

Кальсин Дмитрий Владимирович, магистрант,
кафедра информационных технологий и вычислительных систем,
Московский государственный технический университет «СТАНКИН»,
РФ, г. Москва

Научный руководитель:
Новоселова Ольга Вячеславовна, к.т.н., доц.,
Московский государственный технический университет «СТАНКИН»,
РФ, г. Москва

ИСПОЛЬЗОВАНИЕ КОРРЕЛЯЦИОННОГО АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ СКРЫТЫХ ЗАВИСИМОСТЕЙ

Аннотация: В статье рассматривается использование корреляционного анализа для выявления скрытых зависимостей между метриками, связанными с онбордингом пользователей.

Ключевые слова: Корреляционный анализ, онбординг, адаптация пользователей, пользовательский опыт, оптимизация интерфейса, метрики, прогнозирование поведения пользователей, статистический анализ.

Корреляционный анализ является важным инструментом для анализа и выявления взаимосвязей между различными метриками. В контексте онбординга, например, можно исследовать взаимосвязь между такими метриками, как процент завершения ознакомительного тура и время прохождения тура. Эти метрики представляют собой два аспекта поведения пользователей: как быстро они проходят тур и насколько полно они его завершают. Несмотря на очевидную зависимость, что пользователи, завершившие тур, вероятно, тратят на него определенное время, корреляционный анализ может выявить более глубокие и менее очевидные закономерности.

Корреляция – это статистический показатель, который измеряет степень взаимосвязи между двумя переменными. Например, исследуя, как время прохождения тура влияет на процент завершивших тур, можно получить понимание, как продолжительность тура соотносится с его успешностью. Одной из задач является определение, существует ли «идеальная» продолжительность для различных типов пользователей или же зависимость между временем и завершением тура линейна или имеет нелинейный характер.

В теории, чем меньше времени тратит пользователь, тем выше вероятность завершения тура. Однако на практике может оказаться, что слишком короткий тур может не дать пользователю достаточно информации для принятия решения о дальнейшем использовании продукта, в то время как слишком длинный тур может привести к усталости и снижению конверсии.

Применение кросс-методического анализа для более глубокого понимания сложных взаимосвязей. Например, можно исследовать, как длительность тура и тип контента (например, текстовый или визуальный) взаимодействуют с такими метриками, как процент активных пользователей после завершения тура. Это позволит выделить оптимальные параметры для различных сегментов аудитории. Возможно, для одних пользователей будет оптимально более короткое время, а для других – более продолжительный тур с подробными объяснениями.

Предположим, вы анализируете данные и обнаруживаете, что у пользователей с высоким уровнем вовлеченности время прохождения тура значительно выше, чем у менее вовлеченных пользователей. Это может говорить о том, что для таких пользователей дополнительное время, потраченное на более подробное объяснение, будет полезно. В то же время, если пользователи с низким уровнем вовлеченности тратят много времени на прохождение тура, это может указывать на проблемы с интерфейсом или содержанием тура.



Для улучшения понимания этой взаимосвязи можно добавить дополнительный слой анализа, используя кластеризацию пользователей. Кластеризация позволит сегментировать пользователей на подгруппы, основываясь на различных характеристиках (например, уровень опыта, поведение на сайте), и затем проанализировать, как их поведение в онбординге соотносится с метриками времени и завершения тура.

Для более точной настройки онбординга важно не только выявить зависимость между временем и завершением тура, но и учитывать дополнительные аспекты, такие как темп обучения и структура тура. Например, можно исследовать, как распределение времени по шагам тура влияет на завершение. Важно ли, чтобы каждый шаг был коротким и лаконичным, или стоит больше времени уделить наиболее сложным или важным этапам? Такой подход может значительно повысить точность анализа и прогнозирования пользовательского поведения.

Кроме того, использование методов машинного обучения, таких как регрессия или деревья решений, позволяет не только выявить зависимости, но и строить прогнозные модели. Эти модели могут предсказывать, какой будет вероятность завершения тура в зависимости от его длительности и других факторов, что позволит точно настраивать продукт под различные категории пользователей [1].

Используя результат анализа, можно создать модели прогнозирования, которые позволят с большей точностью предсказать поведение пользователей. Например, на основе корреляционных данных между временем прохождения тура и активностью пользователей можно предсказать, какой тип онбординга будет наиболее эффективен для разных типов пользователей.

Одной из возможных инноваций является создание прогностических моделей, которые бы анализировали реальное поведение пользователей в момент прохождения тура, а не только его результаты. Например, можно следить за первоначальными этапами прохождения тура, чтобы предсказать, завершит ли пользователь его в целом. Это позволяет в режиме реального времени подстраивать подход к пользователю, предоставляя ему адаптивный контент в зависимости от текущего поведения.

В рамках магистерской работы проведен комплексный анализ эффективности онбординговых решений с применением корреляционного и хи-квадрат тестов. Результаты показали статистически значимую связь между взаимодействием с контекстными подсказками (тест "Б") и целевым действием: коэффициент корреляции Спирмена составил 0.435 ($p < 0.001$) для основной подсказки, 0.250 ($p < 0.001$) и 0.187 ($p = 0.004$) для дополнительных, что подтверждается значениями χ^2 (38.619, 3.098 и 6.044 соответственно). В отличие от этого, онбординг в формате многошагового тура (тест "А") продемонстрировала слабую корреляцию с конверсией ($\phi = 0.225$). Полученные данные свидетельствуют о большей эффективности простых решений для продуктов с простой логикой, что согласуется с принципами снижения когнитивной нагрузки [2] и подтверждает выдвинутую гипотезу о нецелесообразности использования сложных туров в подобных сценариях.

Применение корреляционного анализа в сочетании с более сложными статистическими методами и моделями машинного обучения позволяет глубже понять поведение пользователей и оптимизировать онбординг. Включение в анализ дополнительных факторов, таких как темп прохождения тура, тип контента и его структура, позволяет не только повысить эффективность, но и настроить более персонализированный опыт для различных сегментов аудитории.

Таким образом, исследование взаимосвязей между метриками и использование статистических моделей для прогнозирования поведения пользователей имеет значительный потенциал для оптимизации интерфейса и улучшения пользовательского опыта.



Список литературы:

1. Kohavi, R., Tang, D., & Xu, Y. *Online A/B Testing: A Practical Guide* [Книга]. – New York: Springer, 2021. – ISBN: 978-3-030-61476-3. – 225 с. – (дата обращения: 10.12.2024). Текст: непосредственный.
2. Sweller J. Cognitive load during problem solving: Effects on learning [Статья] // *Cognitive Science*. – 1988. – Vol. 12, № 2. – P. 257-285. – ISSN: 0364-0213. – DOI: 10.1207/s15516709cog1202_4. – (дата обращения: 10.12.2024). Текст: электронный.

