

*Андреева Юлия Дмитриевна*

студентка

*Михайлова Елена Андреевна*

студентка

*Бобин Дмитрий Витальевич*

старший преподаватель

ФГБОУ ВО «Чувашский государственный университет

им. И.Н. Ульянова»

г. Чебоксары, Чувашская Республика

## **ОЦЕНКА КРЕДИТОСПОСОБНОСТИ ФИЗИЧЕСКИХ ЛИЦ ПО МОДЕЛИ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ**

*Аннотация:* в статье разработана модель оценки кредитоспособности физических лиц с использованием многомерной логистической регрессии, которая позволяет выявить взаимосвязи между характеристиками заемщиков и классифицировать их как благонадежных или неблагонадежных. Причем точность классификации, как правило, выше, чем при применении одномерных подходов, например скоринговых карт. Для оценки качества классификации применены такие метрики, как точность, специфичность, полнота,  $F$ -мера, AUC. Полученные результаты могут служить основой для вынесения сотрудниками банка решения по выдаче кредита.

*Ключевые слова:* кредитоспособность, логистическая регрессия, кредитный скоринг, кредитный риск, матрица классификации.

Кредитование является неотъемлемой частью сферы банковских услуг. Это обуславливается тем, что основным источником доходов кредитных организаций является предоставление ссуд различным категориям заемщиков, одними из которых являются физические лица.

В целом, можно отметить, что за 2009–2020 г. наблюдалась тенденция к увеличению объема выданных кредитов физическим лицам, за исключением экономического кризиса в 2014–2015 годах, связанного с внешнеполитически-

ми обострениями, которые привели к девальвации рубля (рис. 1). В частности, рост ключевой ставки и инфляция спровоцировали удорожание стоимости кредита, в результате чего произошло снижение спроса на данный банковский продукт.

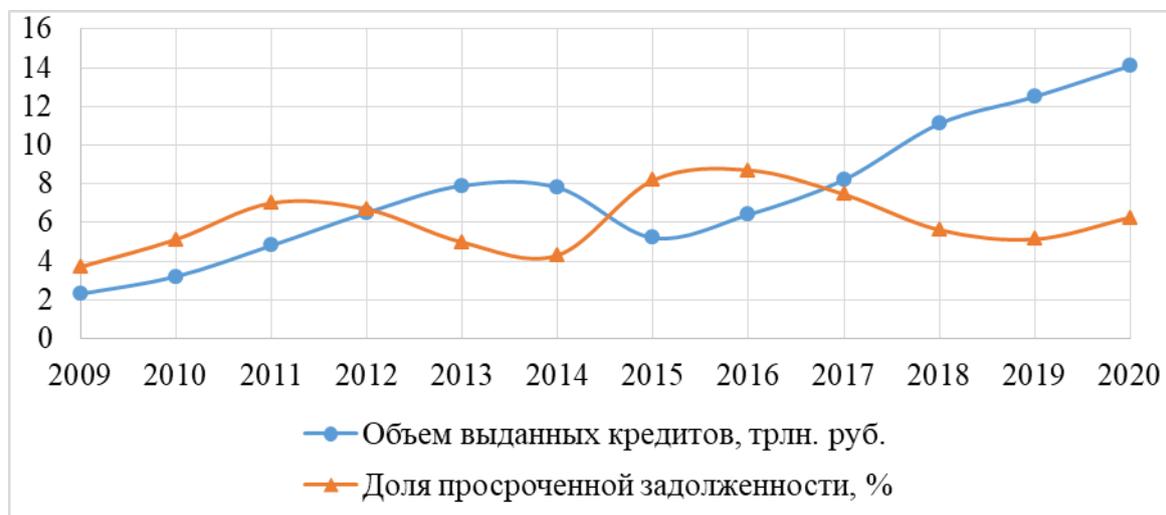


Рис. 1. Объем выданных кредитов и доля просроченной задолженности за 2009–2020 гг.

Серьезное влияние на состояние кредитного портфеля банков оказывают не только количество выданных кредитов и сумма задолженности, но и динамика доли просроченной задолженности. Кредитный риск является наиболее значимым для банка, поэтому его необходимо минимизировать, так как рост просроченной задолженности ведет к увеличению расходов банка по обслуживанию долга и к повышению процентных ставок по ссудам для потенциальных клиентов [2].

Важно отметить, что влияние кризисов 2008–2009 гг. и 2014–2015 гг. на динамику просроченной задолженности заметно через несколько лет (2–3 года), в отличие от объема выданных кредитов, воздействие негативных последствий на которые ощутимо мгновенно. Это можно объяснить следующими причинами:

1) финансовая неграмотность большей части населения. В периоды экономических спадов люди имеют заведомо ложные надежды на то, что нагрузка по кредиту уменьшится за счет ускорения темпов роста инфляции, но при этом заемщики не могут рационально оценить свои финансовые возможности, что ведет к увеличению просроченной задолженности в будущем;

2 <https://phsreda.com>

- 2) повышение процентных ставок по кредитам в периоды кризиса;
- 3) активный рост мошеннических действий в сфере кредитования.

Соответственно, кредитным организациям следует опираться в практике выдачи кредитов на научно-обоснованные алгоритмы машинного обучения, использующие все многообразие количественных и качественных показателей. Описание математических моделей оценки кредитоспособности представлено в трудах Е.А. Воротилкиной [1], В.Ю. Даниловича [3], К.И. Казанцевой [5] А.Ф. Савдеровой [6] и др.

Логистическая регрессия – статистическая модель, которая позволяет оценить вероятность принадлежности объекта к какому-либо классу с помощью логистической функции:

$$f(X) = \frac{1}{1 + e^{-X}}$$

В результате оценки заёмщика получаем значение «1», в случае, когда кредит может быть одобрен, либо «0», когда решение отрицательное. То есть модель имеет бинарную природу и отличается от обычной линейной регрессии.

Для количественного анализа имеется сбалансированная выборка из 104 случаев заёмов, полученная из датасета [7] с помощью случайного сэмплинга. Выборка включает равное количество случаев возврата и невозврата кредита, описанных 8 входными признаками  $X$  и выходной переменной  $Y$ :

$X_1$  – месячный платеж (руб.);

$X_2$  – возраст (лет);

$X_3$  – пол (1 – женский, 0 – мужской);

$X_4$  – семейное положение (1 – холост/не замужем, 0 – женат/замужем);

$X_5$  – количество иждивенцев (чел.);

$X_6$  – заявленный среднемесячный доход (руб.);

$X_7$  – опыт работы (лет);

$X_8$  – недвижимость (кв.м.);

$Y$  – благонадежность заемщика (0 – неблагонадежный, 1 – благонадежный).

В таблице 1 представлены статистические характеристики выборки для числовых признаков:

Таблица 1

## Описательная статистика признаков

Показатели	$X_1$	$X_2$	$X_5$	$X_6$	$X_7$	$X_8$
Среднее значение ( $Y=0$ )	4281,69	34,69	1,46	8288,46	12,81	13,58
Среднее значение ( $Y=1$ )	4073,10	36,21	0,81	18936,54	12,47	12,21
Минимальное значение	1384	23	0	2100	2	4
Максимальное значение	8098	57	4	36700	31	88
Дисперсия	1541,81	7,72	1,10	7801,94	5,13	17,41

С помощью программы Deductor Studio Academic построим модель логистической регрессии. Вследствие мультиколлинеарности признаков такие факторы, как возраст, пол, опыт работы заемщика, наличие у него недвижимости, необходимо исключить, т.к. их использование может привести к недостоверности модели. Итоговая логистическая модель по нашим данным принимает вид:

$$f(X) = \frac{1}{1 + e^{-(1,147 + 0,003X_1 + 3,205X_4 + 5,094X_5 - 0,002X_6)}}$$

Наиболее значимыми признаками оказались ежемесячный платеж, семейное положение, количество иждивенцев, среднемесячный доход. Высокий коэффициент детерминации (0,871) говорит о значимости модели. Пороговое значение данной функции составляет 0,494 (рис. 2). Это означает, что, если вероятность возврата кредита превысит указанную величину, то будет вынесено положительное решение по кредиту.

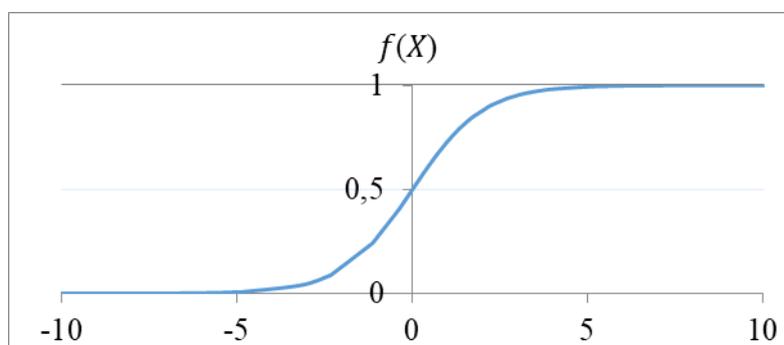


Рис. 2. График функции логистической регрессии

Точность построенной модели оценим по матрице классификации (табл. 2) и метрикам качества:

Матрица классификации заемщиков

Фактическое количество клиентов	Теоретическое количество		Всего
	$f(X) = 1$	$f(X) = 0$	
Благонадежные ( $Y=1$ )	51 ( $TP$ )	1 ( $FN$ )	52
Неблагонадежные ( $Y=0$ )	2 ( $FP$ )	50 ( $TN$ )	52

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = 0,9712,$$

$$precision = \frac{TP}{TP + FP} = 0,9623,$$

$$recall = \frac{TP}{TP + FN} = 0,9808,$$

$$specificity = \frac{TN}{TN + FP} = 0,9615,$$

$$F\_score = \frac{2TP}{2TP + FP + FN} = 0,9714.$$

Все показатели стремятся к 1, поэтому модель можно считать достоверной. Это подтверждает и высокое значение (0,994) AUC – метрики, показывающей площадь под кривой ошибок (ROC-кривая) [4].

В работе была представлена модель оценки кредитоспособности физических лиц, построенная на основе математического механизма логистической регрессии. Ее преимущество перед аналогами состоит в том, что она относится не только к методам классификации, но и к методам регрессионного анализа. Рассмотренная статистическая модель в отличие от других позволяет устранить мультиколлинеарность признаков, т.е. выявляет зависимость как между факторами и результативным показателем, так и между самими факторами. Таким образом, применение многомерной логистической регрессии в процессах банковского кредитования позволяет с большей точностью классифицировать клиентов на благонадежных и неблагонадежных по сравнению со скоринговыми картами и, как следствие, снизить кредитные риски.

### *Список литературы*

1. Воротилкина Е.А. Моделирование кредитоспособности физических лиц и разработка приложения скорингового калькулятора / Е.А. Воротилкина, Д.В. Бобин // Проблемы и перспективы развития социально-экономического потенциала российских регионов: материалы V Всероссийской электронной научно-практической конференции. – Чебоксары: Изд-во Чуваш. ун-та; Пегас, 2016. – С. 369–372.

2. Гусарова О.М. Нейронные сети в кредитном скоринге / О.М. Гусарова, П.И. Комаров, Д.Э. Денисов // Вестник Алтайской академии экономики и права. – 2018. – №5 – С. 107–113.

3. Данилович В.Ю. Скоринговые модели как средство управления кредитными рисками в российских банках / В.Ю. Данилович, Г.С. Курганская // Бизнес-образование в экономике знаний. – 2017. – №1. – С. 29–33.

4. Дьяков О.А. Особенности применения методов Data Mining в скоринговых решениях в коммерческих банках / О.А. Дьяков // Научные записи молодых исследователей. – 2017. – №3. – С. 5–11.

5. Казанцева К.И. Оценка платежеспособности физических лиц / К.И. Казанцева, Д.В. Бобин // Проблемы и перспективы развития социально-экономического потенциала российских регионов: материалы VIII Всероссийской научно-практической конференции. – Чебоксары: Изд-во Чуваш. ун-та; Пегас, 2019. – С. 284–289.

6. Савдерова А.Ф. Совершенствование методов оценки заемщиков как фактор обеспечения стабильности деятельности коммерческого банка / А.Ф. Савдерова, Е.Б. Крылова, К.В. Васильева // Экономическая безопасность как парадигма современной теории и практики управления: сборник материалов Всероссийской научно-практической конференции. – Чебоксары: Изд-во Чуваш. ун-та, 2019. – С. 306–310.

7. Датасет «borrowers» [Электронный ресурс]. – Режим доступа: <https://loginom.ru/sites/default/files/blogpost-files/borrowers.txt> (дата обращения: 20.06.2022).