

ИЗУЧЕНИЕ ОПЫТА ПРОГНОЗИРОВАНИЯ ТУРИСТСКИХ ПОТОКОВ С ПРИМЕНЕНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

С.А. Лочан, Е.Л. Золотарева, Д.И. Коровин, Д.В. Федюнин

Сергей Александрович Лочан* (ORCID 0000-0001-8704-7538), Екатерина Леоновна Золотарева (ORCID 0000-0002-1516-7378), Коровин Дмитрий Игоревич (ORCID 0000-0001-9941-0322)
Финансовый университет при Правительстве Российской Федерации, Ленинградский просп., 49, Москва, 125993, Россия
E-mail: SALochan@fa.ru*, ELZolotareva@fa.ru, DIKorovin@fa.ru

Дмитрий Валерьевич Федюнин (ORCID 0000-0002-8526-3322)
Российский экономический университет им. Г.В. Плеханова, Стремянный пер., 36, Москва, 117997, Россия
E-mail: Fedyunin.DV@rea.ru

В статье для изучения российского и международного опыта прогнозирования туристских потоков с применением алгоритмов машинного обучения была проанализирована поисковая выдача трех крупнейших баз цитирования научной литературы - ScienceDirect (полнотекстовая база данных: доступ к электронным журналам издательства Elsevier), Mendeley (интегрированная с научными платформами программа для управления библиографической информацией, позволяющая хранить и просматривать исследовательские работы) и Elibrary (научная электронная библиотека). В статье отмечено, что в базах ScienceDirect и Mendeley, в основном, обрабатывают научные публикации в англоязычных изданиях, в то время как основной массив записей Elibrary, напротив, составляют публикации на русском языке. В статье указано, что на фоне большого количества публикаций в международных научных изданиях, российский опыт использования машинного обучения для прогнозирования туристических потоков выглядит очень скромно. Одной из основных причин, вероятно, является ограниченный доступ к статистическим данным, необходимым для построения модели машинного обучения. В связи с этим, исследовательская активность в этой сфере концентрируется не в научных кругах, а в корпоративных - внутри специализированных подразделений в компаниях банковского или информационно-телекоммуникационного сектора, владеющих данными о геолокации пользователей, или непосредственно в объектах туристической отрасли. Так, аналитика передвижения населения, в том числе с туристическими целями, ведется в Яндекс, Mail.ru (Profi.Travel), МТС, Билайне, Мегафоне, Сбербанке. Исследования, в основном, носят коммерческий характер и потому отсутствуют в открытом доступе.

Ключевые слова: цифровые технологии, туристский рынок, туристские потоки, моделирование, цифровая экономика, туристская инфраструктура, оптимизационные модели, датасеты, устойчивое развитие, эконометрические модели, дестинация, эффективность.

STUDY OF EXPERIENCE IN FORECASTING TOURIST FLOWS USING MACHINE LEARNING ALGORITHMS

S.A. Lochan, E.L. Zolotareva, D.I. Korovin, D.V. Fedyunin

Sergey A. Lochan (ORCID 0000-0001-8704-7538), Ekaterina L. Zolotareva (ORCID 0000-0002-1516-7378), Dmitry I. Korovin (ORCID 0000-0001-9941-0322)
Financial University under the Government of the Russian Federation, Leningradsky Ave., 49, Moscow, 125993, Russia

Dmitry V. Fedyunin (ORCID 0000-0002-8526-3322)
Plekhanov Russian University of Economics, Stremyanny Lane, 36, Moscow, 117997, Russia
E-mail: Fedyunin.DV@rea.ru

In order to study Russian and international experience in predicting tourist flows using machine learning algorithms, the article analyzed the search output of three major scientific literature citation databases - ScienceDirect (a full-text database: access to electronic journals of Elsevier Publishing), Mendeley (bibliographic information management software integrated with scientific platforms, allowing to store and browse research papers) and Elibrary (a scientific digital library). The article notes that the databases ScienceDirect and Mendeley mainly process scientific publications in English-language editions, while the bulk of the Elibrary records, on the contrary, are publications in Russian. The article points out that against the background of a large number of publications in international scientific journals, Russian experience in using machine learning for forecasting tourist flows looks very modest. One of the main reasons is probably the limited access to statistical data necessary to build a machine learning model. In this regard, research activity in this area is concentrated not in academia, but in corporate - within specialized divisions in companies in the banking or information-telecommunication sector that own user geolocation data, or directly in tourism industry sites. For example, Yandex, Mail.ru (Profi.Travel), MTS, Beeline, Megafon and Sberbank analyze the movement of people, including those travelling for tourism purposes. Studies are mostly commercial in nature and therefore not publicly available.

Keywords: digital technologies, tourist market, tourist flows, modeling, digital economy, tourist infrastructure, optimization models, datasets, sustainable development, econometric models, destination, efficiency.

Для цитирования:

Лочан С.А., Золотарева Е.Л., Коровин Д.И., Федюнин Д.В. Изучение опыта прогнозирования туристских потоков с применением алгоритмов машинного обучения. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2021. № 04(50). С.145-155. DOI: 10.6060/ivecofin.2021504.577

For citation:

Lochan S.A., Zolotareva E.L., Korovin D.I., Fedyunin D.V. Study of experience in forecasting tourist flows using machine learning algorithms. *Ivecofin*. 2021. № 04(50). С.145-155. DOI: 10.6060/ivecofin.2021504.577 (in Russian)

ВВЕДЕНИЕ

В рамках проводимого в Финансовом Университете при Правительстве Российской Федерации исследования, посвященному моделированию туристических потоков внутри России был проведен анализ опубликованных результатов в этой сфере. Для этого были изучены различные базы публикаций.

Для отбора публикаций в системе ScienceDirect использовался следующий поисковый запрос ("machine learning" OR "neural networks" OR "ensembles") AND ("tourism" OR "tourist" OR "passenger") AND ("prediction" OR "forecasting" OR "modeling") с ограничением по году издания – не раньше 2019 г. Объем поисковой выдачи – 5793 записи. В результате анализа первых 100 наиболее релевантных публикаций было отобрано 38 для дальнейшего изучения.

Выборка в системе Mendeley производилась по несколько иному набору ключевых слов (machine learning tourism) и с расширенным временным диапазоном – последние 5 лет.

Также было проанализировано 100 наиболее релевантных публикаций (общий

объем выдачи -1035), из которых были отобраны наиболее подходящие.

Сформированный в итоге массив публикаций для анализа международного опыта прогнозирования туристских потоков с применением алгоритмов машинного обучения насчитывает порядка 70 научных статей, из которых только 6 являются обзорными [1-6].

Для удобства анализа материалы были объединены в базу данных с графическим пользовательским интерфейсом, которая впоследствии может повторно использоваться для быстрого доступа к наиболее интересным материалам, просмотра заметок и комментариев. Поисковая выдача по ключевым словам “машинное обучение туризм” в библиотеке Elibrary содержит только 7 позиций [7-13], что наглядно демонстрирует различие в уровне научной активности в этой отрасли.

Большинство найденных в системе Elibrary.ru статей посвящено перспективам использования современных технологий в туристической отрасли и обсуждению основных трендов. Непосредственно разработке модели машинного обучения посвящены только две статьи, однако и в них не ставится задача прогнозирования туристского потока.

В статье «Европейский туризм в период пандемии коронавирусной инфекции» машинное обучение используется для кластеризации и классификации стран, обладающих схожими характеристиками. В статье «Интеллектуальная система помощи туристу: сервис-ориентированная архитектура и реализация» рассматривается построение интеллектуальной системы помощи туристу. В ее основе лежит рекомендательный сервис, который подбирает достопримечательности согласно пользовательским предпочтениям и строит персонализированный маршрут. Данное исследование является частью проектов, финансируемых Российским фондом фундаментальных исследований и федеральным бюджетом. Примечательно, однако, что в списке работа, на которые автор дает ссылки в обзорной части статьи, отсутствуют отечественные публикации, а источниками данных о достопримечательностях города Санкт-Петербург послужили сведения из Wikipedia, OpenStreetMap и Google Places. Это является еще одним подтверждением вывода о дефиците официальных статистических данных и, как следствие, о практическом отсутствии отечественных научных исследований в области прогнозирования туристского потока с применением алгоритмов машинного обучения.

ОБЗОР МЕЖДУНАРОДНОГО ОПЫТА

Принимая во внимание обозначенные выше проблемы, обзор научных исследований подготовлен на основе анализа публикаций в англоязычных изданиях, что соответствует международному опыту в данном направлении. Можно отметить, что количество публикаций увеличивается с каждым годом (рис. 1), что, в целом, соответствует росту интереса к технологиям искусственного интеллекта в обществе. При этом следует учитывать, что на момент отбора публикаций (15.07.2021) данные по 2021 г. были еще неполными.

Наиболее «продуктивными» журналами в анализируемой области являются журналы *Annals of Tourism Research* и *Tourism Management* (табл. 1), на долю которых приходится 25% от общего количества отобранных статей (12 и 8 соответственно). Кроме того, журнал *Annals of Tourism Research* в 2019 г. запустил цикл статей (*Curated Collection*) на тему прогнозирования туристского спроса.

Оба журнала относятся к первой квантили (Q1) по версии *Scimago Journal and Country Rank*, таким образом, можно предположить, что все наиболее успешные исследования собраны в этих изданиях [14].

Изменение количества статей, посвященных применению машинного обучения в области туризма

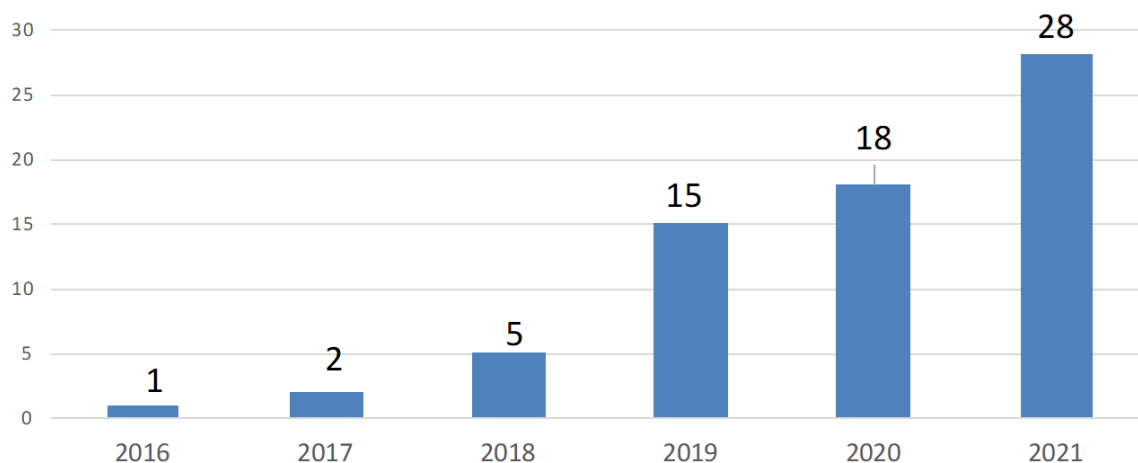


Рисунок 1. Динамика количества публикаций по годам
Figure 1. Dynamics of the number of publications by year

ТЕМАТИКИ ПРИМЕНЕНИЯ МАШИННОГО ОБУЧЕНИЯ

Несмотря на то, что прогнозирование туристского потока (спроса) является основным направлением использования алгоритмов машинного обучения (на него приходится почти половина отобранных статей, 45%), можно выделить

еще ряд задач в сфере туризма, где этот инструмент также является востребованным. В частности, это анализ пользовательских предпочтений и построение рекомендательных систем (около 28% статей), а также моделирование передвижений внутри отдельных локаций (12%) (табл. 2).

Таблица 1. Динамика публикаций в ведущих журналах
Table 1. Dynamics of publications in leading journals

Название журнала	Года			Всего
	2019	2020	2021	
Annals of Tourism Research	3	6	3	12
Tourism Management	4	1	5	8
Всего	7	7	8	20

Таблица 2. Тематики применения машинного обучения в сфере туризма
Table 2. Topics of machine learning applications in tourism

Название тематики	Количество статей
Прогнозирование туристских потоков (спроса)	36
Анализ предпочтений, рекомендательные системы	19
Анализ и моделирование передвижений внутри локации (города)	8
Прочие тематики	6
Всего	69

ПРОГНОЗИРОВАНИЕ ТУРИСТСКИХ ПОТОКОВ (СПРОСА)

Исследования, посвященные прогнозированию туристских потоков, приведены в табл. 3. Среди них также можно выделить несколько подкатегорий.

Так, подавляющее число исследований касается моделирования туристских прибытий в ту или иную дестинацию. При этом следует отметить, что в основном моделируются прибытия в регион, а не передвижения между регионами одной страны, что, очевидно, требует иных требований к сбору данных. В лучшем случае анализируется структура въездного туризма по международным прибытиям в Китай [15] и специальный административный район Гонконг [16], Тайвань [17], Сингапур [18], США [19, 20], Грецию [21] и крупные европейские страны [22] и между аэропортами Азиатско-Тихоокеанского и Латинско-американско-Карибского региона [23].

Также важной особенностью является то, что для целей прогнозирования чаще выбираются

локальные дестинации (например, город, национальный парк), а не регион в целом. Больше всего таких исследований проводится в Китае, однако есть работы, посвященные точкам притяжения в США и Италии. Фокус на локальных дестинациях объясняется тем, что для них легче отобрать релевантные признаки, в частности, ключевые слова в поисковых запросах.

Отдельные исследования туристского потока посвящены предсказанию объема пассажирского потока в аэропорты. Среди других малочисленных подкатегорий – предсказание наполняемости отелей, прогнозирование овертуризма, т.е. негативных последствий массового туризма [24, 25] (оба исследования посвящены Испании) и прогнозирование спроса на круизы в Китае.

Анализ предпочтений, рекомендательные системы

Анализ пользовательских предпочтений предполагает изучение отклика посетителей туристских объектов и сегментацию туристов по каким-либо схожим характеристикам. Этому посвящены исследования [26, 27, 28]. Логическим продолжением анализа структуры туристского потока является построение рекомендательных систем, предлагающих пользователю наиболее подходящие для него точки притяжения, отели, рестораны и прочее. Так, примерно в половине статей из этой тематики, исследование пользовательских предпочтений было предназначено именно для построения рекомендательного сервиса [29, 30]. С точки зрения механизмов моделирования, то исследования в этой области преимущественно основываются на анализе текстового содержания и эмоциональной окраски онлайн-отзывов. Лишь для одного из исследований, посвященных анализу удовлетворенности туристов в Чили и Эквадоре, использовались специально разработанные опросные листы [31].

АНАЛИЗ И МОДЕЛИРОВАНИЕ ПЕРЕДВИЖЕНИЙ ВНУТРИ ЛОКАЦИИ

Еще одно направления применения машинного обучения в туристической сфере – анализ и моделирование передвижений внутри определенной локации, преимущественно внутри городов. Лишь две статьи из этой категории посвящены исследованию передвижений пользователей внутри национального парка. Целью анализа, как правило, является повышения качества оказания туристских услуг, например, выбора оптимального способа передвижения или маршрута, а также распределение нагрузки на городскую транспортную систему.

Таблица 3. Подборка статей по тематике прогнозирования туристского потока
Table 3. A selection of articles about forecasting tourist flow

Категория	Название статьи
Прогнозирование туристских прибытий	A decomposition-ensemble approach for tourism forecasting
	A Neural network enhanced hidden Markov model for tourism demand forecasting
	A paired neural network model for tourist arrival forecasting
	Air passenger forecasting using Neural Granger causal Google trend queries
	Baidu index-based forecast of daily tourist arrivals through rescaled range analysis, support vector regression, and autoregressive integrated moving average
	Bayesian BILSTM approach for tourism demand forecasting
	Daily tourism volume forecasting for tourist attractions
	Forecasting campground demand in US national parks
	Forecasting international tourism demand: a local spatiotemporal model
	Forecasting the number of inbound tourists with Google Trends
	Forecasting tourism demand with denoised neural networks
	Forecasting tourism demand with multisource big data
	Forecasting tourist arrivals using denoising and potential factors
	Forecasting Tourist Arrivals via Random Forest and Long Short-term Memory
	Forecasting tourist arrivals with machine learning and internet search index
	Geo-tagged photo metadata processing method for Beijing inbound tourism flow
	Hierarchical pattern recognition for tourism demand forecasting
	Improving Tourist Arrival Prediction: A Big Data and Artificial Neural Network Approach
	Intelligent forecasting of inbound tourist arrivals by social networking analysis
	International tourism demand forecasting with machine learning models: The power of the number of lagged inputs
	Mode decomposition method integrating mode reconstruction, feature extraction, and ELM for tourist arrival forecasting
	Modelling tourism demand to Spain with machine learning techniques. The impact of forecast horizon on model selection
	Multi-attraction, hourly tourism demand forecasting
	Review of tourism forecasting research with internet data
Stacked autoencoder with echo-state regression for tourism demand forecasting using search query data	
Tourism demand forecasting with time series imaging: A deep learning model	
Tourism demand forecasting: A deep learning approach	
Weight-constrained neural networks in forecasting tourist volumes: A case study	
Прогнозирование пассажиропотока в аэропортах	Forecasting air passenger demand with a new hybrid ensemble approach
	Forecasting air passenger numbers with a GVAR model
	Nonlinear vector auto-regression neural network for forecasting air passenger flow
Прогнозирование спроса на круизы	Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach
Прогнозирование овертуризма	Competitiveness and overtourism: A proposal for an early warning system in spanish urban destinations
	Machine learning techniques as a tool for predicting overtourism: The case of Spain
Прогнозирование наполняемости отелей	Big data from dynamic pricing: A smart approach to tourism demand forecasting
	Machine Learning in Internet Search Query Selection for Tourism Forecasting

Прочие тематики

Прочие тематики включают в себя анализ структуры туристического потока в Южной Корее, прогнозированию онлайн-туризма на примере китайского сервиса, предсказания отмены зарезервированного отеля, исследования ценообразования на примере сервиса Airbnb в Испании, поведенче-

скому анализу туристов из Индии, а также использованию механизмов машинного обучения в индустрии гостеприимства в целом [32].

Источники данных для моделирования

Базой для построения моделей машинного обучения, прогнозирующих туристский поток (целевая переменная), являются временный ряды, описывающие прибытия туристов в те ли иные де-

стинации. Следует обратить внимание на периодичность сбора этих данных официальными статистическими структурами. Так, для стран Европы и США, Сингапура, Китая и Тайваня доступны ежемесячные данные, для основных точек притяжения в Китае – ежедневные, а для достопримечательностей Пекина – почасовые. Для сравнения данные о ночевках и количестве лиц в коллективных средствах размещения в России собираются на квартальной основе, а до 2011 г. – на годовой, что, естественно, снижает объем выборки для моделирования в несколько раз. При этом, в открытом доступе имеются исключительно сведения о пребывании в официально зарегистрированных средствах коллективного размещения, что не отражает полной картины туристского потока в регионах и не дает

информации о туристах, приезжающих в дестинацию без ночевки. Основными источниками данных для объясняющих переменных являются, помимо ретроспективных временных рядов туристских прибытий, история интернет-поиска и сведения из социальных сетей и других аналогичных ресурсов. Отдельные единичные исследования, однако, используют данные мобильной геолокации и Wi-Fi.

История интернет-поиска становится едва ли не основным источником объясняющих переменных для прогнозирования туристского спроса. С каждым годом число исследований в этом направлении растет. Растет и доля работ, использующих данные поисковых запросов, в общем количестве статей по прогнозированию туристского потока (рис. 2).

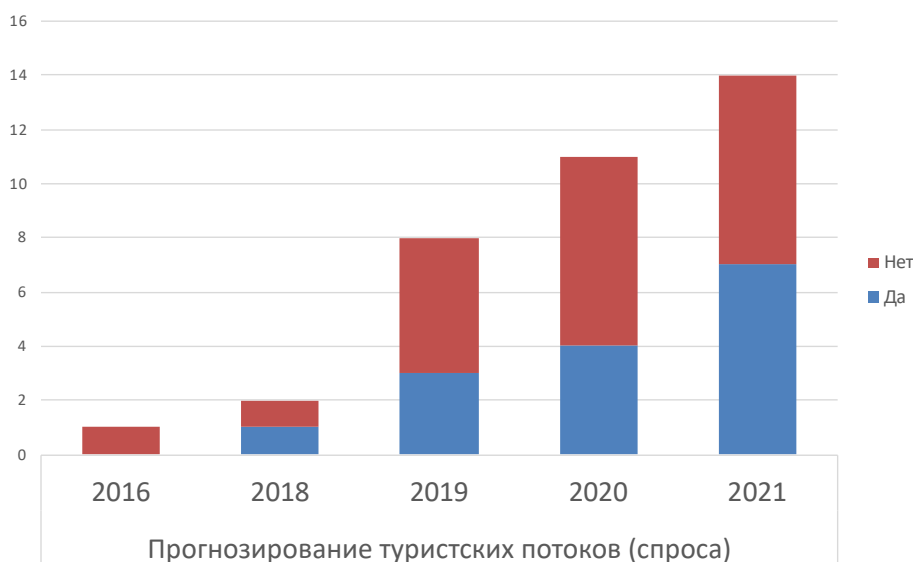


Рисунок 2. Распределение количества исследований по признаку использования данных поисковых запросов (да, нет)

Figure 2. Distribution of the number of studies based on the use of search query data (yes, no)

Еще один выделяющийся тренд – использование онлайн-отзывов, постов в социальных сетях для анализа пользовательских предпочтений и структуры туристического потока. Здесь основными источниками данных служат интернет-ресурсы TripAdvisor, Flickr, Twitter, Strip and Qunar. И если в случае прогнозирования туристского потока использование онлайн-отзывов ограничивается отдельными статьями (ссылки), то для анализа пользовательских предпочтений и построения рекомендательных систем они, как было указано выше, являются основным источником данных (рис. 3).

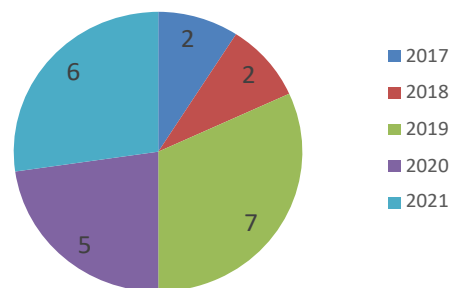


Рисунок 3. Количество статей, посвященных использованию данных из социальных сетей при построении моделей машинного обучения, по годам

Figure 3. Number of articles devoted to the use of social media data in building machine learning models, by year

География прогнозирования туристских потоков

География исследований, посвященных прогнозированию туристского потока с применением алгоритмов машинного обучения, сконцентрирована в Китае, на который приходится более половины научных статей (рис. 4).



Рисунок 4. Распределение количества статей по туристским дестинациям
Figure 4. Distribution of the number of articles by tourist destinations

Основной интерес у исследователей вызывают локальные дестинации в Китае - Пекин, Национальный парк Цзючжайгоу, гора Сигуньян, Гонконг, отдельные работы посвящены прогнозированию туристских прибытий на остров Кулансу, гору Хуаншань, аэропорты Гуанджоу и Пудун, а также в Макао.

Что касается других регионов, Европы и США, то локальные дестинации рассматриваются только в трех случаях – Милан, Италия и Чарльстон, США и национальные парки в США. Все остальные исследования анализируют прибытия на уровне страны. Исключения составляет статья, где моделируется пассажиропоток в крупнейшие аэропорты Азиатско-Тихоокеанского и Латинско-американско-Карибского региона [33]. Следует отметить, что ни в одной работе не ставилась задача моделирования внутреннего туризма между регионами одной страны.

Выбор объясняющих переменных и алгоритма машинного обучения

Алгоритмы, выбираемые авторами исследований для прогнозирования туристских потоков, зависят от объясняющих переменных. Так, традиционно выделяют модели временного ряда. На их долю приходится 15 работ, чуть менее половины от общего количества статей, исследующих туристские прибытия (рис. 5).

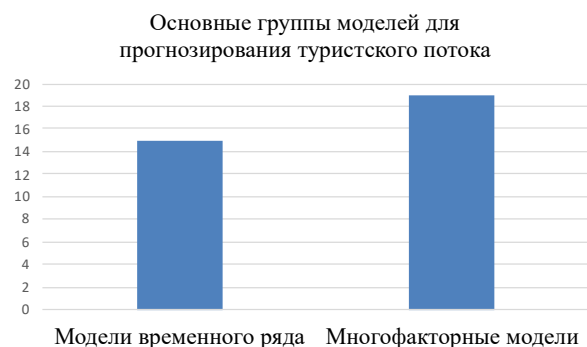


Рисунок 5. Основные группы моделей для прогнозирования туристского потока
Figure 5. The main groups of models for predicting the tourist flow

Модели временного ряда

Признаками в моделях временного ряда являются значения прогнозируемой переменной (то есть пассажиропотока), взятые в исторической перспективе. Другие переменные не используются. Это облегчает поиск исходных данных для модели, но имеет и ряд недостатков, так как на пассажиропоток могут влиять и другие факторы, помимо сведений о прошлом. В большинстве рассмотренных моделей временного ряда используются различные модификации нейронных сетей, в частности, авторегрессионные нейросети, LSTM, ELM и прочие модификации, иногда в комбинации с другими алгоритмами.

Особенный интерес вызывает исследование [34], где временной ряд сначала преобразуется в изображение, которое обрабатывается сверточной нейросетью, а потом передается в сеть LSTM.

Среди остальных алгоритмов (табл. 4) можно выделить модели разложения на эмпирические моды, иерархический метод ближайших соседей и пространственно-временную авторегрессионную модель.

Таблица 4. Алгоритмы, используемые для моделирования временных рядов (кроме нейросетей)

Table 4. Algorithms used for time series modeling (except neural networks)

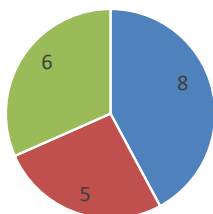
Алгоритм	Название статьи
Ensemble empirical mode decomposition	A decomposition-ensemble approach for tourism forecasting
Hierarchical pattern recognition	Hierarchical pattern recognition for tourism demand forecasting
Spatiotemporal autoregressive models	Forecasting international tourism demand: a local spatiotemporal model

Следует отметить, что практически во всех статьях результаты моделирования сравниваются с классическими эконометрическими (линейными) авторегрессионными моделями (ARIMA, ARIMAX, SARIMA, SARIMAX и проч.) для демонстрации неизменного превосходства более современных и сложных моделей.

Многофакторные модели

Альтернативой моделям временного ряда служат многофакторные регрессионные модели (рис. 6). Они, помимо исторических данных, опираются на признаки, которые, с точки зрения логики, могут оказать влияние динамику целевой переменной. Здесь, как было отмечено выше, абсолютное большинство (13 из 19 статей) составляют исследования, использующие динамику поиска, по ключевым словам, в качестве основных объясняющих переменных – и так называемые, search query data (SQD).

Среди SQD-моделей преобладают различные модификации нейронных сетей (LTSM, ELM и др., однако помимо них используются ансамблевые модели деревьев решений (в частности, случайный лес), машины опорных векторов и другие алгоритмы.



- Многофакторные модели только с SQD
- Многофакторные модели только с SQD и дополнительными переменными
- Многофакторные модели без SQD

Рисунок 6. Распределение исследований в зависимости от используемых переменных
Figure 6. Distribution of studies according to the variables used

В дополнение к переменным SQD, в некоторых исследованиях используются данные о погоде, информация из социальных сетей, экономические индикаторы, что, согласно выводам авторов, обогащает статистику и повышает точность моделирования (табл. 5).

Наконец, самую немногочисленную группу многофакторных моделей составляют исследования, не использующие переменные SQD (табл. 6). В качестве объясняющих переменных выступают экономические показатели, геотеги, показатели конкурентоспособности дестинации и баланс цен спроса и предложения [35].

Таблица 5. Переменные, используемые в дополнение к SQD

Table 5. Variables used in addition to SQD

Название статьи	Переменные, используемые в дополнение к SQD
Forecasting tourism demand with multisource big data	погода, онлайн-отзывы (2 китайских платформы)
Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach	экономические индикаторы
Forecasting tourist arrivals using denoising and potential factors	погода, календарный день, сезонность
Daily tourism volume forecasting for tourist attractions	погода
Intelligent forecasting of inbound tourist arrivals by social networking analysis	Информация из соцсетей, ВВП, ИПЦ, курс доллара США, вместимость отелей и самолетов

ЗАКЛЮЧЕНИЕ

Как правило, современные исследования по моделированию туристских потоков - это исследования, целью которых является не просто прогнозирование туристского потока, а исследование его динамики в зависимости от конкретных показателей. Необходимо отметить, что ни в одной из проанализированных работ не ставилась задача прогнозировать туристский поток под воздействием факторов средового влияния и уровня развития инфраструктуры. Авторы считают это существенным недостатком и предполагают в дальнейшем предоставить свои результаты, которые основываются на учете этих факторов. Необходимо подчеркнуть, что такой подход будет обладать новизной и будет весьма актуален.

В данном обзорном исследовании мы указали не только на отсутствие учета актуальных аспектов, но и на рост интереса к заявленной теме.

Существенным результатом можно считать определение проблемы данных. Действительно, наличие систем сбора данных определяет рост возможностей использования современных методов анализа. Наибольшее число исследований было реализовано в Китае.

Можно сделать вывод о развитии современных методов машинного обучения в Российской Федерации. Актуальные исследования и адекватные результаты наши исследователи смогут получать лишь при наличии открытого доступа к корректно собранным датасетам. Использование датасетов коммерческих компаний в открытых исследованиях на сегодняшний день может носить случайный характер и не может быть

решением некоторой глобальной задачи описания деятельности туристической сферы. Возможно, государственные структуры вынуждены будут создать систему сбора корректной информации, которая могла бы быть использована в открытых исследованиях. В этом случае высокая скорость развития этой сферы (широкий доступ к информации позволит развивать разработку эффективных приемов и методов, отработку методов определения корректности и эффективности алгорит-

мов) будет способствовать и получению более эффективных прогнозов для принятия решений в управленческой деятельности.

Получение данных выводов необходимо признать научной новизной нашего исследования.

Статья написана в рамках НИР «Разработка концепции моделирования рынка туристических услуг России с применением методов экономико-математического моделирования и современных цифровых технологий»

Таблица 6. Многофакторные модели без SQD
Table 6. Multifactor models without SQD

Название статьи	Алгоритм	Переменные
Big data from dynamic pricing: A smart approach to tourism demand forecasting	SARIMA, ARIMA	Цены спроса и предложения
Forecasting air passenger numbers with a GVAR model	Global Vector AutoRegressive	Foreign и domestic variables, country-level proxies
Bayesian BiLSTM approach for tourism demand forecasting	Bayesian Bidirectional Long Short-Term Memory (BBiLSTM) network	Микро и макро показатели
Geo-tagged photo metadata processing method for Beijing inbound tourism flow	NN, SVR, and ELM. RBF (Radial Basis Function) is selected as the kernel function of the SVR	Геотеги фотографий (Flickr)
Competitiveness and overtourism: A proposal for an early warning system in spanish urban destinations	Bayesian estimation, logit model, SVM, BN and CART	Показатели конкурентоспособности региона
Machine learning techniques as a tool for predicting overtourism: The case of Spain	-	Показатели конкурентоспособности региона

ЛИТЕРАТУРА

1. **Li X.** Review of tourism forecasting research with internet data. *Tour. Manag.* 2021. Vol. 83. P. 104245.
2. **Jain P.K., Pamula R., Srivastava G.** A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Comput. Sci. Rev.* 2021. Vol. 41. P. 100413.
3. **Hillel T.** A systematic review of machine learning classification methodologies for modelling passenger mode choice. *J. Choice Model.* 2021. Vol. 38. P. 100221.
4. **Santamaria-Granados L., Mendoza-Moreno J.F., Ramirez-Gonzalez G.** Tourist recommender systems based on emotion recognition—a scientometric review. *Futur. Internet.* 2021. Vol. 13. N 1.
5. **Kirilenko A.P.** Automated Sentiment Analysis in Tourism: Comparison of Approaches. *J. Travel Res.* 2018. Vol. 57. N8.
6. **Parvez M.O.** Use of machine learning technology for tourist and organizational services: high-tech innovation in the hospitality industry. *J. Tour. Futur.* 2020.
7. **Ланина К.С., Ланин М.А.** Технологии машинного обучения как один из новых векторов развития гостиничных предприятий. В сб. «Современные проблемы и перспективы развития туризма и сферы услуг в условиях глобализации». Матер. III Всерос. н.-пр. конф. Владимир: АТЛАС. 2020. С. 227–231.
8. **Ведерникова А.Ю., Нянина Т.А., Гадасина Л.В.** Европейский туризм в период пандемии коронавируса. В сб. «Последствия и вызовы пандемии коронавируса для технологического и социально-экономического развития общества». Матер. III Межд. н.-пр. конф. Ярославль: Ярославский государственный технический университет. 2020. С. 504–510.

REFERENCES

1. **Li X.** Review of tourism forecasting research with internet data. *Tour. Manag.* 2021. Vol. 83. P. 104245.
2. **Jain P.K., Pamula R., Srivastava G.** A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Comput. Sci. Rev.* 2021. Vol. 41. P. 100413.
3. **Hillel T.** A systematic review of machine learning classification methodologies for modelling passenger mode choice. *J. Choice Model.* 2021. Vol. 38. P. 100221.
4. **Santamaria-Granados L., Mendoza-Moreno J.F., Ramirez-Gonzalez G.** Tourist recommender systems based on emotion recognition—a scientometric review. *Futur. Internet.* 2021. Vol. 13. N 1.
5. **Kirilenko A.P.** Automated Sentiment Analysis in Tourism: Comparison of Approaches. *J. Travel Res.* 2018. Vol. 57. N 8.
6. **Parvez M.O.** Use of machine learning technology for tourist and organizational services: high-tech innovation in the hospitality industry. *J. Tour. Futur.* 2020.
7. **Lanina K.S., Lanin M.A.** Machine learning technologies as one of the new vectors of development of hotel enterprises. *Materials of the III All-Russia scientific and practical conference «Modern problems and prospects for the development of tourism and the service sector in the context of globalization»* Vladimir: ATLAS. 2020. P. 227–231. (in Russian).
8. **Vedernikova A.Y., Nianina T.A., Gadasina L.V.** European tourism during the coronavirus pandemic. *Materials of the III International scientific and practical conference «Consequences and causes of the coronavirus pandemic for technological and socio-economic development of society»*. Yaroslavl: Yaroslavl State Technical University. 2020. P.504-510. (in Russian).

9. **Деменев А.** Технологические тренды цифровой трансформации индустрии делового туризма в гостиничном сегменте. В сб. «Разработка и практика внедрения стратегических управленческих проектных решений по развитию гостиничного и туристского бизнеса». Матер. Межд. н.-пр. конф. М.: Русайнс. 2019. С. 176–181.
10. **Михайлов С.А.** Интеллектуальная система помощи туристу: сервис-ориентированная архитектура и реализация. *Научно-технический вестник информационных технологий, механики и оптики*. 2019. Т. 19, № 3. С. 499–507.
11. **Имаметдинова М.Р.** Big Data технологии - основа цифровизации логистической сервисной системы в туристской отрасли. В сб. «Управление в современных системах». Матер. VIII Межд. н.-пр. конф. Челябинск: Южно-Уральский технологический университет. 2018. С. 268–276.
12. **Шпырня О., Коренева М.В.** Новые технологии развития рынка туристских услуг. *Научный вестник Южного института менеджмента*. 2019. № 4 (28). С. 113–116.
13. **Першина Э.С., Дараган С.В.** От больших данных к продвинутой аналитике в индустрии туризма. *Научный вестник МГИИТ*. 2018. № 2(52). С. 60–69.
14. Scimago Journal and Country Rank. <https://www.scimagojr.com/journalsearch.php?q=30718&tip=sid&clean=0>, <https://www.scimagojr.com/journalsearch.php?q=16547&tip=sid&clean=0>.
15. **Xie G., Qian Y., Wang S.** Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach. *Tour. Manag.* 2021. Vol. 82. P. 104208.
16. **Lingyu T., Jun W., Chunyu Z.** Mode decomposition method integrating mode reconstruction, feature extraction, and ELM for tourist arrival forecasting. *Chaos, Solitons & Fractals*. 2021. Vol. 143. P. 110423.
17. **Yuan F.-C.** Intelligent forecasting of inbound tourist arrivals by social networking analysis. *Phys. A Stat. Mech. its Appl.* 2020. Vol. 558. P. 124944.
18. **Kulshrestha A., Krishnaswamy V., Sharma M.** Bayesian BILSTM approach for tourism demand forecasting. *Ann. Tour. Res.* 2020. Vol. 83. P. 102925.
19. **Yao Y., Cao Y.** A Neural network enhanced hidden Markov model for tourism demand forecasting. *Appl. Soft Comput.* 2020. Vol. 94. P. 106465.
20. **Yao Y.** A paired neural network model for tourist arrival forecasting. *Expert Syst. Appl.* 2018. Vol. 114.
21. **Livieris I.E.** Weight-constrained neural networks in forecasting tourist volumes: A case study. *Electron.* 2019. Vol. 8. N 9.
22. **Jiao X., Li G., Chen J.L.** Forecasting international tourism demand: a local spatiotemporal model. *Ann. Tour. Res.* 2020. Vol. 83. P. 102937.
23. **Gunter U., Zekan B.** Forecasting air passenger numbers with a GVAR model. *Ann. Tour. Res.* 2021. Vol. 89. P. 103252.
24. **Perles-Ribes J.F.** Competitiveness and overtourism: A proposal for an early warning system in spanish urban destinations. *Eur. J. Tour. Res.* 2021. Vol. 27.
25. **Perles-Ribes J.F.** Machine learning techniques as a tool for predicting overtourism: The case of Spain. *Int. J. Tour. Res.* 2020. Vol. 22. N 6.
26. **Jain P.K., Pamula R., Srivastava G.** A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Comput. Sci. Rev.* 2021. Vol. 41. P. 100413.
27. **Kirilenko A.P.** Automated Sentiment Analysis in Tourism: Comparison of Approaches. *J. Travel Res.* 2018. Vol. 57. P. 8.
28. **Giglio S.** Using social media to identify tourism attractiveness in six Italian cities. *Tour. Manag.* 2019. Vol. 72.
9. **Demenev A.** Technological trends in digital transformation of business tourism industry in the hotel segment. *Materials of the International scientific and practical conference «Development and practice of implementing strategic management project solutions for the development of hotel and tourism business»*. Moscow: Rusains. 2019. P.176-181. (in Russian).
10. **Mikhailov S.A.** Intelligent system of tourist assistance: service-oriented architecture and implementation. *Scientific and Technical Bulletin of Information Technologies, Mechanics and Optics*. 2019. Vol. 19, N3. P. 499-507. (in Russian).
11. **Imametdinova M.R.** Big data technologies - the basis of digitalization of logistics services in the tourism industry. *Materials of the VIII International scientific and practical conference «Management in modern systems»*. Chelyabinsk: South Ural Technological. 2018. P. 268-276. (in Russian).
12. **Shpyrnia O., Koreneva M.V.** New technologies of tourist services market development. *Scientific Bulletin of the Southern Institute of Management*. 2019. N 4(28). P. 113-116. (in Russian).
13. **Pershina E.S., Daragan S.V.** From big data to advanced analytics in the tourism industry. *Scientific Bulletin MSITI*. 2018. N 2(52). P. 60–69. (in Russian).
14. Scimago Journal and Country Rank. <https://www.scimagojr.com/journalsearch.php?q=30718&tip=sid&clean=0>, <https://www.scimagojr.com/journalsearch.php?q=16547&tip=sid&clean=0>.
15. **Xie G., Qian Y., Wang S.** Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach. *Tour. Manag.* 2021. Vol. 82. P. 104208.
16. **Lingyu T., Jun W., Chunyu Z.** Mode decomposition method integrating mode reconstruction, feature extraction, and ELM for tourist arrival forecasting. *Chaos, Solitons & Fractals*. 2021. Vol. 143. P. 110423.
17. **Yuan F.-C.** Intelligent forecasting of inbound tourist arrivals by social networking analysis. *Phys. A Stat. Mech. its Appl.* 2020. Vol. 558. P. 124944.
18. **Kulshrestha A., Krishnaswamy V., Sharma M.** Bayesian BILSTM approach for tourism demand forecasting. *Ann. Tour. Res.* 2020. Vol. 83. P. 102925.
19. **Yao Y., Cao Y.** A Neural network enhanced hidden Markov model for tourism demand forecasting. *Appl. Soft Comput.* 2020. Vol. 94. P. 106465.
20. **Yao Y.** A paired neural network model for tourist arrival forecasting. *Expert Syst. Appl.* 2018. Vol. 114.
21. **Livieris I.E.** Weight-constrained neural networks in forecasting tourist volumes: A case study. *Electron.* 2019. Vol. 8. N 9.
22. **Jiao X., Li G., Chen J.L.** Forecasting international tourism demand: a local spatiotemporal model. *Ann. Tour. Res.* 2020. Vol. 83. P. 102937.
23. **Gunter U., Zekan B.** Forecasting air passenger numbers with a GVAR model. *Ann. Tour. Res.* 2021. Vol. 89. P. 103252.
24. **Perles-Ribes J.F.** Competitiveness and overtourism: A proposal for an early warning system in spanish urban destinations. *Eur. J. Tour. Res.* 2021. Vol. 27.
25. **Perles-Ribes J.F.** Machine learning techniques as a tool for predicting overtourism: The case of Spain. *Int. J. Tour. Res.* 2020. Vol. 22. N 6.
26. **Jain P.K., Pamula R., Srivastava G.** A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Comput. Sci. Rev.* 2021. Vol. 41. P. 100413.
27. **Kirilenko A.P.** Automated Sentiment Analysis in Tourism: Comparison of Approaches. *J. Travel Res.* 2018. Vol. 57. P. 8.
28. **Giglio S.** Using social media to identify tourism attractiveness in six Italian cities. *Tour. Manag.* 2019. Vol. 72.

29. **Wang M.** Applying Internet information technology combined with deep learning to tourism collaborative recommendation system. *PLoS One*. 2020. Vol. 15. P. 12 December.
30. **Nilashi M.** Analysis of Travellers' Online Reviews in Social Networking Sites Using Fuzzy Logic Approach. *Int. J. Fuzzy Syst.* 2019. Vol. 21. P. 5.
31. **Penagos-Londoño G.I.** A machine learning approach to segmentation of tourists based on perceived destination sustainability and trustworthiness. *J. Destin. Mark. Manag.* 2021. Vol. 19.
32. **Parvez M.O.** Use of machine learning technology for tourist and organizational services: high-tech innovation in the hospitality industry. *J. Tour. Futur.* 2020.
33. **Gunter U., Zekan B.** Forecasting air passenger numbers with a GVAR model. *Ann. Tour. Res.* 2021. Vol. 89. P. 103252.
34. **Bi J.-W., Li H., Fan Z.-P.** Tourism demand forecasting with time series imaging: A deep learning model. *Ann. Tour. Res.* 2021. Vol. 90. P. 103255.
35. **Guizzard A.** Big data from dynamic pricing: A smart approach to tourism demand forecasting. *Int. J. Forecast.* 2021. Vol. 37, N 3. P. 1049–1060.
36. **Astrakhtantseva I., Astrakhtantsev R.** Cryptocurrency as new financial and legal instrument: defining cryptoassets in property law. *SHS Web of Conferences: III International on New Industrialization and Digitalization (NID 2020)*. Ekaterinburg: EDP Sciences. 2021. P. 02002. DOI: 10.1051/shsconf/20219302002.
29. **Wang M.** Applying Internet information technology combined with deep learning to tourism collaborative recommendation system. *PLoS One*. 2020. Vol. 15. P. 12 December.
30. **Nilashi M.** Analysis of Travellers' Online Reviews in Social Networking Sites Using Fuzzy Logic Approach. *Int. J. Fuzzy Syst.* 2019. Vol. 21. P. 5.
31. **Penagos-Londoño G.I.** A machine learning approach to segmentation of tourists based on perceived destination sustainability and trustworthiness. *J. Destin. Mark. Manag.* 2021. Vol. 19.
32. **Parvez M.O.** Use of machine learning technology for tourist and organizational services: high-tech innovation in the hospitality industry. *J. Tour. Futur.* 2020.
33. **Gunter U., Zekan B.** Forecasting air passenger numbers with a GVAR model. *Ann. Tour. Res.* 2021. Vol. 89. P. 103252.
34. **Bi J.-W., Li H., Fan Z.-P.** Tourism demand forecasting with time series imaging: A deep learning model. *Ann. Tour. Res.* 2021. Vol. 90. P. 103255.
35. **Guizzard A.** Big data from dynamic pricing: A smart approach to tourism demand forecasting. *Int. J. Forecast.* 2021. Vol. 37, N 3. P. 1049–1060.
36. **Astrakhtantseva I., Astrakhtantsev R.** Cryptocurrency as new financial and legal instrument: defining cryptoassets in property law. *SHS Web of Conferences: III International on New Industrialization and Digitalization (NID 2020)*. Ekaterinburg: EDP Sciences. 2021. P. 02002. DOI: 10.1051/shsconf/20219302002.

Поступила в редакцию 12.10.2021
Принята к опубликованию 26.10.2021

Received 12.10.2021
Accepted 26.10.2021