

Исследование применимости модели бинарного выбора при анализе оттока в телекоммуникационной отрасли

Панов В. А.

Студент 2-го курса магистратуры, Финансово-экономический факультет
Финансовый Университет при Правительстве РФ
Россия, г. Москва
v.panov2603@gmail.com

Научный руководитель: Борисова О. В. к.э.н., доцент Департамента корпоративных финансов и корпоративного управления
Финансовый Университет при Правительстве РФ

Аннотация. В статье рассматривается актуальность проблемы оттока клиентов в телекоммуникационном секторе, а также преимущества использования эконометрических моделей двоичного выбора, позволяющих предсказывать вероятность оттока абонентской базы.

Ключевые слова: отток клиентов, телекоммуникационная отрасль, двоичный выбор, логистическая регрессия.

STUDY OF APPLICABILITY OF THE BINARY SELECTION MODEL IN ANALYSIS OF OUTFLOW IN THE TELECOMMUNICATION INDUSTRY

Abstract. The article discusses the relevance of the problem of customer churn in the telecommunications sector; as well as the benefits of using econometric binary selection models to predict the probability of customer's base churn.

Keywords: Customer churn, telecommunications industry, binary choice, logistic regression.

Важнейшей современной характеристикой телекоммуникационной отрасли является увеличивающийся отток клиентов по всему миру. Причинами служат многие факторы, среди которых можно отметить сокращающиеся возможности дифференцирования продуктов и услуг, возрастающее ценовое давление, насыщение рынка, сокращение жизненного цикла многих продуктов и услуг. Для примера: в России в 2018 году 39% абонентов сменили оператора мобильной связи [1, с. 41]. Многие телекоммуникационные компании уходят от измерения успешности собственной деятельности исключительно через количество клиентов, уделяя повышенное внимание продолжительности оказания услуг и продажи продуктов. Подобная ситуация приводит к тому, что управление клиентской базой с целью сокращения оттока становится приоритетной работой руководителей телекоммуникационных компаний. Компания McKinsey & Company на основании данных, полученных в рамках взаимодействия с телекоммуникационным сектором по всему миру, утверждает, что тем компаниям, которые осуществляют работу с клиентской базой на основе развернутой аналитики, удастся сократить отток в среднем на 15%. Аналитика должна не просто включать профили клиентов, тарифные планы, специальные предложения, но также учитывать информацию из сети, данные от контактных центров, специфику маркетинговых мероприятий и расходы на них, цены и акции конкурентов.

Ключевое значение в рамках работы по управлению клиентской базой имеет применение современных аналитических методов, при помощи которых становится возможным производить разносторонний анализ значительных объемов разнообразной информации без необходимости осуществления постоянного программирования. Преимуществом телекоммуникационных компаний является наличие у них обширных возможностей для применения передовых аналитических методов, включая имеющиеся колоссальные объемы информации, развитую инфраструктуру. Использование современных подходов к аналитике позволяет выявлять скрытые факторы или их комбинации, помогая предсказывать поведение клиента с целью поиска оптимального персонализированного решения для предотвращения возможного отказа от продуктов и услуг телекоммуникационной компании.

Для целей разработки модели предсказания поведения клиентов в телекоммуникационном секторе рассмотрим применение моделей двоичного (бинарного) выбора, которые позволяют закодировать возможные варианты развития ситуации в виде нуля и единицы. Применение привычной линейной модели в подобной ситуации затруднено потому, что расчетные значения переменных не могут выходить за пределы интервала $[0; 1]$, а также по причине смещенности оценок, гетероскедастичности и неприменимости коэффициента детерминации (R^2). Модели бинарного выбора представляют собой частный случай моделей дискретного выбора, для которых характерным является наличие дискретной зависимой переменной. Среди моделей двоичного выбора внимания заслуживают логит-модель и пробит-модель, которые в настоящее время опробованы и успешно применяются

для определения вероятностей различных событий в таких сферах деятельности, как медицина, экология, социология, экономические исследования, банковский сектор.

Логит-модель (логистическая регрессия; Logit model) представляет собой модель, которая применяется для прогнозирования вероятности какого-либо события в зависимости от значения некоторого количества факторов с применением логистической функции. Логистическая регрессия предназначена для вычисления значения зависимой переменной, которая может принимать значения в интервале $[0;1]$.

Вероятность оцениваемого события в рамках логит-модели определяется на основании формулы [3, с.153]:

$$P_i = F(Z_i) = \frac{1}{1 + e^{-Z_i}}, \quad (1)$$

где Z_i представляет собой линейную комбинацию независимых факторов:

$$Z_i = \beta_1 + \beta_i X_i \quad (2)$$

Логистическая регрессия лишена ключевого недостатка линейной модели (вероятность может выходить за пределы интервала $[0;1]$), из-за которого последняя не может применяться для целей предсказания поведения клиентов. Благодаря тому что в случае, когда $Z \rightarrow \infty$, $e^{-Z} \rightarrow 0$, максимальная вероятность события не может превысить 1. В ситуации, когда $Z \rightarrow -\infty$, $e^{-Z} \rightarrow \infty$, вероятность события не может быть менее 0 [6, с.201].

Оценка параметров логит-модели осуществляется на основании метода максимального правдоподобия, когда оценка регрессионных коэффициентов сводится к получению максимальной вероятности возникновения конкретной выборки. При методе максимального правдоподобия выбираются такие параметры, которые максимизируют значение функции правдоподобия на основе обучающей выборки. Максимизация функции правдоподобия соответствует максимизации ее логарифма.

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_1 + \beta_i X_i, \quad (3)$$

где P_i находится в интервале $[0;1]$, а L_i – в интервале $(-\infty; +\infty)$ и линейна.

При использовании логистической регрессии оценка вероятности наступления оцениваемого события производится в следующей последовательности:

- определение зависимой переменной в качестве линейной комбинации независимых факторов;
- создание уравнения для определения вероятности наступления оцениваемого события;
- нахождение производных с целью оценки кумулятивного, а также предельного воздействия независимых факторов;
- проведения вычислений с использованием метода максимального правдоподобия;
- интерпретация полученных результатов.

Пробит-модель. Вероятность анализируемого события в случае применения данной модели двоичного выбора определяется на основании кумулятивной функции нормального распределения [2, с. 89]:

$$P_i = F(Z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z_i^2}{2}}, \quad (4)$$

где

$$Z_i = \beta_1 + \beta_2 X_2 + \dots + \beta_i X_i, \quad (5)$$

Применение в модели нормального распределения не является ограничением для модели благодаря тому, что возможное ненулевое среднее учитывается в константе. Возможная неединичная дисперсия в пробит-модели учитывается при нормировании коэффициентов.

Оценка параметров пробит-модели производится так же, как и в случае логистической регрессии, на основании метода максимального правдоподобия.

Применение пробит-модели для целей оценки вероятности наступления анализируемого события производится в следующей последовательности [5, с.391]:

- определение зависимой переменной в качестве линейной комбинации независимых факторов;
- создание уравнения для определения вероятности наступления оцениваемого события;
- проведение вычислений с использованием метода максимального правдоподобия;
- расчет предельного эффекта;
- интерпретация полученных результатов.

Рассмотренные модели двоичного выбора (логистическая регрессия и пробит-модель) обладают важной особенностью. Оценки коэффициентов этих моделей в данном случае нельзя рассматривать как показатели, характеризующие связь по причине нелинейности параметров. Однако от оценок коэффициентов можно перейти к показателям, которые будут характеризовать связь отдельных факторов моделей.

Недостатком логит- и пробит-моделей можно считать систематическое завышение регрессионных коэффициентов при небольшом размере выборки [4, с. 91].

При решении аналитических задач, направленных на классификацию поведения клиентов в телекоммуникационном секторе, аналогично описанным выше подходам, можно производить определение вероятности событий не только в ситуации двоичного выбора, но и в случае большего количества возможных исходов (1, 2, 3 и т.д. категории поведения клиентов). Таким образом, логит-модели и пробит-модели могут быть успешно применены и в случае множественного выбора.

Учитывая наличие значительного количества факторов и их сочетаний, способных оказать влияние на принятие решения клиентами об отказе от услуг и продуктов конкретной компании, внедрение передовых методов аналитики в телекоммуникационном секторе должно сочетаться с использованием современных методов Big Data.

Внедрение лучших аналитических практик в телекоммуникационном секторе требует как инвестиций, так и уверенной работы по преодолению инерции в компании и созданию сильных многофункциональных команд. В заключение отметим также, что использование передовых аналитических методов само по себе не способно оказать существенного влияния на снижение клиентского оттока в телекоммуникационных компаниях. Несмотря на то что аналитика позволяет предсказать поведение клиентов, ее значимость становится очевидной только тогда, когда компании действительно меняют поведение клиентов благодаря оперативной разработке, тестированию и реализации новых предложений для различных микросегментов в соответствии с их потребностями и ожиданиями.

Список литературы

1. Построение скоринговых карт с использованием модели логистической регрессии //Науковедение. 2014. Вып. 2 (21). – Режим доступа: <http://znanium.com/catalog/product/487966> (дата обращения: 11.10.2019).
2. Мангус Я.Р., Катыхов П.К., Пресецкий А.А. Эконометрика. Начальный курс: Учеб. – 6-е изд., перераб. и доп. – М.: Дело, 2004. 576 с.
3. Доугерти К. Введение в эконометрику: Учебник для экономических специальностей вузов [текст]/Пер. с англ. Е. Н. Лукаш и др. – М.: ИНФРА-М, 1997. 402 с.
4. Методы эконометрики и многомерного статистического анализа: Учебник [текст]/Н. П. Тихомиров, Т. М. Тихомирова, О. С. Урмаев. – М.: Экономика, 2011. 647 с.
5. Вербик Марно, Путеводитель по современной эконометрике. [текст] Пер. с англ. В. А. Банникова, науч. ред. и предисл. С. А. Айвазяна. – М.: Научная книга, 2008. 616 с.
6. Математическая обработка информации: учебник и практикум для СПО [текст]/М.Ю. Глотова, Е.А. Самохвалова. – М.: Издательство Юрайт, 2016. 344 с.