

Левченко К.К.

Министерство иностранных дел Российской Федерации
Россия, Сочи
e-mail: sochikos@yandex.ru

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СПЕЦИФИКАЦИЙ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ОТМЕН БРОНИРОВАНИЯ В ГОСТИНИЧНОМ БИЗНЕСЕ

Аннотация: В данном исследовании проанализированы спецификации моделей прогнозирования отмены бронирования в гостиничном бизнесе, что особенно актуально в настоящее время в связи с сокращением туристских потоков вследствие внутростанционного давления и глобальных трансформационных процессов. Исследования показывают, что наиболее результативными являются модели прогнозирования, основанные на нейронных сетях, а также. Проведен сравнительный анализ указанных моделей, целью которого является обоснование наиболее результативной по показателям точности количества отмененных бронирований, что позволит оптимизировать прогноз загруженности организации гостиничного бизнеса.

Ключевые слова: отельный бизнес, модели Random Forest, CatBoost и XGBoost, алгоритм, прогнозирование, отмена бронирования.

¹Levchenko K.

¹ Ministry of Foreign Affairs of the Russian Federation
Russia, Sochi
e-mail: sochikos@yandex.ru

Model specifications benchmarking forecasting hotel cancellations

Abstract: This study analyzed the specifications of hotel cancellation prediction models, which is especially important at present due to the reduction in tourist flows due to intra-station pressure and global transformation processes. Studies show that the most productive are prediction models based on neural networks, as well. A comparative analysis of these models was carried out, the purpose of which is to substantiate the most effective number of canceled bookings in terms of accuracy indicators, which will optimize the forecast of the workload of the hotel business organization.

Keywords: hotel business, Random Forest model, CatBoost and XGBoost, algorithm, prediction, kernel reservation.

Левченко К.К.

Ресей Федерациясының Сыртқы істер министрлігі
Россия, Сочи
e-mail: sochikos@yandex.ru

Модельдердің сипаттамаларын салыстырмалы талдау қонақ үй бизнесінде брондаудың күшін жоюды болжау

Аннотация: Бұл зерттеу қонақ үй бизнесінде брондауды жоюды болжау модельдерінің сипаттамаларын талдады, бұл қазіргі уақытта станция ішіндегі қысым мен жаһандық трансформациялық процестерге байланысты туристік ағындардың төмендеуіне байланысты өзекті болып табылады. Зерттеулер нейрондық желілерге негізделген болжау модельдері ең тиімді деп санайды. Көрсетілген модельдерге салыстырмалы талдау жүргізілді, оның мақсаты жойылған брондау санының дәлдігі көрсеткіштері бойынша ең нәтижелі негіздеу болып табылады, бұл қонақ үй бизнесін ұйымдастыру жүктемесінің болжамын оңтайландыруға мүмкіндік береді.

Түйін сөздер: қонақ үй бизнесі, Random Forest, catboost және XGBoost модельдері, алгоритм, болжау, брондауды болдырмау.

Неотъемлемой составляющей любого туристского путешествия является система бронирования номеров в средствах размещения, что предполагает рост востребованности туристских онлайн-

агентств, которые позволяют существенно сокращать затраты времени и средств при непосредственной коммуникации с представителями гостиничного сектора сферы туризма. Кроме того подобная система

предполагает отсутствие штрафов и комиссий в случае отмены брони. Тем не менее, в последние годы наблюдается рост отмен бронирования, вызванных обострением геополитической ситуации в России и переориентированием туристских потоков на внутренние направления. Возрастание числа отмен бронирований приводит к определенным трудностям в планировании загрузки отелей, что напрямую ведет к потере части прибыли. Современные средства размещения в целях минимизации подобных рисков вынуждены использовать более жесткие правила бронирования и овербукинга. Среди современных тенденций развития гостиничного бизнеса, помимо увеличения количества отмен, отмечаются также рост доли прямых бронирований и сокращение окна бронирования, то есть периода времени, когда гостиничное предприятие начинает получать бронирования на определенную дату.

Весной 2022 года с российского рынка ушли такие лидеры каналов продаж, как Booking, Expedia и Agoda. Уход данных онлайн-ресурсов не только ослабил конкуренцию для других интернет-площадок и агентств, но и предоставил возможность отелям повысить эффективность прямых каналов продаж. Поскольку многие туристы не владели информацией об аналогичных известных площадках для бронирования, они чаще стали бронировать через официальный сайт или отдел бронирования отеля, что и спровоцировало рост прямых продаж.

В настоящее время в гостиничном бизнесе многие гостиничные организации используют сервисы-посредники ОТА (Online Travel Agencies), теряя при этом при каждой продаже в среднем 15-20%, выплачивая комиссию ОТА. Расширение прямых каналов продаж также способствует росту доходности, а также позволяет отелю лучше прогнозировать спрос и предложение.

Тенденция к сокращению «окна бронирования» заключается в следующем: если раньше гости планировали свой отдых за несколько месяцев и более, то сейчас, во избежание непредвиденных ситуаций, бронирования могут совершаться за несколько дней до поездки и даже в последний

момент. В связи с этим усложняется прогнозирование дохода отеля, его загрузки и планирование ценовой политики. Вместе с сокращением «окна бронирования» возросло и количество отмен, так как бронирования, сделанные в последний момент, имеют, как правило, импульсивный характер, следовательно, гости могут изменить планы или найти альтернативный вариант размещения. Согласно исследованиям С. Chen, Z. Schwartz, P. Vargas [4], значительная доля отмен бронирований вызвана поиском наиболее предпочтительных для конкретных категорий туристов предложений от средств размещения даже в случае уже наличия брони. То есть турист совершает несколько бронирований, далее делает их отмену кроме одного, которое ему наиболее выгодно.

Отели также используют стратегии избыточного бронирования (овербукинга), чтобы сократить возможный упущенный доход от отмененных бронирований, но данная стратегия имеет много недостатков:

- необходимо постоянно отслеживать количество избыточных бронирований;

- не оправдываются ожидания гостей, которые приводят к плохой репутации отеля, как следствие: негативные отзывы, особенно в социальных сетях, на интернет-площадках для бронирования и снижение лояльности гостей, что может привести к потере их в будущем;

- без прогнозирования и анализа загрузки отеля избыточное бронирование становится слишком рискованным для использования в качестве долгосрочной стратегии;

- зачастую компенсация за овербукинг может оказаться дороже для отеля, так как требует заселения гостя в номер категории выше, чем было им оплачено;

- гости должны быть перенаправлены в другие отели, если в отеле уже нет мест;

- потенциальный риск отказа гостем в услугах отеля;

Появляется необходимость в альтернативных вариантах сокращения отмен бронирований, поскольку они оказывают существенное влияние на менеджмент отеля в туристском бизнесе.

На основе базы данных из системы управления отелем, а также факторов, в

наибольшей степени влияющих на отмену бронирований, сформирован алгоритм, которой позволяет анализировать данные и прогнозировать потенциальные отмены бронирований.

Исследуя варианты прогнозирования отмен бронирований в отелях с помощью алгоритмов машинного обучения, были отмечены три результативные модели для прогнозирования отмены бронирования, в числе которых «Random Forest», «XGBoost» и «CatBoost».

Для выбора конкретной прогностической модели для предприятий гостиничного бизнеса туристской территории проведем сравнительный анализ перечисленных моделей для определения алгоритма с наиболее высокой эффективностью.

1) Модель «Random Forest» (Случайный лес)

Модель «Случайный лес» – это универсальный алгоритм машинного обучения, который используется для решения множества разных задач. В основном, алгоритм применяется в проблемах построения регрессии и классификации.

Поскольку известно, что лес составлен из деревьев, и больше деревьев означает более устойчивый лес. По аналогии с реальным лесом, модель позволяет формировать дерево решений для выборок данных, а затем получать прогноз по каждому из них, выбирая лучшее решение, к примеру, посредством голосования. Несмотря на высокую надежность и достаточную точность данного метода, построение и практическая реализация занимает много времени.

2) Модель CatBoost - это библиотека машинного обучения с открытым исходным кодом, созданная российским поисковой системой Яндекс в 2017 году. По сути дела, CatBoost - это своеобразная платформа, основанная на симметрии дерева решений, которая реализует меньшее количество параметров, поддерживает категориальные переменные и обеспечивает высокую точность.

Преимуществом данной модели является высокая производительность, масштабированность, достаточно низкая трудоемкость. В качестве недостатка следует отметить не достаточную точность осуществления прогноза вследствие использования большого количества случайных величин.

3). Модель XGBoost (Extreme Gradient Boosting), является разновидностью предыдущей модели CatBoost, и представляет алгоритм машинного обучения, однако может работать с большими данными, используя при этом методы оптимизации решений.

CatBoost и XGBoost являются улучшенными вариантами в рамках алгоритма Gradient Boosting Decision Tree (Бустинг от англ. boosting -улучшение). Тем не менее, Яндекс CatBoost считается лучшим алгоритмом, чем XGBoost с точки зрения точности получаемых результатов.

После определения преимуществ, недостатков и особенностей моделей, следует провести сравнительный анализ по основным показателям качества алгоритмов машинного обучения. В таблице 1 приведены средние значения показателей для указанных моделей прогнозирования отмены бронирования.

Показатели	Random Forest	XGBoost	CatBoost
Accuracy	0,864	0,861	0,860
Precision	0,873	0,856	0,858
Recall	0,672	0,711	0,704
F-мера	0,785	0,789	0,782

Таблица 1. Сравнительный анализ показателей моделей
Источник: [1]

В целях визуализации полученные результаты оценок отразим в виде диаграммы (рисунок 1)

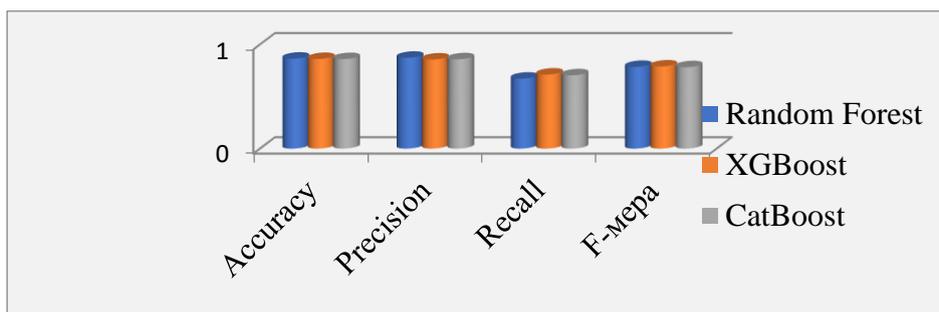


Рисунок 1. Сравнение моделей «Random» Forest, «XGBoost», «CatBoost» по основным показателям оценки

Accuracy - это показатель, который описывает общую точность прогноза модели по всем классам. В практическом прогнозировании в большей степени все же всего используют показатели точности (Precision) и полноты (Recall).

Точность (Precision) показывает количество реальных объектов класса среди всех тех, что классификатор отнес к данному классу, а полнота (Recall) показывает какую долю из общих объектов класса составляют найденные объекты. Поскольку в данном алгоритме может возникнуть риск противоречия, следует использовать усредненную метрику (F-мера), которая представляет собой среднее гармоническое из значений точности и полноты вводимых аналитических параметров. Наилучшей будет классификация, где F-мера соответствует наибольшему значению.

Согласно данным, все модели имеют приблизительно одинаковые результаты. В случае прогнозирования отмены бронирования отелей наилучшей моделью должна быть модель с наибольшей точностью (Precision), то есть отражать точное количество отмененных бронирований, что позволит оптимизировать прогноз загруженности гостиничной организации, внести коррективы в ценовые стратегии. Следовательно, наилучшей моделью исходя из точности прогноза будет модель

случайного леса (Random Forest), в которой имеется максимальное значение Precision и Accuracy.

Модель случайного леса (Random Forest) представляет из себя универсальный алгоритм прогнозирования отмены бронирования, в основе которого заложено применение набора решающих деревьев. Если рассматривать только одно дерево, то высокого качества получить проблематично, а в большом их количестве, то есть в кусте, результатом становится прогноз достаточно высокой точности. Преимуществом модели является также гибкость в решении разноплановых задач.

Порядок действий в алгоритме может быть следующим.

Исходной точкой является загрузка входных параметров методом случайной выборки. Затем в цифровом режиме происходит построение дерева решений. Построение дерева происходит до момента достижения каждым листом не более n объектов или же по достижении определенной высоты.

Следующий шаг – это получение определенного результата прогнозирования по каждому дереву решений, где происходит выбор прогнозируемого результата исходя из лучших признаков. В заключении выбирается результат прогноза с наибольшим количеством лучших параметров, что и будет являться окончательным результатом прогнозирования. Упрощенная схема работы алгоритма представлена на рисунке 2

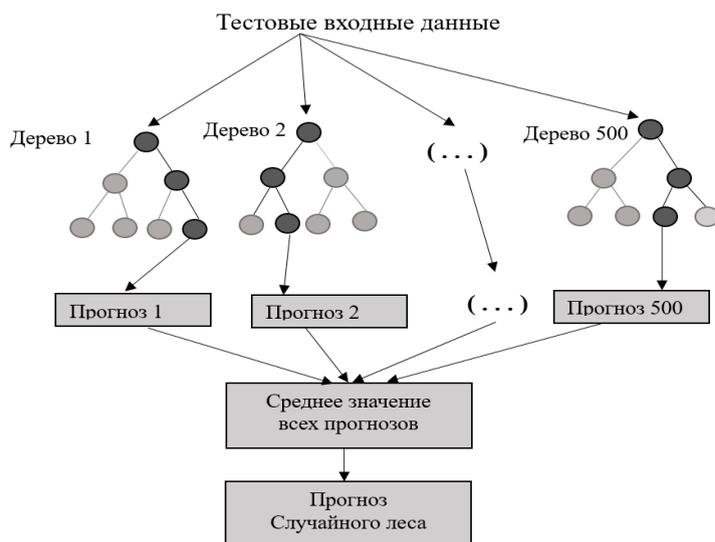


Рисунок 2. Упрощенная схема алгоритма «Случайный лес» (Модель «Random Forest»)

Далее в представленном алгоритме происходит подготовка базы данных.

Проверка корреляций является важной частью процесса анализа набора данных. Этот анализ является одним из методов, используемых для определения того, какие функции больше всего влияют на целевую переменную, и, в свою очередь, используются при прогнозировании этой целевой переменной. Вследствие этого, перед запуском прогностической модели для конкретной гостиничной организации необходимо проанализировать какие факторы влияют на отмену бронирований в большей степени.

На основе базы данных отеля «Сочи

Марриотт Красная Поляна» за 2022 год был проведен корреляционный анализ переменных. Данные собирались непосредственно из базы данных PMS отеля с использованием Microsoft SQL Server. Доля отменённых бронирований за 2022 год составила 15 028 номеров.

Необходимо также оценить структуру базы данных и особенности отеля перед извлечением данных. В таблицах 2 и 3 представлен список всех извлеченных переменных. Эти переменные были выбраны на основе изучения научной литературы по теме отмене бронирований в отелях, но они также учитывали объем данных, доступных в PMS современного отеля.

Показатели	Характеристика
lead_time	Количество дней, прошедших между датой ввода бронирования в PMS и датой прибытия
adr	Средняя дневная ставка, определяемая делением суммы всех транзакций по размещению на общее количество ночей проживания
adults	Число взрослых туристов, которые бронируют номер
children	Число детей (в бронируемом номере)
babies	Количество младенцев / маленьких детей (в бронируемом номере)
Previous bookings not canceled	Количество предыдущих бронирований, которые не были отменены гостем до текущего бронирования
Previous cancellations	Количество предыдущих бронирований, которые были отменены гостем до текущего бронирования
Required car parking spaces	Количество парковочных мест, которые требуются для гостя

Arrival date day of_month	День месяца даты прибытия
Arrival date week number	Номер недели даты прибытия
Arrival date year	Год даты прибытия
booking changes	Количество изменений / дополнений, внесенных в бронирование до момента заселения или отмены
Days in waiting list	Количество дней, в течение которых бронирование находилось в листе ожидания, прежде чем оно было подтверждено гостю.
Stays in weekend nights	Количество ночей в выходные (суббота или воскресенье), которые гость проживал в отеле или забронировал номер для проживания
Stays in week nights	Количество ночей в неделю (с понедельника по пятницу), в которые гость оставался в отеле или забронировал номер для проживания
Total of special requests	Количество особых запросов, сделанных гостем (например, высокий этаж или вид на горы)

Таблица 2. Числовые показатели, выбранные для анализа данных современного отеля

Источник: [5]

Параметры	Характеристика
Is canceled	Переменная, которая показывает, было отменено (1) бронирование или нет (0)
hotel	Тип отеля (Н1 - «Курортный отель» или Н2 -«Городской отель»)
agent	ID туристского агентства, через которое было оформлено бронирование
Arrival date month	Месяц прибытия – 12 значений (Категории: Январь, Февраль, Март и т. д.)
assigned_room_type	Код для типа номера, назначенного для бронирования
company	ID компании/юридического лица, совершившего бронирование или ответственного за оплату
country	Страна проживания гостя
Customer type	Тип бронирования, предполагающий одну из четырех категорий: -Contract; -Group – бронирование группы; -Transient – бронирование не является групповым или совершенным в рамках контракта и не связано с другим временным бронированием; -Transient-party – бронирование является временным, но связано как минимум с другим временным бронированием
Distribution channel	Канал «распространения» бронирования: «ТА» - «Туристические агенты» или «ТО» – «Туроператоры»
Is repeated guest	Значение, указывающее, были ли уже бронирования от этого гостя (1) или нет (0)
Market segment	Сегмент рынка: «ТА» – «Туристические агенты» или «ТО» – «Туроператоры»
meal	Тип забронированного питания: -Undefined/SC – нет определенного питания; BB – только завтрак; НВ – полупансион (завтрак и еще один прием пищи, обычно ужин); FB – полный пансион (завтрак, обед и ужин)
Reservation status	Последний статус бронирования, допускающий одну из трех категорий: Canceled – бронирование было отменено заказчиком; Check-Out – гость зарегистрировался, но уже уехал; No-Show – гость не прошел регистрацию и не проинформировал отель о причине
Reservation status date	Дата, когда был установлен последний статус. Эта переменная может использоваться вместе с reservation_status, чтобы установить, когда было отменено бронирование или, когда гость уехал из отеля
Reserved room type	Код типа забронированного номера

Таблица 3. Категориальные значения параметров, выбранные для анализа данных современного отеля

Источник: [7]

Для отражения всех количественных переменных возможно, построение корреляционной тепловой матрицы, которая показывает наличие прямой или обратной связи между переменной и возможной отменой бронирования.

Модель прогнозирования отмены бронирования в гостиничной организации не

следует использовать обособленно от других систем, более эффективным вариантом является внедрение модели в CRS отеля, это позволит иметь более точные результаты и прогнозы в целом. Использование модели в управлении отелем схематично представлено на рисунке 3.

Рисунок 3. Схематичное изображение использования прогностической модели в системе управления отелем.



Подключив модель прогнозирования напрямую к PMS, CRS может передавать в PMS обновленную информацию по наличию номеров к продаже. Эти данные далее могут быть переданы автоматически, напрямую или

через менеджера каналов, по различным каналам распределения (OTA, GDS, туристические операторы, веб-сайт отеля и т. д.). Эта автоматизация распределения продаж на основе более точного прогноза чистого спроса позволяет отелю немедленно реагировать в случае отмены бронирования или изменения классификации отмены бронирования, корректировать свои запасы для продажи и сообщать об этом по различным каналам сбыта.

При использовании этих моделей необходимо учитывать следующий момент: некоторые переменные-предикторы меняются со временем (например, «LeadTime»), могут

принимать новые значения каждый день, как в случае изменений/поправок к бронированиям (например, «Изменения бронирования» или «Взрослые»). Таким образом, модель следует запускать каждый день, чтобы все внутренние данные и результаты оценивались ежедневно.

Для контроля правильной работы модели в гостиницах и отелях наиболее целесообразным является решение нанять дополнительного сотрудника. Рекомендуется в качестве специалиста выбрать бизнес-аналитика или datascience-специалиста. К должностным обязанностям сотрудника будет относиться обеспечение бесперебойной работы модели, внесение изменений в переменные модели и базу данных, а также формирование на основе прогноза детальные отчеты для руководителя управлением доходов, который будет анализировать полученную информацию и принимать стратегические решения.

Даже самые точные и эффективные модели не остаются эффективными бесконечно. Например, когда отель меняет свои маркетинговые условия и начинает

завоевывать больше рынка у туристских онлайн-агентств, а не у традиционных туроператоров, это может повлиять на многие переменные-предикторы, такие как «Рыночный сегмент», «Канал дистрибуции» и «Время выполнения». Если модель не совершенствовать и не учитывать изменения и возникающие тенденции, ее производительность не останется на прежнем уровне.

Для создания модели отмены бронирования крайне важна качественная база данных и правильный выбор переменных. Как упоминалось ранее и отражено на рисунке, не все функции имеют одинаковый порядок важности, и они не одинаково влияют на прогнозирование отмены бронирования. Это требует выполнения подробного анализа в конкретном отеле. Расположение отеля, услуги, удобства, рынки и каналы сбыта – это те многие факторы, имеющие разный вес для прогнозирования отмены.

Таким образом, знание функций управления доходами отеля и общих знаний в области гостиничного бизнеса в условиях цифровых трансформаций становится недостаточно, чтобы сделать оптимальный выбор. Также важно понимать особенности работы и характеристики каждого отеля. Это может иметь значение с точки зрения окончательной производительности и адекватности модели.

Как и в случае с любой другой проблемой прогнозной аналитики, разработка модели для прогнозирования отмены бронирования требует, чтобы параметры соответствовали всем атрибутам качественных данных: точность, надежность, достоверность, актуальность. Низкое качество базы данных может повлиять на производительность модели. По этой причине отели, которые хотят создавать модели прогнозирования, должны обеспечить наличие политики качества данных.

Эти результаты подтверждают возможность выявления бронирований с высокой вероятностью отмены, что позволит менеджерам отелей принимать меры, чтобы избежать этих потенциальных отмен, таких как специальные тарифы, дополнительные

скидки и предложения. От создания и запуска данной модели прогнозирования можно получить еще большие выгоды. Запуская модель каждый день для всех внутренних бронирований, можно получить еще один важный результат: прогнозируемое количество ночей, которые будут отменены в каждый из следующих дней. Необходимо только добавить прогнозируемое количество бронирований, которые будут отменены за день. Имея более точную информацию о чистом спросе, менеджеры отелей могут разработать более эффективные политику избыточного бронирования и правила отмены бронирования для гостей, что приведет к снижению затрат и снижению риска.

Список литературы:

- 1 Соколов Е.А. Градиентный бустинг. URL: <https://github.com/esokolov/ml-course-hse/blob/master/2020-fall/lecture-notes/lecture-ensembles.pdf> (дата обращения: 15.03.2023).
2. Smith S.J., Parsa H.G., Bujisic M., van der Rest J-P. Hotel cancelation policies, distributive and procedural fairness, and consumer patronage: A study of the lodging industry // Journal of Travel and Tourism Marketing. 2015. № 32 (7). P. 886–906.
3. Talluri K.T., van Ryzin G.J. The theory and practice of revenue management. NY: Kluwer Academic Publishers. 2004. 745 p.
4. Chen C.-C., Schwartz Z., Vargas P. The search for the best deal: How hotel cancellation policies affect the search and booking decisions of deal-seeking customers // International Journal of Hospitality Management. 2011. № 30 (1). P. 129–135.
5. Huang H.-C., Chang A. Y., Ho C.-C. Using artificial neural networks to establish a customer cancellation prediction model // Przegląd Elektrotechniczny. 2013. № 89 (1b). P. 178–180.
6. Yoon M.G., Lee H.Y., Song Y.S. Linear approximation approach for a stochastic seat allocation problem with cancellation and refund policy in airlines // Journal of Air Transport Management. 2012. № 23. P. 41–46.
7. Wingen M. EDA of bookings and ML to predict cancellations. URL: <https://www.kaggle.com/marcuswingen/eda-of-bookings-and-ml-to-predict-cancelations> (дата обращения: 16.02.2023).

References

- 1.Sokolov E.A. Gradient boosting. URL: <https://github.com/esokolov/ml-course-hse/blob/master/2020-fall/lecture-notes/lecture-ensembles.pdf> (дата обращения: 15.03.2023).
2. Smith S.J., Parsa H.G., Bujisic M., van der Rest J-P. Hotel cancelation policies, distributive and procedural fairness, and consumer patronage: A study of the lodging industry // Journal of Travel and Tourism Marketing. 2015. № 32 (7). P. 886–906.
3. Talluri K.T., van Ryzin G.J. The theory and practice of revenue management. NY: Kluwer Academic Publishers. 2004. 745 p.
4. Chen C.-C., Schwartz Z., Vargas P. The search for the best deal: How hotel cancellation policies affect the search and booking decisions of deal-seeking customers // International Journal of Hospitality Management. 2011. № 30 (1). P. 129–135.
5. Huang H.-C., Chang A. Y., Ho C.-C. Using artificial neural networks to establish a customer cancellation prediction model // Przegląd Elektrotechniczny. 2013. № 89 (1b). P. 178–180.
6. Yoon M.G., Lee H.Y., Song Y.S. Linear approximation approach for a stochastic seat allocation problem with cancellation and refund policy in airlines // Journal of Air Transport Management. 2012. № 23. P. 41–46.
7. Wingen M. EDA of bookings and ML to predict cancellations. URL: <https://www.kaggle.com/marcuswingen/eda-of-bookings-and-ml-to-predict-cancelations> (дата обращения: 16.02.2023).

Сведения об авторе

Левченко Константин Константинович

должность: кандидат экономических наук, заместитель руководителя территориального органа Министерства иностранных дел Российской Федерации г. Сочи, Россия

Почтовый адрес: 354000, Россия, г. Сочи, ул. Советская 42/2

Моб. тел: +7 918 408 45 81

Konstantin K. Levchenko

Position: Candidate of Economic Sciences, Deputy Head of the Territorial Body of the Ministry of Foreign Affairs of the Russian Federation, Sochi, Russia

Mailing address: 354000, Russia, Sochi, st. Sovetskaya 42/2

Mob. phone: +7 918 408 45 81