

DOI: 10.37791/2687-0649-2024-19-1-5-27

# Метод прогнозирования оттока клиентов банка на основе ансамблевой модели машинного обучения

А. Ю. Пучков<sup>1\*</sup>, М. И. Дли<sup>1,2</sup>, М. А. Василькова<sup>1</sup>, Н. Н. Прокимнов<sup>2</sup>

<sup>1</sup>Филиал Национального исследовательского университета «МЭИ» в г. Смоленске, Смоленск, Россия

<sup>2</sup>Университет «Синергия», Москва, Россия

\*putchkov63@mail.ru

**Аннотация.** Представлены результаты исследований, целью которых была разработка метода прогнозирования оттока клиентов коммерческого банка на основе применения моделей машинного обучения (в том числе глубоких искусственных нейронных сетей) для обработки клиентских данных, а также создание программных инструментов, реализующих указанный метод. Объектом исследования является коммерческий банк, а предметом исследования – его деятельность в B2C-сегменте, включающая коммерческое взаимодействие бизнеса и частных лиц. Актуальность выбранного направления исследований определяется активизацией деятельности банков в сфере внедрения цифровых сервисов по сокращению неоперационных расходов, связанных, в частности, с удержанием клиентов, так как затраты на привлечение новых клиентов значительно выше, чем на сохранение уже имеющихся. Научную новизну результатов исследований составляет разработанный метод прогнозирования оттока клиентов коммерческого банка, а также алгоритм, положенный в основу программного обеспечения, реализующего предложенный метод. В основе предложенной ансамблевой модели прогнозирования лежит три алгоритма классификации: k-средних, случайный лес и многослойный перцептрон. Для агрегации выходов отдельных моделей предложено использовать обучаемое дерево систем нечеткого вывода типа Мамдани. Обучение ансамблевой модели выполняется в два этапа: сначала обучаются перечисленные три классификатора, а потом, на основе получаемых на их выходах данных, обучается дерево систем нечеткого вывода. Ансамблевая модель в предложенном методе реализует статический вариант прогноза, результаты которого используются в динамическом прогнозе, выполняемом в двух вариантах – на основе рекуррентного метода наименьших квадратов и на основе сверточной нейронной сети. Проведенные модельные эксперименты на синтетическом наборе данных, взятых с сайта Kaggle, показали, что ансамблевая модель имеет более высокое качество бинарной классификации, чем каждая модель в отдельности.

**Ключевые слова:** ансамблевые модели машинного обучения, прогнозирование оттока клиентов, деревья нечеткого вывода

**Для цитирования:** Пучков А.Ю., Дли М.И., Василькова М.А., Прокимнов Н.Н. Метод прогнозирования оттока клиентов банка на основе ансамблевой модели машинного обучения // Прикладная информатика. 2024. Т. 19. № 1. С. 5–27. DOI: 10.37791/2687-0649-2024-19-1-5-27

# A method for predicting bank customer churn based on an ensemble machine learning model

A. Puchkov<sup>1</sup>, M. Dli<sup>1,2</sup>, M. Vasilkova<sup>1</sup>, N. Prokimnov<sup>2</sup>

<sup>1</sup>Branch of the National Research University "MPEI" in Smolensk, Smolensk, Russia

<sup>2</sup>Synergy University, Moscow, Russia

<sup>1</sup>putchkov63@mail.ru

**Abstract.** The results of research are presented, the purpose of which was to develop a method for predicting the outflow of clients of a commercial bank based on the use of machine learning models (including deep artificial neural networks) for processing client data, as well as the creation of software tools that implement this method. The object of the study is a commercial bank, and the subject of the study is its activities in the B2C segment, which includes commercial interaction between businesses and individuals. The relevance of the chosen area of research is determined by the increased activity of banks in the field of introducing digital services to reduce non-operating costs associated, in particular, with retaining clients, since the costs of attracting new ones are much higher than maintaining existing clients. The scientific novelty of the research results is the developed method for predicting the outflow of commercial bank clients, as well as the algorithm underlying the software that implements the proposed method. The proposed ensemble forecasting model is based on three classification algorithms: k-means, random forest and multilayer perceptron. To aggregate the outputs of individual models, it is proposed to use a learning tree of fuzzy inference systems of the Mamdani type. Training of the ensemble model is carried out in two stages: first, the listed three classifiers are trained, and then, based on the data obtained from their outputs, a tree of fuzzy inference systems is trained. The ensemble model in the proposed method implements a static version of the forecast, the results of which are used in a dynamic forecast performed in two versions – based on the recurrent least squares method and based on a convolutional neural network. Model experiments carried out on a synthetic dataset taken from the Kaggle website showed that the ensemble model has a higher quality of binary classification than each model individually.

**Keywords:** ensemble machine learning models, customer churn prediction, fuzzy inference trees

**For citation:** Puchkov A., Dli M., Vasilkova M., Prokimnov N. A method for predicting bank customer churn based on an ensemble machine learning model. *Prikladnaya informatika*=Journal of Applied Informatics, 2024, vol.19, no.1, pp.5-27 (in Russian). DOI: 10.37791/2687-0649-2024-19-1-5-27

## Введение

Цифровизация всех сфер деятельности человеческого общества приводит к необходимости применения инновационных подходов, совершенствования инструментальных средств при реализации бизнес-процессов в различных предметных областях [1]. Особенно ярко

это проявляется в банковской сфере, где внедрение разнообразных и удобных для потребителя новых продуктов и сервисов на основе современных информационных технологий является необходимым условием выживания в агрессивной конкурентной среде.

Актуальным направлением цифровизации банковской сферы является модерни-