

ПРИМЕНЕНИЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРЕДСКАЗАНИЯ ДЕФОЛТА ЗАЕМЩИКОВ

Краева О.С. (Университет ИТМО)

Научный руководитель – кандидат физико-математических наук,
доцент Михайлова Е.Г.
(Университет ИТМО)

Введение. Кредитный скоринг – процесс оценки платежеспособности заемщика на основе анализа его финансовых и социально-демографических данных. Современные модели машинного обучения позволяют в значительной мере повысить точность предсказания дефолта, выявляя скрытые закономерности в больших объемах данных. Их применение способствует большей объективности кредитных решений, снижению рисков для финансовых организаций.

В рамках исследования использованы реальные анонимизированные данные, полученные во время прохождения практики в одной из консалтинговых компаний. Датасет включает 400 000 наблюдений и 15 признаков, 3 из которых качественные (пол, образование, семейное положение). Основными признаками являются количество открытых кредитов, количество просрочек, максимальный баланс для активных счетов, максимальное количество транзакций по активным кредитным картам, максимальная сумма оборота в месяц.

Основная часть. Предсказание дефолта заемщиков требует комплексного подхода, состоящего из следующих элементов: предобработка данных, выбор оптимальных моделей и всесторонняя оценка их качества. На этапе предварительного анализа проводится очистка и обработка данных: устраняются пропуски, нормализуются числовые переменные, кодируются категориальные признаки. Корректная предобработка обеспечивает надежность и стабильность моделей, снижая влияние шумов и выбросов в данных.

Для построения моделей применяются как базовые, так и более сложные алгоритмы. Логистическая регрессия остается одним из базовых методов кредитного скоринга благодаря своей прозрачности и простоте интерпретации результатов. Однако для повышения точности предсказаний также используются ансамблевые методы и нейронные сети, способные выявлять сложные нелинейные зависимости и учитывать широкий спектр факторов, влияющих на вероятность дефолта.

Оценка качества моделей осуществляется с применением различных метрик, включая AUC-ROC, F1-score и др., что позволяет объективно сравнивать их эффективность.

Выводы. Проведенная работа демонстрирует процесс предсказания дефолта заемщиков на основе имеющейся на начальном этапе информации, а также сравнительный анализ эффективности полученных решений. Результаты исследования подтверждают, что использование современных методов машинного обучения позволяет значительно повысить точность прогнозирования и способствует более надежной и справедливой системе кредитования.

Список использованных источников:

1. Tripathi, Diwakar & Edla, Damodar & Bablani, Annushree & Shukla, Alok & Reddy, Ramachandra. Experimental analysis of machine learning methods for credit score classification. 2021.
2. Jia Chong Ong, Lai Soon Lee. Credit Scoring: A Comparison of Machine Learning Models and Their Modifications. AMCI 2025, 14, 57-78.