



Международный журнал информационных технологий и энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ПРИ РАСЧЁТЕ СКОРИНГА КЛИЕНТОВ БАНКА

¹Судаков В. А., ²Семенов Т. П.

МИРЭА - Российский технологический университет, Москва, Россия (119454, г. Москва, пр. Вернадского, 78), e-mail: ¹ sudakov_v@mirea.ru, ² timboy99@mail.ru

В данной статье рассматриваются различные типы и виды машинного обучения, а также понятие самого машинного обучения. Раскрываются основные принципы и особенности функционала машинного обучения при расчёте скоринга клиентов банка.

Ключевые слова: машинное обучение, методы машинного обучения, кредитный скоринг.

MACHINE LEARNING METHODS IN CALCULATION OF BANK CUSTOMER SCORING

¹Sudakov V. A., ²Semenov T. P.

MIREA - Russian Technological University, Moscow, Russia (119454, Moscow, Vernadskogo Ave., 78), e-mail: ¹ sudakov_v@mirea.ru, ² timboy99@mail.ru

This article discusses the various types and types of machine learning, as well as the concept of machine learning itself. The main principles and features of the functional types of machine learning are revealed when calculating the creditworthiness of bank customers.

Keywords: machine learning, machine learning methods, credit scoring.

Кредитный скоринг – это рациональное управление рисками, именно правильное оценивание поможет минимизировать их и даже исключить. Система заключается в том, что строится математическая модель по обработке данных клиентов на основе анкетирования. Есть большая зависимость от различных социальных факторов, напри-мер, семейное положение (женат, замужем, разведён и прочее), количество детей, высшее образование и т.д.. Совокупность этих данных помогает системе спрогнозировать поведение клиента и его возможность вернуть долг. После чего клиенту необходимо предоставить документы и пройти интервью с сотрудником банка. На основе полученных данных будет принято окончательное решение [1].

Данную систему оценки кредитных рисков впервые стали применять американские банки. Это позволило унифицировать процедуры анализа надежности клиентов финансовых организаций. Оценка позволила банкам значительно сократить время и трудозатраты, связанные с ручной проверкой кредитоспособности. Система была предназначена для различных задач:

- определение платёжеспособности клиентов;
- прогнозирование поведения заёмщика;
- распределение клиентов на группы по возможности просрочки возврата долга;
- отслеживание кредитного профиля.

Системы кредитного скоринга могут широко варьироваться в зависимости от целей банка [2]. Но цель редко бывает одинаковой, и поэтому при оценке заемщика кредитор может комбинировать различные типы оценок в анкете. Правда, такой подход используется только для крупных кредитов - например, ипотеки. Зато затраченные ресурсы и время помогают увеличить процент доверенных клиентов.

Машинное обучение представляет собой отдельный раздел искусственного интеллекта, в который входит множество алгоритмов, чтобы осуществлять обучение машин для создания решения задач различного характера. Машинное обучение охватывает различные дисциплины и науки, такие как информатика, статистика и т. д. Однако машинное обучение не всегда было связано с математикой и информатикой, оно касается других областей науки, таких как медицина, биология и другие. В настоящее время сфера применения алгоритмов машинного обучения очень широка, она начала включать в себя знания из разных областей [3].

В настоящее время существует два типа машинного обучения [4]. Первый — индуктивное обучение, задачей которого является поиск закономерностей в разнообразной информации и источниках данных для их систематического систематического описания или прогнозирования будущих данных. Второй — дедуктивное обучение, которое создает общую модель, основанную на экспертных знаниях, которые впоследствии используются для получения выводов, но дедуктивное обучение — это скорее экспертная система, поэтому многие специалисты по индуктивному обучению имеют в виду машинное обучение [5].

Обучение делят на три вида. Первым является обучение с учителем. Этот подкласс очень распространён на сегодняшний день, работает с данными, которые были организованы в виде «объект – метка». В обучении с учителем происходит обучение алгоритма, который ищет зависимость между объектом и меткой. Также обучение с учителем делят на несколько типов:

- классификации;
- регрессии;
- прогнозирования.

В задачах классификации главной целью является правильное отнесение определённого объекта к классу. Уровень выполненной работы алгоритма определяется количество объектов выборки, которые были отнесены к неверному классу. В задачах регрессии метка объекта является действительное число и чего следует то, что количество предполагаемых может быть неограниченно. В задачах прогнозирования объекты – это значения определённого параметра, которые располагаются по оси времени и общее число данных объектов называют временным рядом, а задачу принято называть прогнозированием [6].

Второй – обучение без учителя. В данном подклассе объект является вектором значений определённых признаков или вектором расстояний. Основной целью алгоритма в этом подклассе обучения является поиск между объектами зависимостей, основываясь на данных их признаков.

Существует четыре задачи обучения без учителя:

- кластеризация;

- фильтрация выбросов;
- сокращение размерности;
- заполнение пропусков в данных.

В задачах кластеризации основная идея заключается в разделении выборки объектов на группы так, чтобы объекты сходились по выбранным признакам, а объекты разных групп отличались.

В задачи сокращения размерности входит применение преобразования данных, без потери информации об объектах выборки.

Заполнение пропусков в данных – задача, которая подходит при работе с различными выборками, где пропущено значение.

Третьим подклассом является частичное обучение, в них метки имеются только у некоторых объектов выборки.

Следующий подкласс обучение с подкреплением. В нём объекты — это пары «состояние – действие», а метка – реакция окружающей среды на действие, которая означает правильность действия. Данный вид обучения используется для сложных задач, например, для настраивания системы работы робота [7].

Сегодня существует несколько методов кредитного скоринга, первый – при помощи бинарной логистической регрессии, второй – при помощи деревьев классификации.

Наиболее распространённым методом в кредитном скоринге является бинарная логистическая регрессия, которая состоит в том, что алгоритм применяется для оценки вероятности двоичного ответа, основываясь на нескольких данных [8]. В данной модели логарифм вероятности определённого события является линейной комбинацией независимых переменных. Данная модель рассчитывает, что наличие определённых факторов риска увеличивает вероятность определённого исхода на конкретный фактор.

Другой метод при использовании деревьев классификации состоит в прогнозировании события на основе определённых данных. Этот метод позволяет найти связь между различными переменными. Деревья проверяют различные комбинации полученных переменных для того, чтобы определить каким образом объединяются лучшие переменные для объяснения результата, показывая это графически в узлах, которые отображаются в специальном порядке, в зависимости от их дискриминантной мощности.

Главным отличием деревьев от логистической регрессии является то, что деревья не требуют, чтобы данные имели нормальное распределение. Алгоритмы данного метода могут обрабатывать данные различных типов.

Таким образом, машинное обучение стало очень важным аспектом в развитии многих сфер деятельности человека. В кредитном скоринге алгоритмы машинного обучения являются неотъемлемой частью и помогают избежать лишних рисков и издержек, за счёт работы алгоритмов, которые помогают сотрудникам банка принимать правильные решения касательно вопросов предоставления кредитов.

Список литературы

1. Тукумбетов А.Р., Задорожный Н.С. Кредитный скоринг с помощью методов машинного обучения // Инновации. Наука. Образование. 2021. № 36. 1118 с.

2. Жураев Ж.Д. Использование методов машинного обучения в моделировании кредитного скоринга // Вестник науки. 2021. Т. 2. № 6-1 (39). 87 с.
3. Абдрахманова Э. Р. Машинное обучение: обзор и применение // Техника и технологии: пути инновационного развития – 2019 – 9с.
4. Алханов А.А. Машинное обучение и его применение в современном мире // Проблемы науки – 2021 – № 7 (66) – 25 с.
5. Воробьев Е.А., Шмаль В.Н. Проблемы машинного обучения // Дневник науки – 2021 – № 5 (53) – 18 с.
6. Загайнов М.А., Костенков Е.А., Кузнецов Д.С. Машинное обучение. Области применения технологии и перспективы развития // Colloquium-Journal – 2019 – № 16-1 (40) – 49 с.
7. Згонникова А.О., Прокопенко А.А. Машинное обучение и обучение на протяжении всей жизни // Новые научные исследования. Сборник статей VIII Международной научно-практической конференции – 2022 – 22 с.
8. Баданина Н.Д., Судаков В.А. Модели машинного обучения для классификации отзывов о банках // Препринты ИПМ им. М.В.Келдыша. 2021. No 50. 14 с.

References

1. Tukumbetov A.R., Zadorozhny N.S. Credit scoring using machine learning methods // Innovations. The science. Education. 2021. No. 36. p.1118.
 2. Zhuraev Zh.D. Using machine learning methods in credit scoring modeling. Vestnik nauki. 2021. Vol. 2. No. 6-1 (39) p. 87.
 3. Abdrakhmanova E. R. Machine learning: review and application // Engineering and technology: ways of innovative development - 2019 - p.9.
 4. Alkhanov A.A. Machine learning and its application in the modern world // Problems of Science - 2021 - No. 7 (66) - p.25.
 5. Vorobyov E.A., Shmal V.N. Problems of Machine Learning // Diary of Science - 2021 - No. 5 (53) – p.18.
 6. Zagainov M.A., Kostenkov E.A., Kuznetsov D.S. Machine learning. Application areas of technology and development prospects // Colloquium-Journal - 2019 - No. 16-1 (40) - 49 p.
 7. Zgonnikova A.O., Prokopenko A.A. Machine learning and lifelong learning // New scientific research. Collection of articles of the VIII International Scientific and Practical Conference - 2022 - p.22.
 8. Badanina N.D., Sudakov V.A. Machine Learning Models for Classifying Bank Reviews // IPM Preprints im. M.V. Keldysh. 2021. No 50. p.14.
-