

Егошин Игорь Анатольевич

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

ТЕХНОЛОГИИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ПРОДАЖ В СФЕРЕ ГОСТИНИЧНОГО БИЗНЕСА

***Аннотация.** В данной статье рассматриваются применение технологии машинного обучения для прогнозирования продаж в сфере гостиничного бизнеса.*

***Ключевые слова:** машинное обучение, гостиничный бизнес, прогнозирование.*

Прогнозирование спроса номеров в отелях - ключевой аспект управления доходами в сфере гостиничного бизнеса. Прогнозирование продаж продолжает привлекать значительный и постоянный уровень инвестиций даже среди отелей с относительно зрелыми программами управления доходами, так как неточности прогнозирования ведут к ошибкам в ценообразовании, что негативно повлияет на финансовые показатели гостиницы.

Однако получение достаточно точных прогнозов является сложной задачей в силу ряда причин, таких как внутренняя изменчивость рынка, нехватка персонала, умеющего работать со статистикой и высокая стоимость специализированного программного обеспечения.

Целью работы является разработать модель прогнозирования, предсказывающую продажи номеров в гостинице на каждый день с довольно высокой точностью.

В сфере гостиничного и авиабизнеса уже имеются продукты, использующие технологии машинного обучения в прогнозировании спроса. Система управления доходами от компании "RmsCloud" из Австралии дает возможность формировать отчеты по будущим броням. "Accelya" - ведущий мировой поставщик технологических платформ, программного обеспечения и услуг для индустрии путешествий и транспорта. Их продукт позволяет оптимизировать доступность и цену продажи места в самолетах с помощью прогнозирования с использованием моделей, зависящих от прошлогодней истории продаж билетов и цен билетов у конкурентов.

Для высокой точности прогноза необходимо провести исследования существующих решений по прогнозированию продаж в гостиничном бизнесе и найти подходящие технологии машинного обучения. У рассмотренных приложений закрытая система, и они не раскрывают как реализовывали прогнозирование. Поэтому рассмотреть детали и преимущества похожих на мою работу решений не получится, поэтому рассмотрим некоторые из технологий машинного обучения и алгоритмов, пригодных для прогнозирования.

Линейная регрессия предполагает, что взаимосвязь между ROH (количество занятых размещений) на конкретном DBA (дней до прибытия) и окончательным числом занятых размещений является линейной.

Линейная регрессия может либо рассматривать актуальный ROH, либо может включать ROH по всей кривой бронирования.

У линейной регрессии есть свои недостатки. Для первого варианта, которая учитывает только актуальный ROH, не учитывается динамика изменений продаж номеров. Вместо того, чтобы использовать всю кривую бронирования для составления прогнозов, первая регрессия ближе к прогнозированию с

“точкой бронирования”. Хотя вторая регрессия решает эту проблему за счет включения каждого предсказания в кривую бронирования, она страдает от мультиколлинеарности (скопление ошибок), поскольку RON вычисляется накопительным образом и, таким образом, сильно коррелирует друг с другом. Поэтому коэффициенты второй модели следует интерпретировать с осторожностью.

Нейронные сети могут быть использованы для обучения прогнозированию продаж. Основная логика нейронной сети заключается в извлечении линейных комбинаций входных данных в качестве производных характеристик, а затем моделирование интересующих наблюдений как нелинейной функции подобранных характеристик. Таким образом, нейронную сеть можно рассматривать как многоступенчатую регрессию.

Нейронная сеть превратилась в большое разнообразие моделей и подходов к обучению, но чаще всего это относится к сети с обратным распространением ошибки со скрытыми слоями.

Создание нейронной сети обычно занимает два этапа - создание самой модели и ее обучение. Обычно в конце остается только один выход. Производные признаки генерируются из линейных комбинаций входных данных, а затем интересующее наблюдение моделируется как линейная комбинация функций.

Как и нейронная сеть “Случайный лес” может быть использован в задачах прогнозирования. Случайный лес - это метод классификации, регрессии и некоторых видов прогнозов. Модель на основе “Случайного леса” довольно быстро обучается и имеет высокую эффективность из-за того, что алгоритм создает комитет решающих деревьев, каждое из которых хоть и имеет невысокую точность, но за счет большого их количества результат получается хорошим.

Для каждой из рассмотренных технологий требуются входные данные для прогнозирования. В случае гостиничной сферы это могут быть данные по продажам гостиничных номеров, события, происходящие в городе, погода, цены конкурентов.

В рамках статьи был составлен прототип прогнозирования номеров в гостинице на основе данных по продажам за прошлые 2 года у гостиниц. В качестве технологии машинного обучения использован “Случайный лес”, так как из-за комитета решающих деревьев технология обеспечивает довольно высокую точность.

	CreationDate	Date	TotalSold	TotalAvailable
1	2017-07-10	2017-07-10	4	12
2	2017-10-18	2017-10-18	6	41
3	2017-10-19	2017-10-19	7	39
4	2017-10-20	2017-10-20	5	39
5	2017-10-21	2017-10-21	7	39
6	2017-10-23	2017-10-23	4	39
7	2017-10-24	2017-10-24	5	35
8	2017-10-24	2017-10-25	4	35
9	2017-10-24	2017-10-26	4	35
10	2017-10-25	2017-10-24	5	35
11	2017-10-25	2017-10-25	4	35
12	2017-10-25	2017-10-26	4	35
13	2017-10-26	2017-10-25	4	35
14	2017-10-26	2017-10-26	4	35
15	2017-10-27	2017-10-26	4	35
16	2017-10-28	2017-10-28	11	14
17	2017-10-28	2017-10-29	8	14

Таблица 1. Пример исходных данных по проданным номерам

Как видно по таблице 1, исходные данные состоят из количества забронированных номеров на дату (TotalSold) и самой даты (Date). CreationDate нужен для того, чтобы хранить всю историю изменений продаж в отеле. Количество всех номеров (TotalAvailable) позволит рассчитать загрузку отеля через отношение TotalSold и TotalAvailable.

Модель прогнозирования обучается на основе таких атрибутов как - дата бронирования, день недели, номер месяца и количество продаж. На вход модель принимает дату бронирования, а на выход выдает количество бронирований.

Date	TotalSoldForecast	TotalSold	TotalAvailable	Occupancy	OccupancyForecast
2022-01-01	7	7	11	63.64	63.636.36
2022-01-02	7	7	11	63.64	63.64
2022-01-03	10	10	11	90.91	90.91
2022-01-04	11	11	11	100.00	100.00
2022-01-05	11	11	11	100.00	100.00
2022-01-06	9	10	11	90.91	81.82
2022-01-07	12	11	11	100.00	109.09
2022-01-08	9	8	11	72.73	81.82
2022-01-09	5	5	11	45.45	45.45
2022-01-10	5	4	11	36.36	45.45
2022-01-11	5	4	11	36.36	45.45
2022-01-12	3	4	11	36.36	27.27
2022-01-13	3	2	11	18.18	27.27
2022-01-14	5	6	11	54.55	45.45
2022-01-15	6	6	11	54.55	54.55
2022-01-16	4	4	11	36.36	36.36

Таблица 2. Пример исходных с сравнением действительных и спрогнозированных продаж

Модель была протестирована на основе данных из нескольких гостиниц. Прогнозирование было сделано 31 января 2021 года. Пример результата можно увидеть на таблице 2.

Заключение

Отель	Среднее отклонение загрузки в процентах			
	За 4 дня	За 10 дней	За 18 дней	За 30 дней
Отель 1	5.55	8	12.2	17.46
Отель 2	6.89	6.55	7.66	11.26
Отель 3	0	3.63	5.55	8.48

Таблица 3. Статистика по среднему отклонению загрузки по нескольким отелям

На основе спрогнозированных и действительных данных по продажам была составлена статистика по трем отелям по среднему отклонению загрузки в рамках одного месяца (Таблица 3). По статистике мы можем сказать, что чем больше дней до даты заезда, тем больше процент отклонения прогноза от действительных данных. Результат является вполне логичным, так как чем больше разница между прогнозируемым днем и текущим днем, тем больше ошибка прогноза. Также можно заметить, что Отель 3 имеет довольно низкое отклонение даже за 30 дней до даты заезда. Результат по этому отелю оказался таким точным из-за того, что тенденция по изменению спроса за текущий год у этого отеля такая же как за последние 2 года и другие признаки такие как загруженность конкурентов или праздники не так сильно повлияли на этот спрос.

Таким образом модель прогнозирования имеет неплохую точность в рамках 15 дней до даты заезда, но ее можно улучшить, добавив другие признаки, влияющие на спрос отеля.

Библиографический список

1. Toby Segaran. 2007. Programming Collective Intelligence: Building Smart Web 2.0 Applications.
2. Chris Smith, Mark Koning. 2017. Decision Trees and Random Forests: A Visual Introduction For Beginners: A Simple Guide to Machine Learning with Decision Trees Kindle Edition.
3. David K. Hayes. 2010. Revenue Management for the Hospitality Industry 1st Edition.