

**ОСОБЕННОСТИ ПРИМЕНЕНИЯ МЕТОДА ВЫБОРА  
КОМБИНАЦИЙ ГИПЕРПАРАМЕТРОВ  
ПРИ ИСПОЛЬЗОВАНИИ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ  
ДЛЯ РЕШЕНИЯ УЗКОСПЕЦИАЛИЗИРОВАННЫХ  
БАНКОВСКИХ ЗАДАЧ**

Бурное развитие отрасли информационных технологий, которое наблюдается в современном мире, пожалуй, не имеет аналогов в истории развития человечества. В итоге изменений коснулись все сферы жизни людей. В “Стратегии развития отрасли информационных технологий” отмечено: “Масштаб влияния отрасли информационных технологий на государство значительно превосходит сугубо отраслевые эффекты. Развитие информационных технологий является одним из важнейших факторов, способствующих решению ключевых задач государственной политики Российской Федерации” [1].

Информация сейчас – это одна из самых важных и дорогих сущностей мира. А информационные системы, технологии, процессы, средства и инструменты уже сейчас являются одним из самых определяющих факторов достижения подходов, реализующих инновации, повышающие результативность и эффективность решений, которые необходимо принимать во множестве сфер деятельности - научной, организационной, финансовой, управленческой и многих других.

Бурное развитие сферы ИТ тесно связано с оцифровкой всего наблюдаемого в реальной жизни – всех предметов, явлений, процессов и взаимосвязей между ними. Там, где раньше использовался ручной труд, теперь работают машины и алгоритмы, и с каждым днем все эффективнее.

Важное направление ИТ-отрасли – это автоматизация бизнеса, всех учетных процессов и данных. Это представляет особую коммерческую ценность для автоматизируемых организаций. Там, где раньше информационные технологии были «роскошью» и давали конкурентное преимущество – теперь они являются уже практически жизненной необходимостью.

Важную часть целей автоматизации процессов, например, в банковской сфере, с помощью информационных технологий составляет анализ поведения людей, определение их положительного отношения к организации и выявление нетипичных действий, которые нехарактерны в целом для человека. Вследствие этих особенностей данные вопросы наиболее актуальны в практической плоскости, для конкретного применения в коммерческих предприятиях, банках и т.д.

На сегодняшний день в мире бизнеса, особенно связанного с финансами, активно обсуждается проблема оттока клиентов. Главным показателем в данных процессах является Churn Rate – количество покупателей, прекративших взаимодействие с компанией, относительно числа покупателей, приобретавших в этот период [2]. При этом важным с точки зрения сохранения клиента является установление периода времени, при неактивности клиента, по истечении которого его следует считать утерянным. Очень важно правильно выбрать этот показатель, чтобы избежать неточностей в оценивании оттока. Период измерения оттока должен отражать особенности бизнеса – влияние сезонности, стадии продукта и т. д. Отток за период – важная статистическая метрика, но для постоянного анализа лучше использовать измерение его за период.

С проблемой оттока сталкивается сейчас практически каждая организация. Основная масса денежных средств компании формируется из платежей покупателей, и поэтому клиенты и взаимодействие с ними являются наиболее важными приоритетами для организации. Данная проблема связана с уменьшением клиентской базы и, как следствие, уменьшением выручки. Это очень ощутимо для тех компаний, чья модель монетизации носит подписной и транзакционный характер. Не менее важен при этом фактор стоимости привлечения клиентов – это может обойтись в несколько раз затратнее, чем удерживание уже имеющихся. Этот коэффициент может достигать уровня от 8 до 16 – то есть, чтобы новый клиент стал настолько же лояльным, как старый и узнал столько же о вашем продукте и вашей компании (равно как и вам узнать о нем столько же, сколько знаете о старом клиенте) может понадобиться во столько раз больше денег [3]. Уровень дохода бизнеса напрямую зависит от того, насколько хорошо происходит удержание клиентов. По отчету Гарвардской школы бизнеса увеличение коэффициента удержания всего на 5% может позволить увеличить прибыль на 25-95% [3]. По данным практики, не более 20% организаций создают стратегии по удержанию, а более 80% – фокусируются на привлечении новых.

Поэтому рассматриваемые в рамках данной работы задачи анализа или даже «предсказания» поведения клиента являются важными и актуальными, и могут решаться с помощью нескольких подходов с применением нейронных сетей.

Один из них состоит в использовании стандартной наиболее часто используемой библиотеки машинного обучения, которая наиболее часто применяется на практике. Другой выражает более полное и тщательное исследование данных, закономерностей и взаимосвязей, поиск оптимальных гиперпараметров, лучшую оценку качества и обоснование принятого вывода.

Нужно отметить, что модель машинного обучения имеет несколько параметров, которые не обучаются обучающей выборкой. Эти параметры контролируют точность модели. Их обычно называют гиперпараметрами, настраиваются они заранее и предоставляются вызывающим объектом модели до ее обучения.

Самый простой способ найти хорошую комбинацию гиперпараметров – перебор всех доступных комбинаций. Интуитивно можно создать список возможных значений для каждого гиперпараметра. Затем обучить модель каждой комбинации. Ну и наконец, в завершении оценить каждую комбинацию на протяженном тестовом наборе и сравнить их характеристики, чтобы определить лучшую.

Однако такой метод требует больших вычислительных ресурсов. Предположим, обучение данной модели занимает 3 часа и имеет 2 гиперпараметра. Если протестировать 5 возможных значений для каждого из них, поиск по сетке займет  $(5^2) \cdot 3 = 75$  часов, чтобы оценить все комбинации.

Если модель быстро обучается, имеет мало гиперпараметров или доступны большие вычислительные мощности, то поиск по сетке – это простое в реализации решение. К сожалению, на практике такое случается достаточно редко.

Если нет такого количества времени или вычислительных ресурсов, альтернативой является случайный поиск. Вместо того, чтобы полностью проверять комбинации в сетке, случайный поиск заключается в выборке случайных значений в пространстве гиперпараметров. Хотя он уравнивает вычислительные затраты на поиск по сетке, этот подход неэффективен в том, как он исследует возможные комбинации гиперпараметров. Действительно, алгоритм не учится на ранее проверенных комбинациях.

Взаимным недостатком обоих этих подходов является то, что метод теряет эффективность, если параметры воздействия не известны заранее. Так бывает во многих условиях. С помощью поиска по сетке можно протестировать множество эквивалентных комбинаций, что приведет к потере вычислительных ресурсов и времени.

И так, в рамках представленной работы на основе описанных методов была решена практическая задача – «предсказания» оттока клиентов банка. Примененная на практике методика решения была следующей: в качестве входных данных использовался набор данных со следующими полями: УНП, название организации, дата открытия и закрытия счета, текущий статус, адрес, вид деятельности, причина закрытия, наличия счета в другом банке и иные. Перед началом работы набор данных был подготовлен, а именно проверены на корректность и наличие информации во всех ячейках датасета. Далее удалены неинформативные колонки и сохранены идентификаторы категориаль-

ных колонок в отдельный список. После этого датасет разделен на две части – первая часть предназначалась для обучения нейронной сети, вторая же часть для апробации.

Применив стандартную модель классификации, был получен результат корректности предсказания оттока клиентов банка, равный 64,52%. За счет применения разработанного алгоритма подбора гиперпараметров, а именно использования метода кросс-валидации и стратификации со случайным распределением был получен результат корректности предсказания оттока клиентов банка, равный 74%. Таким образом, модель на базе предложенного подхода решения задачи прогнозирования оттока клиентов, основанная на использовании нейронной сети с предварительным определением гиперпараметров, показала лучший результат, чем классический вариант решения подобных задач, применяемый на практике.

#### ЛИТЕРАТУРА

1. Электронный ресурс, [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_154161/](http://www.consultant.ru/document/cons_doc_LAW_154161/), Распоряжение Правительства РФ от 01.11.2013 N 2036-р (ред. от 18.10.2018) “Об утверждении Стратегии развития отрасли информационных технологий в Российской Федерации на 2014 - 2020 годы и на перспективу до 2025 года”.

2. Электронный ресурс, <https://www.calltouch.ru/glossary/churn-rate>, “Churn rate”.

3. Электронный ресурс, <https://vc.ru/marketing/156703-12-sposobov-nemedlenno-ostanovit-ottok-klientov>, “12 способов немедленно остановить отток клиентов”.

УДК 003.26

Я. Ласык<sup>1</sup>, ассист.; Д.М. Романенко<sup>2</sup>, доц., канд. техн. наук  
П.П. Урбанович<sup>2</sup>, проф., д-р техн. наук

<sup>1</sup>(Люблинский Католический университет, Польша), <sup>2</sup>(БГТУ, г. Минск)

#### **ИСПОЛЬЗОВАНИЕ СЕТЕВЫХ ПРОТОКОЛОВ И СТЕГАНОГРАФИИ ДЛЯ ТАЙНОЙ ПЕРЕДАЧИ ИНФОРМАЦИИ**

В последнее время наблюдается новый и опасный тренд: все больше разработчиков вредоносного ПО и средств кибершпионажа прибегают к использованию стеганографии.

Современные методы компьютерной стеганографии условно можно поделить на 2 класса:

1) методы сокрытия информации в контенте, а также в альтернативных потоках данных (Alternate Data Streams, ADS) файловой си-