

Гайдуков А.И.

Научный руководитель: Лучинин З. С., канд. техн. наук
Поволжский государственный технологический университет

Модуль контролируемого овербукинга на основе машинного обучения.

Аннотация. В данной статье рассматриваются способы применения современных алгоритмов машинного обучения для прогнозирования количества отмен в гостиничной индустрии.

Ключевые слова. Овербукинг, гостиничный бизнес, машинное обучение

Введение

Для успешного ведения бизнеса отели часто применяют стратегию овербукинга: оформление бронирований свыше имеющегося номерного фонда. Причиной этому является высокая вероятность отмены бронирования к моменту заезда гостя в номер. Применение стратегии овербукинга позволяет компенсировать финансовые потери от возможных отмен бронирований, но, в свою очередь, такой подход может привести к ситуации, когда отелю не удалось точно предугадать точное число гостей, которые захотят отменить свой заказ. В таком случае отелю приходится терпеть репутационные и финансовые убытки.

Решением данной проблемы могут являться новые технологии на основе машинного обучения. С их помощью можно реализовать механику прогнозирования отмены бронирования и существенно увеличить эффективность стратегии овербукинга.[1]. Целью данного исследования является создание модели прогнозирования отмены бронирования.

1. Алгоритмы на основе деревьев принятий решений

В области машинного обучения, задачи управляемого прогностического моделирования обычно делятся на два типа задач [2]: регрессия, когда измерение результата является количественным (например, прогнозирование процента отмены бронирований от общего количества бронирований), или как классификация, когда результат является классом/категорией (например, прогнозирование, “отменит” или “не отменит” конкретное бронирование).[3] В рамках данного исследования было принято решение рассматривать проблему прогнозирования отмены бронирования как задачу классификации.

Для реализации классификационных моделей машинного обучения можно использовать алгоритмы, основанные на деревьях принятий решений. Данные алгоритмы извлекают шаблоны из заданного набора данных путем формулирования правил [4]. Эти правила ищут функцию, которая наилучшим образом разделяет классы и данные в соответствии со значениями этой функции [6]. Этот процесс повторяется с новыми подмножествами, так что обучающий набор рекурсивно разбивается на более мелкие подмножества, чтобы найти разделы, содержащие только один класс [5]. Как только процесс завершен, результат может быть представлен в виде древовидной структуры, в которой объекты отображаются в виде узлов, соединенных ветвями, которые представляют его возможные значения.

Таким образом, деревья принятий решений позволяют переводить знания, извлеченные из обучающих данных, в графический инструмент, который может быть легко интерпретирован людьми, в отличие от других методов обработки больших данных, таких как искусственные нейронные сети или SVM, которые действуют как черный ящик.

Одним из наиболее популярных алгоритмов дерева решений является C5.0, обновленная версия предыдущего алгоритма C4.5, который работает быстрее и требует меньшего использования памяти за счет использования

меньшего количества правил [7]. Модель C5.0 применяет максимальную полученную информацию в качестве критериев разделения, которая основана на уменьшении энтропии после разделения набора данных по атрибутам [8]. После построения дерева решений проходит этап последующей обрезки, это означает, что те разделы, которые не предоставляют соответствующей информации, удаляются из древовидной структуры, что снижает сложность модели, а также решает проблемы с переобучением. Таким образом, деревья принятых решений отлично подходят для реализации модели прогнозирования отмены.

2. Модель машинного обучения

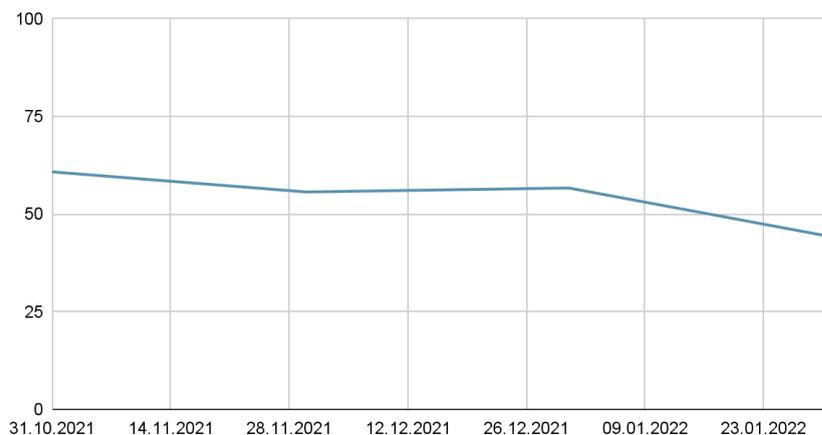
Возможность интеграции с автоматизированными системами управления отелем (АСУ) позволяет получить практически полную информацию о бронировании, что позволяет спроектировать максимально детализированную модель прогнозирования отмены. В рамках данного исследования была спроектирована модель машинного обучения со следующими параметрами:

- **BookingDuration** - длительность бронирования. Вероятности отмены довольно сильно зависят от длительности проживания гостя. Зачастую проживания на один или два дня гораздо чаще отменяются чем более продолжительная бронь. Но, стоит заметить, что такие брони получают меньше изменений, а следовательно меньше влияют на перерасчет прогноза отмены.
- **CheckInDate** - дата заезда бронирования. Данный параметр позволяет отслеживать сезонность бронирования, так как нагрузка у отелей на каждое время года разная, а следовательно и разные показатели отмены.
- **CheckOutDate** - дата выезда бронирования. Данный параметр аналогично параметру **CheckInDate** необходим для учета сезонности бронирований.
- **CancellationReason** - причина отмены бронирования. Один из самых важных параметров, агрегирует все остальные параметры модели, чтобы определить вероятность отмены, а также позволяет учесть нетипичные случаи отмены, например, Covid-19.
- **TotalAmount** - полная стоимость брони с учетом всех включенных услуг и трансферов. Позволяет ввести зависимость отмены от стоимости бронирования в сочетании с остальными параметрами.
- **AdultsAmount** - количество взрослых, проживающих в данном бронировании. Позволяет ввести зависимость от заполняемости номера гостями за одно проживание.
- **ChildrenAmount** - количество детей, проживающих в данном бронировании. Зачастую брони созданные семьей гораздо реже отменяются. Данный параметр позволяет учесть данную особенность.
- **AmountDaysBeforeCancellation** - количество дней до отмены бронирования. Данный параметр позволит динамически отслеживать вероятность отмены текущего бронирования с течением времени.
- **RoomTypeId** - идентификатор тарифного плана отмененной брони. Так как для каждого отеля формирование модели прогнозирования будет индивидуально, то можно сделать зависимость от идентификатора некоторых сущностей, например, тарифный план. Это позволит игнорировать изменения в названии данных сущностей, а также избежать ошибок в обучении из-за одинаковых по имени, но разных по назначению сущностей.
- **RoomId** - идентификатор номера комнаты, на которую было совершено бронирование.

3. Применение алгоритма машинного обучения

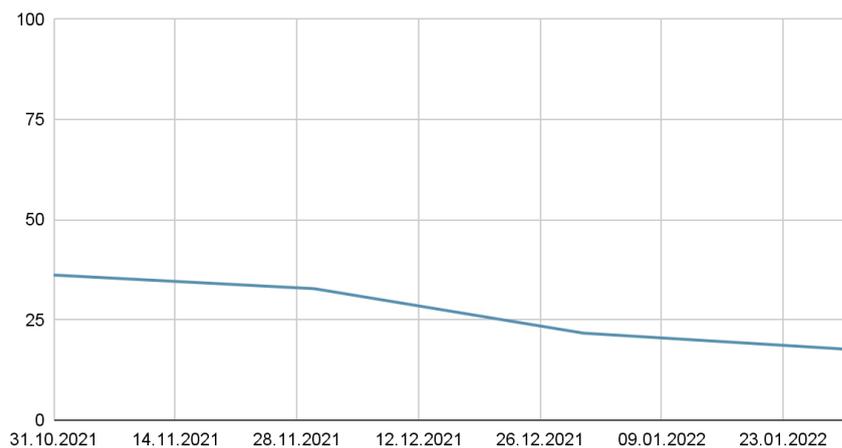
Для демонстрации эффективности модели прогнозирования отмены был разработан программный прототип модуля контролируемого овербукинга. В качестве примера рассмотрим реальные данные отеля по загруженности номеров за период с 01.10.2021 по 31.01.2022.

Загруженность номерного фонда



Как можно видеть из графика загрузки номерного фонда в среднем отелю не удастся реализовать от 40% до 55% номеров в зависимости от времени года, состояния номерного фонда и внешних различных факторов, например, отмена бронирования.

Процент отмененных номеров



На графике отмененных номеров можно видеть, что в среднем у отеля отменяется от 17% до 35% всего номерного фонда. Зная, эту информацию, отель мог бы продать номеров сверх имеющейся квоты и тем самым нивелировать финансовые потери.

Рассмотрим, как сильно изменились показатели отмены после применения прототипа с моделью прогнозирования.

Процент отменных номеров



Обученная модель позволила в среднем снизить процент нереализованных из-за отмененных бронирований номеров на 10% - 15% за счет дополнительной продажи номеров в овербукинг, что в свою очередь позволит существенно повысить доходы отеля.

Заключение

В результате проведенного исследования можно сделать вывод, что современные алгоритмы машинного обучения способны существенно повысить доходы отеля за счет прогнозирования возможных отмен бронирований. Опираясь на эту информацию отели могут существенно повысить эффективность стратегии овербукинга, а также свести к минимуму репутационные потери, связанные с невозможностью заселить гостя.

Целью дальнейшего исследования будет являться повышение точности прогноза количества отмененных бронирований у текущего прототипа за счет улучшения имеющейся модели, а также добавления в систему модели прогнозирования отмены конкретного бронирования. Также необходимо развить имеющийся прототип в полноценный модуль контролируемого овербукинга, представляющего из себя веб-приложение, и реализовать возможность интеграции его с автоматизированными системами управления отеля.

Библиографический список

1. An Automated Machine Learning Based Decision Support System to Predict Hotel Booking Cancellations. Nuno Antonio, Ana de Almeida and Luis Nunes.
2. Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning. Springer series in statistics Springer, Berlin.
3. Predicting hotel booking cancellations to decrease uncertainty and increase revenue.
4. Mingers, J., 1989a. An empirical comparison of pruning methods for decision tree induction. *Mach. Learn.* 4, 17.
5. Mingers, J., 1989b. An empirical comparison of selection measures for decision-tree induction. *Mach. Learn.* 3 (4), 319–342. <https://doi.org/10.1007/BF00116837>.
6. Minz, S., Jain, R., 2003. Rough set based decision tree model for classification. *Data Warehousing Knowledge Discov.* 2737, 172–181. https://doi.org/10.1007/978-3-540-45228-7_18.
7. Pandya, R., Pandya, J., 2015. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *Int. J. Comput. Appl.* 117 (16), 18–21. <https://doi.org/10.5120/20639-3318>.
8. Patil, N., Lathi, R., Chitre, V., 2012. Comparison of C5.0 & CART Classification algorithms using pruning technique. *Int. J. Eng. Res.* 1 (4), 6.