

Кравцова Валерия Павловна,
магистрант II курса направления «Бизнес-информатика» ФГБОУ ВО
«Кубанский государственный университет», г. Краснодар
kravtsova963@gmail.com



Основные аспекты моделирования вероятности отклика

Аннотация. В данной статье рассматриваются некоторые аспекты, связанные с моделированием вероятности отклика клиентов на проводимые кампании по предложению товаров и услуг. Описываются основные проблемы формирования целевой переменной при подготовке данных для разработки скоринга отклика. Рассматривается метод оценки значимости переменных с использованием WoE/IV анализа.

Ключевые слова: скоринг, скоринг отклика, зависимая переменная, независимые переменные, WoE-анализ, IV-анализ, маркетинговые кампании.

Раздел: (4) экономика.

Современный маркетинг все чаще базируется на анализе данных и носит название data driving маркетинг (взаимоотношения с клиентами на основе данных). Data mining (интеллектуальный анализ данных) применяется для ответа на многие вопросы в анализе клиентов организациями не только B2C отраслей, но и государственными структурами. Типичными задачами предиктивной аналитики для digital маркетинга являются:

- прогноз будущего поведения клиента по приобретению продуктов и услуг;
- вероятность ухода клиента;
- вероятность отклика клиента;
- разработка продуктов и услуг под различные сегменты клиентов. Сегменты

выявленные на основе анализа данных о подобных известных организации клиентах.

В финансовой сфере, а именно в ее банковской составляющей, data mining в первую очередь пришел в специфичную для этого бизнеса задачу – задачу управления рисками.

Статистические методы анализа данных позволили Банкам на основе прогнозных моделей оценивать вероятности наступления главного бинарного события – вероятность наступления дефолта по обязательствам. Таким образом, данные методы позволили принимать взвешенное по риску решение о предоставлении кредита. Данный тип задачи называется общим словом – скоринг, от английского scoring – подсчет очков.

Со временем при увеличении каналов коммуникации с клиентами решение задачи оценки вероятности успешности конкретной коммуникации перешло в банковский маркетинг и получило название – скоринг отклика.

Скоринг отклика – это оценка вероятности того, что потенциальный или действующий клиент отреагирует на осуществленное воздействие.

Эффективное решение задачи скоринга отклика позволяет значительно снизить затраты на проведение кампаний по привлечению клиентов на определенные продуктовые предложения.

Технологии стремительно развиваются и даже небольшие компании могут позволить себе организацию массовых воздействий с низкой стоимостью адресного донесения информации по средствам:

- SMS-сообщений;
- e-mail-рассылок;
- телемаркетинга;
- таргетированной рекламы в социальных сетях;
- push-уведомлений.

Снижающаяся стоимость одного адресного воздействия порождает проблему большого количества безрезультативных рекламных сообщений. Что в свою очередь влечет за собой не самое эффективное использование маркетингового бюджета.

С помощью методов анализа данных, а в частности скоринга отклика тот же бюджет можно использовать более эффективно – направив маркетинговое воздействие только на тех клиентов, кто вероятнее всего на него откликнется.

Так что же из себя представляет решение задачи оценки вероятности отклика клиента на предложенный продукт или услугу?

В первую очередь для решения данной задачи необходимы исторические данные по проводимым ранее воздействиям с фиксированием результата прошедшей компании. Это необходимо для формирования зависимой переменной уравнения оценки вероятности. Зависимая переменная содержит в себе информацию об успешности проведенной коммуникации, а так же предыдущий опыт служит для установки точки отсчета в эффективности проводимых воздействий.

Чаще всего предыдущие, исторические воздействия в организациях, которые еще не используют в своей деятельности data mining-инструменты представляют из себя либо беспорядочные, отправлявшиеся всем подряд SMS или e-mail сообщения, либо данные по массовым «холодным» обзвонам. Хуже, если ранее все воздействия проводились с предварительной кластеризацией клиентов методом «пальцем в небо». Например, «отправим всем разведенным мужчинам без детей, но с автокредитом, предложение кредитной карты и депозита». В таком случае, воздействия оказывались на клиентов, которые скорее всего не соответствуют генеральной совокупности клиентской базы организации.

Для оценки зависимой переменной «отклик» необходимы наборы независимых переменных. Независимые переменные – это данные, которые содержат в себе информацию, объясняющую, почему наступает то или иное событие. Почему кто-то платит лучше или хуже, кто-то отреагирует на предложение продукта, а кто-то нет, сработает терапия или нет.

Чаще всего независимые переменные представляют собой набор социально-демографических характеристик, таких как:

- пол;
- возраст;
- семейное положение;
- наличие детей или их количество;
- тип населенного пункта проживания;
- наличие или отсутствие какого-либо имущества;
- и прочие социальные позиции.

Так же важно использовать в наборе данных различные поведенческие паттерны, для скоринга отклика на рекламное предложение в банке это может быть финансовое поведение клиента (накопления, кредитование или траты только собственных средств), характер совершаемых операций, данные указывающие на соблюдение ежемесячного бюджета или, напротив, на склонность к спонтанным тратам.

Банки о своих клиентах скапливают огромное количество информации, с помощью которой можно сформировать огромное количество переменных.

Так какие же переменные включать в набор для разработки модели? Только те, которые несут в себе больше полезной информации. Для оценки предсказательной силы переменных используют метод оценки переменных – *WoE/IV* анализ.

Критерий «информационная ценность/значение»:

$$IV = \sum_i (DistrGood_i - DistrBad_i) \ln \left(\frac{DistrGood_i}{DistrBad_i} \right). \quad (1)$$

где *DistrGood_i* – доля откликнувшихся клиентов *i*-го атрибута от общего числа клиентов;

DistrBad_i – доля проигнорировавших клиентов *i*-го атрибута от общего числа клиентов.

При определении значимости характеристики и выявлении разницы между «плохими» и «хорошими» клиентами, самой распространенной мерой является именно критерий IV (информационная ценность) (см. таблицу).

Статистическая мощность IV

| Значения IV | Статистическая мощность |
|-------------|-------------------------------------|
| Менее 0,02 | Очень слабое статистическое влияние |
| 0,02 – 0,1 | Слабое статистическое влияние |
| 0,1 – 0,3 | Среднее статистическое влияние |
| 0,3 – 0,5 | Сильное статистическое влияние |
| Более 0,5 | Чрезвычайно сильное влияние |

Необходимо помнить о том, что характеристики, имеющие чрезвычайно сильное статистическое влияние, могут подавлять воздействие других, менее информативных характеристик, в связи с чем, их использование при построении скоринговой карты требует особого внимания.

Определив нужные переменные, оказывающие наибольшее статистическое влияние, рассчитывается статистическая значимость каждого атрибута с помощью критерия *WoE* (вес доказательства):

$$WoE_i = \ln \left(\frac{DistrGood_i}{DistrBad_i} \right). \quad (2)$$

Отрицательные значения коэффициентов *WoE* указывают на большую вероятность появления событий в интервале (см. рисунок) [2].

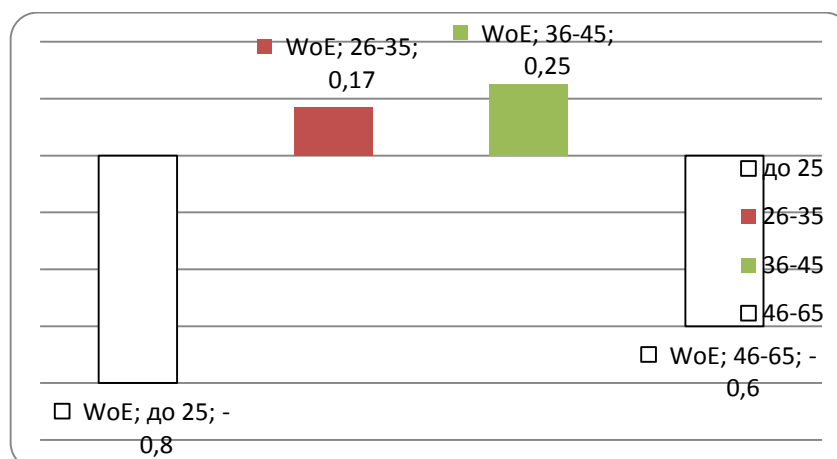


Диаграмма индексов WoE

После того как наиболее значимые переменные найдены, строится модель логистической регрессии.

Логистическая регрессия сегодня остается наиболее популярным инструментом для разработки скоринговых моделей, так как позволяет получать хорошо интерпретируемые балльные скоринговые карты и вероятностные оценки наступления события для конкретного испытуемого (больной/здоровый, возврат кредита/дефолт, отклик/игнорирование и т. д.).

Взаимосвязи между независимой переменной и событием не всегда линейные. Несмотря на это, уравнение логистической регрессии, все равно моделирует линейные зависимости между входами (переменными) и выходами (событиями).

Формирование переменной «события» является наиболее важной задачей при построении модели и зависит от ее целей.

Микрофинансовые организации, строящие скоринговые модели, придерживаются цели максимизировать доход, и событием будет являться возврат займа в срок.

В ритейле, при предотвращении ухода клиента, событие – клиент не совершает покупки. В телекоме – клиент расторгнул контракт.

При формировании маркетинговых воздействий, целью будет являться отклик клиента на персональное предложение, а значит, зависимая переменная – непосредственно сам факт отклика [1].

Со стороны выглядит все не сложно, но если задуматься, появляется ряд вопросов:

- Как фиксировать событие?
- В течение какого временного интервала нужно фиксировать событие?
- Помогает ли выбранное событие и метод его определения решить поставленную бизнес-задачу?

Проведя маркетинговое воздействие, малая доля клиентов, которую ваше предложение заинтересовало, обратится к вам в тот же день. И через день тоже. Так что прежде чем брать за событие «отклик клиента», необходимо выявить временной интервал, в течение которого обращение клиента можно считать откликом на проведенную кампанию.

Разработка скоринговых моделей отклика, и только потом запуск маркетинговых кампаний на определенных клиентов, которые с наибольшей вероятностью от-

реагируют на нее, помогает организациям существенно сократить затраты на воздействия, а так же снизить репутационный риск.

Бизнес движется сегодня в сторону наилучшего понимания своего клиента – его предпочтений, намерений, ожиданий. Коммуникация должна строиться на узких сегментах, собранных по строго заданным признакам.

Ссылки на источники

1. Определение события в скоринговых моделях. – URL: <https://basegroup.ru/community/blogs/learning/6655>.
2. Паклин Н., Афанасьев В. Оптимальное квантование для повышения качества бинарных классификаторов // Искусственный интеллект. – 2013. – № 4.

Valerya Kravtsova,

student of II course of the direction "Business Informatics", Kuban State University, Krasnodar
kravtsova963@gmail.com

The main aspects of modeling the probability of response

Abstract. This article discusses some aspects related to modeling the probability of customer response to the campaign for the supply of goods and services. It describes the main problems of formation of the target variable in the training data for the development of a scoring response. The method of assessing the significance of variables using WoE/IV analysis.

Key words: credit scoring, response scoring, dependent variable, independent variables, WoE-analysis IV-analysis, marketing campaign.

Рекомендовано к публикации:

Бочковой Е. В., кандидатом экономических наук;
 Горевым П. М., кандидатом педагогических наук,
 главным редактором журнала «Концепт»



| | | | |
|---|----------|--|----------|
| Поступила в редакцию <i>Received</i> | 23.05.16 | Получена положительная рецензия <i>Received a positive review</i> | 24.05.16 |
| Принята к публикации <i>Accepted for publication</i> | 24.05.16 | Опубликована <i>Published</i> | 25.05.16 |

www.e-koncept.ru

© Концепт, научно-методический электронный журнал, 2016
 © Кравцова В. П., 2016