

Тематика кейса	Рекламные технологии
	На наш взгляд, кейс имеет научную составляющую, поскольку в рамках данного кейса будет разработано программное решение, позволяющее определять профили аудитории в мобильной среде, что обеспечит максимально точно попадание в социально-демографический профиль аудитории (пол и возраст), интересы и особенности поведения аудитории, при этом сохраняя приватность пользователей. Данное программное решение будет создано Hyper AdTech в рамках осуществляемой исследовательской деятельности
Технологии кейса	Big Data Machine learning
Краткое описание кейса	Разработка технологического решения, обеспечивающего определение профиля аудитории (обезличенные данные по социально-демографическим параметрам, интересам аудитории и особенностям поведения) в мобильных приложениях
Краткое наименование кейса	Hyper AdTech
Описание кейса для сайта конкурса	<p>Мы живем в эру, где определить человека в цифровой среде не составляет особого труда. Но наступает новая эра – эра «cookie less» и «post idfa» где конфиденциальность данных пользователей лежит в основе любого сервиса и продукта. Рекламный рынок уже давно функционирует в тесной привязке к обезличенным данным аудитории и ее поведению в цифровой среде. Для рекламодателей это обеспечивает таргетирование на релевантную аудиторию и проведение эффективных рекламных кампаний. За последнее время со стороны глобальных игроков в цифровой среде начали предприниматься различные меры, которые способствуют ограничению передачи данных аудитории по идентификаторам мобильных устройств (IDFA).</p> <p>Рекламная платформа Hyper AdTech, специализирующаяся на размещении видеорекламы в мобильных приложениях поставила перед собой задачу создать решение, которое позволило бы максимально точно определить профиль аудитории в мобильной среде на основе различных косвенных и исторических аудиторных данных.</p> <p>В рамках решения кейса участникам предлагается разработать программное решение, которое обеспечит определение заданных профилей аудитории в мобильной среде по полу, возрасту, интересам, поведению, месторасположению и другим параметрам.</p>
Расширенное описание кейса	<p>Для решения кейса участникам предоставляется Dataset*.</p> <p><u>Dataset состоит из:</u></p> <ol style="list-style-type: none"> <li>Таргетов. В данном случае это сегменты целевых аудиторий: сегменты целевых аудиторий представляют разные группы пользователей по соцдем профилю и интересам);</li> <li>Признаков трафика. Признаками трафика являются: категория и ID мобильного приложения, дата и время запроса рекламы по Московскому времени, часовой пояс и географическая локация</li> </ol>

	<p>пользователя.</p> <p><u>В предоставленной выборке Dataset данные по 3-м неделям:</u></p> <p>05-11.07.21 02-08.08.21 15-21.09.21</p> <p><u>В рамках кейса необходимо решить следующие задачи:</u></p> <ol style="list-style-type: none"> <li>1. Обучить классификатор предсказывать сегмент по признакам на тренировочной выборке и проверить работу классификатора на тестовой выборке, в которой отсутствуют метки сегментов;</li> <li>2. Используя таблицу признаков, кластеризовать пользователей и сравнить найденные кластеры с сегментами, представленными в Dataset.</li> </ol>														
<p>Диагностика проблемы, определение главной и второстепенных проблем</p>	<p>Помимо очевидных этических и юридических доводов, есть еще и динамика наличия цифровых идентификаторов в реальных устройствах. На данный момент (28 октября 2021 г.) это выглядит так:</p> <table border="1" data-bbox="600 792 1264 1182"> <thead> <tr> <th>ОС</th><th>Процент неидентифицируемых устройств</th></tr> </thead> <tbody> <tr> <td>iOS 15</td><td>76%</td></tr> <tr> <td>iOS 14</td><td>75%</td></tr> <tr> <td>iOS 13</td><td>11%</td></tr> <tr> <td>iOS остальные версии</td><td>9%</td></tr> <tr> <td>Android все версии</td><td>2.5%</td></tr> </tbody> </table> <p>Сравним с состоянием на год назад, для даты 28 октября 2020 г.:</p> <table border="1" data-bbox="600 1289 1264 1370"> <thead> <tr> <th>ОС</th><th>Процент неидентифицируемых устройств</th></tr> </thead> <tbody> </tbody> </table>	ОС	Процент неидентифицируемых устройств	iOS 15	76%	iOS 14	75%	iOS 13	11%	iOS остальные версии	9%	Android все версии	2.5%	ОС	Процент неидентифицируемых устройств
ОС	Процент неидентифицируемых устройств														
iOS 15	76%														
iOS 14	75%														
iOS 13	11%														
iOS остальные версии	9%														
Android все версии	2.5%														
ОС	Процент неидентифицируемых устройств														



	<p>Результат нужно представить в виде таблицы с двумя колонками, где первая колонка - это соответствующий номер записи, а вторая колонка - это предсказанный сегмент.</p> <p>2. Результат по второй задаче: Алгоритм, который на входе получает аналогичную п.1 информацию о признаках трафика, и выводит процентное отношение количества каждого определенного вами кластера. Кроме того необходимо обосновать полученный результат, то есть описать признаки трафика, который принадлежит выделенному вами кластеру.</p> <p>По возможности также нужно обосновать результат; если это предполагает выбранный вами алгоритм, то дополнительно выявленными вами признаками принадлежности к кластеру.</p> <p>3. Требование к презентации результата: Представить процент охвата* по сегментам. Для этого нужно посчитать какой процент каждый из типов предсказанных сегментов занимает по отношению ко всей тестовой выборке.</p> <p>*охват - доля сегмента, или же вероятность получить сегмент в трафике.</p>
Сценарий использования результата	В результате разработанного решения рекламная платформа Hyper AdTech получит технологию, которая обеспечит определение пользователей в мобильной среде по полу, возрасту, интересам, поведенческим данным, месторасположению и другим параметрам, которые необходимы для таргетирования и эффективного проведения рекламных кампаний рекламодателей в мобильной среде
Требования к результату	У наших экспертов будут разные наборы проверочных входных данных , за разные периоды времени и с разным распределением доли охвата сегментов. Есть возможность проверить качество работы вашего решения с помощью специальных методик, принятых в рекламной индустрии.