



Международный журнал информационных технологий и энергоэффективности

Сайт журнала:

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



УДК 004.6

АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ В СОЗДАНИИ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ НА ПРИМЕРЕ СТАТИСТИКИ ПРОДАЖ АВТОМОБИЛЕЙ

Семенов Д.В.

ФГАОУ ВО "САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ", Санкт-Петербург, Россия (190000, город Санкт-Петербург, Большая Морская ул., д.67 лит. а), e-mail: Danya2015rus@yandex.ru

Данная статья рассматривает алгоритмы машинного обучения и их влияние в создании моделей прогнозирования на примере статистики продаж автомобилей.

Ключевые слова: Алгоритм, машинное обучение, статистика, модель, прогнозирование, этапы разработки, данные, анализ.

MACHINE LEARNING ALGORITHMS IN CREATING FORECASTING MODELS BASED ON THE EXAMPLE OF CAR SALES STATISTICS

Semyonov D.V.

ST. PETERSBURG STATE UNIVERSITY OF AEROSPACE INSTRUMENTATION, St. Petersburg, Russia (190000, St. Petersburg, Bolshaya Morskaya str., 67 lit. a), e-mail: Danya2015rus@yandex.ru

This article examines machine learning algorithms and their impact in creating forecasting models using the example of car sales statistics.

Keywords: algorithm, machine learning, statistics, model, forecasting, development stages, data, analysis.

Введение

В современном мире, где информация становится одним из самых ценных ресурсов, способности к обработке и анализу данных играют ключевую роль в принятии бизнес-решений. Одной из самых перспективных областей применения аналитических технологий является машинное обучение, которое позволяет создавать эффективные модели прогнозирования. В этом контексте рассмотрим использование алгоритмов машинного обучения для прогнозирования продаж автомобилей.

1. Введение в машинное обучение

Машинное обучение — это область искусственного интеллекта, которая фокусируется на разработке алгоритмов, способных обучаться на данных и делать прогнозы или принимать решения без явного программирования. Основная идея машинного обучения заключается в том, что алгоритмы могут автоматически улучшаться по мере увеличения объема данных. В контексте продаж автомобилей это может включать предсказания о будущих продажах на основе исторических данных, сезонных колебаний, экономических факторов и других переменных.[1]

Модели машинного обучения в значительной степени зависят от качества данных, на которых они обучаются. Чем больше и разнообразнее данные, тем лучше будет модель. Поэтому важным аспектом является не только выбор алгоритма, но и обеспечение процесса сбора и предварительной обработки данных.

2. Задачи прогнозирования в области продаж автомобилей

Прогнозирование продаж автомобилей может быть разделено на несколько ключевых задач:

1. Прогнозирование общего объёма продаж — оценка, сколько автомобилей будет продано в определённый период (например, месяц, квартал или год). Это особенно важно для планирования производственных мощностей, логистики и управления запасами.

2. Сегментация рынка — анализ, какие сегменты рынка дадут наибольшую отдачу. Например, изучение предпочтений различных групп потребителей позволяет производителям адаптировать свои предложения.

3. Оценка влияния факторов [2] — определение, какие факторы (экономические, социальные, сезонные) имеют наибольшее влияние на продажи. Это может включать анализ влияния цен на нефтепродукты, изменение налогового законодательства, уровней доходов населения и других факторов.

4. Прогнозирование спроса на новинки — оценка, как возможные новинки автомобильного рынка будут восприняты потребителями, основываясь на аналогичных продуктах, вышедших ранее.

5. Анализ времени моделирования — выявление пиковых периодов продаж (например, весна или конец года) и их влияние на общие продажи.

3. Алгоритмы машинного обучения для прогнозирования

Существует множество алгоритмов машинного обучения, которые могут быть применены для прогнозирования продаж автомобилей. Рассмотрим несколько из них подробно:

1) Регрессия

Линейная регрессия и её более сложные версии (например, полиномиальная регрессия) могут использоваться для построения моделей, которые анализируют взаимосвязь между продажами и различными факторами, такими как цена, качество, маркетинг, время года и т.д. [3] Линейная регрессия проста в интерпретации и позволяет быстро выявить тренды, но она может не всегда хорошо работать с нелинейными зависимостями.

2) Деревья решений

Эти алгоритмы разбивают данные на более мелкие подмножества, создавая дерево, которое позволяет делать предсказания на основе базовых условий. Деревья решений просты в интерпретации и наглядны, что особенно полезно для бизнес-аналитики. Однако они могут быть чувствительны к шуму в данных и склонны к переобучению.

3) Случайный лес

Это ансамблевый метод, который объединяет несколько деревьев решений для улучшения точности и устойчивости модели. [4] Случайный лес хорошо подходит для многофакторных данных, таких как данные о продажах автомобилей. Он помогает избежать

проблемы переобучения, позволяя моделям обобщать информацию и лучше описывать сложные зависимости в данных.

4) Градиентный бустинг

Этот метод использует последовательное построение деревьев, улучшая каждое следующее дерево на основе ошибок предыдущего. Это один из самых мощных и гибких подходов к задачам прогнозирования, поскольку он может эффективно обрабатывать как линейные, так и нелинейные зависимости и обеспечивать высокую предсказательную силу.

5) Нейронные сети

Современные нейронные сети, особенно глубокие, могут обрабатывать сложные зависимости и различия в данных. [5] Например, рекуррентные нейронные сети (RNN) могут быть полезны для временных рядов, анализируя последовательности данных во времени, тогда как сверточные сети (CNN) могут помочь в анализе изображений и других форматов данных.

4. Этапы разработки модели

Разработка модели прогнозирования продаж автомобилей с использованием машинного обучения включает несколько этапов:

1) Сбор данных

Включает сбор исторических данных о продажах, ценах, экономических условиях, рекламе и т.д. Для этого могут использоваться как внутренние источники (продажные отчеты, данные CRM), так и внешние (экономические показатели, данные о потребителях, исследования рынка). Качественные данные являются основой для построения успешной модели.

2) Предварительная обработка данных

Чистка данных (удаление пропусков, аномалий) и их последующее преобразование (нормализация, кодирование категориальных переменных) обязательно перед обучением модели. [6] Например, могут быть использованы техники, такие как one-hot encoding для преобразования категориальных данных и стандартное min-max нормализация для числовых.

3) Разделение данных

Данные разделяются на обучающую и тестовую выборки. Обычно это соотношение составляет 70-80% данных для обучения и 20-30% для тестирования. Это позволяет оценить качество модели на новых данных, которые не использовались в процессе обучения.

4) Обучение модели

Выбор алгоритма и его обучение на обучающей выборке. Важно помнить, что для разных типов задач могут быть предпочтительнее разные алгоритмы, поэтому может потребоваться предварительное тестирование.

5) Оценка и валидация [7]

Использование тестовой выборки для оценки точности модели. Метрики, такие как RMSE (среднеквадратическая ошибка), MAE (средняя абсолютная ошибка) или R^2 (коэффициент детерминации), помогают определить её эффективность. Также может использоваться кросс-валидация для более надежной оценки производительности.

6) Оптимизация модели

Подбор гиперпараметров и использование методов кросс-валидации для повышения качества прогнозирования. Подбор параметров может включать алгоритмы, такие как Grid Search или Random Search, и может значительно повысить производительность модели.

7) Внедрение и мониторинг

После внедрения модели в производственный процесс её необходимо регулярно мониторить и обновлять по мере изменения рынка. [8] Например, если происходит значительное изменение в экономической ситуации или потребительских предпочтениях, модель должна быть адаптирована для учета этих изменений, чтобы поддерживать её актуальность.

5. Реальные примеры применения

Для более глубокого понимания важности машинного обучения в прогнозировании продаж автомобилей рассмотрим несколько примеров реальных компаний:

- Ford использует сложные модели машинного обучения для анализа тенденций в потребительских предпочтениях. С помощью динамической оценки предпочтений клиентов и логистики, они могут адаптировать свои маркетинговые стратегии и производственные процессы, чтобы оптимально отвечать спросу.
- Toyota применяет методы машинного обучения для прогнозирования спроса на свои новые продукты, учитывая сезонные колебания, экономические условия и даже социальные факторы (например, влияние окружения на выбор модели автомобиля). Это помогает компании более точно управлять запасами и улучшать удовлетворенность клиентов.
- Tesla активно использует данные о продажах и поведении пользователей, собирая информацию о предпочтениях и стиле вождения владельцев своих автомобилей. Это позволяет им не только улучшать продукты, но и адаптировать маркетинговые стратегии, базируясь на реальных данных.

Заключение

Применение алгоритмов машинного обучения для прогнозирования продаж автомобилей открывает новые горизонты для бизнеса. Эффективные модели прогнозирования помогают компаниям лучше понимать рынок, оптимизировать запасы и принимать стратегические решения на основе данных. Однако для получения качественных результатов необходимо не только правильно выбрать алгоритм, но и уделить внимание всем этапам разработки модели.

Современные модели машинного обучения позволяют не только предсказывать продажи, но и интегрировать множественные факторы в процессе принятия решений. В условиях динамичного рынка автомобилестроения это особенно важно и актуально. Успешное применение машинного обучения в этой области может стать значительным конкурентным преимуществом для компаний, стремящихся как к повышению продаж, так и к улучшению качества обслуживания клиентов, что, в конечном счете, способствует устойчивому развитию и росту бизнеса.

Список литературы

1. Вьюгин, В.В. Математические основы машинного обучения и прогнозирования / В.В. Вьюгин. - М.: Московский центр непрерывного математического образования (МЦНМО), 2018. - 844 с.
2. Анализ данных и процессов. - М.: БХВ-Петербург, 2009. - 512 с.
3. Афанасьев, В.Н. Анализ временных рядов и прогнозирование: Учебник / В.Н. Афанасьев, М.М. Юзбашев. - М.: Финансы и статистика, Инфра-М, 2012. - 320 с.
4. Петрухин, А. В. (2021). "Основы машинного обучения: теории и практики". Научно-издательский центр "Логос". — Введение в теории машинного обучения и его практические приложения.
5. Бачурин, А.А. Планирование и прогнозирование деятельности автотранспортных организаций / А.А. Бачурин. - Вологда: Инфра-Инженерия, 2011. - 272 с.
6. Бишоп, К. М. (2006). Распознавание образов и машинное обучение. — Обширное введение в методы машинного обучения и статистического анализа.
7. Алехин, А. И. (2018). Машинное обучение и его применение в экономике и бизнесе. ГЭОТАР-Медиа. — Описание основных методов машинного обучения и их применения в экономике.
8. Тихомиров, Н. П. Демография. Методы анализа и прогнозирования / Н.П. Тихомиров. - М.: Экзамен, 2013. - 256 с.

References

1. Vyugin, V.V. Mathematical foundations of machine learning and forecasting / V.V. Vyugin. - Moscow: Moscow Center for Continuing Mathematical Education (MCNMO), 2018. - 844 p.
 2. Analysis of data and processes. - M.: BHV-Petersburg, 2009. - 512 p.
 3. Afanasyev, V.N. Time series analysis and forecasting: Textbook / V.N. Afanasyev, M.M. Yuzbashev. - M.: Finance and Statistics, Infra-M, 2012. - 320 p.
 4. Petrukhin, A.V. (2021). "Fundamentals of machine learning: theory and practice". Scientific publishing center "Logos". — Introduction to the theory of machine learning and its practical applications.
 5. Bachurin, A.A. Planning and forecasting of the activities of motor transport organizations / A.A. Bachurin. Vologda: Infra-Engineering, 2011. - 272 p.
 6. Bishop, K. M. (2006). Pattern recognition and machine learning. — An extensive introduction to machine learning and statistical analysis methods.
 7. Alekhine, A. I. (2018). Machine learning and its application in economics and business. GEOTAR Media. — Description of the main methods of machine learning and their application in economics.
 8. Tikhomirov, N. P. Demography. Methods of analysis and forecasting / N.P. Tikhomirov. - M.: Exam, 2013. - 256 p.
-