

# «ЦИФРОВОЙ СЛЕД», КАК ИНСТРУМЕНТ ПОВЫШЕНИЯ КАЧЕСТВА ИСХОДНЫХ ДАННЫХ СКОРИНГ-МОДЕЛЕЙ ПОТЕНЦИАЛЬНОГО ЗАЁМЩИКА

А.Б. Андреева

Научный руководитель: А.О. Савельев

Томский политехнический университет

aba17@tpu.ru

## Введение

Цифровой след (англ. digital footprint) — данные о посещениях и вкладе пользователя во время пребывания в цифровом пространстве. Цифровой след обеспечивает базу данных и фактов, имеющих отношение к работе кого-либо в цифровой среде. Это могут быть личные профили и учетные записи в социальных сетях, информация о посещаемых сайтах, открытые, созданные файлы, личные сообщения, комментарии, видео, фотографии, персональные данные пользователя.

Цифровые следы являются предметом сбора и исследования некоторых организаций, например, занимающихся онлайн-опросами. Также они могут использоваться правоохранительными органами для получения информации. Социальные медиа используют цифровой след для анализа личных интересов пользователей Интернета. Такие данные могут быть собраны и проанализированы без ведома пользователя, чтобы воссоздать его личность [1].

В банковской сфере применение технологии Digital Footprint на сегодняшний день встречается редко. Однако ее использование сможет способствовать более эффективному

функционированию системы банковского кредитования – повысить качество исходных данных для скоринг-моделей потенциальных заемщиков.

Скоринг представляет собой математическую или статистическую модель, с помощью которой на основе информации, известной о заемщике, банк определяет вероятность того, что потенциальный заемщик вернет кредит в срок. Задачи, решаемые с помощью кредитного скоринга:

1. Выявление кредитоспособных клиентов и формирование стратегии по их сохранению.
2. Предсказание будущего поведения существующих должников, выделение недобросовестных клиентов, уменьшение вероятности возникновения проблемных ссуд.
3. Выбор оптимальных схем поведения для сокращения числа должников.

Таким образом, задача кредитного скоринга состоит не только в выявлении платежеспособности заемщика, а также в привлечении добросовестных клиентов[2]. Процесс принятия решения о выдаче займа с использованием технологии «Digital footprint» представлен на рисунке 1.

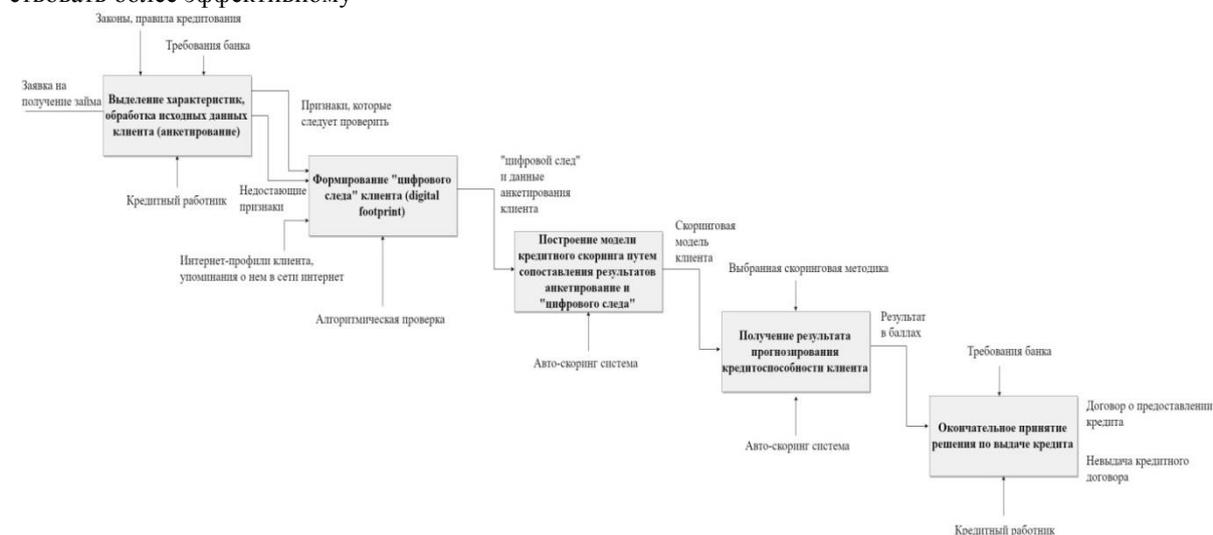


Рис. 1. IDEF-диаграмма процесса принятия решения о выдаче займа

## Задачи каждого из основных этапов процесса принятия решения о выдаче займа

*Обработка исходных данных клиента:*

- 1) Выделение характеристик клиентов и их признаков, которые принимают данные переменные.

*Формирование «цифрового следа» клиента:*

- 1) Восполнение недостающей информации о клиенте, формирование цифрового следа.
- 2) Проверка информации известной о клиенте из

анкетирования.

*Построение модели кредитного скоринга путем сопоставления результатов анкетирования и результатов «цифрового следа»:*

- 1) Построение скоринг-модели определения кредитного рейтинга заемщика.
- 2) Нахождение закономерности между данными профилей и целевыми переменными банка, которые влияют на решение.

- 3) Выявление попыток мошенничества.
- 4) Определение рисков, связанных с осуществлением сделки с конкретным потенциальным заемщиком, на основании оценки определенного набора его характеристик.

*Получение результата прогнозирования кредитоспособности клиента:*

- 1) Определение вероятности того, что конкретный потенциальный заемщик вернет кредит в срок.
- 2) Расчет конечного показателя.

*Окончательное решение по выдаче кредита:*

- 1) Принятие правильного решения относительно предоставления кредита, снижение рисков.
- 2) Проведение повторной проверки клиента, если принять решение не дается.

Скоринг-модель потенциального заемщика составляется на основе его «цифрового следа» и данных анкетирования. Скоринговая система оценки клиентов банка, как правило, содержит три раздела: информация по кредиту, сведения о клиенте, финансовое положение клиента.

Каждый клиент  $\{Z_i\}$ ,  $i = 1, \dots, n$ , имеет множество наборов данных полученных с помощью технологии Digital footprint:

$$DF = (I_1, I_2, I_3, \dots), \text{ где}$$

$I_1$ : Анализ данных взятых из социальных сетей клиента (возраст, образование, стаж и т.п.)

$I_2$ : Анализ лайков, групп, друзей и т.п.

$I_3$ : Анализ «языка» постов и сообщений.

$I_4$ : Анализ баз данных государственных служб.

$I_5$ : Поиск упоминаний в сети интернет.

Все наборы данных о клиенте характеризуются  $p$  – мерным вектором признаков  $X_i = (x_{i1}, \dots, x_{ip})^T$ . Признаки определяются требованиями банка. Известна принадлежность каждого клиента  $Y_i$  к одному из двух классов кредитоспособности:

$Y = 0$  – клиент некредитоспособен;

$Y = 1$  – клиент кредитоспособен.

На основе данной выборки необходимо описать процедуры, с помощью которых можно было бы с наибольшей точностью отнести клиентов  $\{W_j\}$ ,  $j = 1, \dots, m$  к одному из классов  $k \geq 2$ ,  $X_i = (x_{i1}, \dots, x_{ip})^T$ , описывающих их.

Так как клиенты могут характеризоваться и количественными, и качественными признаками, возникает задача их классификации в пространстве разнотипных признаков. Набор признаков  $x_{i1}, \dots, x_{ip}$  представляет собой данные из анкет, которые заполняются при подаче заявки на кредит. Основной набор признаков: ФИО, дата рождения, паспортные данные, образование, адрес, семейное положение, ФИО и дата рождения родственников, сведения об основной работе, ежемесячные доходы

и расходы, информация об имеющемся имуществе, информация об имеющихся долгах и обязательствах, запрашиваемая сумма и т.д. Кроме данных из анкеты, если клиент уже брал кредит в прошлом или имеет кредит в настоящее время, в качестве признаков могут также выступать данные о кредитной истории, полученные из кредитного бюро посредством запроса [3]. Для решения поставленной задачи применимы такие модели классификации как, дискриминантный анализ, логистическая регрессия, метод опорных векторов и т.д.

### **Заключение**

В данной работе исследован подход прогнозирования кредитоспособности клиентов банка, основанный на использовании технологий digital footprint. Подход потенциально позволяет найти недостающую информацию о клиенте, либо выявить ложную, которая может кардинальным образом повлиять на его скоринговый балл, рейтинг. Преимущество инструмента «цифрового следа» состоит в том, что он может использоваться как для извлечения требуемой личной информации, так и для проведения анализа его поведения в социуме. Открытая информация в сети Интернет является основным источником данных. Однако следует учесть, что клиент может влиять на свою сетевую идентичность, полностью контролировать и корректировать ее по своему усмотрению. Содержимое и метаданные цифрового следа способны воздействовать на конфиденциальность в Интернете, доверие, безопасность, цифровую репутацию. Правовая сторона вопроса использования данной технологии в РФ не проработана [1][4].

### **Список использованных источников**

1. Digital footprint definition. [Электронный ресурс]. – Режим доступа: [https://techterms.com/definition/digital\\_footprint/](https://techterms.com/definition/digital_footprint/), свободный (дата обращения: 20.11.2018).
2. В.В. Никаненкова. Кредитный скоринг как инструмент оценки кредитоспособности заемщиков, 2012. С. 2-3.
3. Юлия Сергеевна Шунина, Венера Арифзяновна Алексеева, Владимир Николаевич Клячкин. Прогнозирование кредитоспособности клиентов на основе методов машинного обучения. Издательский дом «ФИНАНСЫ и КРЕДИТ», 2015. С. 2-4.
4. Кирсанова Светлана Олеговна, Калинин Алина Александровна. Виртуальные следы: понятие, сущность, проблемы. // Вопросы студенческой науки 2018. Вып.3. С. 16.