-технологій «УАБС»
Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

на тему: «МОДЕЛЮВАННЯ ПРОЦЕСУ ОЦІНЮВАННЯ
КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКІВ БАНКУ»

Виконала студентка 2 курсу, групи Ек.м.-81а

(номер курсу)

(шифр групи)

Спеціальності 051 «Економіка» («Економічна
кібернетика»)

Баранова А.О.

(прізвище, ініціали студента)

Керівник професор, д.е.н. Олійник В.М.

(посада, науковий ступінь, прізвище, ініціали)

ЗМІСТ

ВСТУП.....	7
РОЗДІЛ 1 СУЧАСНІ ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ ТА МОЖЛИВОСТІ ЇХ ВИКОРИСТАННЯ ДЛЯ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА БАНКУ	9
1.1 Сутність оцінювання кредитоспроможності позичальника банку як складової кредитної діяльності банку.....	9
1.2 Сучасний стан моделювання оцінювання кредитоспроможності позичальника та обґрунтування доцільності застосування кредитного скорингу.....	19
1.3 Змістовна постановка задачі моделювання.....	26
РОЗДІЛ 2 РОЗРОБКА СКОРИНГОВОЇ МОДЕЛІ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА БАНКУ	28
2.1 Загальні вимоги до моделі оцінювання кредитоспроможності позичальника банку.....	28
2.2 Ознаковий простір моделі оцінювання кредитоспроможності позичальника банку.....	30
2.3 Математичний опис моделі оцінювання кредитоспроможності позичальника банку.....	39
РОЗДІЛ 3 ПРАКТИЧНА РЕАЛІЗАЦІЯ СКОРИНГОВОЇ МОДЕЛІ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА БАНКУ	47
3.1. Програмна реалізація моделі та апробація на реальних даних	47
3.2 Аналіз отриманих результатів та перевірка адекватності розробленої моделі.....	53
3.3 Оцінка очікуваних ефектів від впровадження скорингової моделі	59
ВИСНОВКИ	61
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	63
ДОДАТКИ	67

ВСТУП

Останнім часом в Україні спостерігається певний бум банківського споживчого кредитування. Причому відбувається зростання не лише абсолютних величин кредитів, наданих банками фізичним особам, а й питомої ваги споживчих кредитів у загальній сумі банківських кредитів. Це свідчить про позитивні тенденції у сфері банківського споживчого кредитування. Разом із тим слід зазначити, що вітчизняні банки зіткнулись із проблемою неповернення населенням отриманих кредитів. Це підкреслює актуальність розроблення методик оцінювання кредитоспроможності фізичних осіб і ризику банків при наданні споживчих кредитів. Адекватним заходом, що знижує ризик банку та дозволяє оптимально вирішувати завдання, є кредитний скоринг, що являє собою математичну або статистичну модель, за допомогою якої банк визначає, наскільки велика ймовірність, що цей потенційний позичальник поверне кредит у встановлений строк.

З огляду на це, метою даного дослідження є розробка моделі оцінювання кредитоспроможності позичальника-фізичної особи на основі підходу кредитного скорингу.

Об'єктом дослідження є кредитна діяльність банку.

Предметом дослідження є методи і моделі оцінювання кредитоспроможності позичальника-фізичної особи.

Для досягнення поставленої мети необхідно виконати ряд завдань:

– розглянути сутність оцінювання кредитоспроможності позичальника банку як складової кредитної діяльності банку;

– провести огляд сучасного стану моделювання оцінювання кредитоспроможності позичальника та обґрунтувати доцільність застосування кредитного скорингу;

– сформулювати змістову постановку задачі моделювання;

- сформувати загальні вимоги до розроблюваної моделі оцінювання кредитоспроможності позичальника банку;
- описати ознаковий простір моделі, описати вхідні та вихідні дані моделювання;
- описати математичну модель оцінювання кредитоспроможності позичальника банку;
- зробити практичну реалізацію скорингової моделі оцінювання кредитоспроможності позичальника банку;
- проаналізувати адекватність, точність та чутливість розробленої моделі;
- розробити пропозиції з використання розробленої моделі та шляхів її подальшого удосконалення.

РОЗДІЛ 1 СУЧАСНІ ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ ТА МОЖЛИВОСТІ ЇХ ВИКОРИСТАННЯ ДЛЯ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА БАНКУ

1.1 Сутність оцінювання кредитоспроможності позичальника банку як складової кредитної діяльності банку.

Розвиток позик для фізичних осіб збігається зі створенням та розвитком банківської справи, відіграє важливу роль у розвитку громадянських відносин на ринку фінансової безпеки та споживання та виконує свою первинну функцію щодо розширення можливостей населення та рівня життя.

З огляду на специфічний характер роздрібної позики та характер позики в цілому, можна зробити висновок, що банківська позика фізичним особам є типом економічних відносин між банком та позичальником, що включає надання банком тимчасових коштів на умовах повернення, забезпечення та платності [29].

В Україні популярність кредитування фізичних осіб в останні роки дуже висока. Попит на фінансові ресурси, запозичені громадянами, та їх постачання банківськими фінансовими установами зростає.

В умовах добре функціонуючої економіки міжбанківська конкуренція є жорсткою для кредитних послуг клієнтів. Можна виділити наступні особливості банківських позик для фізичних осіб:

- позичальник вільний вибрати банк, у якому він хотів би отримати кредит. Особа має право одночасно брати позики у різних банків;
- кредитний банківський кредит надається після закінчення терміну дії кредитного договору;
- будь-які проблеми, що виникають із позик, вирішуються шляхом прямої угоди між банком та кредитором;

- за умови домовленості, кожен суб'єкт господарювання у кредитних відносинах бере на себе конкретні обов'язки;
- якщо вони позичають позикодавцю більше 10% власного капіталу («великі позики»), комерційний банк повідомляє про такі випадки Національний банк України.

Основними видами банківських позик фізичним особам є позики на придбання житла (заставу), на придбання транспортних засобів (автомобілів) та на придбання товарів тривалого користування (споживача) [25].

За даними НБУ, за 9 місяців 2019 року загальний обсяг кредитів, виданих банками українським позичальникам, виріс на 4,5 мільярда гривень або на 2,4% – до 192,7 мільярда гривень. Для порівняння: за 2018 рік обсяг кредитів українців зменшився на 11 мільярдів або 5,5%. Більше того – досі, з 2015 року, обсяг кредитування фізичних осіб щорічно скорочувався (рис. 1.1).

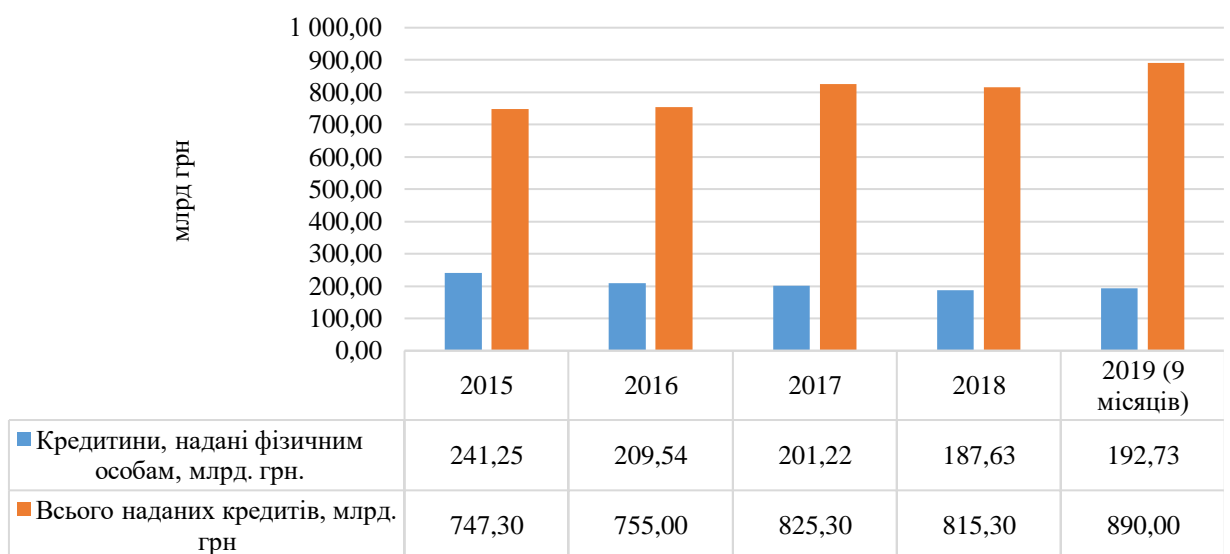


Рисунок 1.1 – Динаміка порівняння обсягів всього наданих кредитів та кредитів, наданих фізичним особам, млрд грн.

При цьому різні напрямки кредитування фізичних осіб (споживче, житлове та позики на авто) демонстрували різну, а іноді і зовсім протилежну

динаміку. НБУ констатує, що споживче кредитування продовжує зростати та зберігає надзвичайну привабливість для банків. Воно вже становить понад половину кредитного портфеля деяких із них. Наведемо дані, які вказують на серйозне зростання чистих кредитів, виданих фізособам. Якщо наприкінці 2016 року їх загальний обсяг у країні становив понад 50 млрд грн, то за кілька років вже перевищив рівень 100 млрд грн. Так у 2017 році цей показник сягнув позначки 87 млрд грн, що на 7,8 % більше ніж за минулий рік. У 2018 році обсяг споживчих кредитів дорівнює 111 млрд грн. Зазначимо, що за чотири перші місяці 2019 року цей показник упритул підібрався до позначки 120 млрд грн, а обсяг іпотеки, за той же період за даними Національного банку України, мав такі показники. У 2016 році обсяг наданих кредитів на житло дорівнював 1,2 млрд грн, у 2017 році даний показник зменшився та сягнув позначки у 0,9 млрд грн. Станом на 2018 рік обсяги іпотечного кредитування збільшилися на 8.1% (порівняно з 2017 роком). За 2018 рік було видано нових іпотечних кредитів на суму, що перевищує 2,4 мільярди гривень. Для того, щоб доцільно порівняти показники оберемо дані обсягу іпотек за період перших чотирьох місяців 2019 року, а саме 1,2 млрд грн (рис. 1.2).

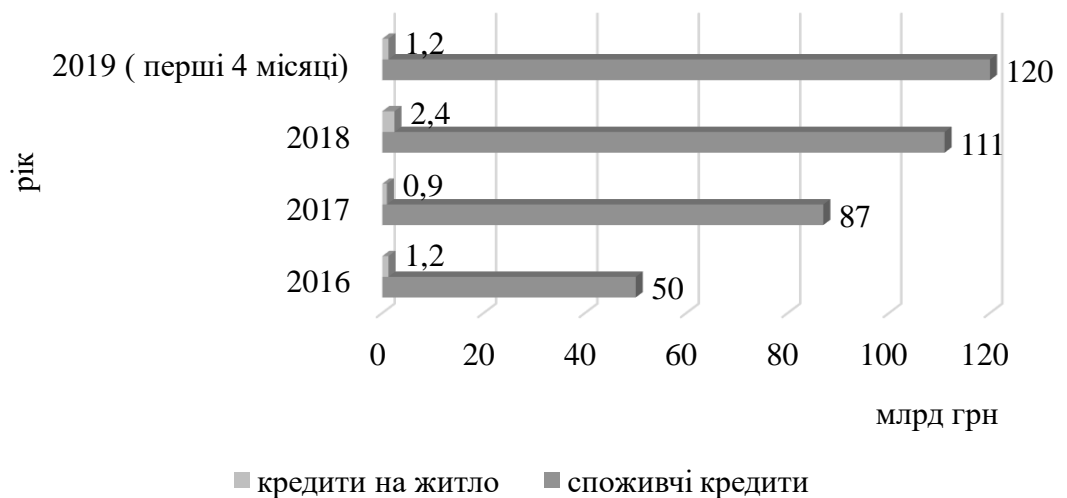


Рисунок 1.2 – Обсяг кредитування фізичних осіб українськими банками, млрд грн.

У 2018 році кредити на поточні потреби становлять 70% роздрібного портфеля (ще 20% – на придбання нерухомості, 7% – автокредити та 3% – на інші цілі). Одразу після кризи зростання кредитування значною мірою пояснювалось відкладеним споживчим попитом. Наразі цей фактор не настільки вагомий. Визначальними є високі споживчі настрої та жваве зростання доходів населення. Перші спонукають населення витратити майбутні доходи на задоволення поточних потреб. Другі створюють впевненість, що кредит не обтяжить сімейний бюджет. Високий попит на кредитні кошти дає змогу банкам зберігати високі кредитні ставки. Через поширену кризову ситуацію в банківській системі України кредитна ситуація змінилася. Це стосується всіх видів позик для фізичних осіб: іпотечних, автокредитів та споживчих кредитів [7].

Утім, стандарти кредитування теж пом'якшилися: згідно з опитуваннями про умови кредитування після кризи, банки значно знизили вимоги до позичальників. Від середини 2018 року вони знову збільшили вимоги до позичальників, але досі значно м'якші, ніж наприкінці 2015 року.

Банки збільшили аванси, збільшили терміни подання заявок та зменшили суми кредитів. Значно більші вимоги до визначення рівня доходів позичальника - тепер банки потребують офіційного підтвердження доходу. Крім того, збільшилися суми, зібрані з комісій.

Існуючі правила не сприяють повному врегулюванню відносин між кредиторами та позичальниками – фізичними особами, по-перше, через відсутність універсального нормативно-правового акту, який регулює позики, а, по-друге – спеціального регулюючого акту у сфері позик фізичним особам [13].

Факторами, що перешкоджають розвитку кредитів для фізичних осіб в Україні, є зниження довіри до банків та зниження попиту з боку населення, високі процентні ставки за кредитами через приховані витрати, збільшення вартості позичальника, зростання інфляції та знецінення, правові способи захисту прав кредитора та кредитора.

Аналізуючи кредитування фізичних осіб, слід також зазначити, що вдосконалення його організаційного, фінансового та економічного підходу дозволить скористатись великими можливостями цього виду кредитування для здійснення та підвищення соціального добробуту населення, особливо при купівлі житло, а саме:

- розширення та вдосконалення організаційної, правової, фінансової та економічної бази банківських позик для фізичних осіб;
- забезпечення прозорості кредитних схем, їх дискреційну динаміку та підвищує надійність позик за допомогою заходів, що сприяють розвитку цього виду позик;
- розроблення методів оцінки кредитної вартості кредитора та зменшити кредитний ризик для фізичних осіб.

Реалізація цих заходів, а також заходів щодо імплементації конкретного законодавства сприятимуть розширенню масштабів та якості банківського кредитування як ключового засобу прискорення темпів соціально-економічного розвитку країни [2].

Найбільш поширений в банках є підхід до зниження кредитного ризику - це оцінка кредитної вартості кредитора. У банківській практиці метод, заснований на оцінці кредитора, стає все більш широким. Критерії, за якими кредитор зазвичай оцінюється індивідуально для кожного банку, ґрунтуються на їх практичному досвіді та регулярно переглядаються.

Усі банки зобов'язані оцінювати кредитоспроможність кредиторів при видачі кредиту. Для кредитора ця кредитна оцінка також важлива, оскільки вона залежить від того, на скільки надана позика та чи вона взагалі надана.

Результатом систематично неправильно здійснюваного оцінювання кредитоспроможності позичальників може бути погіршення якості кредитного портфеля банку, яке, до того ж, спричинює ще й необхідність зайвого резервування [1].

Аналіз кредитоспроможності як головного критерію в побудові оперативних кредитних відносин між банком і позичальником полягає в наступному [24]:

- оцінювання кредитоспроможності позичальників повинно проводитися, виходячи зі сумісних інтересів суб'єктів кредитних відносин. Занадто високі вимоги до позичальників кредитних ресурсів зі сторони банку призводять до втрати клієнтів;

- спільність мети позичальника та кредитора в процесі формування кредитних відносин виражається в прагненні досягти ними обома максимальної прибутковості;

- кредитоспроможність позичальників міцно пов'язана з вартістю надання кредитних послуг, видами позик, формами кредитів, залежить від рівня розвитку інфраструктури даних ринків, попиту і пропозиції на кредитні ресурси на ринках.

Існують різні підходи щодо визначення поняття кредитоспроможності.

Так, О. В. Дзюблук розглядає кредитоспроможність як спроможність позичальника в повному обсязі й у визначений кредитною угодою термін розраховуватися за своїми борговими зобов'язаннями винятково грошовими коштами [26].

Л. П. Белих зазначає, що кредитоспроможність – це спроможність компанії або приватної особи залучати позичковий капітал і в майбутньому належним чином обслуговувати свій борг [30].

В. Галасюк вважає, що кредитоспроможність – це здатність позичальника за конкретних умов кредитування в повному обсязі й у визначений кредитною угодою термін розрахуватися за своїми борговими зобов'язаннями виключно грошовими коштами, що генеруються позичальником у ході звичайної діяльності. Отже, для банку, який надає позику, важливо, щоб позичальник був не лише платоспроможним, а й кредитоспроможним [28].

Між кредитоспроможністю і платоспроможністю існує істотна відмінність. Кредитні зобов'язання позичальника перед банком, на відміну від інших зобов'язань, повинні бути погашені виключно грошовими коштами.

Тому для банків, які надають кредит, недостатньо того, щоб позичальники були лише платоспроможними, також йому необхідно бути кредитоспроможним. У зв'язку з цим виникає необхідність використання поняття кредитоспроможності для характеристики здатності позичальника сплачувати свої зобов'язання перед кредитором-банком лише грошовими коштами [9].

Платоспроможність – це можливість позичальника своєчасно та в повній мірі здійснювати розрахунки за всіма видами своїх зобов'язань.

Якщо розглядати такі поняття як платоспроможність та кредитоспроможність, то слід зазначити, що вони мають ряд відмінностей.

Дані відмінності представлені на рисунку 1.3.

На законодавчому рівні визначення поняттю «кредитоспроможність» надано у Положенні про порядок формування та використання банками України резервів для відшкодування можливих втрат за активними банківськими операціями №23 від 25.01.2012 [31].

Згідно з Положенням кредитоспроможність - наявність у позичальника банку передумов для отримання кредиту та його здатність повернути кредит і проценти за ним у повному обсязі та в обумовлені договором строки. До основних передумов для отримання кредиту відносяться юридична самостійність суб'єктів господарювання, дієздатність з правового погляду фізичних осіб, отримання позичальником регулярних доходів, наявність заставного майна.

Для банку оцінка кредитоспроможності позичальника є одним з методів мінімізації ризику в кредитній діяльності. Конкретні висновки та пропозиції за результатами оцінки кредитоспроможності позичальників дають змогу ухилитися від невиправданих ризиків при здійсненні кредитних операцій. Достовірність оцінки є важливою також і для позичальника, саме

від неї буде залежати рішення про отримання кредиту та про його можливий обсяг [8].



Рисунок 1.3 – Відмінності понять «кредитоспроможність» і «платоспроможність» [18] .

Якісна оцінка кредитоспроможності є тим напрямом, оптимізація якого дасть максимальний результат щодо мінімізації кредитного ризику [19].

Розглянемо види кредитоспроможності за основними ознаками. Дані представлено у таблиці 1.1.

Необхідно зауважити, що до основних класифікаційних властивостей можна віднести: строк запозичених коштів, статус кредитора та позичальника, регіональну та галузеву приналежність, масштаб діяльності. Така класифікація дозволяє врахувати специфіку різних аспектів кредитоспроможності при прийнятті конкретного кредитного рішення.

Таблиця 1.1 – Види кредитоспроможності за основними ознаками [20] .

За строком запозичення коштів	комерційна кредитоспроможність
	інвестиційна кредитоспроможність
За статусом кредитора	кредитоспроможність, що оцінюється приватною особою
	кредитоспроможність, що оцінюється підприємством
	кредитоспроможність, що оцінюється кредитною організацією
	кредитоспроможність, що оцінюється державою
За регіональною приналежністю позичальника	кредитоспроможність національного позичальника
	кредитоспроможність іноземного позичальника
За статусом позичальника	кредитоспроможність корпоративного позичальника
	кредитоспроможність індивідуального позичальника
За масштабом діяльності позичальника	кредитоспроможність великого позичальника
	кредитоспроможність середнього позичальника
	кредитоспроможність малого позичальника
За сферою діяльності корпоративного позичальника	кредитоспроможність виробничого підприємства
	кредитоспроможність підприємства, працюючого у сфері торгівлі та послуг
	кредитоспроможність фінансової організації

Інформаційне забезпечення як юридичних осіб, так і фізичних осіб представляє собою процес підбору необхідних показників для проведення достовірної оцінки та аналізу всіх напрямків діяльності позичальника [34].

Аналіз кредитоспроможності клієнтів здійснюється з метою отримання даних для того, щоб прийняти рішення про можливість й умови позики. Глибина оцінки залежить від відсутності або наявності кредитних відносин банківської установи з конкретним клієнтом, від результату надання ним достовірної та в повному обсязі інформації. В умовах ринкової економіки оцінка кредитоспроможності позичальника є базовою характеристикою, що розкриває можливість встановлення кредитних відносин між банком та клієнтом [35].

На сьогодні в економіці України спостерігаються негативні тенденції щодо розвитку банківського кредитування. Надання кредитів та проведення кредитної політики є найпоширенішими і найризикованішими операціями банківських установ, хоча і приносять їм основну частину доходу. Однак аналіз ситуації, що склалася у банківській сфері свідчить, що світова економічна криза суттєво вплинула на процес кредитування і принесла із

собою низку проблем, що стосуються надання кредитів як фізичним, так і юридичним особам по всій території України. Однією із таких проблем можна вважати підвищену зацікавленість в отриманні кредиту з боку шахраїв. Застереженням таким вчинкам можна вважати єдину інформаційну систему «Реєстр позичальників» (далі – ЄІС «Реєстр позичальників»), яка була створена за ініціативою Національного банку України з метою зменшення ризику неповернення комерційним банкам кредитних коштів [6].

ЄІС «Реєстр позичальників» створений для зміцнення довіри до банків, зміцнення довіри до банків та банківської системи, зменшення ризиків, що виникають внаслідок обслуговування клієнтів, та діє добровільно (контракт).

Основне завдання системи – забезпечити облік тих позичальників, які мають прострочені борги перед банками. Це запобігає навмисному несплаті кредиту, що у свою чергу значно знижує кредитний ризик. Ця система забезпечує збереження банківської інформації клієнтів та доступ до неї уповноваженого банківського персоналу. Однак ця система має ряд недоліків і не працює на повну потужність.

«Реєстр кредиторів» ЄІС допомагає працівникам банку у вирішенні питання про затвердження чи відхилення кредитування фізичних та юридичних осіб. Ця база даних є найбільш повною і має найбільшу кількість боржників.

Основна мета банківської установи - знати, як отримати прибуток. Це означає, що банкам потрібно збільшувати свої кредитні записи, видаючи позики. Бажання залучити нових клієнтів, відігравати більшу роль на кредитному ринку, не збільшуючи їх кредитний ризик, може бути найважливішою метою банку. Тому виникають суперечності, оскільки необхідність скоротити час їх обслуговування привертає все більше клієнтів, що може призвести до поганої оцінки кредитної вартості клієнта. Наслідком такої ситуації є підвищений ризик запозичення.

У банківській практиці застосування методів кредитного скорингу набуло поширення при оцінці кредитоспроможності як нового клієнта, так і уже співпрацюючого з банком позичальника.

Кредитний скоринг – це метод оцінки кредитоспроможності позичальника, який включає систему балів для аналізу якісних та кількісних показників його діяльності.

Важливою особливістю системи оцінювання кредитів є те, що вона не може застосовуватись на основі одного і того ж паттерну, а повинна бути розроблена на основі особливостей банківської установи та клієнтів банку з урахуванням банківського законодавства та національних традицій. Тобто систему кредитного скорингу слід регулярно переглядати та змінювати.

1.2 Сучасний стан моделювання оцінювання кредитоспроможності позичальника та обґрунтування доцільності застосування кредитного скорингу.

Найважливіший крок управління кредитним ризиком - це оцінити його під час позики, що є фазою прихованого ризику. Необхідність оцінки кредитного ризику пов'язана з прагненням кредитора мінімізувати ризик.

Існують прямі та непрямі методи аналізу надійності клієнтів у банківській практиці.

Прямі прийоми застосовуються рідко. Згідно з ними сума балів, зароблених замовником, фактично дорівнює сумі отриманої ним позики.

Непрямі методи поширені. Його суть полягає у наданні певної ваги різним показникам оцінки, а результатом оцінки є зняття класу кредитоспроможності клієнта.

Сьогодні кредити в Україні - одна з найважливіших галузей економіки країни. Корпоративні позики - найпоширеніший засіб фінансування у разі браку фінансів. При цьому банківські позики завжди пов'язані з ризиком неповернення позикових коштів та процентних ставок за їх використання.

Кредитний аналіз є одним із факторів, який допомагає банку зменшити кредитний ризик. Однак комерційні банки в Україні не мають єдиного методологічного підходу до аналізу кредитоспроможності позичальників [36, 37, 38].

Серед існуючих зарубіжних методів і моделей оцінки кредитоспроможності виділяють наступні:

- класифікаційні (статистичні) методи оцінки, до яких належать рейтингові (бальні) системи оцінки та моделі прогнозування банкрутства;
- моделі комплексного аналізу (на основі “напівемпіричних” методів, тобто які базуються на експертних оцінках аналізу економічної доцільності надання кредиту): “правило п’яти Сі”, PARTS, CAMPARI, MEMO RISK,
- PARSER, система 4FC.

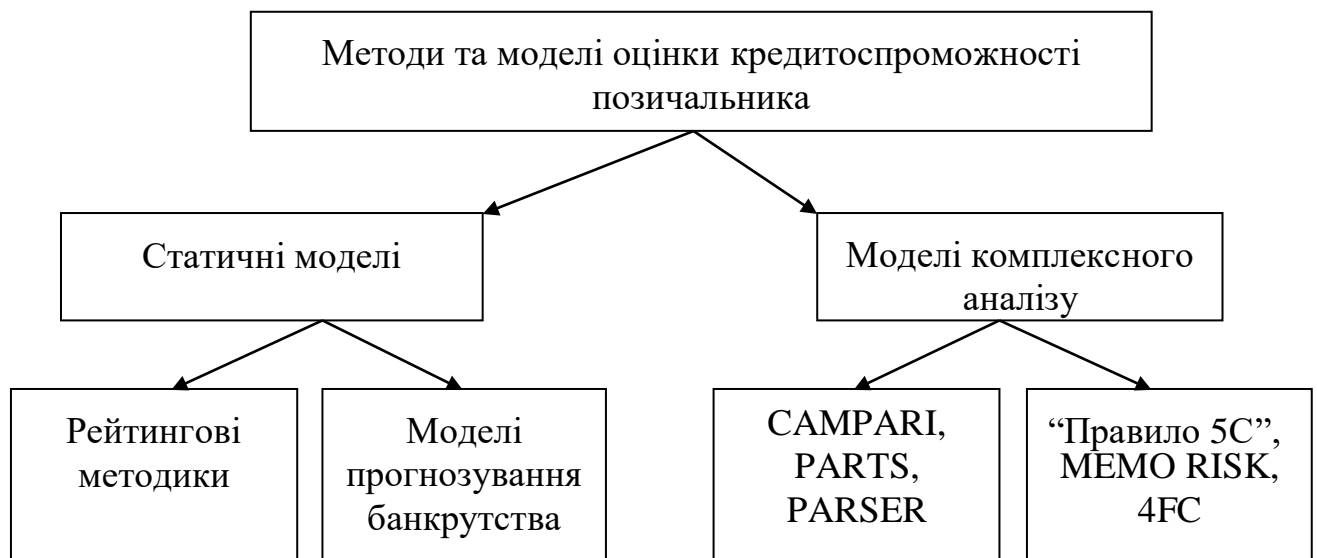


Рисунок 1.4 – Класифікація методів і моделей оцінки кредитоспроможності позичальників.

Одним із найпоширеніших методів, що застосовуються в міжнародному банкінгу для визначення кредитоспроможності позичальника, є система рейтингу, яка враховує якісні та кількісні характеристики клієнта.

Надійність вимірюється за допомогою синтезованого показника - рівень класифікації ризику позичальника залежить від рейтингу та кількості

набраних балів. Системи класифікації, насамперед, забезпечують вибір та обґрунтування системи індикаторів та її класифікацію. Ці групи класифікуються відповідно до важливості кредитоспроможності банку в оцінці банку. Рейтингові системи відображають специфічний підхід банку до оцінки якості кредиту і можуть суттєво відрізнятися.

Основні вимоги щодо видачі кредитів англійськими банками, сконцентровані у слові PARTS: Purpose – мета кредиту; Amount – розмір кредиту; Repayment – оплата; Term – термін; Security – застава [1, 2, 4].

Крім того, банки роблять оцінку можливого ризику неплатежу по кредиту із використанням методики PARSEL, що є досить схожою і включає: Person – репутація позичальника; Amount – обґрунтування суми кредиту; Repayment – погашення; Security – забезпечення; Expediency – доцільність кредиту; Remuneration – винагорода банку (% ставка) за ризик надання кредиту.

Також досить популярним є принцип С.А.М.Р.А.Р.І. Методика CAMPARI полягає у почерговому виділенні з кредитної заявки і фінансових документів, що додаються, найсуттєвіших чинників, які визначають діяльність клієнта в їх оцінці та уточненні після особистої зустрічі з клієнтом.

У практиці американських банків застосовують «правило п'яти Сі»:

- 1С – customer's character (характер позичальника) – репутація позичальника, ступінь відповідальності, готовність і бажання погашати борг. Джерелами інформації є бесіда з позичальником, досьє з архіву, консультації з іншими банками і фірмами та ін.;

- 2С – capacity to pay (фінансові можливості) – припускає ретельний аналіз доходів і витрат позичальника та перспектив зміни їх у майбутньому. Аналізуються касові надходження, рух активів, обсяг заборгованості, динаміка кредиторської заборгованості, зміни товарних запасів;

- 3С – capital (капітал, майно) – аналіз власного (акціонерного) капіталу фірми, його структури, співвідношення з іншими статтями активів і пасивів;

- 4C – collateral – забезпечення позики, достатність, якість і ступінь реалізованості застави у випадку непогашення позики;
- 5C – current business conditions and goodwill (загальні економічні умови) – визначають діловий клімат в країні і впливають на стан як банку так і позичальника (конкуренція на ринку, податки, ціни на сировину та ін.).

У Франції для загальної оцінки підприємства на предмет кредитування використовується методика, до якої входить фінансово-економічна оцінка, що включає характер діяльності і тривалість функціонування підприємства та фактори виробництва: трудові ресурси, виробничі ресурси, фінансові ресурси, економічне середовище [5].

Німецькі банки використовують метод кредитного рейтингу, включаючи оцінку кредитного рейтингу компанії та його гарантій. В одному з німецьких банків кредитний рейтинг клієнта визначається за 17 критеріями, які поділяються на 5 груп:

- менеджмент: якість управління, грамотність у фінансових документах;
- ринок (галузь): ринковий і промисловий розвиток, вплив ринкової динаміки, географія покупців та постачальників, ризики експорту та імпорту, сильна конкуренція, асортимент продукції та стандарти;
- взаємовідносини з клієнтом: ведення записів, готовність клієнта та інформація про його діяльність;
- фінансові умови: річний баланс, оцінка загального майнового стану;
- перспектива розвитку компанії: розвиток, планування виробництва, планування рентабельності та майбутнє обслуговування капіталу, конкретні виробничі ризики після останнього балансу.

Існує чимало методик аналізу фінансового стану клієнта та його надійності з точки зору вчасного повернення боргу. В Україні НБУ також розробляє для комерційних банків рекомендації щодо визначення фінансового стану і кредитоспроможності позичальників. Згідно з "Положенням про порядок формування та використання банками України

резервів для відшкодування можливих втрат за активними банківськими операціями" від 25.01.2012 р. № 23 [31], критерії оцінки фінансового стану позичальника встановлюються кожним банком самостійно його внутрішніми положеннями щодо проведення активних операцій, у яких мають бути визначені ґрунтовні, технічно виважені критерії економічної оцінки фінансової діяльності позичальників. Методика проведення оцінки фінансового стану позичальника, яка розроблена банком, є невід'ємним додатком до внутрішньобанківського положення про кредитування [27].

Тим не менш, на сьогодні вітчизняні комерційні банки не мають стабільної і дієвої методики визначення кредитоспроможності позичальників. У той же час західні банки широко практикують скорингові системи для визначення кредитоспроможності та прийняття рішення про надання йому позики [11].

В таблиці 1.2 представлено переваги та недоліки найпоширеніших методик оцінки кредитоспроможності.

Отже, розглянуті підходи до визначення оцінки кредитоспроможності є досить оригінальними, проте їм притаманний описовий характер, що обмежує їх практичне застосування. Ми будемо використовувати кредитний скоринг.

Скоринг – це математична модель у вигляді зваженої суми певних характеристик, за допомогою якої на основі минулого досвіду банк намагається з'ясувати ймовірність того, що конкретний позичальник не поверне кредит.

В основі оцінки лежить вивчення кредитної історії позичальників, які вже брали позики, класифікація та визначення характеристик надійних і бідних клієнтів для погашення кредитного боргу.

Таблиця 1.2 – Переваги та недоліки методик оцінювання кредитоспроможності.

Назва методики	Сутність методики	Переваги	Недоліки
Класифікаційні (статистичні методики)	Оцінюються всі показники, які характеризують кредитоспроможність позичальника. Ці показники можуть бути проаналізовані, як в даний момент, так і за певний період (в динаміці).	Широка вживаність та об'єктивність.	Необхідність значного масиву вихідних даних. Метод не враховує слабоформалізовані якісні чинники кредитоспроможності та умов кредитування. Неможливість застосування до умов української економіки.
Назва методики	Сутність методики	Переваги	Недоліки
Комплексного аналізу	Визначається значимість кожного з показників кредитоспроможності позичальника в балах.	Дозволяє швидко і з мінімальними витратами праці обробити великий обсяг кредитних заявок. За рахунок цього суттєво зменшуючи рівень операційних банківських витрат, пов'язаних з кредитною діяльністю.	Використання пов'язане із деякими труднощами: бальні системи оцінки кредитоспроможності клієнта вимагають постійного оновлення інформації, що може бути коштовним для банку. Орієнтовані на великі банки, які можуть дозволити собі розробку власних моделей аналізу кредитоспроможності. Стандартний характер цих моделей не враховує специфічних особливостей окремих позичальників [33].

Вперше в 1941 році американський економіст Д. Дюран розробив технологію кредитування. Він виділив групу факторів, які, на його думку, могли б достатньо визначити рівень кредитного ризику при отриманні споживчого кредиту. Кожна банківська установа самостійно вибере перелік найважливіших факторів та призначить їм відповідні коефіцієнти. Результатом реалізації методу є інтегрований показник, який порівнюється з

певним пороговим числом або роздільною лінією, яка є прямою. Клієнтам з інтегрованою оцінкою вище цього порогу надається кредит, а клієнтам, які мають бали нижче рівня рівноваги, відмовляють у позиках. Все здається дуже простим, але важко визначити, які особливості слід включати в модель і які ваги повинні відповідати їм.

При кредитуванні фізичних осіб також проводиться процедура оцінки їх кредитоспроможності, яка може здійснюватися на підставі рівня доходу позичальника, вивчення його кредитної історії, а також стандартизованої скорингової оцінки (особисті якості позичальника) [39].

Безпосередній розрахунок інтегрованої скорингової оцінки позичальників банку передбачає розробку скорингової карти. В карту включають усі основні кількісні та якісні показники, які визначають кредитоспроможність позичальників.

Ключові переваги від впровадження скорингової системи:

1. Скорочення термінів затвердження рішення про позику. Збільшення кількості та швидкості опрацювання запитів, щоб мінімізувати рахунок, пропонуючи кредити приватним клієнтам, як найважливіший спосіб забезпечити прибутковість роздрібних кредитів.

2. Ефективна оцінка та порівняльна оцінка конкретних позицій.

3. Зменшення поточних суб'єктивних факторів. Пропонування функціональної оцінки, кредитних інспекторів з усіх філій та відділень банку.

4. Фізичні особи в цілому банку, включаючи його відділення, зараховуються за оцінку та управління відповідним портфелем. Останні зареєстровані позики однакові і зберігаються в кредитному портфелі.

5. Запровадження унікального методу оцінки позичальника за різними видами банківських позичкових продуктів (експрес-позики, кредитні картки, споживчі кредити, автокредити, іпотека).

6. Коригування параметрів кредиту в межах позитивного продукту (коригування кредитного продукту).

7. Різні розширення, пов'язані з коригуванням кредиту, складним та чітким кредитом.
8. Зменшення кількості персоналу, фінансів персоналу нижчих професій.
9. Усі етапи переглядаються програмами.
10. Можливість закріплення методології комітології централізована і негативно реалізована у всіх відділах [32].

Проте, незважаючи на значні перспективи, використання та зручність, скорингові системи оцінки кредитоспроможності мають і недоліки. Можна виділити дві основні проблеми, що пов'язані з використанням скорингу.

Перша стосується того, що для розробки моделі використовуються лише дані про клієнтів, яким надали кредит. Невідомо, як повів би себе клієнт, якому у кредиті було відмовлено.

Друга полягає в тому, що з часом змінюються люди і соціально-економічні умови, що впливають на їхню поведінку. Тому скорингові моделі повинні постійно коригуватись і вдосконалюватись на основі нової вибірки позичальників. Цим і обумовлена мета даного дослідження, а саме – розробка моделі оцінювання кредитоспроможності позичальника-фізичної особи на основі підходу кредитного скорингу.

1.3 Змістовна постановка задачі моделювання.

Мета розробки моделі, яка б оцінювала рівень кредитоспроможності фізичної особи, полягає в зниженні ризику неповернення коштів банку, який виникає при кредитуванні фізичних осіб.

У сучасній світовій практиці оцінки рівня кредитоспроможності набуває широкого використання такий інструмент, як скорингова модель. Модель об'єднує у систему визначені критерії, створюючи своєрідний «портрет ідеального позичальника», що може бути використаний для ефективності оцінки кредитоспроможності.

Отже, в результаті проведення даного дослідження необхідно розробити модель, яка б дозволяла оцінювати рівень кредитоспроможності позичальників банку, а саме фізичних осіб.

Користувачами розроблюваної моделі можуть бути співробітники банку під час прийняття рішення про видачу кредиту, що дозволить застосовувати уніфікований та позбавлений суб'єктивізму підхід у всіх відділеннях банку та зменшити ризик неповернення кредитних коштів.

Результатом застосування даної моделі для конкретного позичальника-фізичної особи буде визначення його рівня кредитоспроможності.

У нашому випадку доцільним є моделювання методами логістичної регресії. Кредитоспроможність оцінюють якісними та кількісними показниками, тому звичайні арифметичні операції застосувати не можна. Виходячи з цього ми обрали один з підходів до оцінки кредитоспроможності – скорингову модель.

РОЗДІЛ 2 РОЗРОБКА СКОРИНГОВОЇ МОДЕЛІ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА БАНКУ

2.1 Загальні вимоги до моделі оцінювання кредитоспроможності позичальника банку.

Як вже згадувалося раніше, досить нагальним завданням комерційних банків є прийняття рішення про можливість кредитування за короткий час. Одним з можливих інструментів отримання кредитного рейтингу позичальника є модель балів.

Успіх моделі балів визначається декількома важливими факторами:

- неупередженість в оцінці (бали виключають суб'єктивну оцінку);
- стандартизація кредитних рейтингів;
- автоматизація;
- контроль (через стандартизацію кредитних операцій контролювати ефективність кредитних рішень не дуже важко);
- підвищення прибутковості (автоматизація процесів знижує витрати на ручну обробку заявок на позику).

Концептуальна схема оцінювання кредитоспроможності позичальника банку представлена на рис. 2.1.

Основою для прийняття рішення про видачу кредиту клієнту є характеристики останнього. Залежно від конкретних значень виокремлених характеристик проводиться оцінка клієнта і формується відповідний позитивний або негативний висновок про можливість надання чи не надання кредиту.



Рисунок 2.1 – Концептуальна схема банку моделювання оцінки кредитоспроможності позичальників банку (фізичних осіб).

Схема, що представлена на рисунку 2.1, може бути реалізована за допомогою різних інструментів моделювання, але у будь-якому випадку побудована модель має відповідати певним вимогам.

Вимоги до моделі можна поділити на дві групи: загальні і спеціальні.

До загальних вимог відносять:

- адекватність. Модель вважається адекватною, якщо відбиває задані властивості об'єкта з прийнятною точністю. Точність визначається як ступінь збігу значень вихідних параметрів моделі й об'єкта;
- універсальність. Модель вважається універсальною, якщо вона може бути застосована до будь-якого об'єкту з об'єктів певного класу;
- економічність. Економічність моделі характеризується витратами обчислювальних ресурсів для її реалізації, що включають витрати машинного часу T_M і пам'яті P_M .

Специфічними вимогами моделі оцінювання кредитоспроможності позичальника банку є:

- адекватне відображення реально існуючих залежностей між відомими характеристиками клієнта та його кредитоспроможністю;
- можливість використання моделі оцінки кредитоспроможності позичальника у різних банківських установах.

До вхідних даних моделі також висувається ряд вимог:

- бути легко доступними для користувачів моделі;
- відображати реальний стан досліджуваного явища на конкретний період;
- не допускати «пробілів» (відсутності даних на певних інтервалах часу);
- вноситись в однаковій розмірності.

Модель повинна давати відповідь на запитання: чи є позичальник кредитоспроможним?

2.2 Ознаковий простір моделі оцінювання кредитоспроможності позичальника банку.

Загальновідомо, що на якість економіко-математичної моделі суттєво впливає правильне визначення її ознакового простору. Відомо, що всі об'єкти в економіці є багатовимірними, тобто описуються великою кількістю елементарних і складних ознак [1].

При визначенні ознакового простору моделі оцінювання кредитоспроможності позичальника банку слід визначити:

- інформаційну базу, за якою буде будуватися скорингова модель;
- залежну змінну;
- незалежні змінні.

Розробка скорингової моделі ґрунтується на аналізі попереднього кредитного досвіду. Достатній обсяг інформації – одна з головних передумов

побудови якісної моделі. Кількість даних може варіюватися залежно від конкретної математичної бази моделі, але в цілому набір вхідних даних має задовольняти вимогу статистичної значущості і випадковості.

Ідеально, якщо моделі скорингу застосовуються в тих самих умовах, які були основою її побудови: одні й ті самі кредитні продукти, сектори ринку, однакова економічна ситуація тощо. Наприклад, відомості про позичальників, що скористались споживчим кредитом, не можуть використовуватись для розробки скорингової моделі оцінки позичальника з автокредитування.

Крім цього, вхідні дані для побудови моделі мають відповідати принципу достатньої історичної давнини. Цей принцип визначає період, за який збираються дані для побудови скорингової моделі. Так, наприклад, дані по автокредитах, що видані 4 місяці тому, не підійдуть для побудови моделі, а інформація по кредитах, що видані 3 роки тому, скоріш за все буде корисною; дані десятирічної давнини будуть вважатися досить застарілими.

Історичний період даних для побудови моделі визначається, як правило, видом скорингу і видом кредитування, а також вимогами наглядаючи органів.

Крім зазначеного, дані про певні типи клієнтів слід виключити з початкової інформаційної бази. Це можуть бути нетипові клієнти: шахраї, співробітники банку, VIP-клієнти, неповнолітні тощо. Також краще виключити кредити з аномально великими сумами, нестандартними умовами погашення, нетиповими цілями позики тощо.

Окремо слід вирішити питання щодо пропущених вхідних даних в історичній вибірці. Великі масиви фінансових даних, як правило, містять пропущені значення різних величин, які слід ретельно аналізувати. Причиною пропусків можуть бути помилки при зборі та введенні даних, неможливість отримання інформації, відмова позичальника давати відповідь на деякі питання тощо. Спостереження з пропусками у значеннях окремих змінних або виключають, або замінюють пропущені значення їх середніми

величинами. Вважається, що наявність пропусків не більше ніж у 5% вимірювань є випадковою. У протилежному випадку слід виявити причину неповноти інформаційної бази і поновити пропущені значення або виключити певні змінні з аналізу.

Після формування вхідної інформаційної бази переходять до визначення залежної та незалежних змінних.

На етапі визначення залежної змінної позичальників поділяють на три групи: «погані», «гарні» і «невизначені». Для шахраїв, банкрутів і безнадійних кредитів критерій визначення «поганого» позичальника однозначний. По інших позичальниках таким критерієм може бути кількість днів прострочки по платежах.

До групи «невизначених» клієнтів відносяться, як правило, позичальники з недостатньою кредитною історією, або ті, хто має невелику прострочку платежів тощо [23].

При побудові скорингової карти використовується інформація тільки по тих клієнтах, які визначені як «гарні» і «погані».

Досить часто в окрему категорію виділяють «відхилених» клієнтів, тобто позичальників, яким було відмовлено в кредиті. Врахування «невизначених» і «відхилених» клієнтів дозволяє врахувати у навчальній вибірці пропорції генеральної сукупності [14].

Процес побудови скорингової карти досить часто проходить у два етапи:

- 1) побудова первинної моделі за даними «поганих» і «гарних» клієнтів без врахування «відхилених» клієнтів;
- 2) побудова кінцевої моделі з врахуванням аналізу відхилених замовлень.

Однак експерти у галузі кредитного скорингу відзначають, що аналіз відхилених замовлень клієнтів потребує значних ресурсів і не завжди призводить до покращення якості кінцевої скорингової моделі [15].

Вибір залежної змінної визначається метою побудови скорингової моделі, а також наявним масивом вхідних даних. Мета може бути загальною, наприклад, скорочення втрат за новими кредитними рахунками, або конкретною – скорочення кількості дефолтів за виданими кредитами протягом першого року кредитування.

Залежна змінна може приймати як кількісні, так і якісні значення. Прикладом кількісної залежної змінної є значення середньої суми, яку внесе позичальник за простроченим кредитом.

Відповідно до Базеля II, дефолт боржника вважається таким, що відбувся, якщо має місце одна з наступних подій:

- банк вважає, що боржник не в змозі повністю погасити власні кредитні зобов'язання перед Банком без прийняття Банком таких заходів, як реалізація забезпечення (якщо таке у наявності);
- боржник більше ніж на 90 днів прострочив погашення всіх суттєвих кредитних зобов'язань перед Банком.

У таблиці 2.1 наведено характеристики, які містяться у вхідному масиві даних, який буде використано для побудови моделі.

Таблиця 2.1 – Характеристики вхідного масиву позичальників.

Назва характеристики	Позначення	Одиниці вимірювання	Шкала вимірювання
Позичальник, який протягом останніх 2 років не повернув кредит	SeriousDlqin2yrs	1 – так; 0 – ні	Номінальна
Загальний залишок по кредитних картах і особистих кредитних лініях (окрім нерухомості і автокредитування), поділений на суму кредитних лімітів і помножений на 100	RevolvingUtilizationOfUnsecuredLines	Відсотки	Абсолютна

Продовження таблиці 2.1

Вік позичальника у роках	Age	Натуральні числа	Абсолютна
Кількість прострочок позичальника по платежах за останні 2 роки терміном від 30 до 59 днів	NumberOfTime3059DaysPastDueNotWorse	Натуральні числа і нуль	Абсолютна
Щомісячні сплати боргів, аліментів, інших витрат, поділені на місячний валовий дохід і помножені на 100	DebtRatio	Відсотки	Абсолютна
Щомісячний дохід	MonthlyIncome	Дійсні числа	Абсолютна
Кількість відкритих кредитів (розстрочка, авто, іпотека) і кредитних ліній (кредитних карток)	NumberOfOpenCreditLinesAndLoans	Натуральні числа і нуль	Абсолютна
Кількість прострочок позичальника по платежах за останні 2 роки терміном більше 90 днів	NumberOfTimes90DaysLate	Натуральні числа і нуль	Абсолютна
Кількість кредитів з нерухомості (у тому числі іпотечних)	NumberRealEstateLoansOrLines	Натуральні числа і нуль	Абсолютна
Кількість прострочок позичальника по платежах за останні 2 роки терміном від 60 до 89 днів	NumberOfTime6089DaysPastDueNotWorse	Натуральні числа і нуль	Абсолютна
Кількість утриманців у сім'ї : жінка, діти тощо(за виключенням позичальника)	NumberOfDependents	Натуральні числа і нуль	Абсолютна

За залежну було обрано змінну *SeriousDlqin2yrs*, яка приймає тільки два значення: 1 – «поганий» позичальник, 0 – «гарний» (табл. 2.1).

Для моделювання значень такої змінної ідеально підходить логістична регресія [16].

Зауважимо, що слід відрізнити залежну змінну, яка буде використовуватись при побудові скорингової моделі, від вихідної змінної модельних розрахунків.

Якщо враховувати, що для розрахунків використовуватиметься логістична регресія, то з математичної точки зору результатом побудованої моделі буде залежність логарифма шансу дефолту за кредитом (логіта) від лінійної комбінації незалежних змінних [16, с. 17]:

$$\ln\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1x_i^{(1)} + b_2x_i^{(2)} + \dots + b_nx_i^{(n)} + \varepsilon_i \quad (2.1)$$

де p_i – ймовірність настання дефолту за кредитом i -го позичальника;

x_i^j – значення i -ї незалежної змінної для позичальника з номером j ,

$j = 1..n$;

b_j – параметри моделі;

ε_i – випадкова компонента.

Для інтерпретації отриманих коефіцієнтів моделі логістичної регресії можна використовувати експоненційну форму запису моделі:

$$p_i = \frac{1}{1 + \exp(-(b_0 + b_1x_i^{(1)} + b_2x_i^{(2)} + \dots + b_nx_i^{(n)} + \varepsilon_i))} \quad (2.2)$$

При інтерпретації коефіцієнтів моделі логістичної регресії слід також враховувати тип відповідної незалежної змінної. Якщо він буде неперервним, то коефіцієнти формули (2.2) будуть показувати, на скільки у середньому за сукупністю змінюватимуться шанси настання дефолту при зміні відповідної незалежної змінної на одиницю свого виміру за умови незмінності інших факторів. Для бінарної незалежної змінної коефіцієнт логістичної регресії

буде показувати співвідношення шансу прояву дефолту за наявності фактора, що відображається бінарною змінною, у порівнянні з його відсутністю.

Дані, що визначаються як незалежні змінні, бувають різних типів:

- соціально-демографічні;
- інформація про кредит, який замовлено;
- маркетингові дані;
- кредитна історія позичальника.

Наведемо для кожного з наведених типів коротку характеристику. Так до соціально-демографічних даних про позичальника відносять стать, сімейний стан, вік, посаду, загальний стаж роботи, стаж роботи на останньому місці, термін проживання за вказаною адресою, наявність дітей, рівень освіти, дохід позичальника і дохід сім'ї тощо.

За вхідні дані може бути обрана така інформація про кредит, який замовлено: термін погашення кредиту, сума кредиту, розмір щомісячних платежів, розмір першого внеску, мета кредитування, наявність забезпечення тощо.

Маркетингові дані, такі як джерело реклами, мотив вибору банку тощо, досить рідко використовуються як вхідні дані скорингової моделі і скоріше збираються для оцінки наявної маркетингової політики.

Внутрішня кредитна історія позичальника характеризується поточним балансом рахунку, поточною заборгованістю, кількістю рахунків, наявністю і обсягом заощаджень, кількістю попередніх кредитів у банку, найбільшим значенням суми заборгованості за попередніми кредитними рахунками, наявністю прострочених платежів, наявністю інших банківських продуктів і послуг, регулярністю виплат за попередніми боргами за всіма зобов'язаннями тощо. Для нових клієнтів банку ця інформація невідома. У такому випадку джерелом аналогічних даних можуть бути дані Бюро кредитних історій: рейтинг позичальника, детальна інформація про кредити в інших банках, детальна інформація про прострочені або сплачені попередні кредити тощо.

Отже, незалежні змінні можуть бути представлені у різних шкалах виміру. Скорингові карти можуть бути побудовані тільки з кількісними типами незалежних змінних, тільки з якісними або з кількісними і категоріальними змінними одночасно.

Незалежні змінні, що обрані нами для побудови скорингової карти, можна віднести до двох типів:

– дані з кредитного замовлення (вік клієнта у роках, щомісячний дохід, кількість утриманців у сім'ї);

– внутрішня кредитна історія позичальника (загальний залишок по кредитних картах і особистих кредитних лініях (окрім нерухомості і автокредитування), поділений на суму кредитних лімітів і помножений на 100; кількість прострочок позичальника по платежах за останні 2 роки терміном від 30 до 59 днів; кількість прострочок позичальника по платежах за останні 2 роки терміном від 60 до 89 днів; кількість прострочок позичальника по платежах за останні 2 роки терміном більше 90 днів; кількість відкритих кредитів (розстрочка, авто, іпотека) і кредитних ліній (кредитних карток); кількість кредитів з нерухомості (у тому числі іпотечних).

Слід відмітити, що всі характеристики, які наведені у таблиці 2.1, кількісні, а залежна змінна є бінарною. Отже, скорингова карта буде побудована тільки за кількісними типами незалежних змінних.

Наведемо характеристики вхідних та вихідних змінних моделі побудови логістичної регресії (табл. 2.2).

Таблиця 2.2 – Опис вхідних даних.

Показник (позначення)	Економічний зміст	Шкала вимірювання	Допустимі значення
Y_i	значення наявності або відсутності дефолту за кредитом для позичальника номеру i у навчальній вибірці	номінальна	{0; 1}
x_i^j	значення незалежних змінних для позичальника номер i у навчальній вибірці	абсолютна	(0;∞)

Продовження таблиці 2.2

i	номер позичальника у навчальній вибірці	номінальна	{1; ...; K}
K	кількість позичальників у навчальній вибірці	порядкова	(0;65535)
j	номер параметру позичальника у навчальній вибірці	номінальна	{1; ...; n}
n	кількість параметрів позичальника у навчальній вибірці	порядкова	(0;16384)

Зазначимо, що припустимі значення для кількості позичальників і кількості параметрів позичальника ґрунтуються на максимально припустимій кількості рядків і стовпчиків робочого листа Excel. Звичайно, що жодна вибірка не буде містити кілька тисяч характеристик позичальника, тому це обмеження дещо умовне.

Вихідною змінною даної моделі буде імовірність настання дефолту за кредитом (табл. 2.3).

Таблиця 2.3 – Опис вихідних даних.

Показник (позначення)	Економічний зміст	Шкала вимірювання	Допустимі значення елементів матриць
$p_i = (i = 0..n)$	Імовірність настання дефолту	відносна	[0; 1]

У подальшому за коефіцієнтами логістичного регресійного рівняння можна буде розрахувати скоринговий бал позичальника та прогнозувати ймовірність настання дефолту за кредитом. Більше детально це описано у наступному пункті.

2.3 Математичний опис моделі оцінювання кредитоспроможності позичальника банку.

У даному пункті наведемо основні математичні формули та співвідношення, що використовуються при побудові моделі оцінювання кредитоспроможності позичальника банку.

До основних етапів побудови скорингових моделей відносять наступні.

1. Підготовка вхідних даних для побудови скорингової моделі.
 - 1.1. Формування вхідної інформаційної бази.
 - 1.2. Визначення залежної змінної.
 - 1.3. Визначення незалежних змінних.
 - 1.4. Формування навчальної і тестової вибірок.
 - 1.5. Визначення необхідного обсягу вибірки.
2. Аналіз і коригування змінних для побудови моделі.
 - 2.1. Оцінка мультиколінеарності між кількісними змінними.
 - 2.2. Оцінювання взаємозв'язку скорингових змінних та настання дефолту.
3. Побудова скорингової моделі за логістичною регресією.
4. Оцінка якості побудованої скорингової моделі.
5. Переведення коефіцієнтів моделі у скорингову карту.

Розглянемо зміст кожного етапу.

На етапі підготовки вхідних даних для побудови скорингової моделі формування вхідної інформаційної бази і визначення залежної та незалежних змінних детально розглянуто у попередньому пункті. Перейдемо до формування навчальної і тестової вибірок [21].

Навчальна вибірка – та частина історичної вибірки, на якій безпосередньо будується скорингова модель. Контрольна (тестова) вибірка – спостереження, для яких відомо значення залежної змінної, але вони не беруть участі у побудові моделі, а використовуються для перевірки точності передбачень моделі. Навчальна і контрольна вибірки формуються за

механізмами випадкового вибору, як правило, у співвідношенні 70-80% до 30-20% відповідно від обсягу історичної вибірки [22].

Тестові вибірка застосовується після побудови скорингової моделі для перевірки її достовірності. Цей етап ще називають валідацією. У випадку кредитного скорингу, достовірність моделі – це її здатність відрізнити «гарних» позичальників від «поганих». Валідна модель повинна давати коректні прогнози не тільки на навчальній вибірці, але й на тестовій. Якщо показники точності, які отримані на тестовій і навчальній вибірці, співпадають, то вважається, що на практиці скорингова модель буде працювати аналогічно.

У даній роботі навчальна і тестова вибірка формувались випадково у співвідношенні 70% на 30% за допомогою спеціального модуля Statistica.

Наступний крок — визначення обсягу вибірки — є вагомим методологічним питанням. Опрацювання наукової літератури дозволило виокремити наступні рекомендації щодо мінімального обсягу вибірки і кількості предикторів (незалежних змінних) у моделі.

Так, вибір мінімального обсягу вибірки залежить від рівномірності розподілу значень залежної змінної. При нормальному розподілі необхідно мати не менше 10 спостережень на 1 предиктор. Однак чим більше розподіл залежної бінарної змінної зміщено на користь однієї з категорій, тим більше спостережень слід брати на один предиктор.

Інший підхід пов'язаний з так званим правилом 20EPV [17]. За ним, слід визначити N_p – кількість спостережень в історичній вибірці, що відповідає категорії залежної змінної з найменшим обсягом. У нашому випадку – це кількість «поганих» позичальників у вихідному файлі. Величину N_p слід поділити на кількість незалежних змінних (m), що включені у модель. Обсяг вибірки буде вважатися достатнім, якщо отримана частка не менше 20:

$$\frac{N_p}{m} \geq 20 \quad (2.3)$$

де Np – кількість спостережень в історичній вибірці, що відповідає категорії залежної змінної з найменшим обсягом; m – кількість незалежних змінних моделі.

Однак у кредитному скорингу розподіл залежної змінної завжди зміщений у бік «гарних» позичальників, тому більш доречним буде визначення мінімального обсягу вибірки на основі максимально припустимої помилки в оцінці співвідношення «поганих» і «гарних» позичальників у генеральній сукупності.

Виходимо з тієї вимоги, що навчальна вибірка позичальників має відображати генеральну сукупність всіх потенційних позичальників з невідомими частками «гарних» і «поганих» кредитів. Припустимо, що ми маємо бути впевнені на 95%, що співвідношення «гарних» і «поганих» кредитів відображає генеральну популяцію позичальників. Тоді мінімально необхідний обсяг навчальної вибірки можна розрахувати за формулою:

$$n = \frac{z_y^2 w(1-w)}{\Delta_w^2} \quad (2.4)$$

де n – мінімально необхідний обсяг вибірки;

z_y – значення стандартного нормального закону розподілу, що залежить від обраного рівня надійності;

w – частка «поганих» клієнтів в історичній вибірці;

Δ_w – максимально припустима гранична помилка оцінки частки «поганих» позичальників.

При визначенні мінімального обсягу навчальної вибірки можна скористатися і формулою оцінки генеральної долі:

$$w - z_y \sqrt{\frac{w(1-w)}{n}} \leq P \leq w + z_y \sqrt{\frac{w(1-w)}{n}}, \quad (2.5)$$

де P – доля «поганих» позичальників у генеральній сукупності, яку оцінюють;
 w – частка «поганих» клієнтів у тестовій вибірці;
 n – обсяг початкової історичної вибірки;
 z_y – значення стандартного нормального закону розподілу, що залежить від обраного рівня надійності u .

На етапі аналізу і коригування змінних для побудови моделі першочергово слід перевірити незалежні змінні на мультиколінеарність. Пропонується застосувати для цього кореляційну матрицю факторних змінних. Якщо у ній будуть присутні коефіцієнти парної кореляції більші за 0,8 за модулем, вважатимемо, що у моделі присутня мультиколінеарність.

У випадку наявності мультиколінеарності необхідно знайти оптимальні варіанти виключення тісно корелюючих між собою факторів моделі. Іншим способом позбавлення мультиколінеарності є збільшення обсягу початкової вибірки. Слід зазначити, що за наявності вибірок великого обсягу (у кілька тисяч або десятків тисяч позичальників) мультиколінеарність не дуже викривляє оцінки параметрів моделі. Ще один кардинальний метод позбавлення від мультиколінеарності – використання головних компонент замість вхідних значень змінних – використовувати не дуже доречно, тому що це дуже ускладнює подальший розрахунок скорингових балів на основі агрегованих показників.

Подальший аналіз змінних передбачає оцінку взаємозв'язку незалежних змінних і даних про настання дефолту з метою зменшення кількості предикторів, що включені у модель.

Основним методом оцінки наявності зв'язку між залежною бінарною змінною і незалежними кількісними змінними є розрахунок критерія χ^2 .

При використанні зазначеного критерію висувають гіпотезу H_0 : розподіл між «поганими» і «гарними» позичальниками за незалежними змінними однаковий.

За таблицями спряженості між залежною і кожною незалежною змінними окремо розраховують статистику χ^2 – критерія за формулою:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad (2.6)$$

де f_{ij} – фактичні частоти;

e_{ij} – очікувані частоти;

m і k – кількість рядків і стовпчиків у таблиці спряженості.

Якщо розрахункове значення χ^2 буде перевищувати критичне значення χ^2 , то наявність взаємозв'язку між конкретною незалежною змінною і даними про настання дефолту доведена. Зауважимо, що критичне значення χ^2 знаходять з відповідних таблиць розподілу χ^2 з заданим рівнем значущості α і кількістю ступенів свободи $\nu = (m-1)(k-1)$.

Відмітимо, що у пакеті Statistica для відповідних розрахунків використовується спеціальний модуль «Отсеивание признаков».

Побудова скорингової моделі за логіт-регресійною моделлю. Логістична регресія – це різновид множинної регресії, остання задається формулою:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2.7)$$

де x_i – значення незалежних змінних;

b_i – коефіцієнти регресії.

Множинну регресію на пряму неможна застосовувати для передбачення бінарного результату, тому що може видавати передбачувані значення більші за 1 і менші за 0. Однак такі значення не припустимі у постановці нашої задачі.

Для рішення означеної проблеми замість передбачення бінарної змінної можна моделювати неперервну змінну на відріжку $[0; 1]$. Це досягається за рахунок застосування логіт-перетворення [19, с. 107]:

$$P = \frac{1}{1+e^{-y}} \quad (2.8)$$

де P – ймовірність того, що настане дефолт за кредитом;
 e – основа натурального логарифму;
 y – стандартне регресійне рівняння (2.7).

Якщо для P буде отримане значення менше за 0,5, можна припустити, що дефолт за кредитом не настане.

Існує кілька алгоритмів знаходження коефіцієнтів логістичної регресії: метод спряжених градієнтів, методи градієнтної метрики тощо. Найчастіше використовується метод максимальної правдоподібності. Його сутність у наступному: за «найбільш правдоподібне» значення параметра обирають значення θ , що максимізує ймовірність отримати при n спробах дану вибірку $Y_1, Y_2 \dots Y_k$. Значення параметру θ і є шуканою оцінкою.

Введемо позначення P_i – ймовірність появи одиниці (настання дефолту за кредитом); $P_i = \text{Prob}(Y_i = 1)$. Ця ймовірність буде залежати від $y_i = Bx_i$:

$$P_i = F(Bx_i) \quad (2.9)$$

$$F(y) = \frac{1}{1 + e^{-y}}$$

де x_i – рядок матриці незалежних змінних;
 B – вектор коефіцієнтів регресії.

Роль шуканого значення θ у формулах (2.9) відіграє вектор коефіцієнтів регресії B .

Алгоритм методу правдоподібності для знаходження коефіцієнтів логістичної регресії містить такі кроки:

а) складається логарифмічна функція правдоподібності:

$$L = \sum_{i \in I_1} \ln P_i(B) + \sum_{i \in I_0} \ln(1 - P_i(B)) = \sum_{i=1}^k [Y_i \ln P_i(B) + (1 - Y_i) \ln(1 - P_i(B))] \quad (2.10)$$

де I_0 та I_1 – множини спостережень, для яких $Y_i=0$ і $Y_i=1$ відповідно;

б) знаходяться похідні за елементами шуканого вектору коефіцієнтів регресії B і прирівнюються до нуля:

$$\frac{\partial L}{\partial B} = 0 \quad (2.11)$$

в) розв’язується отримана на попередньому кроці система відносно невідомих – елементів вектору коефіцієнтів регресії B .

Зазначимо, що сучасний розвиток комп’ютерних програм дозволяє не проводити описані вище кроки вручну. У даній роботі для знаходження коефіцієнтів логістичної регресії було використано вузол Logit Regression with Deployment інструменту DataMiner.

Схематично структура скорингової моделі оцінювання кредитоспроможності позичальників банку представлена на рисунку 2.2.

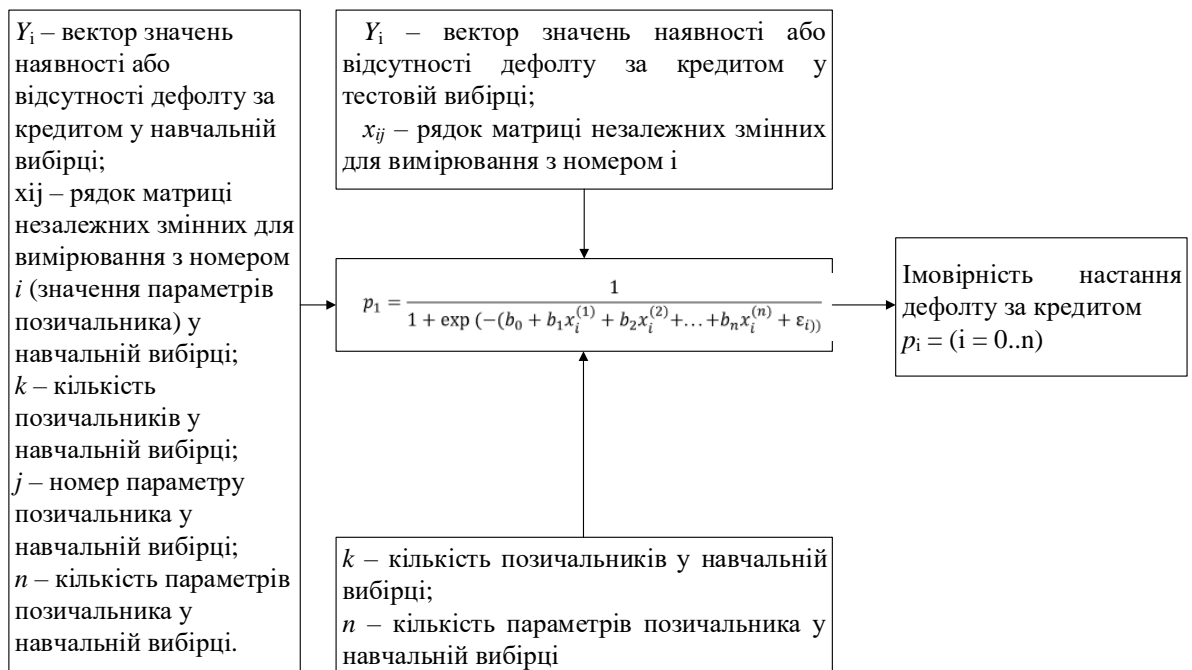


Рисунок 2.2 – Схема структури скорингової моделі оцінювання кредитоспроможності позичальників банку.

Наступним кроком модельних розрахунків є оцінка якості побудованої скорингової моделі, для якої існує досить значна кількість тестів. Опишемо ті, що використовувались у даній роботі:

- тест Вальда для перевірки статистичної значущості окремих параметрів логістичної регресії;
- аналіз меж довірчих інтервалів для підтвердження статистичної значущості окремих параметрів логістичної регресії.

РОЗДІЛ 3. ПРАКТИЧНА РЕАЛІЗАЦІЯ СКОРИНГОВОЇ МОДЕЛІ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА БАНКУ

3.1 Програмна реалізація моделі та апробація на реальних даних.

Реалізація скорингової моделі оцінювання кредитоспроможності позичальника банку здійснювалась із застосуванням пакету Statistica. Для побудови скорингових карт даний програмний засіб містить окремий модуль Scorecard Preparation, однак він не надається для користування у вільному доступі, тому ми застосовували доступні нам інструменти [4].

Інформаційна база для побудови моделі була взята з платформи проведення змагань з аналізу даних, де проводилося змагання з data science «Give Me Some Credit :: Competition Data» [40]. Вхідний файл містить 65535 записів з такими відомостями по кредитах:

- відомості про повернення чи неповернення позичальником кредиту протягом останніх 2 років;
- загальний залишок по кредитних картах і особистих кредитних лініях (окрім нерухомості і автокредитування), поділений на суму кредитних лімітів (у відсотках);
- вік позичальника у роках;
- кількість прострочек позичальника по платежах терміном від 30 до 59 днів (за останні 2 роки);
- щомісячні сплати боргів, аліментів, інших витрат, поділені на місячний валовий дохід (у відсотках);
- щомісячний дохід;
- кількість відкритих кредитів (розстрочка, авто, іпотека) і кредитних ліній (кредитних карток);

– кількість прострочек позичальника по платежах терміном більше 90 днів (за останні 2 роки).

На першому етапі побудови скорингової моделі перевірялась наявність пропущених даних в історичній вибірці і вилучення їх у тому випадку, якщо вимірювань, буде достатньо для побудови скорингової карти [12].

На етапі підготовки, очистки і перетворення даних застосовувався вузол *Replace Missing Data* для вилучення рядків з пропущеними значеннями. Його налаштування по вкладках наведено у Додатку Б.

Загальний вигляд відповідного процесу, побудованого за допомогою DataMiner, наведений на рисунку 3.1.

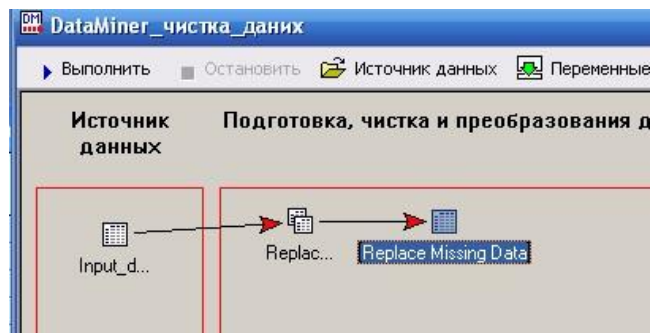


Рисунок 3.1 – Побудова процесу чистки даних у DataMiner.

Результатом виконання заданого процесу є автоматично сформований файл даних пакету Statistica «Replace missing data», який далі аналізувався за допомогою описових статистик (рис. 3.2).

Таблица частот: SeriousDlqin2yrs (Replace_missing_data)						
K-S d=.53842, p<.01 ;Лиллиефорса p<.01						
Группа	Частота	Кумул. частота	Процент допуст.	Кумул. % допуст.	% всех наблюд.	Кумул. % от всех
-.200000<x<=.722E-15	48986	48986	93,17711	93,1771	93,17711	93,1771
,722E-15<x<=.2000000	0	48986	0,00000	93,1771	0,00000	93,1771
,2000000<x<=.4000000	0	48986	0,00000	93,1771	0,00000	93,1771
,4000000<x<=.6000000	0	48986	0,00000	93,1771	0,00000	93,1771
,6000000<x<=.8000000	0	48986	0,00000	93,1771	0,00000	93,1771
,8000000<x<=1,000000	3587	52573	6,82289	100,0000	6,82289	100,0000
Пропущ.	0	52573	0,00000		0,00000	100,0000

Рисунок 3.2 – Результат виконання модуля описових статистик по залежній змінній за даними історичної вибірки (таблиця частот).

За даними таблиці частот для змінної *SeriousDlqin2yrs* (позичальник, який за останні два роки не погасив кредит) встановлено, що таблиця скоригованих вхідних даних «Replace missing data» містить 52573 записи. З них кількість «поганих» відгуків становить 3587, а «гарних» 48986, пропущених значень немає. Відсоток «поганих» відгуків складає 6,8%.

Наступним кроком обраховувалась мінімально необхідна кількість вхідних даних для побудови скорингової моделі.

Припустимо, що історична вибірка, на якій будуватиметься модель, має містити відхилення між «гарними» і «поганими» клієнтами не більше ніж у 5% випадків у порівнянні з генеральною сукупністю. Відповідно, табличне значення стандартного нормального закону розподілу дорівнює 1,96. Тоді гранична помилка оцінки частки становитиме $0,068 * 0,05 = 0,0034$. Отже, за формулою (2.4), мінімально необхідний розмір історичної вибірки – 20985 вимірювань. Файл «Replace missing data» містить 52573 записи, навіть при розбитті вхідних даних у співвідношенні 70% на 30% обсяг навчальної вибірки становитиме 34290 вимірювань, що значно більше, ніж 20985. Можна зробити висновок, що довжина історичних даних, яку ми використовуватимемо для побудови скорингової моделі, цілком задовільна.

Результати розрахунків, на яких ґрунтуються описані вище висновки, наведено у Додатку В.

У подальшому історична вибірка була поділена на навчальну і тестову. При цьому було створено новий файл DataMiner з файлом вхідних даних «Replace missing data». Це обумовлено тим, що кожен раз при запуску обчислювального процесу, створеного у DataMiner, розрахунки починаються з самого початку, тобто, з першого вузла, що уповільнює отримання остаточного результату.

Для отримання навчальної і тестової вибірок використовувався вузол *Split Input Data into Training and Testing Samples*. Його налаштування показані у Додатку Г. Діаграма відповідного обчислювального процесу зображена на рисунку 3.3.

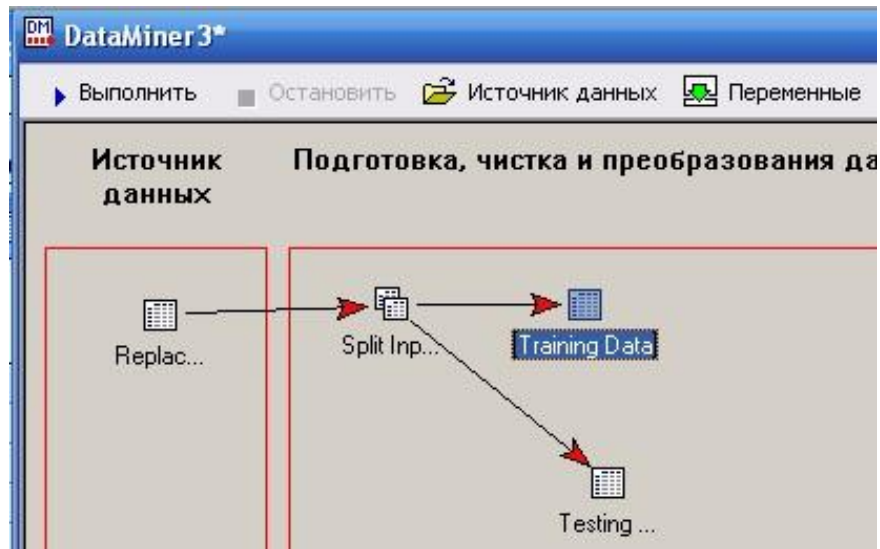


Рисунок 3.3 – Діаграма розбиття історичної вибірки на навчальну і тестову.

Як показує рисунок 3.3, в результаті виконання вузла *Split Input Data into Training and Testing Samples* утворилося два файли вхідних даних «Training Data» і «Testing Data».

Як зазначено раніше, співвідношення між «гарними» і «поганими» відгуками у навчальній і тестовій вибірках мають співпадати з історичними. Перевіримо це за допомогою інструмента описових статистик. Результати зображено на рисунку 3.4-3.5.

Таблица частот: SeriousDlqin2yrs (Split Input Data into Training and Testing)
K-S d=,53868, p<,01 ;Лиллиефорса p<,01

Группа	Частота	Кумул. частота	Процент допуст.	Кумул. % допуст.	% всех наблюд.	Кумул. % от всех
-,200000<x<=,722E-15	29001	29001	93,32282	93,3228	93,32282	93,3228
,722E-15<x<=,2000000	0	29001	0,00000	93,3228	0,00000	93,3228
,2000000<x<=,4000000	0	29001	0,00000	93,3228	0,00000	93,3228
,4000000<x<=,6000000	0	29001	0,00000	93,3228	0,00000	93,3228
,6000000<x<=,8000000	0	29001	0,00000	93,3228	0,00000	93,3228
,8000000<x<=1,000000	2075	31076	6,67718	100,0000	6,67718	100,0000
Пропущ.	0	31076	0,00000		0,00000	100,0000

Рисунок 3.4 – Результат виконання модуля описових статистик по залежній змінній за даними навчальної вибірки (таблиця частот).

Таблица частот: SeriousDlqin2yrs (Split Input Data into K-S d=.53920, p<.01 ;Лиллиефорса p<.01						
Группа	Частота	Кумул. частота	Процент допуст.	Кумул. % допуст.	% всех наблюд.	К:о
-,200000<x<=,722E-15	12493	12493	93,63664	93,6366	93,63664	
,722E-15<x<=,2000000	0	12493	0,00000	93,6366	0,00000	
,2000000<x<=,4000000	0	12493	0,00000	93,6366	0,00000	
,4000000<x<=,6000000	0	12493	0,00000	93,6366	0,00000	
,6000000<x<=,8000000	0	12493	0,00000	93,6366	0,00000	
,8000000<x<=1,000000	849	13342	6,36336	100,0000	6,36336	1
Пропущ.	0	13342	0,00000		0,00000	1

Рисунок 3.5 – Результат виконання модуля описових статистик по залежній змінній за даними тестової вибірки (таблиця частот).

Складемо таблицю для порівняння відсотків «гарних» і «поганих» відгуків у всіх наявних вибірках (табл. 3.1).

Таблиця 3.1 – Відсотки «гарних» і «поганих» відгуків в історичній, навчальній і тестові вибірках.

Тип вибірки	% «гарних» відгуків	% «поганих» відгуків
Історична	93,18	6,82
Навчальна	93,32	6,68
Тренувальна	93,64	6,36

Як видно з таблиці 3.1, відсотки «гарних» і «поганих» відгуків у всіх вибірках приблизно однакові, що дає змогу використовувати їх у подальшому моделюванні.

Для подальшої перевірки наявності мультиколінеарності серед незалежних змінних застосуємо блок парних і часткових кореляцій пакету Statistica (рис. 3.6).

Корреляции (Replace_missing_data) Отмеченные корреляции значимы на уровне p <.05000 N=44418 (Построчное удаление ПД)										
Переменная	nOfUn	age	daysl	btRat	thlyInco	reditLi	mes9C	stateLc	DaysP	fDep
RevolvingUtilizationOfUnsecuredLines	1,00	-0,26	0,23	-0,00	-0,10	-0,17	0,32	-0,08	0,21	0,08
age	-0,26	1,00	0,06	-0,01	0,09	0,17	-0,11	0,04	-0,06	-0,23
NumberOfTime3059DaysPastDueNotWorse	0,23	-0,06	1,00	-0,00	-0,04	-0,04	0,61	-0,04	0,56	0,03
DebtRatio	-0,00	-0,01	0,00	1,00	-0,08	0,01	-0,01	0,03	-0,00	0,01
MonthlyIncome	-0,10	0,09	0,04	-0,08	1,00	0,27	-0,07	0,36	-0,04	0,19
NumberOfOpenCreditLinesAndLoans	-0,17	0,17	0,04	0,01	0,27	1,00	-0,12	0,43	-0,06	0,04
NumberOfTimes90DaysLate	0,32	-0,11	0,61	-0,01	-0,07	-0,12	1,00	-0,09	0,56	0,04
NumberRealEstateLoansOrLines	-0,08	0,04	0,04	0,03	0,36	0,43	-0,09	1,00	-0,05	0,13
NumberOfTime6089DaysPastDueNotWorse	0,21	-0,06	0,56	-0,00	-0,04	-0,06	0,56	-0,05	1,00	0,02
NumberOfDependents	0,08	-0,23	0,03	0,01	0,19	0,04	0,04	0,13	0,02	1,00

Рисунок 3.6 – Результат виконання модуля парних і часткових кореляцій для незалежних змінних моделі.

За даними рисунку 3.6, мультиколінеарність у масиві незалежних змінних відсутня, тому що всі коефіцієнти парної кореляції за модулем менші за 0,8.

Для перевірки наявності зв'язку між залежною бінарною змінною і незалежними кількісними змінними використовувався модуль «Отсеивание признаков», результат зображено на рисунках 3.7-3.8.

		Наилучшие предикторы для кате	
		Chi-квадрат	p-значение
IC	NumberOfTimes90DaysLate	5543,576	0,000000
1E	RevolvingUtilizationOfUnsecuredLines	3949,705	0,000000
4L	NumberOfTime3059DaysPastDueNotWorse	3773,578	0,000000
	NumberOfTime6089DaysPastDueNotWorse	3575,084	0,000000
	age	547,404	0,000000
	MonthlyIncome	274,409	0,000000
	NumberOfOpenCreditLinesAndLoans	247,799	0,000000
	NumberRealEstateLoansOrLines	197,150	0,000000
	NumberOfDependents	72,420	0,000000
	DebtRatio	8,745	0,067797

Рисунок 3.7 – Результати оцінки зв'язку між настанням дефолту за кредитом і незалежними змінними за критерієм χ^2 .

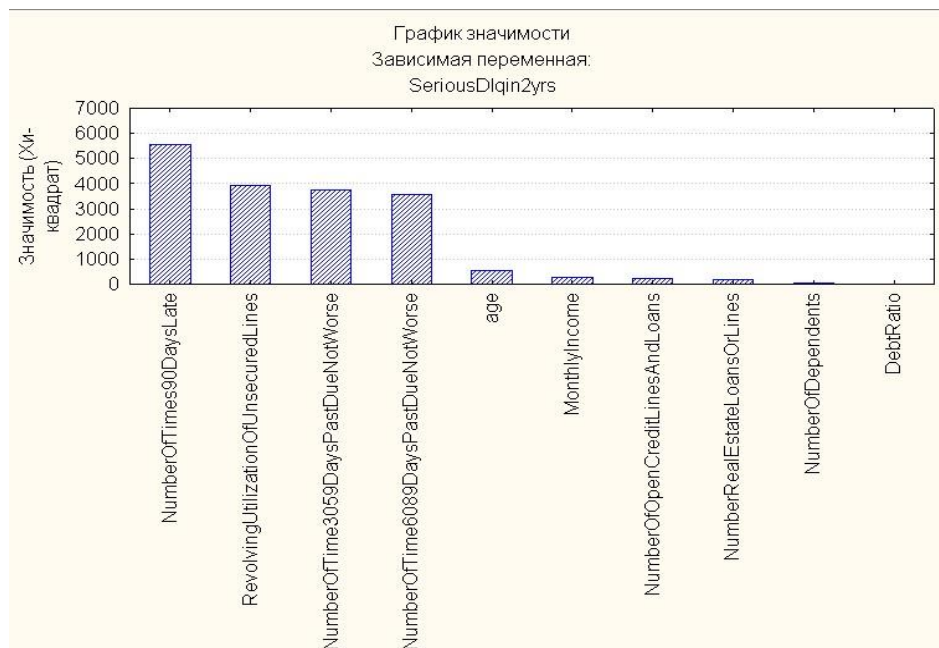


Рисунок 3.8 – Графічна оцінка зв'язку між настанням дефолту за кредитом і незалежними змінними за критерієм χ^2 .

За даними рисунку 3.7 і 3.8 можна побачити, що зв'язок між настанням дефолту за кредитом і змінною *DebtRatio* (частка щомісячної сплати боргів, аліментів, інших витрат у місячному валовому доході у %) відсутній. Отже, дану змінну можна викинути з подальшого аналізу. Однак подальші розрахунки показали, що це робити недоречно, так як ускладнюються внутрішні обчислювальні процеси модуля побудови логістичної регресії і не вдається отримати результат. Тому вирішено було залишити всі незалежні змінні.

3.2 Аналіз отриманих результатів та перевірка адекватності розробленої моделі.

Як зазначалось раніше, у результаті виконання вузла побудови логістичної регресії з'явилося два нових файли – Testing_PMML_Logit16 і Logit Regression with Deployment (у розділі «Отчеты»). У першому з них містяться результати перевірки якості побудованої моделі за тестовою вибіркою, у другому – результати побудови логістичної регресії.

Для оцінки якості побудованої моделі логістичної регресії відкріємо у розділі «Отчеты» вузол з результатами Logit Regression with Deployment (рис. 3.9-3.10).

SeriousDlqin2yrs - Параметры оценивания (Split Input Data into Распределение : Биномиальное Функция связи: Логит						
Эффект	Уровень Эффект	Столбец	Оценка	Стандарт Ошибка	Вальда стат-ка	p
Своб. чл.		1	3,49906	0,116758	898,1025	0,000000
RevolvingUtilizationOfUnsecuredLines		2	-1,92661	0,070592	744,8739	0,000000
age		3	0,01455	0,001960	55,1105	0,000000
NumberOfTime3059DaysPastDueNotWorse		4	-0,41163	0,024065	292,5741	0,000000
DebtRatio		5	0,00009	0,000064	1,9386	0,163819
MonthlyIncome		6	0,00004	0,000007	39,0806	0,000000
NumberOfOpenCreditLinesAndLoans		7	-0,04298	0,005611	58,6739	0,000000
NumberOfTimes90DaysLate		8	-0,81447	0,036376	501,3370	0,000000
NumberRealEstateLoansOrLines		9	-0,05854	0,027441	4,5509	0,032902
NumberOfTime6089DaysPastDueNotWorse		10	-0,73005	0,052025	196,9205	0,000000
NumberOfDependents		11	-0,03917	0,020606	3,6129	0,057334

Рисунок 3.9 – Результати побудови логістичної регресії (вкладка «Параметры оценивания»).

Як видно з рисунку 3.9, коефіцієнти логістичної регресії при змінній *DebtRatio* (частка щомісячної сплати боргів, аліментів, інших витрат у місячному валовому доході, виражена у %) і *NumberOfDependents* (кількість відкритих кредитів і кредитних ліній) є статистично незначущими (відповідні *p*-значення більші за рівень значущості $\alpha=0,05$).

		SeriousDlqin2yrs - Доверительные интервалы		
		Распределение : Биномиальное		
		Функция связи: Логит		
Эффект	Уровень Эффект	Столбец	Нижн. 95, %	Верх. 95, %
Своб. чл.		1	3,27021	3,72790
RevolvingUtilizationOfUnsecuredLines		2	-2,06497	-1,78825
age		3	0,01071	0,01839
NumberOfTime3059DaysPastDueNotWorse		4	-0,45880	-0,36446
DebtRatio		5	-0,00004	0,00021
MonthlyIncome		6	0,00003	0,00006
NumberOfOpenCreditLinesAndLoans		7	-0,05397	-0,03198
NumberOfTimes90DaysLate		8	-0,88577	-0,74318
NumberRealEstateLoansOrLines		9	-0,11232	-0,00476
NumberOfTime6089DaysPastDueNotWorse		10	-0,83202	-0,62809
NumberOfDependents		11	-0,07955	0,00122

Рисунок 3.10 – Результати побудови логістичної регресії (вкладка «Доверительные интервалы оценивания»).

Аналогічні висновки можна зробити і з аналізу меж довірчих інтервалів для коефіцієнтів рівняння логістичної регресії. Для цих самих змінних ліві і праві границі довірчих інтервалів мають різні знаки. Це вимагає видалення відповідних змінних з рівняння регресії і потребує перерахунку процесу DataMiner.

Для цього подвійним клацанням відкриємо вікно налаштувань таблиці вхідних даних *Replace_missing_data* і по-новому задамо залежні та незалежні змінні моделі (рис. 3.11). Далі запустимо процес на виконання.

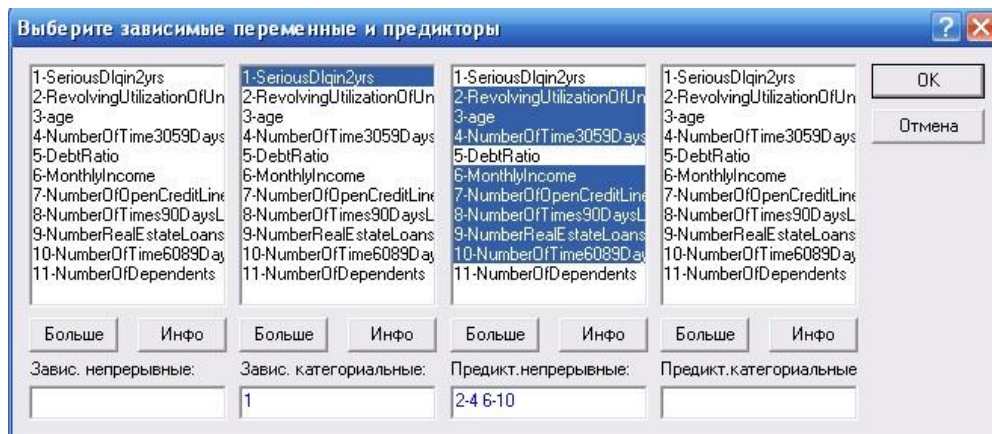


Рисунок 3.11 – Завдання нових змінних для обчислювального процесу.

Оновлені результати аналізу якості побудованої моделі логістичної регресії та статистична значущих параметрів розрахованої моделі (рис 3.12).

		SeriousDlqin2yrs - Параметры оценивания (Split Input Data into Распределение : Биномиальное Функция связи: Логит					
Эффект	Уровень Эффект	Столбец	Оценка	Стандарт Ошибка	Вальда стат-ка	p	
Своб. чл.		1	3,40843	0,114208	890,6760	0,000000	
RevolvingUtilizationOfUnsecuredLines		2	-1,86709	0,071032	690,9179	0,000000	
age		3	0,01574	0,001951	65,0969	0,000000	
NumberOfTime3059DaysPastDueNotWorse		4	-0,43925	0,024011	334,6455	0,000000	
MonthlyIncome		5	0,00004	0,000007	29,3689	0,000000	
NumberOfOpenCreditLinesAndLoans		6	-0,03748	0,005684	43,4796	0,000000	
NumberOfTimes90DaysLate		7	-0,78698	0,035771	484,0103	0,000000	
NumberRealEstateLoansOrLines		8	-0,06937	0,027662	6,2880	0,012156	
NumberOfTime6089DaysPastDueNotWorse		9	-0,67595	0,051253	173,9346	0,000000	
Масштаб			1,00000	0,000000			

Рисунок 3.12 – Оновлені результати побудови логістичної регресії (вкладка «Параметры оценивания»).

		SeriousDlqin2yrs - Доверительные интервалы Распределение : Биномиальное Функция связи: Логит			
Эффект	Уровень Эффект	Столбец	Нижн. 95, %	Верх. 95, %	
Своб. чл.		1	3,18459	3,63228	
RevolvingUtilizationOfUnsecuredLines		2	-2,00631	-1,72787	
age		3	0,01192	0,01956	
NumberOfTime3059DaysPastDueNotWorse		4	-0,48631	-0,39219	
MonthlyIncome		5	0,00002	0,00005	
NumberOfOpenCreditLinesAndLoans		6	-0,04862	-0,02634	
NumberOfTimes90DaysLate		7	-0,85709	-0,71687	
NumberRealEstateLoansOrLines		8	-0,12358	-0,01515	
NumberOfTime6089DaysPastDueNotWorse		9	-0,77640	-0,57549	

Рисунок 3.13 – Оновлені результати побудови логістичної регресії (вкладка «Доверительные интервалы оценивания»).

За даними рисунків 3.13-3.14 можна стверджувати, що оновлені коефіцієнти рівняння логістичної регресії є статистично значущими, отже, можуть бути використані у математичній формі запису моделі і у подальших прогнозних розрахунках.

Коефіцієнти логістичної регресії можна виписати з рисунку 3.12, стовпчик «Оценка». Отже, рівняння логістичної регресії за формулою (2.2) має вигляд:

$$p = \frac{1}{1 + e^{(3.4 - 1.86x_1 + 0.01x_2 - 0.43x_3 + 0.00004x_5 - 0.37x_6 - 0.79x_7 - 0.07x_8 - 0.68x_9)}} \quad (3.1)$$

де x_1 – *RevolvingUtilizationOfUnsecuredLines* (загальний залишок по кредитних картах і особистих кредитних лініях (окрім нерухомості і автокредитування), поділений на суму кредитних лімітів і помножений на 100);

x_2 – *Age* (вік позичальника у роках);

x_3 – *NumberOfTime30-59DaysPastDueNotWorse* (кількість прострочек позичальника по платежах за останні 2 роки терміном від 30 до 59 днів);

x_4 – *MonthlyIncome* (щомісячний дохід);

x_5 – *NumberOfOpenCreditLinesAndLoans* (кількість відкритих кредитів - розстрочка, авто, іпотека - і кредитних ліній - кредитних карток);

x_6 – *NumberOfTimes90DaysLate* (кількість прострочок позичальника по платежах за останні 2 роки терміном більше 90 днів);

x_7 – *NumberRealEstateLoansOrLines* (кількість кредитів з нерухомості, у тому числі іпотечних);

x_8 – *NumberOfTime60-89DaysPastDueNotWorse* (кількість прострочек позичальника по платежах за останні 2 роки терміном від 60 до 89 днів).

Результатом обчислення за формулою (3.1) є число від 0 до 1, яке характеризує ймовірність дефолту за кредитом. Якщо $p > 0,5$, то вважається, що ймовірність повернення кредиту досить висока, інакше – скоріш за все, кредит не буде повернено у повному обсязі. Отже, якщо для нової позики розраховане значення p більше за 0,5, то приймається рішення про видачу кредиту позичальнику, у протилежному випадку клієнт отримує відмову.

З аналізу рівняння (3.1) також можна зробити висновки, що із зростанням таких показників позичальника, як загальний залишок по кредитних картах і особистих кредитних лініях, кількість прострочок позичальника по платежах за останні 2 роки терміном від 30 днів, кількість відкритих кредитів, кредитних ліній, кредитів з нерухомості ймовірність настання дефолту по кредиту зростає. Навпаки, при збільшенні таких показників як вік позичальника у роках і щомісячний дохід ймовірність неповернення кредиту зменшується.

Модель вважається адекватною, якщо відображає задані властивості об'єкта. На підставі результатів порівняння експериментальних і обчислених за моделлю значень з використанням методів математичної статистики можна відповісти на питання про адекватність моделі.

Для моделей з бінарним відгуком, зокрема, скорингових моделей, перевіряється:

- статистична значущість окремих параметрів логістичної регресії;
- наявності взаємозв'язку між вхідними даними про настання дефолту за кредитом і отриманими відповідними прогнозними даними;
- % неправильних відгуків, отриманих при застосуванні побудованої моделі на даних тестової вибірки.

Для перевірки наявності взаємозв'язку між вхідними даними про настання дефолту за кредитом і отриманими відповідними прогнозними даними відкриємо таблицю `Testing_PMML_Logit16` і застосуємо до неї модуль «Качество підгонки, классификации, прогноза» (рис. 3.14).

	Итоги по качеству подгонки (Logit F Наблюдаемая переменная: Serious GeneralizedLinearModelПредск
Статистика Хи-квадрат	5706,820
Статистика G-квадрат	4627,118
Процент несогласия	6,907

Рисунок 3.14 – Результати оцінки адекватності моделі логістичної регресії за даними тестової вибірки.

За даними рисунку 3.14, побудована модель логістичної регресії дає гарні результати. Це засвідчується високими значеннями статистик χ^2 і G-квадрат, значно більшими за асимптотичне значення χ^2 , що дорівнює 1185. Кількість помилкових прогнозів становить 6,91%, що є доволі непоганим результатом.

Для оцінки чутливості і специфічності побудованої скорингової моделі складемо таблицю порівняння результатів класифікації моделі з реальними результатами. Вхідні дані для цієї таблиці взято з файлу Testing_PMML_Logit16, отриманого після виконання процесу побудови логістичної регресії.

Таблиця 3.2 – Таблиця співвідношення результатів класифікації моделі з реальними результатами.

Класифікація на основі моделі	Фактично	
	Позитивно	Негативно
Позитивно	14694	911
Негативно	126	192

Розрахунки, що проведені за формулами (2.15) і (2.16), показали, що значення чутливості $Se = \frac{14694}{14694+126} * 100\% = 99,15\%$, а специфічності $Sp = \frac{192}{192+911} * 100\% = 17,41\%$. Відповідні електронні розрахунки наведено у Додатку Є.

Значення чутливості значно вище за значення специфічності, що свідчить про досить ризиковану позицію банку у видачі кредитів. У даному

випадку банком мінімізуються комерційні ризики, що пов'язані з втраченою вигодою.

Якщо керівництво банку вирішує обрати консервативну кредитну політику, слід будувати скорингові моделі з високою специфічністю. У даному випадку основною стратегією буде мінімізація кредитного ризику, пов'язаного з втратами власних коштів, виданих у позику, відсотків і додаткових доходів з обслуговування кредиту.

3.3 Оцінка очікуваних ефектів від впровадження скорингової моделі.

У науковій літературі описані різні напрями оцінки ефективності від впровадження математичних моделей. Ми вибрали ті, які, на наш погляд, відносяться до оцінки ефективності скорингових моделей:

- інноваційний ефект (новий спосіб, метод, технологія тощо);
- економічний ефект (збільшення фінансових надходжень, економія фінансових і матеріальних засобів, попередження економічного збитку);
- управлінський (організаційний) ефект;
- соціальний ефект (покращення житлових і культурно-побутових умов населення; покращення рівня здоров'я в регіоні; збільшення кількості робочих місць);
- науково-технічний ефект (приріст інформації, отримання нових знань).

Інноваційний ефект від впровадження скорингової моделі проявляється у впровадженні нового способу оцінки кредитоспроможності клієнта банку, навіть якщо в установі вже діяла певна аналогічна модель. Справа в тому, що скорингові моделі мають перебудовуватись раз на 2-3 роки завдяки швидкій мінливості економічних і політичних процесів у суспільстві.

Економічний ефект проявляється неоднозначно. Виявлення на попередньому етапі «поганих» клієнтів призводить до попередження економічного збитку, однак, з іншого боку, якщо «гарний» клієнт випадково

потрапить у розряд «поганих», то це буде певною втратою банком можливого прибутку.

У цьому розумінні з економічним ефектом тісно пов'язаний управлінський ефект. Побудована скорингова модель не тільки дає змогу передбачати ймовірність дефолту, але й дозволяє обрати певну кредитну політику. Якщо керівництво вирішує мінімізувати комерційні ризики, воно має обрати для використання модель з високим значенням чутливості. Навпаки, при мінімізації ризиків, пов'язаних з втратами власних коштів, слід прийняти модель з високим значенням специфічності. Отже, виважене застосування моделі кредитного скорингу дозволить приймати обґрунтовані управлінські рішення.

Науково-технічний ефект проявляється в отриманні нових знань щодо портрету «гарного» позичальника, виявленні тих його характеристик, які є значущими при оцінці кредитоспроможності.

Соціальний ефект виражається через цілі затребуваної позики і може призводити до покращення:

- житлових умов (іпотечні кредити, кредити на придбання нерухомості);
 - побутових умов (кредити на ремонт, придбання техніки, автокредитування тощо);
 - рівня здоров'я в регіоні (кредити на путівки для оздоровлення, спортивне обладнання, абонементи у спортклуби тощо);
 - збільшення кількості робочих місць (кредити для розвитку бізнесу);
- хоча в даному випадку може прослідковуватись і обернений ефект, пов'язаний зі зменшенням кількості робочих місць завдяки впровадженню у виробничий процес нових технологій.

Враховуючи все вище наведене, впровадження скорингової моделі оцінки кредитоспроможності позичальника є доцільним.

ВИСНОВКИ

В ході виконання дипломної роботи було розроблено скорингову модель оцінювання кредитоспроможності позичальника банку. Висунуто вимоги до моделі оцінювання кредитоспроможності позичальника банку, визначено ознаковий простір моделі.

В процесі було сформовано основну мету, а саме розробка моделі оцінювання кредитоспроможності позичальника-фізичної особи на основі підходу кредитного скорингу.

В свою чергу для досягнення мети моделювання були вирішені наступні завдання:

- розглянуто сутність оцінювання кредитоспроможності позичальника банку як складової кредитної діяльності банку;
- проведено огляд сучасного стану моделювання оцінювання кредитоспроможності позичальника та обґрунтовано доцільність застосування кредитного скорингу;
- сформовано змістову постановку задачі моделювання;
- сформувано загальні вимоги до розроблюваної моделі оцінювання кредитоспроможності позичальника банку;
- описано ознаковий простір моделі, вхідні та вихідні дані моделювання;
- описано математичну модель оцінювання кредитоспроможності позичальника банку;
- зроблено практичну реалізацію скорингової моделі оцінювання кредитоспроможності позичальника банку;
- проаналізовано адекватність, точність та чутливість розробленої моделі;

– розроблено пропозиції з використання моделі та шляхів її подальшого удосконалення.

Практична значущість побудованої моделі полягає у можливості використання її у банках та кредитних установах для оцінки кредитоспроможності позичальника і зменшення ризиків неповернення кредитів. Користувачами розроблюваної моделі можуть бути робітники банку під час прийняття рішення про видачу кредиту, що дозволить застосовувати уніфікований та позбавлений суб'єктивізму підхід у всіх відділеннях банку та зменшити ризик неповернення кредитних коштів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Berg, T., M. Puri, and J. Rocholl. Loan Officer Incentives, Internal Rating Models and Default rates, *Working Paper*. 2017. 52 p.
2. Dorfleitner, G, C. Priberny, S. Schuster, J. Stoiber , M. Weber, I. de Castro , and J. Kammler. Descriptiontext related soft information in peer-to-peer lending – Evidence from two leading European platforms. *Journal of Banking & Finance*. 2016. №64. P 169-187.
3. Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. Predictably Unequal? The Effects of Machine Learning on Credit Markets, *Working Paper*. 2017. 72 p.
4. Fuster, A., M. Plosser, P. Schnabl, and J. Vickery. The Role of Technology in Mortgage Lending, *The Review of Financial Studies*. 2018. vol 32(5). P 1854-1899.
5. Gao, Q., M. Lin, and R. Sias. Word Matters: The Role of Texts in Online Credit Markets, *Working Paper*. 2017. 58 p.
6. Hertzberg, A., A. Liberman, and D. Paravisini. Adverse Selection on Maturity: Evidence from On-Line Consumer Credit, *Working Paper*. 2016. 66 p.
7. Hildebrandt, T., M. Puri, and J. Rocholl. Adverse Incentives in Crowdfunding. *Management Science*. 2017. 63(3). P 587-608.
8. Iyer, R., A. Khwaja, E. Luttmer, and K. Shue. Screening Peers Softly: Inferring the Quality of Small Borrowers, *Management Science*. 2016. 62(6). P 1554-1577.
9. Kawai, K., K. Onishi, and K. Uetake. Signaling in Online Credit Markets, *Working Paper, Yale University*. 2016. 54 p.
10. Puri, M., J. Rocholl, and S. Steffen: What do a million observations have to say about loan defaults? Opening the black box of relationships, *Journal of Financial Intermediation*. 2017 31, 1-15 p.

11. Vallee, B., Yao Zeng: Marketplace Lending: A New Banking Paradigm? Working Paper. 2018. 54 p.
12. Luo, C. A comprehensive decision support approach for credit scoring, *Industrial Management & Data Systems*. 2019. <https://doi.org/10.1108/IMDS-03-2019-0182>
13. Малярець Л. М. Вимірювання ознак об'єктів в економіці: методологія та практика. Харків: ХНЕУ, 2006. 382 с.
14. Allison, P.D. Logistic regression using the SAS system: theory and application. Cary, NC: SAS Institute, 1999. 303 p.
15. Anderson R. The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. New York: Oxford University press, 2007. 790 p.
16. Сорокин А. С. Построение скоринговых карт с использованием модели логистической регрессии. *Интернет-журнал «Наукоедение»*. Вып. 2 2014. URL: <http://naukovedenie.ru /PDF/180EVN214.pdf>
17. Harrell F. Regression Modeling Strategies URL: <http://www.springer.com/gb/book/9783319194240>.
18. Интервальное оценивание генеральной доли URL: <http://math.semestr.ru/group/interval-estimation-share.php>.
19. Жилина Е. В. Использование бинарной логистической регрессии для оценки качества адаптивного теста. *Вестник Томского государственного университета*. 2010. С. 106 – 109.
20. Паскар В. В. Методики визначення якості рейтингових моделей оцінки кредитоспроможності. URL: http://www.rusnauka.com/13_EISN_2012/Economics/8_109839.doc.htm.
21. Паклин Н. Логистическая регрессия и ROC-анализ — математический аппарат. URL: <http://www.basegroup.ru/library/analysis/regression/logistic/>.

22. Гідулян А. Актуальні питання поліпшення методики оцінки кредитоспроможності позичальників банками України. *Вісник НБУ*. січень 2012. С. 50-53.
23. Engelmann B. Testing rating accuracy / В. Engelmann, E. Hayden, D. Tasche. *Risk magazine*. January 2003. P. 82-86.
24. Васильчак, С.В. Оцінка кредитоспроможності позичальника як один із методів забезпечення економічної безпеки банку. *Національний лісотехнічний університет України*. 2012. № 22.2. С. 153-161.
25. Андрушків, Т. В. Проблеми оцінки кредитоспроможності позичальників в управлінні кредитним ризиком банку. *Актуальні питання теорії і практики банківської діяльності*. 2008. № 2(15). С. 113-118.
26. Дзюблюк, О. В. Особливості оцінювання грошового потоку підприємства в системі банківського аналізу кредитоспроможності позичальника. *Банківська справа*. 2001. №1. С. 8-12.
27. Аналіз і оцінка кредитоспроможності позичальника: науково – практичний посібник / Д. А. Єндовицький, І. В. Бочарова. Москва: КНОРУС, 2005. – 272 с.
28. Галасюк, В. В. Оцінка кредитоспроможності позичальників – що оцінюємо? URL: <http://www.galasyuk.com.ua/downloads/publications/credit1.pdf>.
29. Слав`янська, Н. Г. Банківські операції: підручник / Н. Г. Слав`янська ; ред. В. І. Міщенко. Київ : Знання, 2006. 727 с.
30. Бордюг, В.В Теоретичні основи оцінки кредитоспроможності позичальника банку. *Вісник університету банківської справи національного банку України*. 2008. №3. С. 112115.
31. Про порядок формування та використання банками України резервів для відшкодування можливих витрат за активними банківськими операціями: положення затверджене постановою Правління Національного

банку України від 15 лютого 2012 року № 23. URL: <http://document.ua/pro-zatverdzhennja-polozhennja-proporjadok-formuvannja-ta-v-doc89633.html>.

32. Остафіль, О. М. Комплексна оцінка кредитоспроможності позичальника як інструмент управління кредитним ризиком банку / О. Остафіль, М. Рубаха. *Формування ринкової економіки в Україні*. 2009. Вип.19. С. 387-396.

33. Оцінка кредитоспроможності та інвестиційної привабливості суб'єктів господарювання : монографія / А. О. Єпіфанов, Н. А. Дехтяр, Т. М. Мельник, І. О. Школьник та ін.; ред. д-ра екон. наук А. О. Єпіфанова. Суми : УАБС НБУ, 2007. 286 с.

34. Шемшученко, Г. Ю. Банківський кредит у системі кредитних відносин. Держава і право : зб. наук. пр. Юридичні і політичні науки. Київ : Ін-т держави і права ім. В. М. Корецького НАН України, 2010. Вип. 10. С. 45–47.

35. Бібліотека економіста. Операції комерційних банків URL: <http://library.if.ua/book/89/6252.html>.

36. Манако, О.В. Системний підхід до оцінок кредитоспроможності підприємства : збі. наук. праць МННЦ ІТіС. 2007. №12. С. 83-95.

37. Левицький, В.В. Особливості оцінки кредитоспроможності позичальника : журн. європ. економіки. 2011. №13. С. 239-242

38. Некрашевич, В.Н. Оцінка кредитоспроможності підприємства. *Наука й економіка*. 2010. №14. С. 114117.

39. Гаряга Л. О. Моніторинг кредитного ризику з використанням інтегрального показника URL: <http://www.nbu.gov.ua>.

40. Kaggle: The Home of Data Science URL: <https://www.kaggle.com/brycecf/give-me-some-credit-dataset>.

ДОДАТКИ

ДОДАТОК А

SUMMARY

Baranova A.O. Modeling the process of assessing the creditworthiness of bank borrowers. – Qualification master’s work. Sumy State University, Sumy, 2019.

The master’s thesis focuses on the essence of the credit borrower's creditworthiness assessment as a component of bank credit activity is investigated and the expediency of using credit scoring is grounded. The main aim of this research is to develop a model for assessing the creditworthiness of a borrower-individual based on a credit scoring approach.

Keywords: creditworthiness, modeling, commercial bank, credit scoring, borrower.

АНОТАЦІЯ

Баранова А.О. Моделювання процесу оцінювання кредитоспроможності позичальників банку. – Кваліфікаційна магістерська робота. Сумський державний університет, Суми, 2019.

У роботі досліджено сутність оцінювання кредитоспроможності позичальника банку як складової кредитної діяльності банку та обґрунтовано доцільність застосування кредитного скорингу. Основною метою цього дослідження є розробка моделі оцінювання кредитоспроможності позичальника-фізичної особи на основі підходу кредитного скорингу.

Ключові слова: кредитоспроможність, моделювання, комерційний банк, кредитний скоринг, позичальник.

ДОДАТОК Б

Налаштування вузла Replace Missing Data для вилучення пропущених даних з історичної вибірки

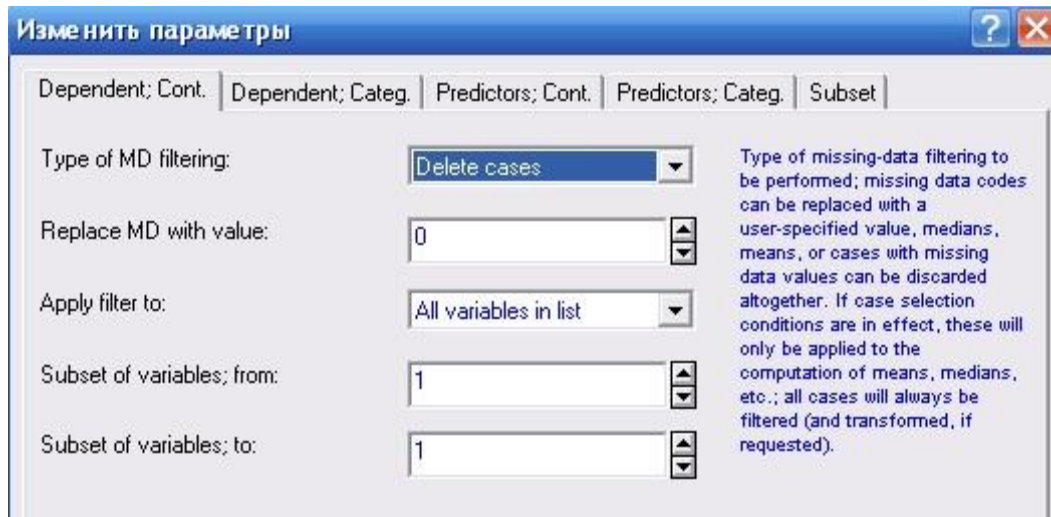


Рисунок Б.1 – Налаштування вкладки «Dependent; Cont»

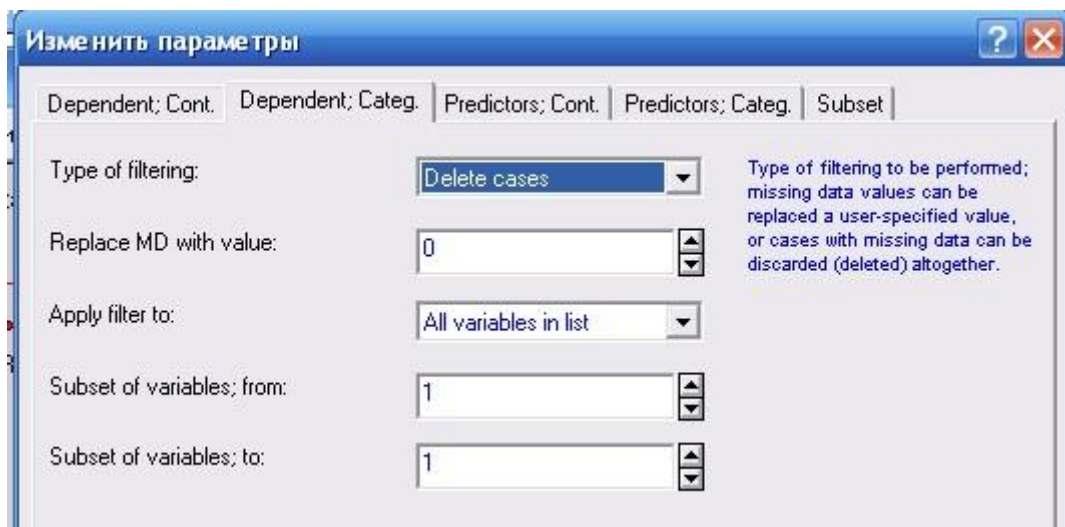


Рисунок Б.2 – Налаштування вкладки «Dependent; Categ»

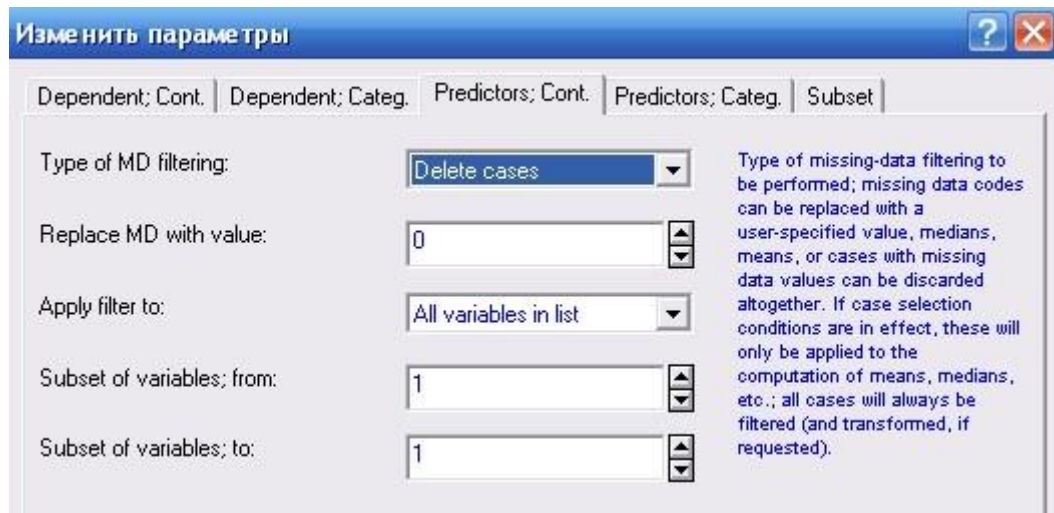


Рисунок Б.3 – Налаштування вкладки «Predictors; Cont»

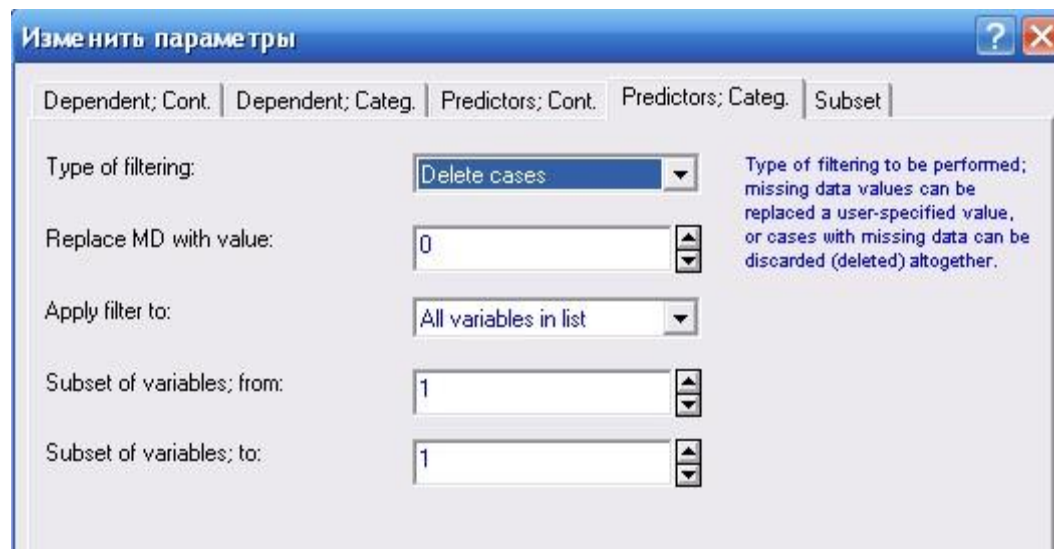


Рисунок Б.4 – Налаштування вкладки «Predictors; Categ»

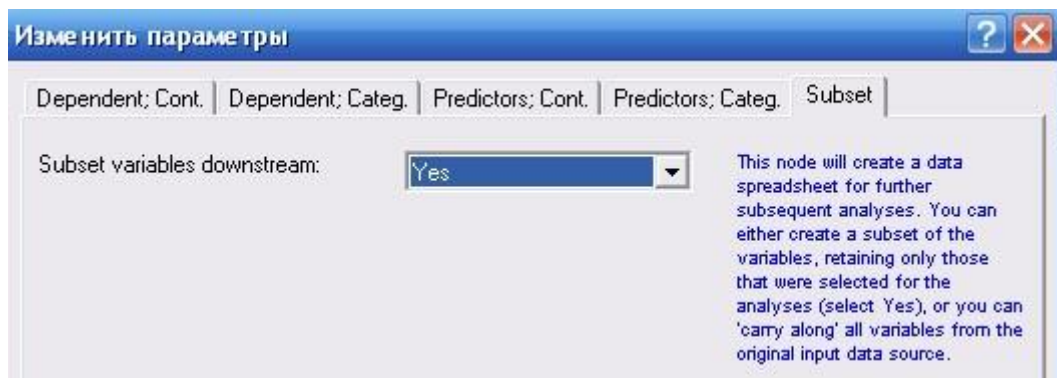


Рисунок Б.5 – Налаштування вкладки «Subset»

ДОДАТОК В

Розрахунки для визначення мінімально необхідного обсягу навчальної вибірки

	A	B	C	D
1	3587	кількість дефолтів		
2	48986	кількість позитивних відгуків		
3	52573	заг к-сть відгуків		
4				
5	0,07	частка "поганих" відгуків		
6	0,93	частка "гарних" відгуків		
7	0,05	рівень значущості		
8	1,96	z-табличне значення		
9	0,0034	гранична помилка оцінки частки		
10				
11	n=	20985	<i>мінімальна кількість вимірювань у навчальній вибірці</i>	
12				
13	34290	<i>навчальна вибірка</i>		
14	15772	<i>тестова вибірка при розбитті 70% на 30%</i>		
15				

Рисунок В.1 – Результати розрахунку мінімально необхідного обсягу навчальної вибірки (виконано в MS Excel)

	A	B	C
1	3587	кількість дефолтів	
2	48986	кількість позитивних відгуків	
3	=A1+A2	заг к-сть відгуків	
4			
5	=A1/A3	частка "поганих" відгуків	
6	=A2/A3	частка "гарних" відгуків	
7	0,05	рівень значущості	
8	1,96	z-табличне значення	
9	=A5*A7	гранична помилка оцінки частки	
10			
11	n=	=A8^2*A5*(1-A5)/A9^2	<i>мінімальна кількість вимірювань у навчальній вибірці</i>
12			
13	=A2*0,7	<i>навчальна вибірка</i>	
14	=A3*0,3	<i>тестова вибірка при розбитті 70% на 30%</i>	
15			

Рисунок В.2 – Формули MS Excel, за якими проводились розрахунки

ДОДАТОК Г

Налаштування вузла Split Input Data into Training and Testing Samples
для розбиття історичної вибірки на навчальну і тестову

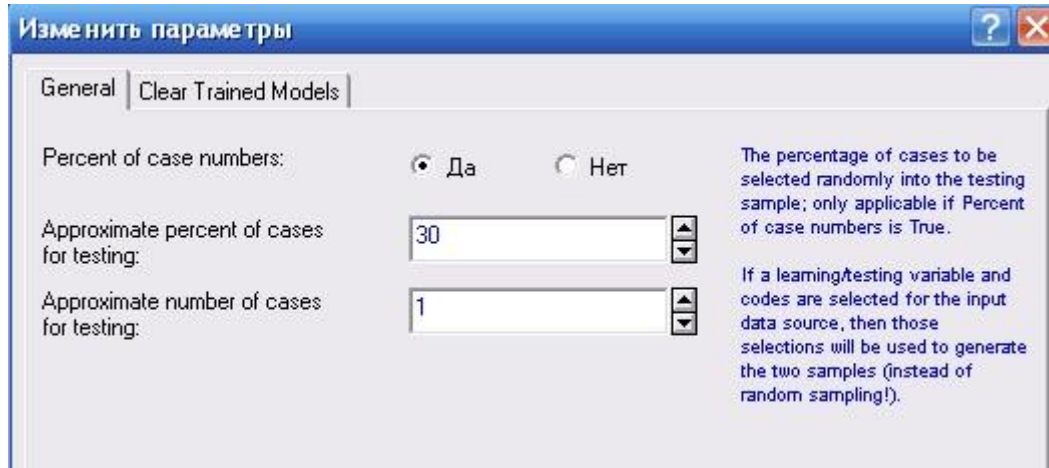


Рисунок Г.1 – Налаштування вкладки «General»

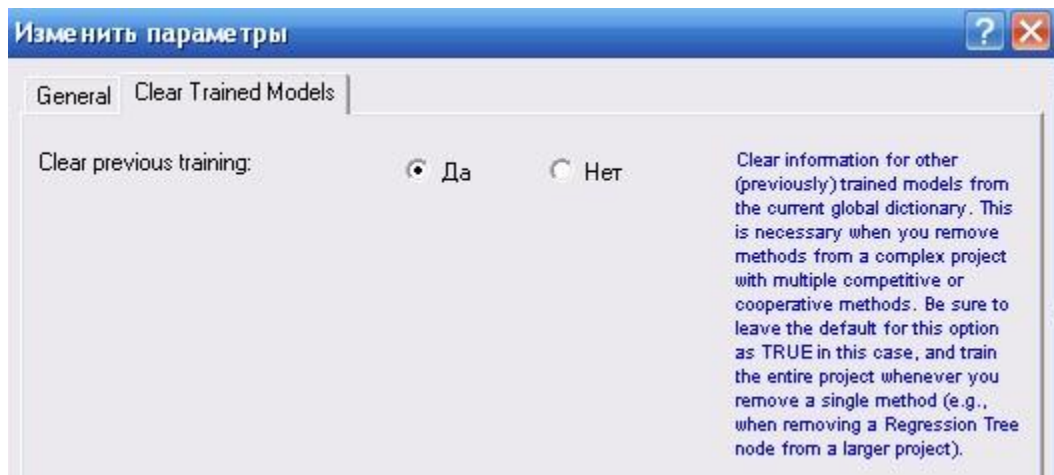


Рисунок Г.2 – Налаштування вкладки «Clear Trained Models»

ДОДАТОК Д

Налаштування вузла Logit Regression with Deployment для побудови моделі логістичної регресії

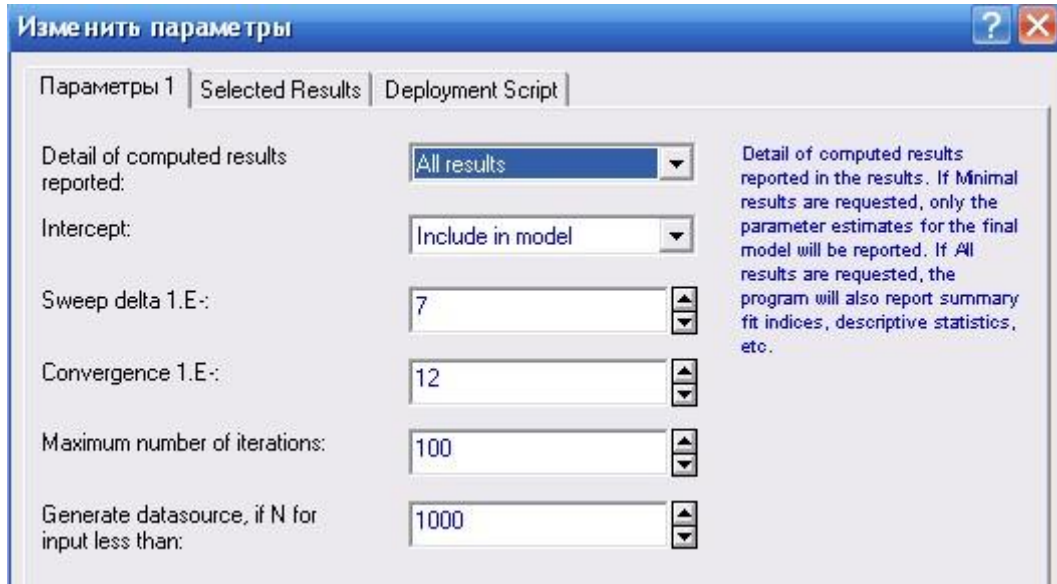


Рисунок Д.1 – Налаштування вкладки «Параметры 1»



Рисунок Д.2 – Налаштування вкладки «Selected Results»

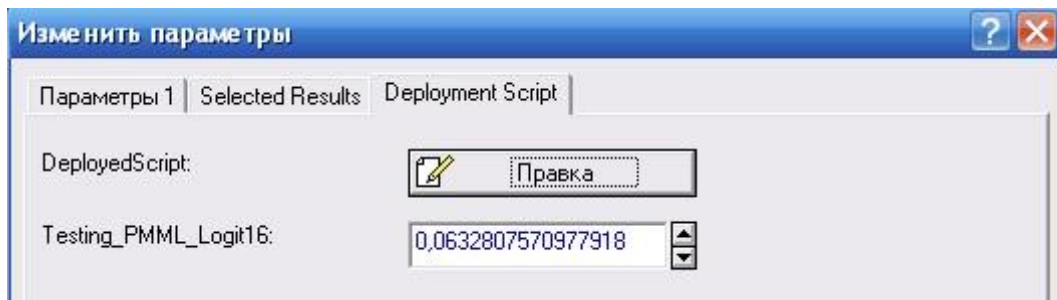


Рисунок Д.3 – Налаштування вкладки «Deployment Script»

ДОДАТОК Е

Розрахунки чутливості та специфічності побудованої скорингової моделі

	А	В	С
1	Класифікація на основі моделі	Фактично	
2		Позитивно	Негативно
3	Позитивно	14694	911
4	Негативно	126	192
5			
6			
7	Чутливість=	99,15	
8	Специфічність=	17,41	
а			

Рисунок Е.1 – Результати розрахунку чутливості і специфічності скорингової моделі (виконано в MS Excel)

	А	В	С
1	Класифікація на основі моделі	Фактично	
2		Позитивно	Негативно
3	Позитивно	14694	911
4	Негативно	126	192
5			
6			
7	Чутливість=	=B3/(B3+B4)*100	
8	Специфічність=	=C4/(C3+C4)*100	
а			

Рисунок Е.2 – Формули MS Excel для розрахунку чутливості і специфічності скорингової моделі