

ИССЛЕДОВАНИЕ КРЕДИТНОГО РИСКА ПОТЕНЦИАЛЬНОГО КЛИЕНТА СРЕДСТВАМИ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUCTOR

Ханжин С.В., Ханжина Н.В.

Оренбургский государственный университет

Являясь наиболее распространенным видом финансового риска, кредитный риск представляет собой элемент неопределенности при выполнении контрагентом своих договорных обязательств, связанных с возвратом заемных средств. Уменьшить кредитный риск позволяет особо скрупулезный отбор потенциальных заемщиков, предварительный анализ обстоятельств предоставления кредита, регулярный мониторинг финансового состояния клиента, его способности и готовности вернуть кредит. Выполнение этих условий дает гарантии для успешного проведения одной из важнейших банковских операций - предоставления кредитов [1].

Кредитный скоринг — система оценки кредитоспособности (кредитных рисков) лица, основанная на численных статистических методах. Как правило, он используется в потребительском экспресс-кредитовании на небольшие суммы [2].

Цель исследования состоит в выработке правил оценки кредитного риска потенциального клиента кредитной организации.

Исходными данными для исследования является информация из скоринговых карт клиентов кредитных организаций. Целевой атрибут – потенциальный заемщик – может принимать два значения {«плохой», «хороший»}, предсказывающих атрибутов – одиннадцать [3].

В платформу Deductor Studio Academic импортируем данные по 3085 клиентам кредитных организаций. Для исследования используем атрибуты: соотношение по кредитным картам и кредитным линиям, возраст, количество последних займов, платежеспособность, ежемесячный доход, количество открытых ссуд, количество просрочек 90 и более дней, наличие ипотечных кредитов, количество займов с просрочкой 60-89 дней, количество иждивенцев. Средствами платформы анализируем данные на пригодность для обработки, результаты анализа качества данных представлены на рисунке 1

№	Столбец	Тип данных	Вид данных	Пропуски		Выбросы		Экстремальные		Колво уникальных	Качество данных	Резюме
				Колво	Действие	Колво	Действие	Колво	Действие			
✓ 1	ID	9.0 Вещественный	— Непрерывный								0,8081	Пригоден
2	Соотношение бал...	9.0 Вещественный	— Непрерывный								0,8085	Пригоден
3	Возраст	9.0 Вещественный	— Непрерывный								0,9563	Пригоден
4	Количество посл...	9.0 Вещественный	... Дискретный			32	Заменять медиа...			5	0,3617	Предобработка
5	Платежеспособн...	9.0 Вещественный	— Непрерывный	1	Заменять мед...	82	Ограничивать	5	Ограничивать		0,1884	Предобработка
6	Ежемесячный до...	9.0 Вещественный	— Непрерывный			54	Ограничивать				0,7585	Предобработка
7	Количество откр...	9.0 Вещественный	— Непрерывный			46	Ограничивать				0,8632	Предобработка
8	Просрочка 90 и б...	9.0 Вещественный	... Дискретный			38	Заменять медиа...			4	0,1991	Предобработка
9	Ипотечные креди...	9.0 Вещественный	... Дискретный							5	0,7950	Пригоден
10	Количество займ...	9.0 Вещественный	... Дискретный							3	0,2271	Пригоден
11	Количество ижди...	9.0 Вещественный	... Дискретный							5	0,7009	Пригоден

Рисунок 1 Результаты анализа данных

Применяя инструменты Deductor Studio Academic «Заполнение пропущенных значений» к атрибуту платежеспособность и «Редактирование выбросов и экстремальных значений» к пяти атрибутам, получаем данные, качество которых представлено на рисунке 2.

№	Столбец	Тип данных	Вид данных	Пропуски		Выбросы		Экстремальные		Кол-во уникальных	Качество данных	Резюме
				Колво	Действие	Колво	Действие	Колво	Действие			
1	ID	9.0 Вещественный	— Непрерывный								0,8081	Пригоден
2	Соотношение бал...	9.0 Вещественный	— Непрерывный								0,8085	Пригоден
3	Возраст	9.0 Вещественный	— Непрерывный								0,9563	Пригоден
4	Количество посл...	9.0 Вещественный	... Дискретный					32	Заменять медиа...	3	0,3617	Предобработка
5	Платежеспособн...	9.0 Вещественный	— Непрерывный			215	Ограничивать				0,3071	Предобработка
6	Ежемесячный до...	9.0 Вещественный	— Непрерывный								0,8952	Пригоден
7	Количество откр...	9.0 Вещественный	— Непрерывный								0,8929	Пригоден
8	Просрочка 90 и б...	9.0 Вещественный	... Дискретный							2	0,0958	Пригоден
9	Ипотечные креди...	9.0 Вещественный	... Дискретный							5	0,7950	Пригоден
10	Количество займ...	9.0 Вещественный	... Дискретный							3	0,2271	Пригоден
11	Количество ижди...	9.0 Вещественный	... Дискретный							5	0,7009	Пригоден

Рисунок 2 Качество данных после первичной предобработки

Повторное применение инструмента «Редактирование выбросов и экстремальных значений» к атрибутам платежеспособность и количество последних займов позволяет получить данные пригодные для исследования в Deductor Studio Academic

№	Столбец	Тип данных	Вид данных	Пропуски		Выбросы		Экстремальные		Кол-во уникальных	Качество данных	Резюме
				Колво	Действие	Колво	Действие	Колво	Действие			
1	ID	9.0 Вещественный	— Непрерывный								0,8081	Пригоден
2	Соотношение бал...	9.0 Вещественный	— Непрерывный								0,8085	Пригоден
3	Возраст	9.0 Вещественный	— Непрерывный								0,9563	Пригоден
4	Количество посл...	9.0 Вещественный	... Дискретный							2	0,4917	Пригоден
5	Платежеспособн...	9.0 Вещественный	— Непрерывный								0,2872	Пригоден
6	Ежемесячный до...	9.0 Вещественный	— Непрерывный								0,8952	Пригоден
7	Количество откр...	9.0 Вещественный	— Непрерывный								0,8929	Пригоден
8	Просрочка 90 и б...	9.0 Вещественный	... Дискретный							2	0,0958	Пригоден
9	Ипотечные креди...	9.0 Вещественный	... Дискретный							5	0,7950	Пригоден
10	Количество займ...	9.0 Вещественный	... Дискретный							3	0,2271	Пригоден
11	Количество ижди...	9.0 Вещественный	... Дискретный							5	0,7009	Пригоден

Рисунок 3 Качество данных после повторной предобработки

Для принятия решения о выдаче кредита или отказе воспользуемся методом дерева решений.

Дерево решений – это метод изображения правил в последовательной иерархической структуре. База такой структуры – ответы вида “Да” или “Нет” как в случае алгоритма CART, представляющего из себя бинарные деревья, или ответы на последовательность вопросов, определяемых динамически, как в случае деревьев с произвольным числом потомков, например в алгоритме C 4.5.

В отличие, например, от результата работы алгоритмов основанных на нейронных сетях, представляющих собой “черные ящики”, результат работы алгоритмов создания деревьев решений, достаточно легко интерпретируется

исследователем. Данная особенность деревьев решений не только существенна при причислении к определенному классу нового объекта, но и важна при интерпретации модели классификации в целом. Модель, показанная в виде дерева решений, делает более простым понимание решаемой задачи и является интуитивной. Дерево решений дает возможность объяснить и понять, почему данный объект имеет отношение к тому или иному классу. Деревья решений позволяют получать правила из совокупности данных на естественном языке.

В случае моделей, образованных при помощи деревьев решений, их точность сравнима с другими способами получения классификационных моделей (нейронные сети, статистические методы). Существует ряд масштабируемых алгоритмов, которые могут быть использованы для построения деревьев решения на очень больших базах данных; в этом случае понятие масштабируемости показывает, что с ростом числа записей или примеров базы данных время, затрачиваемое на обучение, т.е. на создание деревьев решений, растет линейно.

На данный момент существует немалое количество алгоритмов, используемых при построении деревьев решений: ITrule, NewId, CHAID, CN2, C4.5, CART и другие. Если дерево имеет произвольное число потомков, т.е. является одной из разновидностей деревьев решений, то оно, благодаря своим свойствам, дает более универсальный метод решения задач классификации [4].

В аналитической платформе Deductor реализован алгоритм построения деревьев решений, впервые предложенный Р. Куинленом (R. Quinlan), лежащий в основе алгоритма C 4.5.

После обработки данных в Deductor Studio Academic получили значимость предсказывающих атрибутов, представленную в таблице 1

Таблица 1 – Значимость предсказывающих атрибутов

Предсказывающие атрибуты	Значимость атрибутов в %
Платежеспособность	48,891
Ежемесячный доход	23,617
Соотношение по кредитным картам и кредитным линиям	12,967
Возраст	10,774
Количество иждивенцев	1,355
Наличие ипотечных кредитов	0,885
Количество займов с просрочкой 60-89 дней	0,683
Количество открытых ссуд	0,416
Количество последних займов	0,321
Количество просрочек 90 и более дней	0,101

Полученные правила с достоверностью 100 % указывают на то, что потенциальный заемщик должен быть отнесен к классу «плохой», если его

доход меньше 4300 рублей. Если платежеспособность потенциального заемщика больше 0,9 и ежемесячный доход более 7200 рублей, то с достоверностью 92 % он может быть отнесен к классу «хороший».

Результатом исследования являются сформулированные правила оценки потенциального заемщика и возможность использования для этих целей встроенных инструментов предварительной обработке исходных данных и алгоритма построения деревьев решений аналитической платформы Deductor.

Список литературы

1. Андреева, Е. [Электронный ресурс]. / Е. Андреева // *Российская бизнес газета*. – Москва, 2013. - № 930(1).- Режим доступа: <http://www.rg.ru/2014/01/14/kredity.html>

2. Литвинова, С.А. Скоринговые системы как средство минимизации кредитного риска банка// *Аудит и финансовый анализ*. – 2010. - №2 с.141.

3. Ханжин, С.В. Математическая модель кредитного скоринга потенциальных клиентов банка / С.В. Ханжин // *Фундаментальные и прикладные исследования: проблемы и результаты: сборник материалов XX Международной научно-практической конференции* / Под общ. ред. С.С. Чернова. - Новосибирск: Издательство ЦРНС, 2015. – с. 184 – 188.

4. Ханжин, С.В. Исследование кредитного скоринга на основе метода анализа иерархий / С.В. Ханжин // *Фундаментальные и прикладные исследования: проблемы и результаты: сборник материалов XXV Международной научно-практической конференции* / Под общ. ред. С.С. Чернова. - Новосибирск: Издательство ЦРНС, 2016. – с. 143 – 148.