

Цифровая экономика

ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ТЕКСТОВ ДЛЯ ОЦЕНКИ ПОТРЕБИТЕЛЬСКОГО ОПЫТА НА ПРИМЕРЕ РЕСТОРАНОВ САНКТ-ПЕТЕРБУРГА

Статья рекомендована к публикации главным редактором Т. В. Ершовой 10.03.2025.

Бурков Иван Павлович

Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербургская школа экономики и менеджмента, аспирант

Университет ИТМО, факультет технологического менеджмента и инноваций, старший преподаватель

Санкт-Петербург, Россия

ivan.p.burkov@gmail.com

Аннотация

В условиях цифровизации интернет-ресурсы оказывают существенное воздействие на потребительские предпочтения, а онлайн-отзывы формируют ожидания и поведенческие установки потребителей. В данном исследовании предложен методологический подход, сочетающий структурно-тематическое моделирование и регрессионный анализ, что позволило выявить ключевые факторы удовлетворенности клиентов в ресторанной индустрии Санкт-Петербурга. Результаты показали, что качество обслуживания и атмосфера становятся ключевыми в формировании дальнейших поведенческих намерений. Предложенная методология может быть адаптирована для анализа онлайн-отзывов в различных сферах, таких как гостиничный бизнес, туризм, онлайн-ритейл и услуги. Исследование обосновывает применимость интеллектуального анализа текстов в управлении клиентским опытом и стратегическом маркетинге, расширяя существующие подходы к изучению потребительского поведения.

Ключевые слова

онлайн-отзывы, интеллектуальный анализ текстов, потребительское поведение, клиентский опыт, структурно-тематическое моделирование

Введение

В современном мире интернет-ресурсы играют ключевую роль в формировании потребительского поведения. Они оказывают значительное влияние на принятие решений о покупке, особенно в сфере гостеприимства и туризма [4, 7]. Онлайн-отзывы становятся важным источником информации для потребителей, так как они представляют собой форму пользовательского контента и могут быть более релевантными, чем рекомендации близких людей [3]. Потребители доверяют отзывам, которые помогают им сделать осознанный выбор места посещения.

Одним из ключевых аспектов успешного бизнеса является понимание и учёт особенностей поведения потребителей среди различных групп населения. Научные исследования активно используют онлайн-отзывы для анализа моделей поведения потребителей и намерений в различных отраслях. Такие данные помогают предпринимателям лучше понимать своих клиентов, формировать поведенческие намерения и достигать коммерческой эффективности [5]. Понимание предпосылок, предшествовавших этим намерениям, открывает новые возможности для улучшения маркетинговых стратегий и управления в индустрии гостеприимства и туризма [4]. Онлайн-отзывы оказывают значительное влияние на эффективность работы и стратегии: они служат важным источником обратной связи, формируют репутацию заведения и способствуют распространению информации посредством эффекта сарафанного радио, которое влияет на бизнес, предоставляя

© Бурков И. П., 2025

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «С указанием авторства - С сохранением условий версии 4.0 Международная» (Creative Commons Attribution – ShareAlike 4.0 International; CC BY-SA 4.0). См. <https://creativecommons.org/licenses/by-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2025_06_12

данные о потребительских предпочтениях и способствуя созданию имиджа бренда и привлечению новых клиентов [7]. Эти данные помогают предприятиям улучшать свои услуги и продукты, а также разрабатывать более эффективные маркетинговые кампании [4].

С точки зрения теории сигналов, онлайн-отзывы представляют собой форму сигнала, способного служить эталоном качества и надежности для потенциальных потребителей, особенно в условиях информационной асимметрии между предприятиями и клиентами [12]. Основываясь на теории вовлеченности, степень вовлеченности потребителей может снижать влияние онлайн-отзывов о их намерениях вернуться [2]. Потребители с низкой вовлеченностью могут быть более подвержены эмоциональным настроениям, чем функциональным деталям, упомянутым в обзорах. В то же время, согласно теории планируемого поведения и когнитивно-аффективно-конативной модели, удовлетворенность является наиболее значимым предиктором [1, 10]. Таким образом, степень вовлечённости потребителей может снижать влияние отзывов на их решения, при этом удовлетворенность от посещения является наиболее значимым фактором, определяющим намерения вернуться.

Анализ онлайн-отзывов является важным инструментом для повышения эффективности бизнеса, что особенно актуально в сфере гостеприимства и туризма и в ресторанном бизнесе, в частности [2, 5, 12]. Он помогает компаниям лучше понимать клиентов, улучшать услуги и продукты, а также укреплять доверие и репутацию. В условиях высокой конкуренции и быстро меняющихся потребительских предпочтений, использование онлайн-отзывов становится необходимым условием для успешного ведения бизнеса. Данное исследование направлено на выявление факторов, влияющих на удовлетворенность клиентов на примере ресторанной индустрии Санкт-Петербурга. Исследовательский вопрос сформулирован следующим образом: Какие факторы влияют на удовлетворённость клиентов в ресторанной индустрии Санкт-Петербурга, и существуют ли различия в формировании потребительского опыта среди разных категорий потребителей? Данное исследование направлено на разработку методологии оценки предыдущего потребительского опыта через призму конкретного действия – желания поделиться впечатлениями, написав отзыв. Это придаёт исследованию научную новизну, поскольку существующие работы основаны на оценке потребительского опыта посредством прямого взаимодействия с потребителями [14]. Учитывая, что факторы удовлетворённости могут значительно различаться между различными группами населения, выявление этих различий приобретает особую важность [15].

1 Методология исследования

Для проведения исследования при помощи пакета “RSelenium” в R были собраны отзывы с платформы TripAdvisor, что позволило автоматизировать процесс и охватить всю генеральную совокупность ресторанов города. В результате было собрано 62,319 текстов. Собранный датасет включает как информацию об отзыве – текст, оценку; так и об авторе – пол, возраст, город проживания, уровень пользователя на платформе, который зависит от количества опубликованных отзывов. Дополнительно были посчитаны длины отзывов и индекс удобочитаемости.

Предыдущие исследования потребительского поведения чаще всего применяли качественные методы анализа, например интервью или кейс-стади [9], либо количественно оценивали факторы при помощи анкетирования и дальнейшего структурного моделирования [7]. Такие методы могут быть нерепрезентативными и анализ онлайн отзывов оказывается более надежным и признается полезным для более глубокого анализа поведенческих особенностей [8]. Для работы с текстами часто применяется метод анализа тональности, но он может быть ненадежным и вводить в заблуждение, так как опирается на анализе отдельных слов вне их контекста [13]. В данном исследовании предлагается выявление отдельных тем, которые в совокупности могут описать потребительский опыт и повлиять на дальнейшие намерения. Для этого было применено структурно-тематическое моделирование (STM). STM позволяет анализировать тексты отзывов, выявляя темы, называемые «мешками слов» [6, 11]. Количество тем (K) для STM было определено статистически с помощью метода оценки информативности. Оптимальное количество тем составило 20, которые наиболее полно описывали содержание отзывов. Перед проведением STM данные были предварительно обработаны, включая очистку текста от слов без смысла и пунктуации, нормализацию и токенизацию. Итоговый набор данных был представлен в виде матрицы, где каждая строка соответствует отзыву, а столбцы – темам,

выявленным в процессе моделирования. Таким образом, был получен вес каждой темы в каждом отзыве для дальнейшего определения взаимосвязей между другими переменными.

Статистический анализ был проведен для определения факторов, влияющих на переменную «оценка» (намерение делиться положительными эмоциями) с использованием обобщенных линейных моделей. Анализ включал оценку двух моделей регрессии Пуассона для изучения взаимосвязи между оценкой и набором независимых переменных. Первая модель включала текстовые характеристики (длина текста, средний индекс удобочитаемости), демографические характеристики пользователей (статус постоянного жителя или туриста, пол, возраст, уровень пользователя) и переменные долей тем в текстах отзывов. Сокращенная вторая модель исключала тематические переменные, учитывая только текстовые функции и демографические данные пользователей. Для оценки потенциальной мультиколлинеарности между независимыми переменными был проведен анализ коэффициента инфляции дисперсии с установленным пороговым значением VIF > 5. Высоких VIF не наблюдалось, следовательно, модели не обладала мультиколлинеарностью. Результаты моделей представлены в таблице 1.

Таблица 1. Результаты построения регрессионных моделей

	Модель 1	Модель 2
Длина текста	0.0003*** (0.00004)	-0.001*** (0.00004)
Индекс удобочитаемости	-0.0004 (0.0005)	0.003*** (0.0004)
Статус резидента Петербурга	-0.020*** (0.004)	-0.040*** (0.004)
Пол пользователя	-0.005 (0.004)	-0.025*** (0.004)
Возраст 18-24	0.028 (0.057)	0.049 (0.057)
Возраст 25-34	0.017 (0.057)	-0.048 (0.057)
Возраст 35-49	0.022 (0.057)	-0.041 (0.057)
Возраст 50-64	0.025 (0.057)	-0.019 (0.057)
Возраст 65+	0.004 (0.061)	-0.050 (0.061)
Критик 2-го уровня	-0.001 (0.019)	-0.055*** (0.013)
Критик 3-го уровня	0.001 (0.017)	-0.051*** (0.010)
Критик 4-го уровня	-0.002 (0.016)	-0.052*** (0.010)
Критик 5-го уровня	-0.006 (0.016)	-0.067*** (0.010)
Критик 6-го уровня	-0.015 (0.016)	-0.067*** (0.009)
Тема 1: Антисанитария и плохое обслуживание	-1.677*** (0.052)	
Тема 2: Бюджетные обеды	-0.236*** (0.034)	

Тема 3: Семейные рестораны и развлечения	0.050	
	(0.037)	
Тема 4: Пабы и пиво	0.002	
	(0.038)	
Тема 5: Петербург	0.260***	
	(0.042)	
Тема 6: Праздники и благодарности	0.300***	
	(0.042)	
Тема 7: Бронирование и обслуживание клиентов	-0.687***	
	(0.047)	
Тема 8: Дружеские встречи	0.055	
	(0.037)	
Тема 9: Музыка и ночная жизнь	0.018	
	(0.042)	
Тема 10: Изысканная подача блюд	0.179***	
	(0.044)	
Тема 11: Грузинская кухня	0.175***	
	(0.043)	
Тема 12: Кофе и десерты	0.108***	
	(0.036)	
Тема 13: Азиатская кухня	-0.129***	
	(0.045)	
Тема 14: Популярные заведения	0.042	
	(0.042)	
Тема 15: Медленное обслуживание	-0.930***	
	(0.045)	
Тема 16: Мясные блюда	0.156***	
	(0.048)	
Тема 17: Итальянская кухня	-0.207***	
	(0.045)	
Тема 18: Приятная атмосфера	0.198***	
	(0.034)	
Тема 19: Русская кухня	-1.041***	
	(0.163)	
Тема 20: Общие впечатления	0.771***	
	(0.171)	
Constant	1.481***	1.597***
	(0.065)	(0.058)
Наблюдения	62,319	62,319
Логарифм правдоподобия	-92,141.610	-110,413.100
Akaike Inf. Crit.	184,351.200	220,856.200

Примечание:

*p<0.1; **p<0.05; ***p<0.01

2 Результаты

В рамках данного исследования были построены две модели для анализа оценок отзывов о ресторанах в Санкт-Петербурге. Первая модель (полная) включает текстовые характеристики, демографические переменные и тематические факторы, в то время как вторая модель (сокращённая) учитывает только текстовые и демографические характеристики, исключая тематические переменные. Длина текста положительно влияет на оценку в первой модели, но оказывает отрицательное влияние во второй модели. Это может указывать на эффект зависимости от тематики. Индекс удобочитаемости (ARI) не является значимым в первой модели, но оказывает положительное влияние во второй модели. Это говорит о важности данного параметра в отсутствии тематических переменных. Статус резидента оказывает отрицательное влияние на оценку в обеих моделях, что свидетельствует о меньшей удовлетворенности или критическом отношении местных жителей. Пол пользователя незначим в первой модели, но оказывает негативное влияние во второй модели, что может говорить о различиях в ожиданиях между полами. Возрастные группы в целом не демонстрируют значимых эффектов, кроме того, что пользователи 25–34 лет во второй модели оценивают ниже. В первой модели уровни критиков не оказывают значимого влияния. Во второй модели пользователи с более высоким уровнем (критик 2–6 уровня) демонстрируют отрицательные коэффициенты, что указывает на более низкие оценки от опытных пользователей. Первая модель показывает лучшее соответствие данным, чем вторая модель, что подтверждает важность учета тематических переменных.

Анализ тематических факторов показал, что наиболее негативное влияние на оценки пользователей оказывают проблемы с антисанитарией и плохим обслуживанием, медленное обслуживание, сложности с бронированием, а также темы, связанные с русской и азиатской кухней. Это свидетельствует о том, что потребители особенно критично относятся к качеству сервиса и организации процесса обслуживания. В то же время, положительное влияние на оценки оказывают темы, связанные с атмосферой, изысканной подачей блюд, праздниками, грузинской кухней, кофе и десертами, а также общими впечатлениями. Это указывает на то, что высокие оценки чаще связаны с эмоциональной составляющей посещения, эстетическим и гастрономическим удовлетворением. Таким образом, для повышения уровня пользовательских оценок заведениям важно не только обеспечивать высокий уровень сервиса, но и создавать приятную атмосферу и акцентировать внимание на положительных эмоциональных аспектах опыта посетителей.

Заключение

В данном исследовании предложен методологический подход к анализу онлайн-отзывов, сочетающий структурно-тематическое моделирование (STM) и регрессионный анализ, что позволило выявить ключевые факторы, влияющие на удовлетворенность клиентов ресторанных бизнесов. Результаты позволяют выявить различия в восприятии ресторанных услуг между различными группами потребителей, а также определить ключевые факторы, влияющие на удовлетворенность. Анализ тематического моделирования показал, что восприятие ресторанных услуг существенно варьируется в зависимости от статуса посетителя, что подчеркивает необходимость более детализированного подхода к изучению клиентских предпочтений. В отличие от традиционных методов, таких как анализ тональности или анкетирование, данный подход позволяет учитывать контекст пользовательских отзывов, выявлять тематические особенности и оценивать их влияние на пользовательские оценки [6, 8, 13].

Помимо прикладного значения для ресторанных бизнесов, предложенная методология может быть полезна для других сфер, зависящих от пользовательских отзывов, включая гостиничный бизнес, сферу туризма, онлайн-ритейл и услуги. Возможность выявления ключевых драйверов удовлетворенности на основе анализа больших массивов текстовых данных делает данный метод ценным инструментом для управления репутацией и стратегического маркетинга.

Научная новизна исследования заключается в разработке комплексного метода анализа потребительских отзывов, который может быть адаптирован для различных отраслей. Будущие исследования могут быть направлены на расширение модели, включая анализ многоканальных данных (например, отзывов на разных платформах), учет сезонных факторов и применение методов глубокого обучения для повышения точности тематического моделирования.

Литература

1. Ajzen I. The theory of planned behavior // *Organizational Behavior and Human Decision Processes*. 1991. № 2 (50). C. 179–211.
2. Aureliano-Silva L., Leung X., Spers E. E. The effect of online reviews on restaurant visit intentions: applying signaling and involvement theories // *Journal of Hospitality and Tourism Technology*. 2021. № 4 (12). C. 672–688.
3. Duffy A. Friends and fellow travelers: comparative influence of review sites and friends on hotel choice // *Journal of Hospitality and Tourism Technology*. 2015. № 2 (6). C. 127–144.
4. Guo B., Zhou S. Understanding the impact of prior reviews on subsequent reviews: The role of rating volume, variance and reviewer characteristics // *Electronic Commerce Research and Applications*. 2016. (20). C. 147–158.
5. Gupta S., McLaughlin E., Gomez M. Guest Satisfaction and Restaurant Performance // *Cornell Hotel and Restaurant Administration Quarterly*. 2007. № 3 (48). C. 284–298.
6. Hannigan T. R. [и др.]. Topic Modeling in Management Research: Rendering New Theory from Textual Data // *Academy of Management Annals*. 2019. № 2 (13). C. 586–632.
7. Jeong E., Jang S. Restaurant experiences triggering positive electronic word-of-mouth (eWOM) motivations // *International Journal of Hospitality Management*. 2011. № 2 (30). C. 356–366.
8. Kirilenko A. P., Stepchenkova S. O., Dai X. Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply? // *Tourism Management*. 2021. (83). C. 104241.
9. Ng R., Prendergast G. P., Chang L. M. K. The Features and Effectiveness of Chinese Language Online Recommendations // *Journal of International Consumer Marketing*. 2021. № 4 (33). C. 371–385.
10. Oliver R. L. Satisfaction: A Behavioral Perspective on the Consumer / R. L. Oliver, 2-е изд., New York: Routledge, 2014.
11. Roberts M., Stewart B., Tingley D. stm: An R Package for Structural Topic Models // *Journal of Statistical Software*. 2019. (91).
12. Seghezzi A., Winkenbach M., Mangiaracina R. On-demand food delivery: a systematic literature review // *The International Journal of Logistics Management*. 2021. № 4 (32). C. 1334–1355.
13. Tao S., Kim H.-S. Online customer reviews: insights from the coffee shops industry and the moderating effect of business types // *Tourism Review*. 2022. № 5 (77). C. 1349–1364.
14. Tsaur S.-H. [и др.]. Festival attachment: antecedents and effects on place attachment and place loyalty // *International Journal of Event and Festival Management*. 2019. № 1 (10). C. 17–33.
15. Vidal Rua S. Perceptions of tourism: a study of residents' attitudes towards tourism in the city of Girona // *Journal of Tourism Analysis: Revista de Análisis Turístico*. 2020. № 2 (27). C. 165–184.

APPLICATION OF TEXT MINING METHODS FOR EVALUATING CONSUMER EXPERIENCE: CASE OF ST. PETERSBURG'S RESTAURANTS

Burkov, Ivan

*HSE University, Saint-Petersburg school of economics and management, PhD student
ITMO University, faculty of technological management and innovations, senior lecturer
Saint-Petersburg, Russia
ivan.p.burkov@gmail.com*

Abstract

In the context of digitalization, Internet resources have a significant impact on consumer preferences, and online reviews form expectations and behavioral attitudes of consumers. In this study, we propose a methodological approach that combines structural topic modeling and regression analysis, which allowed us to identify key satisfaction factors in the St. Petersburg restaurant industry. The results indicated that the quality of service and the atmosphere become key in forming further behavioral intentions. The proposed methodology can be adapted for analyzing online reviews in various fields, such as hotel business, tourism, online retail and services. The study substantiates the applicability of text mining in customer experience management and strategic marketing, expanding existing approaches to studying consumer behavior.

Keywords

online reviews, text mining, consumer behavior, consumer experience, structural topic modeling

References

1. Ajzen I. The theory of planned behavior // *Organizational Behavior and Human Decision Processes*. 1991. № 2 (50). P. 179–211.
2. Aureliano-Silva L., Leung X., Spers E. E. The effect of online reviews on restaurant visit intentions: applying signaling and involvement theories // *Journal of Hospitality and Tourism Technology*. 2021. № 4 (12). P. 672–688.
3. Duffy A. Friends and fellow travelers: comparative influence of review sites and friends on hotel choice // *Journal of Hospitality and Tourism Technology*. 2015. № 2 (6). P. 127–144.
4. Guo B., Zhou S. Understanding the impact of prior reviews on subsequent reviews: The role of rating volume, variance and reviewer characteristics // *Electronic Commerce Research and Applications*. 2016. (20). P. 147–158.
5. Gupta S., McLaughlin E., Gomez M. Guest Satisfaction and Restaurant Performance // *Cornell Hotel and Restaurant Administration Quarterly*. 2007. № 3 (48). P. 284–298.
6. Hannigan T. R. [et al.]. Topic Modeling in Management Research: Rendering New Theory from Textual Data // *Academy of Management Annals*. 2019. № 2 (13). P. 586–632.
7. Jeong E., Jang S. Restaurant experiences triggering positive electronic word-of-mouth (eWOM) motivations // *International Journal of Hospitality Management*. 2011. № 2 (30). P. 356–366.
8. Kirilenko A. P., Stepchenkova S. O., Dai X. Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply? // *Tourism Management*. 2021. (83). P. 104241.
9. Ng R., Prendergast G. P., Chang L. M. K. The Features and Effectiveness of Chinese Language Online Recommendations // *Journal of International Consumer Marketing*. 2021. № 4 (33). P. 371–385.
10. Oliver R. L. *Satisfaction: A Behavioral Perspective on the Consumer* / R. L. Oliver, 2-е изд., New York: Routledge, 2014.
11. Roberts M., Stewart B., Tingley D. *stm: An R Package for Structural Topic Models* // *Journal of Statistical Software*. 2019. (91).
12. Seghezzi A., Winkenbach M., Mangiaracina R. On-demand food delivery: a systematic literature review // *The International Journal of Logistics Management*. 2021. № 4 (32). P. 1334–1355.
13. Tao S., Kim H.-S. Online customer reviews: insights from the coffee shops industry and the moderating effect of business types // *Tourism Review*. 2022. № 5 (77). P. 1349–1364.
14. Tsaur S.-H. [и др.]. Festival attachment: antecedents and effects on place attachment and place loyalty // *International Journal of Event and Festival Management*. 2019. № 1 (10). P. 17–33.
15. Vidal Rua S. Perceptions of tourism: a study of residents' attitudes towards tourism in the city of Girona // *Journal of Tourism Analysis: Revista de Análisis Turístico*. 2020. № 2 (27). P. 165–184.