# РОЗДІЛ 1. CКОРИНГОВІ СИСТЕМИ ЯК МЕТОД УПРАВЛІННЯ КРЕДИТНИМИ РИЗИКАМИ

## Кредитний ризик та методи боротьби з ним. Постановка задачі магістерської роботи.

## Основні поняття регулюванням кредитного ризику.

Основним завданням цього підрозділу - є пояснення причини виникнення та впровадження кредитного скорингу, щоб, в подальшому, провести оцінку автоматизованої системи кредитного скорингу.

Кредитний ризик — ([англ.](https://uk.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D1%96%D0%B9%D1%81%D1%8C%D0%BA%D0%B0_%D0%BC%D0%BE%D0%B2%D0%B0) *credit risk*)ймовірність відхилення від запланованих (очікуваних) показників через невиконання позичальником зобов'язань перед банком. Кредитний ризик доцільно розділяти на індивідуальний (конкретний контрагент банку) та портфельний (сукупна заборгованість) ризики[1].

Щоб правильно оцінити платоспроможність потенційного клієнта, використовують скоринг.

Насамперед, скоринг - це система оцінки позичальників, яка заснована на математикостатистичних методах. Мета скорингу – це оцінити рівень платоспроможності клієнта по деяким факторам і відібрати потенційних позичальників. Кредитні менеджер, під час співбесіди з позичальником, збирає максимум інформації про нього, після чого скорингова система обробляє інформацію, нараховуючи бали за кожен сприятливий фактор [2].

Найперша скорингова модель з'явилася в банках США, коли, під час Другої світової війни, практично всі кредитні інспектори були покликані захищати батьківщину. За проханням банків та інших кредитних організацій, перед відходом, кредитні аналітики розробили модель для прийняття рішень про видачу кредиту. Цією моделлю могли користатися навіть неспеціалісти.

Наступний крок у розвитку скорингу - це консалтингова компанія Fair Issac Corporation (FIC), яка займалася розвитком скорингових моделей. Сьогодні FIC продовжує розробляти скорингові системи для банків. А використовувана шкала FICO (Таблиця 1), яку створила компанія FIC, застосовується в більшості банків США. До уваги приймаються наступні складові:

* якість кредитної історії,
* наявність і розмір поточних боргів,
* тривалість відносин з кредиторами,
* співвідношення кількості поданих заявок і виданих кредитів,
* типи виданих кредитів [3].

|  |  |
| --- | --- |
| ***Бали*** | ***Опис*** |
| ***750+*** | Відмінний кредитний бал. Позичальнику надається кредит на самих вигідних умовах |
| ***700–750*** | Хороший кредитний бал. Немає проблем з видачею кредиту. Позичальник отримує кредит під хороший  відсоток. |
| ***640–700*** | Середній кредитний бал. Банк може видати позичальникові кредит, але на не дуже вигідних умовах.  Також можливий відмова банку. |
| ***580–640*** | Слабкий кредитний бал. Довгий процес прийняття рішення про видачу кредиту позичальникові і високі  процентні ставки |
| ***>580*** | Поганий кредитний бал. відмова у видачі кредиту позичальникові. |

Таблиця 1. Шкала FICO - показник якості позичальника [4]

На розвиток скорингу вплинуло ще два фактора. У США був прийнятий закон, про рівні отримання кредиту, який зобов'язав банки розглядати всі заявки (боротьба з расовою дискримінацією). Наступний фактор – це розвиток інформаційних технологій, які могли обробляти велику кількість надходження кредитних заявок.

Скорингові системи діляться на кілька видів, самі популярні з них:

• скоринг на підставі кредитної історії, який враховує кредитну історію потенційного позичальника (позики різних банків, прострочення (якщо вони є), спроби взяти позику у банку, наявність кредитних карт).

Головний недолік такого підходу очевидний. Вибірка класифікується тільки по тим клієнтам, яким вже давали кредит. Залишалося невідомим, як повели б себе ті клієнти, яким було відмовлено в кредиті або які за ним навіть не зверталися[3].

• Соціально-демографічний скоринг – це оцінка позичальника на підставі таких показників як вік або гендерні показники, сімейне положення, стаж роботи, професія.

Також скоринг враховує заробітню плату, як правило, за останні 6 місяців [3].

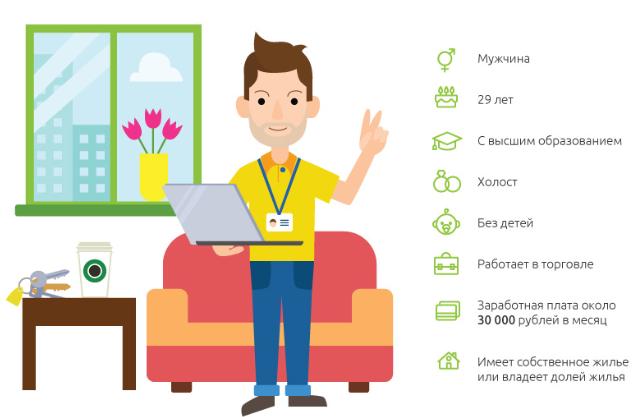


Рис. 1.1.1. Соціально-демографічні дані про клієнта

Кредитний менеджер проводить співбесіду і анкетування з потенційним позичальником, після чого вносить дані в програму. На підставі цих даних скорингова система присвоює бали за кожен фактор, а в кінці процедури відносить позичальника до певної «групи ризику» і віддає висновок про можливість надання кредиту.

Незважаючи на скоринговую систему, в кожному банку існує власна система

оцінки потенційних позичальників. Наприклад, є певні пороги за рівнем заробітної плати, якщо значення нижче поставленого порога, клієнт отримує відмову банку ще на рівні прескорінга.

В умовах кризи, для мінімізації фінансових ризиків деякі банки ввели мораторій на видачу позички непрацюючим пенсіонерам, чий дохід потенційно не зможе забезпечити повного погашення кредиту. Крім цього, співробітники можуть візуально оцінювати потенційного контрагента, чия неадекватна поведінка або неприйнятний зовнішній вигляд може стати перешкодою до отримання кредиту.

Після всіх, вищевикладених методів оцінки, заявка повинна пройти етап андеррайтингу і отримати схвалення у департаменту аналізу ризиків та служби

безпеки. У разі, якщо всі етапи пройдено, це ще не гарантія отримання кредиту. В умовах нестабільної економіки в регіонах, внутрішній контроль має право відмовити клієнту, проте оформлення страхового договору, дає високі

шанси на успіх позичальника [5].

Управління кредитними ризиками займає окреме місце в ефективному менеджменті будь-якого банку. Під кредитним ризиком мається на увазі невиконання контрагентом своїх кредитних зобов'язань. Найбільш поширений кредитний ризик - дефолт позичальника, коли контрагент не виконує зобов'язання перед банком щодо повернення грошових коштів згідно з умовами кредитного угоди в силу економічної неспроможності або небажання.

Фактори кредитного ризику носять як зовнішній характер, так і внутрішній. Зовнішні фактори пов'язані з можливістю реалізації кредитного ризику через нездатність позичальника погасити заборгованість перед банком. В той час

як, внутрішні фактори пов'язані з помилками кредитних менеджерів, департаменту аналізу ризиків або інших співробітників, які були допущені в ході оформлення заявки або оцінки позичальника [6].

Таблиця 2.Внутрішні чинники кредитного ризику

|  |  |
| --- | --- |
| **Внутрішні фактори** | **Характеристика факторів кредитного ризику** |
| *Фактори,* | * Зміст і умови комерційної діяльності позичальника |
| *пов'язані з* | * Кредитоспроможність позичальника |
| *діяльністю* | * Рівень менеджменту позичальника |
| *позичальника* | * Банкрутство позичальника |
|  | * Шахрайство з боку позичальника |
|  | * Репутація позичальника |
| *Фактори,* | * Адекватність вибору кредитної політики |
| *пов'язані з* | * Помилкові дії кредитних працівників |
| *діяльністю*  *банка- кредитора* | * Якість технологій * Кваліфікація персоналу * Тип ринкової стратегії |
|  | * Здатність розробляти і просувати |
|  |  |

1.1.2 Проблема кредитного скорингу в Україні

Проблема кредитного скорингу в Україні досить актуальна на сьогодні. У кожного банку своя кредитна політика, але спільне завдання - мінімізувати ризики при видачі кредитів.

Зараз банки, в основному, вибирають консервативну політику, що, на наш погляд, є правильним рішенням, особливо після кризи 2008 року, коли в США іпотечні менеджери видавали позику практично всім підряд. В українській практиці існує два основних підходи оцінки ризику кредитування, які, зазвичай, застосовуються в поєднанні один з одним [7]:

• Суб'єктивне укладення кредитних менеджерів

• Автоматизовані системи скорингу.

Основним завданням цих двох підходів є з'ясування ймовірності повернення кредиту контрагентом. Скорингові системи в Україні більш зосереджені на використанні оцінки ризиків при кредитуванні фізичних осіб, ніж юридичних. Автор зазначає, що це пов'язано, перш за все, з труднощами оцінки фінансового стану підприємств, так як всі компанії різні за своєю діяльністю, працюють в різних секторах, і у кожної компанія різні масштаби. Для визначення ризиків кредитування юридичних осіб, крім скорингових систем, використовується моніторинг фінансового стану компанії шляхом оцінки вартості бізнесу та його активів [7]. Для роздрібного кредитування все набагато простіше, так як потенційні позичальники сегментируются на «хороших» і «поганих». В Україні банки стали використовувати скорингові моделі на початку нового століття. З високою конкуренцією і активним розвитком споживчого кредитування, без скорингових систем, банкам не вдасться конкурувати на ринку [8]. Кредитний скоринг розділяється на кілька типів. Далі ми розглянемо, які типи актуальні для українських банків:

• Application-скоринг. Application-скоринг - оцінка кредитоспроможності позичальників для отримання банківської позики по даних із заяви на кредит. Частка прострочення досягає 4,3% і продовжує зростати. Таким чином, можна сказати, Application-скоринг найбільш актуальний тип скорингу для України.

• Collection-скоринг. Останнім часом українські банки наголошують на необхідності збільшити використання Collection-скорингу. Цей тип скорингу допомагає проводити планомірну роботу з простроченою заборгованістю до моменту її передачі в колекторський відділ. Досвід показує, що значну частину заборгованості в ході цієї роботи вдається ліквідувати [9].

• Behavioral-скоринг. Behavioral-скоринг - це оцінка динаміки стану кредитного рахунку позичальника. Така модель може спрогнозувати зміни платоспроможності контрагента на підставі різних факторів. Поведінковий скоринг дозволяє визначити оптимальні для позичальника строки погашення, суми та ліміти.

• Fraud-скоринг. Спільно з вищеописаними типами скорингу і службою безпеки, банки використовують Fraud-скоринг, який дозволяє визначити ймовірність шахрайських дій з боку клієнта [9].

Хочеться відзначити, що з розвитком інформаційних технологій і появою автоматизованих скорингових систем, процес прийняття рішень про видачу кредиту став набагато зручніше, безпечніше і швидше. В умовах конкурентної боротьби на ринку кредитування неможливо уявити банк, який не використав би скорингові системи і інші методи управління кредитними ризиками. В даний час, підвищений попит позичальників на кредити зобов'язує банки використання політики управління кредитними ризиками .

Всякий ризик - це ймовірність, як сприятливого, так і негативного результату. Щоб не зазнати збитків банкам необхідно мінімізувати ризики, а скорингові системи дозволяють зробити це. Тому завдання магістерської роботи – освоїти методи побудови скорингу та покращити точність скорингової моделі.

* 1. Аналіз існуючих у світі способів її вирішення.

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

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

## Побудова скорингової моделі проходить такі етапи:

## **C:\Users\User\Desktop\практика\Схема.PNG**

Рис.1.2.1. Етапи побудови скорингової моделі

Щоб побудувати скорингову модель, нам потрібно класифікувати клєнтів.

Тому розлянемо основні методи класифікації.

* + 1. Методи класифікації

Класифікація - це віднесення об'єктів до певного класу по набору ознак. Наприклад, розпізнавання номерів машин, або в медицині діагностика захворювань, або кредитний скоринг в банківській сфері. Цеодин з розділів машинного навчання, що присвячений вирішенням наступного завдання. Нехай ми маємо безліч об'єктів (ситуацій), розділених деяким чином на класи. Дано скінченну кількість об'єктів, для яких відомо, до яких класів вони належать. Ці об’єкти називається навчальною вибіркою. Класова приналежність інших об'єктів невідома. Потрібно побудувати алгоритм, здатний класифікувати довільний об'єкт.

Класифікувати об'єкт - значить, вказати номер (або найменування класу), до якого належить даний об'єкт.

На сьогодні існує 4 типи задач[10]:

* Навчання з учителем;
* Навчання без вчителя;
* Навчання з частковим залученням вчителя;
* Навчання з підкріпленням.

**Навчання з вчителем**

Навчання з учителем - це навчання на тренувальному наборі даних, шляхом підгонки результатів навчання на тренувальний набір даних.

Основне завдання - знайти найбільш оптимальні параметри моделі для прогнозування. Якщо результат прогнозу є дійсним числом, то це завдання регресії. Якщо має обмежену кількість значень, де ці значення є неврегульованими, то це завдання класифікації.

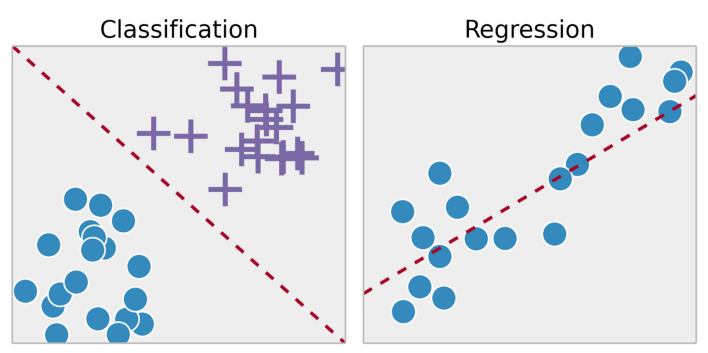


Рис1.2.2. Навчання із вчителем

**Навчання без вчителя**

У навчанні без контролю у нас менше інформації про об'єкти. Зокрема, тренувальний набір не має маркованих даних, що відносяться до певного класу заздалегідь зумовлених даних. Яка наша мета зараз? Спостерігати деяку схожість між групами об'єктів і включати їх до відповідних кластерів.

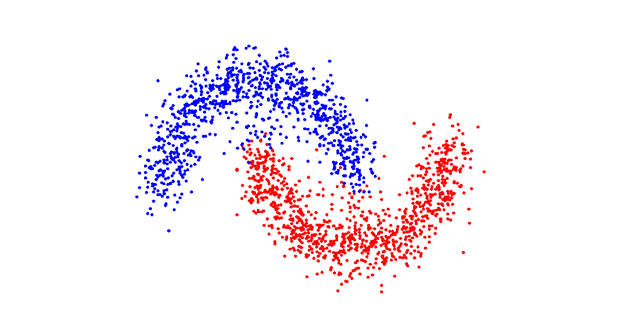


Рис1.2.3. Навчання без вчителя

**Навчання з частковим залученням вчителя**

Навчання з частковим залученням вчителя включає обидві проблеми, які ми описали раніше: вони використовують марковані і немарковані дані. Це відмінна можливість для тих, хто не може розподілити маркувати свої дані. Цей метод дозволяє значно підвищити точність, оскільки ми можемо використовувати немарковані дані в тренувальному наборі даних з невеликою кількістю маркованих даних.

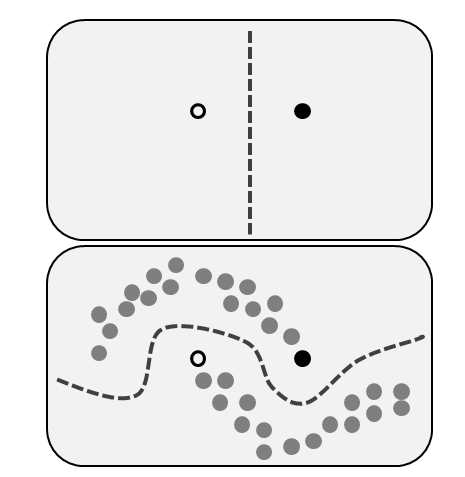


Рис1.2.4. Навчання з частковим залученням вчителя

**Навчання з підкріпленням**

Навчання з підкріпленням не схоже на будь-яку з наших попередніх завдань, тому що тут ми не маємо в своєму розпорядженні ні зумовленими маркованими даними, ні немаркованими наборами даних. Навчання з підкріпленням - область машинного навчання, пов'язана з тим, як агенти програмного забезпечення повинні вживати дії в деякому середовищі, щоб максимізувати деяке поняття кумулятивної нагороди.

Ідея навчання з підкріпленням полягає в тому, що система буде вчитися в середовищі, взаємодіяти з нею і отримувати винагороду за виконання дій.

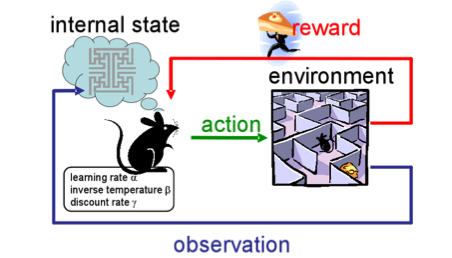


Рис.1.2.5. Навчання з підкріпленням

Уявіть, що ви робот в якомусь дивному місці. Ви можете виконувати дії і отримувати нагороди від навколишнього середовища за них. Після кожної дії поведінка навколишнього середовища стає більш складнішою тому ви тренуєтеся, щоб вести себе найбільш ефективним способом на кожному кроці. У біології це називається адаптацією до природного середовища.

Завдання класифікації відноситься до розділу навчання з учителем.

Основними методами класифікації для побудови скорингової моделі серед банків України:

- Дерева прийняття рішень.

- Логістична регресія;

1.2.3. Дерева прийняття рішень (С4.5, CART)

Алгоритм С4.5

Алгоритм C4.5 будує класифікатор в формі дерева рішень.

Припустимо, що у нас є набір даних - це дані про групу потенційних клієнтів банку. Ми знаємо різні параметри кожного пацієнта: вік, освіта, наявність страхування наявність нерухомості, історію сім'ї і так далі (це параметри). На підставі цих параметрів ми хочемо передбачити, чи може пацієнт отримати кредит. Пацієнт може потрапити в один з 2 класів: отримає та не отримає кредит. Алгоритму C4.5 повідомляє клас кожного клієнта використовуючи набір параметрів пацієнта і відповідний клас, C4.5 будує дерево рішень, здатне передбачити клас для нових пацієнтів на підставі їх параметрів.

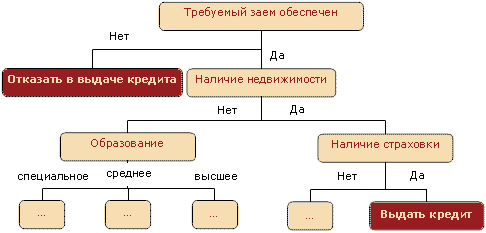


Рис.1.2.6. Дерево рішень

Класифікація методом дерева рішень створює своєрідні блок-схеми для розподілу нових даних. У кожній точці блок-схеми задається питання про значимість того чи іншого параметра, і в залежності від цих параметрів він або вона [клієнти] потрапляють в певний клас.

Задля того, щоб визначити найбільш інформативну змінну(змінна, що буде стояти на початковому етапі відбору) використовують принцип припливу інформації(Entropy and Information Gain).

Ентропія (Entropy) – це те, як багато інформації вам не відомо про стан об'єкта (або джерела інформації)[10].

Використовується, щоб визначити скільки інформації наявно у певному джерелі інформації.

У загальному випадку, якщо опиратися на теорію ймовірності, джерело інформації – це є ансамблі таких станів K = {k1, k2, ..., kN} , тобто ймовірність цих станів - {Р(k1), Р(k2), ..., Р(kN)} при цьому сума ймовірностей станів рівна 1.

Кількість інформації, яку ми можемо отримати з певного джерела інформації розраховується за формулою, яку запропонував К. Шеннон в 1946 році:

d

B(k) = -Σ P(k\_d) \* log2(P(k\_d)).

1

d - к-сть змінних,

P(k\_d) – ймовірність класу k змінної d.

Розрахуємо ентропію та знайдемо найбільш інформативні змінні.

Нехай клієнти банку розподілені таким чином:

Таблиця 1.2.1 Розподіл клієнтів банку

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Клієнт | Наявність  забезпечення | Наявність нерухомості | Наявність страхування |
| 1 | Good | yes | no | yes |
| 2 | Bad | no | yes | yes |
| 3 | Good | yes | yes | no |
| 4 | Bad | no | no | no |
| 5 | Good | no | no | no |
| 6 | Good | yes | no | yes |
| 7 | Bad | no | yes | no |

Обрахуємо ентропію в загальному випадку, тільки на наданих клієнта – «Good»/«Bad»:

B(Клієнт) = -p('yes')log2(p('yes')) - p('no')log2(p('no'))

B(Клієнт) = -(4/7)log2(4/7) - p(3/7)log2(3/7)

B(Клієнт) = 0.985228136034

Тож значення загальної ентропії - 0.985.

Щоб дізнатися порядок змінних у дереві застосовують критерій - Information Gain.

Gain(variable) = total\_entropy - remainder(variable)

branch\_d

remainder(variable) = Σ P(variable\_branch\_d)\*B(branch)

branch

total\_entropy – загальна ентроія;

B(branch) – ентропія змінної;

P(variable\_branch\_n) – ймовірність кожного класу змінної.

Обрахуємо Information Gain для змінної «Наявність забезпечення».

remainder(Наявність забезпечення) = P('yes, Good')\*B('yes') + P(‘yes, Bad')\*B('yes') + P('no, Good')\*B('no') + P('no, Bad')\*B('no').

remainder(Наявність забезпечення) = (2/7)\*B('yes') + (1/7)\*B('yes') + (2/7)\*B('no') + (2/7)\*B('no').

remainder(Наявність забезпечення) = 0.463587.

Gain(Наявність забезпечення) = 0.985228 - 0.463587 = 0.521641.

Для порівняння обраховуємо Information Gain для інших змінних:

Gain('Наявність страхування') = 0.020244;

Gain('Наявність нерухомості') = 0.128085.

Отже, змінна «Наявність забезпечення» буде розташована на «вершині» дерева.

Відмінності C4.5 від інших систем, що використовують дерева рішень:

o По-перше, C4.5 використовує приплив інформації, при створенні дерева рішень.

o По-друге, щоб уникнути перенавчання(overfiting) вибірки використовують метод 'pruning' - відсікання гілок, які збільшують похибку моделі.

o По-третє, C4.5 може працювати з дискретними і неперервними значеннями. Обмежуючи діапазони і встановлюючи пороги даних, трансформуючи неперервні дані в дискретні.

Ймовірно, найбільшим гідністю дерев рішень є їх проста інтерпретація, також вони мають досить високу швидкість роботи, а отримані дані легкі для розуміння.

Алгоритм CART

CART (classification and regression trees) - це абревіатура, що позначає методи класифікації і регресії з використанням дерева рішень. Це методика навчання, заснована на деревах рішень, яка повертає класифікаційні або регресивні дерева.

Наприклад, знову візьмемо набір даних про клієнта. Ви можете спробувати передбачити, чи отримає клієнт кредит. Тут можливе використання двох класів: «Good» - отримає, «Bad» - не отримає.

Задля того, щоб визначити найбільш інформативну змінну(змінна, що буде стояти на початковому етапі відбору) у методі CARD використовують Gini Impurity .

d

G(k) = Σ P(k\_d) \* (1 - P(k\_d))

i=1

d - к-сть змінних,

P(k\_d) – ймовірність класу k змінної d.

Застосуємо Gini Impurity на нашому прикладі:

G(Клієнт) = P('yes') \* (1 - P('yes')) + P('no') \* (1 - P('no'));

G(Клієнт) = 4 / 7 \* (1 - 4/7) + 3 / 7 \*( 1 - 3/7);

G(Клієнт) = 0.489796.

Gini Impurity - це показник того, як часто рандомним чином вибраний елемент з набору буде неправильно позначений, тобто у нашому випадку неправильно класифікованих об`єктів буде ~ 49%.

Визначення найбільш інформативної змінної робиться аналогічним чином, як обчислення інформації для ентропії, за винятко, що замість того, щоб взяти зважену суму ентропій кожної змінної, ми беремо зважену суму Gini Impurity.

Gini\_Gain(variable) = total\_impurity - impurity\_remainder(variable)

branch\_d

remainder(variable) = Σ P(attribute\_branch\_d)\*G(branch)

branch

total\_impurity – загальний показник Gini Impurity;

G(branch) – Gini Impurity змінної;

P(variable\_branch\_d) – ймовірність кожного класу змінної.

Таблиця 1.2.2. Співвідношення CART і C.4.5

|  |  |
| --- | --- |
| **C4.5** | **Cart** |
| Для оцінки точності моделі в процесі класифікації використовується приплив інформації(Entropy and Information Gain) до сегменту даних. | Для оцінки точності моделі в процесі класифікації використовується критерій Джині(Gini Impurity) |
| Використовує однопрохідний метод проріджування, щоб зменшити перенавчання. | Починаючи з низу дерева, CART оцінює помилку класифікації в вузлі і поза вузла. Якщо похибка перевищує граничну, то гілка відкидається |
| Вузли дерева рішень можуть мати дві або більше гілок | Вузли рішення мають дві гілки |

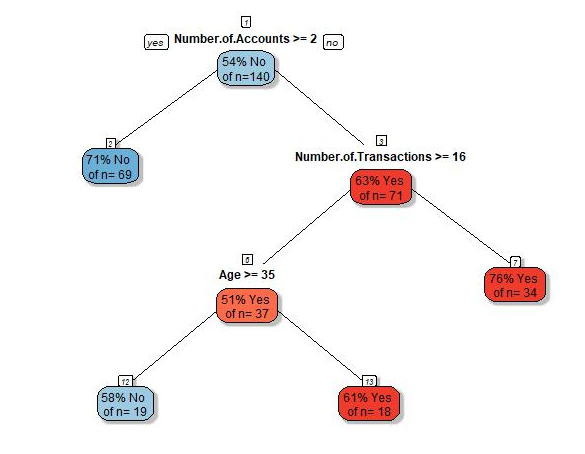


Рис.1.2.7. приклад побудови дерева рішень на основі алгоритму CART

Як і C4.5, CART досить швидкий та отримані дані легкі для розуміння.

Логарифмічна регресія

Лінійна регресійна модель не завжди здатна якісно передбачати значення залежної змінної. Вибираючи для побудови моделі лінійне рівняння, ми природним ніяк не накладаємо ніяких обмежень на значення залежної змінної. А такі обмеження можуть бути істотними. Наприклад, при проектуванні оптимальної довжини шахти ліфта в новій будівлі необхідно врахувати, що ця довжина не може перевищувати висоту будівлі взагалі.

Лінійна регресійна модель може дати результати, несумісні з реальністю. З метою вирішення даних проблем корисно змінити вид рівняння регресії і підстроїти його для вирішення конкретного завдання.

Логарифмічна регресія - це різновид множинної регресії, призначення якої полягає в аналізі зв'язку між декількома незалежними змінними (предикторами) і залежною змінною. За допомогою бінарної логістичної регресії можна оцінювати вірогідність того, що подія настане для конкретного випробуваного (повернення кредиту / дефолт).

Математична основа логістичної регресії

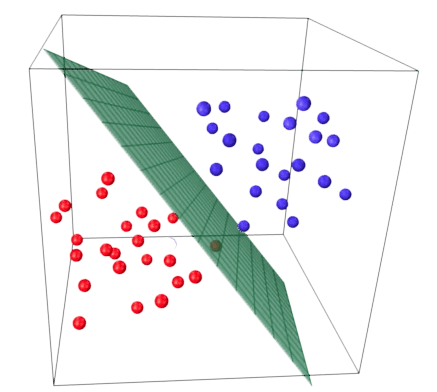
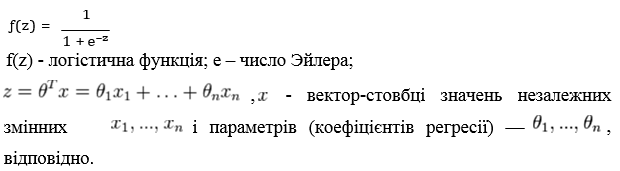


Рис.1.2.8. Класифікація за допомогою логарифмічної регресії

Завдання регресії може бути сформульована інакше: замість передбачення бінарної змінної ми передбачаємо неперервну змінну зі значеннями на відрізку [0,1] при будь-яких значеннях незалежних змінних. Це досягається

застосуванням наступного регресійного рівняння (логіт-еквіваленті):



Залежність, що зв'язує ймовірність події і величину y, показана в наступній діаграмі :

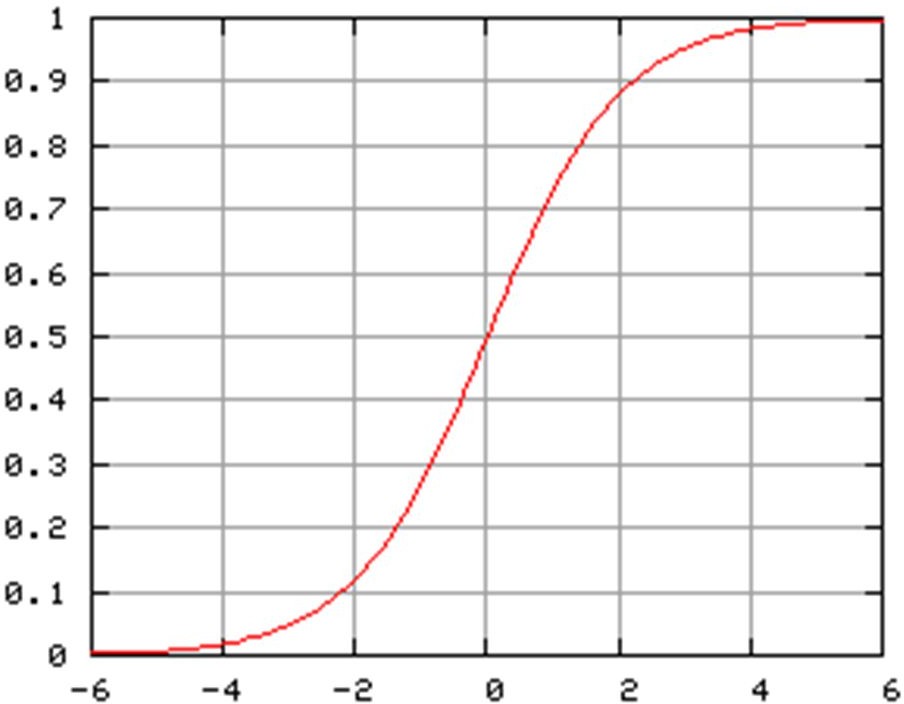


Рис1.2.9. Логарифмічна функція

Легко побачити, що незалежно від регресійних коефіцієнтів чи величин х, передбачені значення (у) в цій моделі завжди будуть лежати в діапазоні від 0 до 1.

Логістична регресія – один з найпопулярніших методів класифікації, адже цей метод досить точно і швидко працює з великим об’ємом вибірки. Також при вирішенні задач класифікації об'єкти можна розділяти на кілька груп. Наприклад, не тільки - (0 - поганий, 1 - хороший), а й кілька груп (1, 2, 3, 4 групи ризику).

Проте є недоліки при роботі із малою кількістю спостережень у вибірці. Коли розмірніcть вибірки(кількість спостережень) < 500, то можливе завищення оцінки коефіцієнтів регресії.

Присутні обов’язкові умови до вибірки при побудові моделі. Потрібно мінімально 10 спостережень на кожну незалежну змінну (рекомендоване значення 30-50):

Наприклад, смерть пацієнта. Якщо 50 пацієнтів з 100 вмирають, то максимальне число незалежних змінних в моделі = 50/10 = 5.

* 1. Визначення вхідних даних для вирішення поставленої задачі.

Вибір і побудова моделі, впровадження кредитного скорингу та його застосування, безумовно, складна і трудомістка задача. В умовах існування кредитних бюро завдання дещо спрощується щодо позичальників, які ще не

були клієнтами конкретного банку, однак існують заявники, які не зверталися до послуг кредиторів раніше. Тому цілком природно існує практика окремо оцінювати заявників, що мають і не мають кредитну історію.

Особливу увагу варто звернути на дані, за якими навчається і оцінюється модель. З одного боку, вибірка повинна бути репрезентативними, тому в них відображені як хороші, так і погані кредитні ризики.

Для прикладу була використана тестова вибірка даних. За початкові дані були взяті репрезентативна вибірка банку АТ «Альфа-Банк».

Вибірка складається із 20 змінних:

1. Стан поточного рахунку
2. Тривалість кредиту
3. Кредитна історія
4. Мета кредиту
5. Сума кредиту
6. Ощадний рахунок / облігації
7. Стаж поточної роботи
8. Ставка внеску (у відсотках від наявного доходу)
9. Сімейний статус і стать
10. Відповідальна особа
11. Нинішня резиденція
12. Тип власності
13. Вік
14. Заявник має інший кредит на розстрочку
15. Тип власності на дім
16. Кількість кредитів у цьому банку
17. Тип зайнятості
18. Кількість людей, які несуть відповідальність за кредит
19. Номер телефону(Є/ Немає)
20. Іноземний працівник (Так / Ні)

1.4. Висновки

На даний момент ринок інформаційних рішень скоринга ще не досить розвинутий в Україні.

Логістична регресія - достатньо поширений метод в кредитному скорингу.

Вона значно менше чутлива до розміру вибірки у порівнянні з багатьма іншими методами, що застосовуються для класифікації взагалі та кредитного скоринга, зокрема. Наразі існує кілька методів її вдосконалення – лассо регресія, яка «регулює» змінні, які входять у модель.

Також для класифікації плохих і хороших кредитів використовуються дерева рішень. В залежності від алгоритму породження дерева (C4.5, CART) метод має деякі переваги, серед яких простота в інтерпретації і розумінні, менша необхідність в передобробці даних (може працювати одночасно зі змінними, які оцінені за різними шкалами, не вимагає нормування або заповнення пропусків в даних). Проте зараз існують нові методи, основані на ідеї дерева рішень, які показують кращу точність – метод випадкоаих лісів та градієнтний бустинг.