

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК [004.021:004.85]:336.77

Фёдоров Илья Васильевич

Алгоритм классификации клиентов банка для создания скоринговой карты

**АВТОРЕФЕРАТ**

на соискание степени магистра технических наук  
по специальности 1-40 80 02

Системный анализ, управление и обработка информации

---

Научный руководитель  
Севернёв Александр Михайлович  
кандидат технических наук доцент,

---

Минск 2020

## ВВЕДЕНИЕ

С целью повышения доходности кредитных операций и уменьшения риска выдачи кредита заёмщику, который не в состоянии выплатить кредит, в банковской сфере используется кредитный скоринг (*credit scoring*). Кредитный скоринг представляет собой один из современных подходов к минимизации кредитного риска, в рамках которого используются автоматизированные системы принятия решений по выдаче банковских кредитов и условиям кредитования.

Скоринг также находит свое применение и в маркетинге. В маркетинге выделяют следующие виды скоринга: скоринг отклика (*response scoring*) и скоринг сохранения/потерь (*retention/attrition scoring*). Скоринг отклика – модели, которые оценивают реакцию потребителей на почтовую рекламную рассылку информации о товаре. Скоринг сохранения/потерь – модели, предсказывающие возможное поведение клиента – дальнейшее использование продукта или прекращение его использования.

В данной диссертации рассматриваются модели кредитного скоринга с использованием алгоритмов машинного обучения, которые собираются банком с помощью анкеты и уже имеющейся кредитной истории клиента.

Результат анализа, подкреплённый экспериментами с различными алгоритмами, позволит выбрать наиболее подходящую модель для кредитного скоринга.

В данной диссертации строятся модели с помощью следующих алгоритмов машинного обучения: дерево решений, случайный лес, логистическая регрессия.

В эксперименте представлены результаты работы алгоритмов *C4.5*, *CART*, случайного леса и логистической регрессии в среде программирования *Python*.

Актуальность данной диссертации подтверждается отсутствием в открытом доступе используемых алгоритмов для классификации клиентов при кредитном скоринге финансовыми организациями в виду их экономической и научной значимости.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность данной работы заключается в необходимости создания системы автоматизированного кредитного скоринга в виду отсутствия в открытом доступе соответствующих алгоритмов, используемых финансовыми организациями. Данное исследование так же может быть применено, например, в сфере страхования, безопасности, при трудоустройстве и других сферах.

Целью магистерской диссертации является анализ различных алгоритмов кредитного скоринга, возможностей их реализации при помощи машинного обучения, а также сравнение данных алгоритмов.

В ходе работы над магистерской диссертацией были решены следующие задачи:

- показана актуальность исследования в области кредитного скоринга;
- проанализированы существующие алгоритмы кредитного скоринга;
- проведено сравнение алгоритмов машинного обучения для выполнения поставленной задачи;
- описан процесс кредитного скоринга;
- изучены исследования в рассматриваемой области.

Результаты данной магистерской диссертации были опубликованы в сборниках 27-ой Международной научно-практической конференции «*Advances in science and technology*» [1–А.] и 28-ой Международной научно-практической конференции «*Eurascience*» [2–А.].

Общий объём магистерской диссертации составляет 60 страниц, включая 13 иллюстраций, 18 таблиц, библиографический список из 31 наименования, 2 приложения.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В главе 1 «Анализ научной литературы» проанализирована существующая информация о текущих разработках в сфере кредитного скоринга, рассмотрены их недостатки и преимущества, а также данные, которые использует определенный алгоритм. Был произведен обзор появления кредитного скоринга, перехода от использования ручного труда к автоматизированным алгоритмам. Был сделан вывод о необходимости поиска оптимального алгоритма классификации клиентов для создания кредитного скоринга.

В главе 2 «Алгоритмы машинного обучения для построения модели кредитного скоринга» исследуются существующие методы классификации данных в разрезе классификации клиентов, рассматривается общая процедура по подготовке и анализу данных. Дается математическое описание алгоритмам (логистическая регрессия, дерево решений и другим) и способу их оценки (*ROC-кривая*).

В главе 3 «Анализ моделей кредитного скоринга на основе анкетных данных и кредитной истории» производится практическое применение рассмотренных в главе 2 алгоритмов, проводится их сравнительный анализ с последующим выбором наиболее оптимального алгоритма.

Магистерская диссертация выполнена самостоятельно, проверена в системе «Антиплагиат». Процент оригинальности соответствует норме, установленной кафедрой. Цитирования обозначены ссылками на публикации, указанные в «Списке использованных источников».

## ЗАКЛЮЧЕНИЕ

В диссертации рассмотрены алгоритмы построения модели кредитного скоринга с использованием алгоритмов машинного обучения на основе анкетных данных, собираемых банком.

Результат анализа, подкреплённый экспериментами с различными алгоритмами, позволил выбрать наиболее подходящую модель для кредитного скоринга.

В диссертации также рассмотрено, как методы больших данных и машинного обучения изменяют отрасль кредитного скоринга.

В настоящее время в свободном доступе отсутствуют алгоритмы, используемые крупными компаниями для кредитного скоринга, что повышает актуальность данной диссертации. Данное исследование так же может быть применено, например, в сфере страхования, безопасности, при трудоустройстве и других сферах.

В данной работе получены следующие основные результаты:

– дан общий анализ проблемы разработки моделей кредитного скоринга на основе методов машинного обучения;

– для задачи кредитного скоринга, рассматриваемой как задачи статистической классификации, описаны используемые в работе алгоритмы логистической регрессии, *C4.5*, *CART*, *Random Forest*;

– построены модели кредитного скоринга с помощью алгоритмов *C4.5*, *CART*, *Random Forest* на языке программирования *Python* для данных, собранных банком с помощью анкеты;

– проведен сравнительный анализ моделей кредитного скоринга для различных алгоритмов классификации.

Оказалось, что наиболее эффективным алгоритмом классификации клиентов при кредитном скоринге является случайный лес (*Random Forest*), использующий данные, собираемые банком с помощью анкеты.

## СПИСОК ПУБЛИКАЦИЙ АВТОРА

[1–А.] Фёдоров, И.В. Алгоритм классификации клиентов банка для создания скоринговой карты / И.В. Фёдоров, И.В. Нестерчук // XXVII Международная научно-практическая конференция «*Advances in science and technology*» – Москва, 2020 – С. 85–88.

[2–А.] Фёдоров, И.В. Алгоритм классификации клиентов банка для создания скоринговой карты / И.В. Фёдоров, И.В. Нестерчук // XXVIII Международная научно-практическая конференция «*Eurasiascience*» – Москва, 2020. – С. 89–92.

Библиотека БГУИР