

## МЕТОДЫ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ВЕРОЯТНОСТИ НАСТУПЛЕНИЯ СОБЫТИЙ

*Филатов Илья Игоревич, младший научный сотрудник учебно-научной лаборатории «Искусственный интеллект в АПК» ФГБОУ ВО «Российский государственный аграрный университет-МСХА имени К.А. Тимирязева», [filtovilya@mail.ru](mailto:filtovilya@mail.ru)*

*Демичев Вадим Владимирович, кандидат экономических наук, доцент кафедры статистики и кибернетики ФГБОУ ВО «Российский государственный аграрный университет-МСХА имени К.А. Тимирязева», [demichev\\_v@rgau-msha.ru](mailto:demichev_v@rgau-msha.ru)*

***Аннотация:** Данная статья исследует методы моделирования машинного обучения, которые используются для прогнозирования вероятности наступления событий. В контексте растущей важности прогнозирования вероятности событий в различных областях, таких как финансы, медицина, промышленность и другие, понимание и применение эффективных методов моделирования является ключевым.*

***Ключевые слова:** машинное обучение, методы, прогнозирование*

Технологии современного мира непрерывно развиваются, появляются все более усовершенствованные средства и системы решения различного рода задач, но также возникают и новые проблемы, решение которых выходит на первый план. Примером такой проблемы является случай, когда человеку необходимо обрабатывать огромные потоки информации, анализировать их и совершать определенные умозаключения, но в связи с ограниченностью человеческих возможностей совершать подобного рода операции крайне неэффективно по времени и продуктивности. Одним из распространенных на сегодняшний день методов решения и оптимизации такой проблемы является машинное обучение.

Машинное обучение как класс методов является ответвлением от методов искусственного интеллекта, сама суть машинного обучения заключается не в прямом решении поставленной задачи, а в обучении, где применяется множество решений схожих задач средствами математического анализа, численных методов, теории вероятности, математической статистики и других различных техник работы с данными в цифровом виде.

Данная проблема является актуальной, поскольку повсеместно появляются задачи, требующие обработки большого объема данных, но при этом требующие и большой отлаженности и структурированности, которые человек зачастую не в состоянии самостоятельно выполнить.

В последние годы машинное обучение стало неотъемлемой частью различных сфер деятельности, таких как финансы, медицина, транспорт и

многие другие. Одной из важных задач машинного обучения является прогнозирование вероятности наступления событий. Это позволяет принимать решения на основе предсказаний и управлять рисками более эффективно. В данной статье мы рассмотрим некоторые методы моделирования машинного обучения, которые широко применяются для прогнозирования вероятности событий [1,2,3].

#### 1. Логистическая регрессия

Логистическая регрессия является одним из наиболее распространенных методов для прогнозирования вероятности бинарных событий. Она основана на логистической функции, которая преобразует линейную комбинацию входных признаков в вероятность наступления события. Логистическая регрессия может быть расширена для прогнозирования вероятности многоклассовых событий с использованием методов, таких как one-vs-all или softmax.

#### 2. Деревья принятия решений

Деревья принятия решений представляют собой графическую модель, состоящую из узлов и ребер, где каждый узел представляет тест на определенный признак, а каждое ребро соответствует возможному значению этого признака. Деревья принятия решений могут быть использованы для прогнозирования вероятности событий путем оценки доли положительных классов в каждом листовом узле. Они могут быть расширены до случайных лесов или градиентного бустинга для повышения точности предсказаний.

#### 3. Нейронные сети

Нейронные сети являются мощными моделями машинного обучения, способными моделировать сложные взаимосвязи между входными признаками и целевой переменной. Для прогнозирования вероятности событий можно использовать различные архитектуры нейронных сетей, такие как многослойные перцептроны, сверточные нейронные сети или рекуррентные нейронные сети. Глубокое обучение и методы передачи обучения также могут быть применены для улучшения результатов прогнозирования.

#### 4. Методы градиентного бустинга

Методы градиентного бустинга, такие как градиентный бустинг деревьев решений (Gradient Boosting Decision Trees) и градиентный бустинг над нейронными сетями (Gradient Boosting over Neural Networks), объединяют слабые модели в сильную композицию для прогнозирования вероятности событий. Они работают путем последовательного обучения моделей, при этом каждая следующая модель исправляет ошибки предыдущих. Методы градиентного бустинга обычно позволяют достичь высокой точности предсказаний.

#### 5. Метод опорных векторов (SVM):

Метод опорных векторов (Support Vector Machines) является мощным алгоритмом машинного обучения, который может быть использован для прогнозирования вероятности событий. Он основан на поиске оптимальной

разделяющей гиперплоскости, которая максимизирует отступ между классами. SVM может быть расширен для задачи прогнозирования вероятности с помощью калибровки результатов.

#### 6. Гауссовские процессы:

Гауссовские процессы (Gaussian Processes) являются вероятностным методом машинного обучения, который может быть использован для моделирования вероятности событий. Они представляют собой непараметрический подход, основанный на распределении Гаусса, и могут обеспечивать не только точечные прогнозы, но и оценки неопределенности.

#### 7. Байесовская сеть:

Байесовская сеть (Bayesian Network) - это вероятностная модель, которая использует графическое представление для описания зависимостей между переменными. Она может быть применена для прогнозирования вероятности событий путем комбинирования априорных знаний и наблюдаемых данных.

#### 8. Решающие деревья и случайные леса:

Решающие деревья (Decision Trees) представляют собой древовидную структуру, где каждый узел представляет тест на признак, а каждое ребро соответствует значению этого признака. Они могут быть использованы для прогнозирования вероятности событий, оценивая долю положительных классов в листовых узлах. Случайные леса (Random Forests) являются ансамблем решающих деревьев и могут значительно повысить точность прогнозирования.

#### 9. Эволюционные алгоритмы:

Эволюционные алгоритмы - это класс алгоритмов, вдохновленных принципами биологической эволюции. Они могут быть применены для оптимизации моделей машинного обучения, включая задачи прогнозирования вероятности событий. Эволюционные алгоритмы могут итеративно улучшать модель путем применения генетических операторов, таких как скрещивание и мутация[4,5,6?7].

Комбинация различных методов моделирования машинного обучения может привести к лучшим результатам прогнозирования вероятности наступления событий. Важно учитывать особенности задачи, доступные данные и требования к точности и интерпретируемости модели при выборе соответствующего подхода.

Методы моделирования машинного обучения предоставляют широкий набор инструментов для прогнозирования вероятности наступления событий. Логистическая регрессия, деревья принятия решений, нейронные сети, методы градиентного бустинга, SVM, гауссовские процессы, байесовские сети, решающие деревья, случайные леса и эволюционные алгоритмы - все они предлагают различные подходы к моделированию и прогнозированию вероятности событий.

Выбор конкретного метода зависит от особенностей задачи, доступных данных, требований к точности и интерпретируемости модели. Важно

провести анализ данных, выбрать подходящую модель и правильно настроить ее параметры для достижения оптимальных результатов прогнозирования вероятности событий. Кроме того, применение техник предварительной обработки данных, калибровки результатов и ансамблирования моделей также может улучшить качество прогнозов.

Однако, необходимо помнить, что прогнозирование вероятности событий является сложной задачей, особенно при наличии неопределенности и шума в данных. Важно учитывать ограничения и предпосылки моделей, а также проводить оценку и валидацию результатов.

Благодаря развитию методов моделирования машинного обучения, мы можем получать более точные прогнозы вероятности наступления событий, что помогает в принятии информированных решений и эффективном управлении рисками.

### Библиографический список

1. Гладких А.В. Методы машинного обучения. Издательство БХВ-Петербург, 2018.
2. Воронцов К.В. "Машинное обучение". Издательство "БХВ-Петербург", 2018.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
4. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
5. Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
6. Кузнецов, М.П. Машинное обучение и анализ данных / М.П. Кузнецов. – Москва: Издательство «ДМК Пресс», 2016.
7. Khoruzhy, L.I., Katkov, Y.N., Romanova, A.A. Cloud Technologies in the Accounting Information System of Interorganizational Cooperation, Innovation, Technology and Knowledge Management [this link is disabled](https://www.scopus.com/authid/detail.uri?authorId=57221331639), 2023, pp. 25–37 <https://www.scopus.com/authid/detail.uri?authorId=57221331639> (Scopus)
8. Курс социально-экономической статистики: Учебник для вузов / В. Л. Соколин, М. Р. Ефимова, А. Л. Кевеш [и др.]. – Москва : Финстатинформ, 2002. – 976 с. – ISBN 5-7866-0021-1. – EDN TDCCKD.
9. Состояние социально-трудовой сферы села и предложения по ее регулированию: Ежегодный доклад по результатам мониторинга 2011 г. / Л. В. Бондаренко, А. В. Турьянский, Т. И. Наседкина [и др.]; Ответственные за подготовку доклада: Д.И. Торопов. Том Выпуск 13. – Москва: Российский научно-исследовательский институт информации и технико-экономических исследований по инженерно-техническому обеспечению агропромышленного комплекса, 2012. – 220 с. – ISBN 978-5-7367-0903-8.