

АЛГОРИТМЫ АНАЛИЗА ПОЛЬЗОВАТЕЛЬСКОЙ АКТИВНОСТИ

В мире разработки современного программного обеспечения понятие аналитической утилиты давно стало бизнес-требованием, ведь анализ и дальнейшее прогнозирование пользовательской активности в тех или иных частях продукта помогает явным образом влиять на оценочную стоимость всей системы.

Если обобщить понятие аналитики современных веб и мобильных приложений, то оно формулируется как процесс анализа, прогнозирования и выявления тенденций измеряемых критериев на основании аналитических данных. Зачастую источником аналитических данных является сам продукт или приложение, в которое встраивается дополнительное программное обеспечение целью которого является сбор и передачи информационных блоков. Подобный информационный блок может содержать информацию о способе взаимодействия пользователя с приложением, времени совершения того или иного события, а также дополнительные нагрузочные параметры, определенные бизнес-требованиями.

Поскольку одной из ключевых целей сбора аналитической информации является дальнейшее прогнозирование пользовательской активности, то есть выявление паттернов и алгоритмов взаимодействия пользователя с программным продуктом, то одним из наиболее важных этапов настройки аналитических утилит является выбор актуальных для данной конкретной системы критериев анализа. Среди общих критериев можно отметить:

1. Количество активных пользователей на заданном этапе, как правило это дневные, месячные или квартальные показатели;
2. Коэффициент удержания или цифровое представление уровня конкурентоспособности продукта;
3. Местоположения и места пребывания основной аудитории и непосредственных пользователей.

Основная часть аналитических утилит доступных для использования предоставляет возможность работы с приведенными критериями, но каждая из такого рода платформ имеет свои особенности.

Среди наиболее популярных стоит отметить Google Firebase, Mixpanel а также Flurry. Каждая платформа предоставляет широкий набор аналитических инструментов, но тем не менее основным критерием в пользу выбора той или иной являются уникальные особенности система. Например, Google Firebase станет отличным решением

если программное средство поддерживает или взаимодействует с другими элементами платформы, Mixpanel может стать оптимальным решением если требованием является максимально гибкая настройка, а ориентированность Flurry на мобильные операционные системы позволяет продуктивно её использовать в мобильных решениях.

В общем случае модель любой аналитической системы строится на основании одного или нескольких алгоритмов анализа. Как правило данные алгоритмы реализуются на уровне хранилища аналитических данных. Так, например, программное средство SQL Server, поддерживает следующие алгоритмы:

1. Алгоритмы классификации, осуществляющие прогнозирование одной или нескольких дискретных переменных на основе других атрибутов в наборе данных;

2. Регрессивные алгоритмы, которые могут осуществлять прогнозирование одной или нескольких непрерывных числовых переменных, например, прибыли или убытков, на основе других атрибутов в наборе данных;

3. Алгоритмы сегментации, необходимые для деления данных на группы или кластеры элементов, имеющих схожие свойства;

4. Алгоритмы взаимосвязей, с помощью которых можно осуществлять поиск корреляции между различными атрибутами в наборе данных. Наиболее частым применением этого типа алгоритма является создание правил взаимосвязи, которые могут использоваться для анализа потребительской корзины;

5. Алгоритмы анализа последовательностей, которые обобщают часто встречающиеся в данных последовательности, такие как серия переходов по веб-сайту или событий, зарегистрированных в журнале перед ремонтом оборудования.

Ввиду специфичности задачи анализа пользовательской активности, на практике часто используются комбинации нескольких методов, то есть в общем случае идеальных решений в данной области нет. Тем не менее, на основании комбинаций алгоритмов анализа данных можно сформировать формулу расчета коэффициентов для данного специфичного продукта или системы. Данная формула может содержать неограниченное количество входных параметров, для анализа определения которых могут использоваться любые системы.