

*М. М. Дульчевский, магистрант  
Э. М. Аксень, д-р экон. наук, доцент  
БГЭУ (Минск)*

## МЕТОДИКА АВТОМАТИЗИРОВАННОЙ ОЦЕНКИ КРЕДИТНОГО РИСКА ФИЗИЧЕСКИХ ЛИЦ

В последние годы в стране наблюдается интенсивный рост рынка кредитования и в частности сектора кредитования физических лиц, что приводит к увеличению кредитных рисков, которые принимает на себя банковская система страны. По мере увеличения объемов розничного кредитования традиционные методы оценки кредитоспособности физических лиц теряют свою эффективность, что заставляет белорусские банки более серьезно задуматься над вопросом применения автоматизированной оценки кредитного риска физических лиц, а именно скоринга новых клиентов.

Процесс скоринга можно условно разделить на три этапа: 1) сбор, анализ и обработка данных; 2) построение модели; 3) оценка качества модели. В первую очередь следует определить анализируемый период — данные должны быть актуальными. Следует также учитывать, что у кредита к моменту разработки скоринговой карты должен быть достаточный период жизни. После нахождения баланса между периодом жизни кредита и актуальностью данных следует определить понятие «плохого» заемщика. Критерий «плохого» заемщика для каждого случая выбирается индивидуально в зависимости от программы кредитования и других факторов. Для подготовки к статистическому анализу данные необходимо очистить от выбросов и восстановить пропуски. Далее историческую совокупность на основе механизма случайного отбора следует разделить на две подвыборки: обучающую и тестовую. Обычно выбирается соотношение 70:30 и 80:20 % от объема исторической выборки.

Первым этапом построения модели является категоризация количественных переменных, или процесс бининга. Данные по анализируемому фактору на основе равных процентилей разделяются на  $n$  групп (бинов); для каждого бина находятся доли «плохих» и «хороших» заемщиков, вес категорий предиктора  $WoE$ , а также информационное значение  $IV$ ; полученные показатели весов категорий анализируются, происходит объединение соседних категорий и пересчет показателей  $WoE$ . Понятие близости значений  $WoE$  достаточно субъективно. Весь процесс бининга опирается не только на статистику, но и на логику, целесообразность и экспертное мнение разработчика.

Для построения модели используются только категориальные переменные, значения которых заменены на показатели  $WoE$ . Итеративной процедурой построения модели логистической регрессии из дальнейшего анализа исключаются факторы,  $p$ -Value которых лежит за пределами уровня значимости, равного 10 %. Далее для оценки мультиколлинеарности между предикторами строится корреляционная матрица, на основании которой исключаются факторы, между которыми наблюдается сильная мультиколлинеарность. На оставшихся предикторах строится итоговая модель. В основе оценки качества модели лежит достоверность, характеризующаяся способностью одинаково хорошо различать и «хороших», и «плохих» заемщиков. Для этих целей строится ROC-кривая, из которой по соотношению площадей определяется показатель  $AUC$ . Способность одинаково хорошо различать «хороших» и «плохих» заемщиков основывается на распределении скоринговых баллов. В связи с этим для оценки качества прогнозирования рассчитывается тест Колмогорова–Смирнова как максимальная разница между кумулятивными функциями распределений «плохих» и «хороших» заемщиков. Заключительным этапом разработки скоринговой карты является перевод коэффициентов логистической регрессии в скоринговые баллы с использованием масштабирования.