



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ _____ Фундаментальные науки
КАФЕДРА _____ Математическое моделирование

Научно-исследовательская работа

НА ТЕМУ:

“Кредитный скоринг”

Студент _____
ФН12-11М
(Группа)

(Подпись, дата)

А. А. Кононенко

(И. О. Фамилия)

Руководитель научной работы

(Подпись, дата)

М. С. Виноградова

(И. О. Фамилия)

2023 г.

Оглавление

Введение	3
1. Постановка задачи	4
2. Первичный анализ данных для возможности предоставления кредита	5
3. Вывод	7
Список использованных источников	8

Введение

Кредитный скоринг (от англ. score «оценка») — система оценки кредитоспособности (кредитных рисков) лица, основанная на численных статистических методах. Кредитный скоринг широко используется как крупными банками, микрофинансовыми организациями, так и в потребительском (магазинном) экспресс-кредитовании на небольшие суммы. Также возможно его использование в бизнесе сотовых операторов, страховых компаний и т. д. Скоринг заключается в присвоении баллов по заполнению некой анкеты, разработанной оценщиками кредитных рисков андеррайтерами. По результатам набранных баллов системой автоматически принимается решение об одобрении или отказе в выдаче кредита.

Основными источниками информации для построения скоринговых моделей являются:

- кредитная история;
- анкета заемщика;
- собственная информация кредитора (например, банки используют информацию о движении финансовых средств своего клиента).

Является упрощённой системой анализа заёмщика, что позволяет упразднить субъективизм принятия решения кредитного инспектора, снизить уровень внутреннего мошенничества, увеличить скорость принятия решения по кредиту. Аналогичным способом скоринговая модель может позволять рассчитывать индивидуальную ставку по страховому продукту, устанавливая толерантность к риску.

1. Постановка задачи

Задача кредитного скоринга состоит в определении вероятности возврата кредита заемщиком. Для этого необходимо разработать модель, которая будет оценивать риски, связанные с выдачей кредита каждому конкретному заемщику на основе имеющихся данных о нем.

Цель задачи:

- Прогнозирование вероятности возврата кредита для каждого заемщика;
- Минимизация рисков для кредитора;
- Максимизация выдачи кредитов для заемщиков с высокой вероятностью возврата.

Для достижения этой цели необходимо выполнить следующие задачи:

1. Исследование и подготовка данных:
 - Сбор данных о заемщиках и кредитной истории;
 - Удаление дубликатов, заполнение пропусков, преобразование категориальных переменных в числовые;
 - Анализ корреляции между переменными и удаление излишне коррелирующих переменных.
2. Построение модели кредитного скоринга:
 - Разделение данных на обучающую и тестовую выборки;
 - Выбор и обучение подходящей модели (например, логистическая регрессия, случайный лес, градиентный бустинг);
 - Оценка модели на тестовой выборке с использованием соответствующих метрик (например, ROC-AUC, точность, полнота, F-мера).
3. Определение порогового значения вероятности возврата кредита:
 - Оценка бизнес-требований и снижение рисков с учетом пороговой вероятности;
 - Определение точности и рентабельности портфеля кредитов.
4. Применение модели в практических условиях:
 - Оценка вероятности возврата кредита для новых заемщиков;
 - Принятие решений о выдаче или отказе в кредите на основе вероятности возврата.
5. Мониторинг и обновление модели:
 - Следить за эффективностью модели в течение времени;
 - Регулярное обновление модели на основе новых данных;
 - Анализ результатов и внесение изменений, если требуется.

Таким образом, задача кредитного скоринга заключается в разработке модели, которая будет предсказывать вероятность возврата кредита заемщиком на основе имеющихся данных о нем. Это позволит кредитору принимать более обоснованные решения о выдаче кредита и минимизировать риски связанные с невозвратом кредита.

2. Первичный анализ данных для возможности предоставления кредита

Был выбран датасет <https://www.kaggle.com/datasets/laotse/credit-risk-dataset/data>. При поиске источника данных приоритет отдавался таким факторам как: адекватность, разнообразие, наполненность данных, большое количество категорий. Им удовлетворял выбранный датасет, в котором присутствует 12 критериев: возраст клиента, годовой доход, формат владения недвижимостью (личная недвижимость, ипотека, аренда, другое), продолжительность трудовой деятельности, причина займа (личные причины, образование, медицина, венчурные, обустройство дома, консолидация задолженностей), класс получения кредита (A – D), сумма кредита, процентная ставка, статус займа (1 – по умолчанию, 0 – не по умолчанию), процентный доход, историческое значение по умолчанию (было ли одобрение в прошлом), длина кредитной истории.

	person_age	person_income	person_home_ownership	person_emp_length	loan_intent	loan_grade	loan_amnt	loan_int_rate	loan_status	loan_percent_income	cb_person_default_on_file	cb_persc
2	22	59000	RENT	123.0	PERSONAL	D	35000	16.фев	1	0.59	Y	3
3	21	9600	OWN	5.0	EDUCATION	B	1000	ноя.14	0	0.1	N	2
4	25	9600	MORTGAGE	1.0	MEDICAL	C	5500	дек.87	1	0.57	N	3
5	23	65500	RENT	4.0	MEDICAL	C	35000	15.23	1	0.53	N	2
6	24	54400	RENT	8.0	MEDICAL	C	35000	14.27	1	0.55	Y	4
7	21	9900	OWN	2.0	VENTURE	A	2500	июл.14	1	0.25	N	2
8	26	77100	RENT	8.0	EDUCATION	B	35000	дек.42	1	0.45	N	3
9	24	78956	RENT	5.0	MEDICAL	B	35000	11.ноя	1	0.44	N	4
10	24	83000	RENT	8.0	PERSONAL	A	35000	08.сен	1	0.42	N	2
11	21	10000	OWN	6.0	VENTURE	D	1600	14.74	1	0.16	N	3
12	22	85000	RENT	6.0	VENTURE	B	35000	окт.37	1	0.41	N	4
13	21	10000	OWN	2.0	HOMEIMPROVEMENT	A	4500	авг.63	1	0.45	N	2
14	23	95000	RENT	2.0	VENTURE	A	35000	07.сен	1	0.37	N	2
15	26	108160	RENT	4.0	EDUCATION	E	35000	18.39	1	0.32	N	4
16	23	115000	RENT	2.0	EDUCATION	A	35000	07.сен	0	0.3	N	4
17	23	500000	MORTGAGE	7.0	DEBTCONSOLIDATION	B	30000	окт.65	0	0.06	N	3
18	23	120000	RENT	0.0	EDUCATION	A	35000	07.сен	0	0.29	N	4
19	23	92111	RENT	7.0	MEDICAL	F	35000	20.25	1	0.32	N	4
20	23	113000	RENT	8.0	DEBTCONSOLIDATION	D	35000	18.25	1	0.31	N	4

Рис. 1. Исходная таблица с данными (32582 строки).

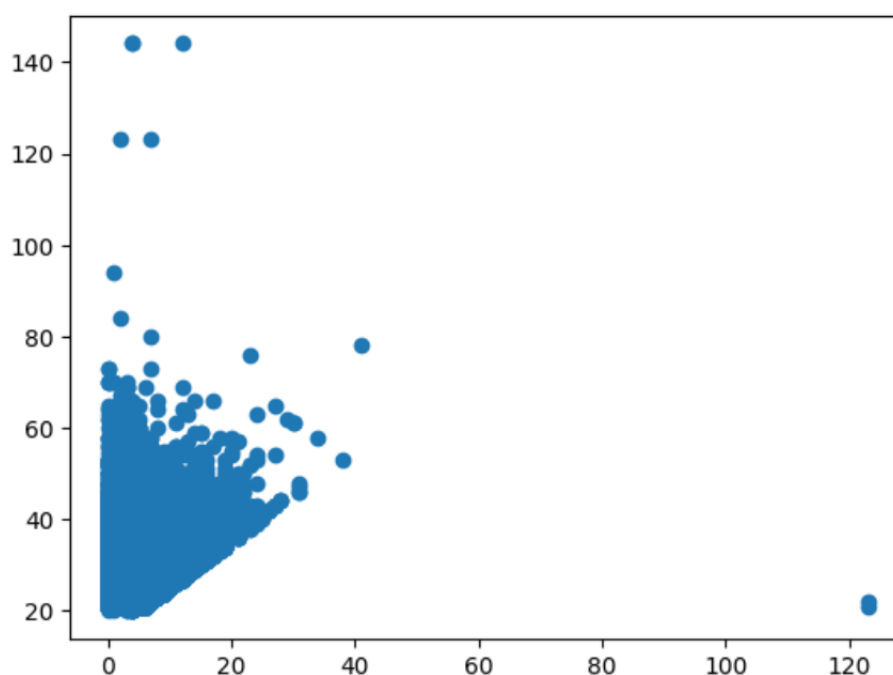


Рис. 2. График возраста и стажа работы клиентов до преобразования данных.

Чтобы написать базу данных и серию специальных запросов для быстрого принятия решений необходимо было сделать первичный анализ и изменение данных для возможности предоставления кредита. Подготовка данных заключалась в их очистке от выбросов и пропущенных значений, а также преобразовании данных при необходимости, факторизация категориальных значений. Вся работа с данными производилась на языке программирования *Python*, визуализация осуществлялась с помощью библиотеки *pandas* и *matplotlib* (модуль *Pyplot*).

В ходе обработки данных были

- удалены пользователи с аномально большим возрастом: пять значений возраста 144 (строки 83, 185 и 32299) и 123 (строки 577 и 749), что является выбросом. В силу того, что максимальный возраст человека составляет 122 года, выброс является либо ошибкой записи при составлении анкеты респондентом либо ошибкой при вводе значений в таблицу, так как старше 100 лет только 2 различных значения, повторяющиеся несколько раз. Пользователей с возрастом меньше 18-ти лет не найдено (их и не должно было существовать в таблице);
- удалены пользователи со стажем работы более 100 лет. Таких было двое респондентов (строки 0 и 210) с ошибочным стажем работы 123 года. Учитывая, что такой продолжительный стаж работы невозможен, что число 123 возникало как возраст других двух респондентов и что до 123 лет максимальный стаж работы был 41 год, делается вывод, что эти данные ошибочные;
- избавились от строк с пустыми ячейками (среди них отсутствовал стаж работы у 895 человек, а статус кредита у 3116 человек, некоторые люди не имели ни статуса кредита, ни стажа работы);
- произведена факторизация переменных по 4-м столбцам, значения переменных в которых были не числовые.

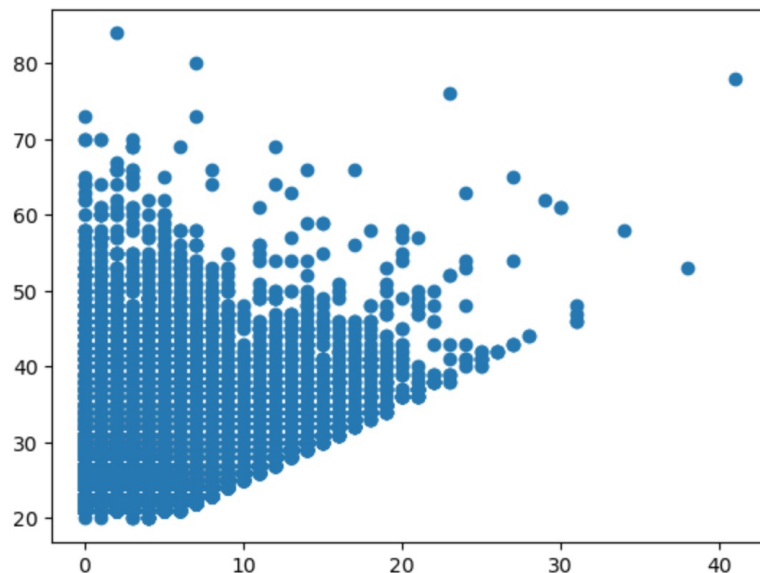


Рис. 3. График возраста и стажа работы клиентов после преобразования данных.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		person_age	erson_incom	home_own	son_emp_len	loan_intent	loan_grade	loan_amnt	loan_int_rate	loan_status	percent_incon	son_default	on_cred_hist
2	1	21	9600	1	5	1	1	1000	11,14	0	0,1	1	2
3	2	25	9600	2	1	2	2	5500	12,87	1	0,57	1	3
4	3	23	65500	0	4	2	2	35000	15,23	1	0,53	1	2
5	4	24	54400	0	8	2	2	35000	14,27	1	0,55	0	4
6	5	21	9900	1	2	3	3	2500	7,14	1	0,25	1	2
7	6	26	77100	0	8	1	1	35000	12,42	1	0,45	1	3
8	7	24	78956	0	5	2	1	35000	11,11	1	0,44	1	4
9	8	24	83000	0	8	0	3	35000	8,9	1	0,42	1	2
10	9	21	10000	1	6	3	0	1600	14,74	1	0,16	1	3
11	10	22	85000	0	6	3	1	35000	10,37	1	0,41	1	4
12	11	21	10000	1	2	4	3	4500	8,63	1	0,45	1	2
13	12	23	95000	0	2	3	3	35000	7,9	1	0,37	1	2
14	13	26	108160	0	4	1	4	35000	18,39	1	0,32	1	4
15	14	23	115000	0	2	1	3	35000	7,9	0	0,3	1	4
16	15	23	500000	2	7	5	1	30000	10,65	0	0,06	1	3
17	16	23	120000	0	0	1	3	35000	7,9	0	0,29	1	4
18	17	23	92111	0	7	2	5	35000	20,25	1	0,32	1	4
19	18	23	113000	0	8	5	0	35000	18,25	1	0,31	1	4
20	19	24	10800	2	8	1	1	1750	10,99	1	0,16	1	2

Рис. 4. Подготовленная таблица (28633 строки).

3. Вывод

Таким образом был произведён первичный анализ содержания датасета кредитного скоринга, визуализация и обработка данных. Суммарно было удалено 3949 строк с выбросами, пустыми значениями или данными, в действительности невозможными. Удалённые неподходящие данные составляют 0.12 процентов от исходных. Из первоначального анализа сделан вывод, что значение в столбце *loan_percent_income* (процентный доход) больше всего влияет на решение о выдаче кредита.

Список использованных источников

1. Данилович В. Ю., Курганская Д. С. Скоринговые модели как средство управления кредитными рисками в российских банках. Иркутский Государственный Университет, г. Иркутск, 2007.
2. Котляр В. П. Антипова Е. А. Разработка скоринговой модели. Методы классификации заёмщиков. Орловский государственный университет, г. Орёл, 2013.
3. Артем Румянцев. Скоринговые системы: наука помогает бизнесу. Журнал "Финансовый Директор ISSN 1680 - 1148" №7/2006. — Описание принципов работы наиболее распространенных скоринговых систем.
4. Михаил Чуклин. Что такое кредитный скоринг? Финансовый информационный проект *Fin.zone*. — Методы оценки. Принципы работы автоматического скоринга. Скоринг баллы.