

АНАЛИЗ СОЦИАЛЬНЫХ МЕДИА И БОЛЬШИХ ДАННЫХ: ИЗУЧЕНИЕ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ И ТРЕНДОВ

Голубничий А.С., Семёнов Д.В.

Уфимский университет науки и технологий, Уфа, Россия

Ключевые слова: социальные медиа, большие данные, поведенческий анализ, персонализация контента, сегментация аудитории, визуализация данных.

Аннотация. Данная статья посвящена вопросам анализа социальных медиа и больших данных с целью изучения поведения пользователей и выявления трендов. В ней рассматриваются современные методы и технологии, используемые в анализе данных, полученных из социальных платформ, а также подходы к обработке больших объемов информации. Основные темы, затронутые в статье, включают важность аналитики социальных медиа для бизнеса, маркетинга и научных исследований, подчеркивая роль выделения ключевых трендов и паттернов в пользовательском поведении.

SOCIAL MEDIA AND BIG DATA ANALYSIS: STUDY OF USER BEHAVIOR AND TRENDS

Golubnichiy A.S., Semenov D.V.

Ufa University of Science and Technology, Ufa, Russia

Keywords: social media, big data, behavioral analysis, content personalization, audience segmentation, data visualization.

Abstract. This article addresses the issues of social media and big data analysis aimed at studying user behavior and identifying trends. It explores contemporary methods and technologies used in analyzing data derived from social platforms, along with approaches to processing large volumes of information. Key topics covered in the article include the significance of social media analytics for business, marketing, and scientific research, emphasizing the role of identifying key trends and patterns in user behavior.

Введение

Социальные медиа и большие объемы данных, порождаемые этими платформами, стали основополагающими элементами современной цифровой эры. Исследование поведения пользователей и анализ трендов в социальных медиа не только отражают изменяющуюся динамику взаимодействия в онлайн-среде, но и являются ключевым инструментом для понимания цифровой культуры, влияния социальных платформ и формирования общественного мнения.

Глубокий анализ данных, генерируемых пользователями в социальных медиа, требует умения выявлять не только количество публикаций и активности, но и суть взаимодействия, предпочтений и тенденций пользователей. Такой анализ становится ключом к разгадыванию закономерностей и механизмов, лежащих в основе формирования контента и воздействия на аудиторию.

Исследование направлено на углубленное изучение социальных медиа и больших данных с целью не только изучить поведение пользователей, но и проанализировать актуальные тренды. В процессе этого анализа мы уделяем внимание теоретическим основам анализа данных в социальных медиа, моделям поведения пользователей и ключевым аспектам анализа трендов.

Подход, основанный на теории и практике, позволит не только понять механизмы взаимодействия пользователей в онлайн-пространстве, но и выявить влияние социальных медиа на общественные процессы, формирующие наше цифровое окружение.

Важным этапом является предоставить полное понимание теоретических основ и методологий анализа социальных медиа и больших данных, а также продемонстрировать их практическое применение для изучения поведения пользователей и выявления актуальных трендов в онлайн-среде.

1. Основные принципы анализа данных в социальных медиа

Анализ социальных медиа и больших данных является многогранным процессом, требующим понимания основных теоретических концепций в области социальных наук, информатики и маркетинга для эффективного понимания взаимодействия пользователей и данных.

2. Теоретические модели поведения пользователей в социальных медиа

Модели поведения пользователей в социальных медиа часто основываются на теории социальных сетей. Эта теория рассматривает индивидуумов в контексте их связей и взаимодействий, позволяя понять, как пользователи формируют свои социальные связи, коммуницируют и воздействуют друг на друга в сетевом окружении. Моделирование поведения в социальных медиа также использует теорию социального влияния. Она объясняет, как одни пользователи могут влиять на других через убеждения, нормы и информацию, что является ключевым аспектом формирования предпочтений и решений в сетевом сообществе [1].

3. Аспекты анализа трендов в больших данных социальных медиа

Первоначальным этапом анализа трендов является сбор больших объемов информации из различных источников социальных медиа. Это может включать текстовые сообщения, изображения, видео, комментарии и другие формы контента, которые создают пользователи. Полученные данные требуют обработки и фильтрации для удаления шума и извлечения существенной информации. Этот этап включает в себя очистку данных от дубликатов, удаление несущественной информации и преобразование в удобный для анализа формат. Для выявления трендов в больших данных социальных медиа широко применяются алгоритмы машинного обучения. Эти алгоритмы позволяют выявлять закономерности, группировать данные, определять ключевые слова или темы, а также прогнозировать будущие тренды на основе анализа прошлых данных.

4. Этические аспекты анализа трендов в больших данных социальных медиа

При осуществлении анализа данных в социальных медиа и работе с обширными объемами информации необходимо пристально обращать внимание на вопросы этики. Это включает в себя не только законные аспекты сбора и обработки информации, но и более глубокие моральные и социальные вопросы, связанные с использованием этих данных. Одним из ключевых аспектов является защита личной информации пользователей. Необходимо строго соблюдать приватность и обеспечивать безопасность данных, избегая их неправомерного использования или передачи третьим лицам без согласия владельцев информации. Важность справедливости и прозрачности в процессе анализа данных не может быть недооценена. Это связано с необходимостью обеспечения объективности и минимизации возможной предвзятости при использовании алгоритмов и методов анализа. Объективность играет решающую роль, поскольку она гарантирует, что данные анализируются без искажений или влияния субъективных предположений. Минимизация предвзятости важна для того, чтобы исключить возможность создания и распространения искаженной информации, которая может повлиять на общественное мнение или решения, основанные на этих данных.

5. Основные преимущества анализа социальных медиа в бизнесе

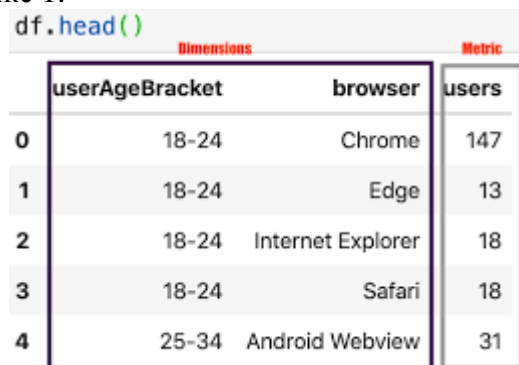
Анализ социальных медиа предоставляет бизнесу уникальные инсайты, оказывающие значительное влияние на стратегии маркетинга и взаимодействие с аудиторией. Одним из основных преимуществ является доступ к огромному объему данных, которые пользователи сами генерируют в реальном времени. Этот поток информации позволяет компаниям не только понимать текущие предпочтения и потребности своей аудитории, но и отслеживать изменения в их поведении и трендах [2].

6. Python в анализе социальных медиа и больших данных

Использование Python в анализе социальных медиа и больших данных тесно связано с его гибкостью, мощными инструментами анализа данных и активным сообществом разработчиков. Python начал активно применяться в анализе социальных медиа в начале 2000-х, когда появилась необходимость в обработке и анализе огромных объемов данных, генерируемых пользователями социальных платформ [3]. С появлением библиотек, таких как NumPy и Pandas, Python стал одним из предпочтительных инструментов для работы с данными. Он предоставляет широкий функционал для обработки, анализа и визуализации информации из социальных медиа, позволяя исследователям и бизнесам извлекать ценные инсайты из огромного объема данных. В развитие Python в анализе больших данных внесли значительный вклад библиотеки машинного обучения, такие как Scikit-learn, TensorFlow и PyTorch [4]. Эти инструменты позволили строить сложные модели анализа поведения пользователей и прогнозирования трендов на основе данных социальных медиа.

7. Методы сбора и анализа данных о поведении пользователей

Исследование поведения пользователей в социальных медиа требует использования разнообразных методов сбора и анализа данных, чтобы получить полное представление об их взаимодействии с платформами, предпочтениях и реакциях на контент. Методы сбора данных - один из основных методов сбора данных в социальных медиа заключается в использовании API социальных сетей для доступа к информации о профилях пользователей, их активности, сообщениях и других данных [5]. Это позволяет получить структурированные данные, которые затем анализируются для извлечения ключевых показателей поведения. Примером такого API является Google Analytics API. Он позволяет собирать данные о посетителях сайта. Выводимые данные зависят от настроек API. Чаще всего собираются данные о используемом браузере, для вывода этих данных используется библиотека pandas для Python, пример выводимой таблицы изображен на рисунке 1.



The image shows a terminal window displaying the output of a pandas DataFrame's head() method. The output is a table with 5 rows and 3 columns. The columns are labeled 'userAgeBracket', 'browser', and 'users'. The rows are indexed from 0 to 4. The data is as follows:

	Dimensions	Metric	
	userAgeBracket	browser	users
0	18-24	Chrome	147
1	18-24	Edge	13
2	18-24	Internet Explorer	18
3	18-24	Safari	18
4	25-34	Android Webview	31

Рис. 1. Результат сбора данных

Анализ данных о поведении – полученные данные о поведении пользователей требуют детального анализа для выявления основных тенденций и паттернов. Основные инструменты для анализа данных о поведении включают в себя методы статистического анализа, визуализации данных и машинного обучения. Анализ поведения пользователей в Python поддерживается библиотеками Pandas и NumPy, обеспечивающими удобную обработку и работу с данными. Они позволяют проводить предварительную обработку данных, фильтрацию, группировку и агрегацию, что важно для подготовки данных к дальнейшему анализу. Интеграция Python с инструментами визуализации, такими как Matplotlib, Seaborn и Plotly, обеспечивает возможность создания графиков, диаграмм и визуализаций результатов анализа данных. Это значимо для визуального представления трендов и позволяет лучше понимать данные о поведении пользователей. Пример визуализации данных с библиотекой matplotlib на Python изображен на рисунке 2.

Статистический анализ помогает выявить корреляции и зависимости между различными переменными, такими как время активности пользователя

и его взаимодействие с контентом. Визуализация данных позволяет создать понятные графики и диаграммы, облегчающие восприятие основных тенденций. Методы машинного обучения могут быть применены для прогнозирования поведения пользователей на основе имеющихся данных, что может быть важным инструментом для понимания будущих трендов [6]. Пример корреляционного анализа на Python с использованием библиотек pandas и seaborn изображен на рисунке 3.

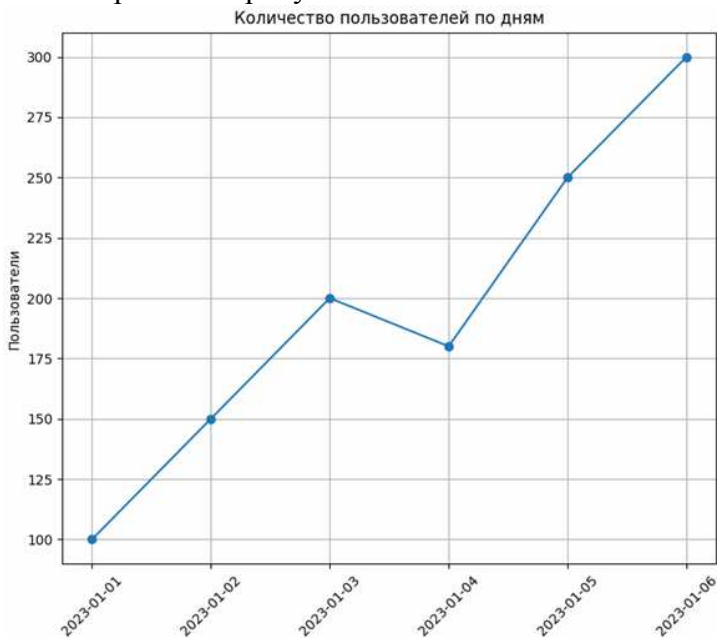


Рис. 2. Количество пользователей по дням

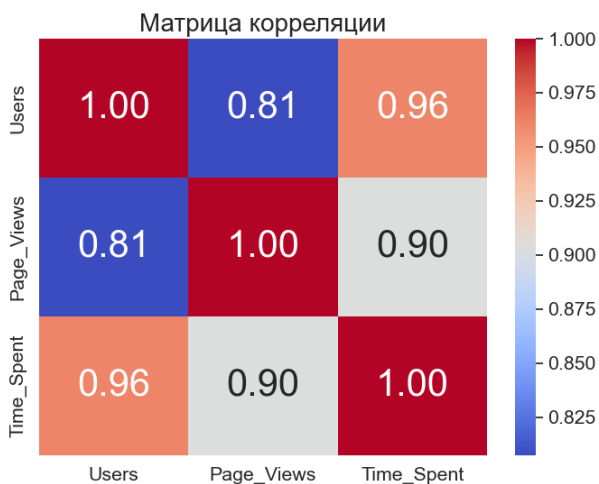


Рис. 3. Матрица корреляции

Например, для изучения эффективности рекламной кампании можно использовать методы анализа данных, чтобы оценить вовлеченность пользователей в зависимости от времени публикации или вида рекламного контента.

Для исследования воздействия публикаций в социальных медиа на мнения пользователей можно провести анализ тональности комментариев и реакций на определенный контент.

Python также широко используется в разработке алгоритмов анализа текста и обработки естественного языка (Natural Language Processing, NLP), что полезно для анализа комментариев, обзоров или других текстовых данных пользователей, содержащих важную информацию о их предпочтениях и поведении [7].

С помощью Python можно автоматизировать процессы сбора, обработки и анализа данных о поведении пользователей, что позволяет компаниям и исследователям быстрее и эффективнее извлекать ценные инсайты из больших объемов информации.

Заключение

Социальные медиа и большие данные играют ключевую роль в современном информационном ландшафте, влияя на поведение, мнения и взаимодействие пользователей. В данной работе были рассмотрены основные аспекты анализа социальных медиа и больших данных, связанные с изучением поведения пользователей и трендов. Сравнительный анализ поведения пользователей на различных социальных медиа-платформах позволяет выявить различия в типах контента, реакциях аудитории и алгоритмах работы, что существенно для адаптации стратегий воздействия в зависимости от платформы. Исследование влияния социальных медиа на формирование общественного мнения и поведения пользователей показывает, что они оказывают значительное воздействие на восприятие информации, распространение контента и даже формирование социальных стереотипов. Следующие шаги в данной области могут включать в себя более глубокий анализ влияния социальных медиа на социальные процессы, в том числе на политические предпочтения и общественные движения, а также разработку инновационных методов анализа для прогнозирования трендов и поведения пользователей. В целом, изучение социальных медиа и больших данных в контексте анализа поведения пользователей и трендов представляет собой важный инструмент для понимания современного информационного пространства и воздействия на общественные процессы.

Список литературы

1. Мустафина С.И., Акимов А.А. Технологии нейронных сетей в моделировании цифровых двойников социально-значимых объектов // Современные физика, математика, цифровые и нанотехнологии в науке и образовании. Сборник тезисов всероссийской молодежной школы-конференции, посвященной 100-летию со дня рождения А.Д. Сахарова. – Уфа: Башкирский государственный педагогический университет им. М. Акмуллы, 2022. – С. 111.
2. Акимов А.А., Валитов Д.Р., Кубряк А.И. Предварительная обработка данных для машинного обучения // Научное обозрение. Технические науки. –2022. – № 2. – С. 26-31.

3. Бизли Д. Python. Подробный справочник / Пер. с англ. – СПб.: Символ-Плюс, 2010. – 864 с.
4. Прохоренок Н.А. Python 3 и PyQt. Разработка приложений. – СПб.: БХВ-Петербург, 2012. – 704 с.
5. Натан Марц, Джеймс Уоррен. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени. – М.: Вильямс. – 2016. – 292 с.
6. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистика. 2003. – 352 с.
7. Овечкин Г.В. Компьютерное моделирование: Учебник. – М.: Academia, 2017. – 379 с.

References

1. Mustafina S.I., Akimov A.A. Neural network technologies in modeling digital twins of socially significant objects // Modern Physics, Mathematics, Digital and Nanotechnologies in Science and Education. Collection of abstracts of the All-Russian Youth School-Conference dedicated to the 100th anniversary of A.D. Sakharov's birth. – Ufa: Bashkir State Pedagogical University named after M. Akmulla, 2022. – P. 111.
2. Akimov A.A., Valitov D.R., Kubryak A.I. Data preprocessing for machine learning // Scientific Review. Technical Sciences. 2022, no. 2, pp. 26-31.
3. Beazley D. Python. Detailed Reference. Translated from English. – SPb.: Symbol-Plus, 2010. – 864 p.
4. Prokhorenok, N.A. Python 3 and PyQt. Application Development. – SPb.: BHV-Petersburg, 2012. – 704 p.
5. Nathan Marz, James Warren. Big Data. Principles and Practices of Building Scalable Real-Time Data Processing Systems. – М.: Williams, 2016. – 292 p.
6. Dubrov A.M., Mkhitarian V.S., Troshin L.I. Multivariate Statistical Methods. – Moscow: Finance and statistics, 2003. – 352 p.
7. Ovechkin G.V. Computer Modeling: Textbook. – М.: Academia, 2017. – 379 p.

Голубничий Антон Сергеевич – магистр	Golubnichiy Anton Sergeevich – master student
Семёнов Дмитрий Валентинович – аспирант	Semyonov Dmitry Valentinovich – postgraduate student
ohsenpai1337@gmail.com	

Received 14.01.2024