

Применение интеллектуального анализа данных в задаче оптимизации контекстной рекламы

Б.О. Непряхин
Самарский национальный
исследовательский университет им.
академика С.П. Королева
Самара, Россия
ceo@bjjobs.ru

Аннотация — В статье рассматривается применение методов интеллектуального анализа данных для оптимизации рекламных кампаний в Яндекс Директ. Проведен анализ данных пяти рекламных кабинетов в сфере недвижимости, включающий 7 миллионов записей о пользовательских сессиях с 314 характеристиками, описывающими интересы пользователей второго и более глубоких уровней. В работе предложен подход к оптимизации контекстной рекламы, основанный на последовательном процессе анализа данных, выделении ключевых факторов конверсии и последующей настройке параметров рекламных кампаний. Рассмотрены основные этапы реализации последовательного процесса интеллектуального анализа. Полученные результаты позволили выделить важнейшие признаки пользователей и сформировать рекомендации по эффективному целевому охвату.

Ключевые слова — машинное обучение, интернет-реклама, анализ данных, кластеризация, предсказание конверсий, оптимизация рекламы, data mining.

I. ВВЕДЕНИЕ

В условиях высокой конкуренции в интернет-рекламе важно эффективно управлять кампаниями для минимизации затрат и максимизации конверсий. Контекстная реклама в системе Яндекс Директ позволяет провести глубокую аналитику, однако традиционные методы часто не позволяют выявить скрытые закономерности в поведении пользователей [1].

Основные проблемы в анализе интернет-рекламы:

- Большой объём информации и гетерогенность данных.
- Высокая динамичность изменений.
- Сложность учета множества факторов, влияющих на конверсии.
- Недостаточная регулируемость базовых методов оптимизации контекстной рекламы.

Для решения этих проблем в работе предложен подход к оптимизации контекстной рекламы, включающий предобработку, анализ данных, сегментацию аудитории и динамическую настройку ставок. Отличительная особенность подхода — мультимодальность, то есть использование спектра алгоритмов для классификации и кластеризации пользователей.

II. ПОДГОТОВКА К ИССЛЕДОВАНИЮ

Исследуемая выборка включает 7 миллионов записей из пяти рекламных кабинетов. Каждый пользователь описывается 314 характеристиками, включающими возраст, пол, временные параметры сессий и категории интересов второго и более глубоких уровней.

Этапы подготовки данных:

- Удаление дубликатов и пропущенных значений.
- Преобразование категориальных признаков (One-Hot Encoding, Label Encoding).
- Разделение данных на обучающую и тестовую выборки (80/20).
- Анализ значимости признаков (SHAP, корреляционный анализ, PCA) [2].

III. ПРОВЕДЕНИЕ ИССЛЕДОВАНИЯ

В рамках исследования проведена классификация данных с целью выделения целевых признаков, на основе которых выполнена последующая кластеризация. Для анализа применялись следующие методы:

Классификация:

- 1) *Decision Tree* — выявлены ключевые признаки конверсии, но модель склонна к переобучению.
- 2) *Random Forest* — получены стабильные результаты при этом переобучение устранено.
- 3) *Градиентный бустинг (XGBoost, LightGBM, CatBoost)* — подтверждены выявленные значимые признаки и обеспечил наибольшую точность классификации.

Исследование и настройка модели в процессе классификации позволяет выделить те параметры, которые в большей степени влияют на конверсию. Основным результатом этапа классификации состоит в том, что мы имеем возможность отобрать ключевые параметры, которые будут передаваться на этап кластеризации данных.

Кластеризация:

- 1) *KMeans* — разделение пользователей на сегменты с чёткими границами.
- 2) *DBSCAN* — выделение аномальных групп и плотных кластеров.
- 3) *Иерархическая кластеризация* — выявление вложенных сегментов пользователей [3].

Кластеризация позволила осуществить сегментирование пользователей с похожими характеристиками. За счет сегментации стала возможной оптимизация рекламных стратегий для разных сегментов пользователей и решен вопрос обнаружения аномальных пользовательских сегментов, выявлены скрытые паттерны в данных. Проведенная сегментация позволила сфокусировать ставки рекламы на более релевантных аудиториях.

IV. ЗАКЛЮЧЕНИЕ

Проведённое исследование подтвердило, что методы интеллектуального анализа данных могут значительно повысить эффективность рекламных кампаний в Яндекс Директ. Используемый подход оптимизации контекстной рекламы включает структурированную последовательность действий — от анализа данных до динамической настройки рекламных ставок.

Отличительной особенностью работы является мультимодальный подход, использующий сочетание классификационных моделей и методов кластеризации. Это позволяет точнее сегментировать аудиторию и более эффективно перераспределять рекламный бюджет.

Перспективным направлением дальнейших исследований является расширение применения нейросетевых моделей и учет временных зависимостей в поведении пользователей. Разработка более точных моделей прогнозирования конверсий и анализ динамики пользовательской активности позволит повысить эффективность рекламных кампаний и предложить персонализированные решения для аудитории.

ЛИТЕРАТУРА

- [1] Михайлов Д.В. Анализ практики применения алгоритмов машинного обучения в российских розничных сетях для развития операционной деятельности и разработки маркетинговой стратегии предприятия / Д.В. Михайлов // Сборник научных статей по материалам XII Международной научно-практической конференции. Уфа. – 2023. – С. 122–141.
- [2] Панская М.А. Использование искусственного интеллекта в маркетинге / М.А. Панская // Новое в науке и образовании. Материалы Международной ежегодной научно-практической конференции. Ростов-на-Дону. – 2023. – С. 326 – 332.
- [3] Кричевский М.Л. Использование методов машинного обучения для оценки инвестиционной деятельности различных регионов России / М.Л. Кричевский, Ю.А. Мартынова // Вопросы инновационной экономики. – 2019. – Т. 9(4). – С. 1557 – 1572.