

Асоян Мария Вачагановна
студент бакалавриата
Самарский государственный социально-педагогический университет
(г. Самара)

ИНТЕРПРЕТИРУЕМАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ОТТОКА КЛИЕНТОВ ЦИФРОВОГО СЕРВИСА ДЛЯ ПОДДЕРЖКИ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ

Аннотация. В статье разработана интерпретируемая математическая модель прогнозирования оттока клиентов цифрового сервиса. В качестве базового метода использована логистическая регрессия, позволяющая оценивать индивидуальную вероятность прекращения использования сервиса и объяснять влияние факторов на результат. На имитационной выборке выполнены обучение и проверка модели, рассчитаны метрики качества, а также определен экономически целесообразный порог отбора клиентов для удерживающей кампании. Показано, что порог, максимизирующий управленческий эффект, может отличаться от стандартного значения 0,5.

Ключевые слова: бизнес-информатика, прикладная математика, прогнозирование оттока, логистическая регрессия, поддержка принятия решений.

Для цифровых сервисов удержание действующих клиентов является задачей, находящейся на пересечении управления, анализа данных и проектирования информационных систем. Сам по себе прогноз оттока не создает ценности: результат модели должен быть преобразован в понятное управленческое действие, например в персональное предложение, звонок специалиста или изменение сценария обслуживания. Поэтому математическая модель должна одновременно обеспечивать приемлемую точность, интерпретируемость и возможность связать прогноз с экономическим эффектом [1, 3].

В исследовании рассматривается бинарная целевая переменная y_i , принимающая значение 1, если клиент прекращает пользоваться сервисом в

заданном периоде, и 0 в противном случае. Для демонстрации сформирована имитационная выборка из 1500 наблюдений, не содержащая персональных данных. В модель включены пять признаков: срок использования сервиса, число пользовательских сессий в месяц, количество обращений в поддержку, наличие задержки оплаты и размер абонентской платы. Выборка разделена на обучающую и тестовую части в отношении 70:30 с сохранением доли целевого класса.

Вероятность оттока оценивается с помощью логистической функции. Для i -го клиента линейная комбинация факторов задается выражением:

$$z_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im}.$$

Индивидуальная вероятность оттока определяется формулой [1]:

$$p_i = P(y_i = 1 | x_i) = 1 / (1 + e^{-z_i}).$$

Клиент относится к группе риска, если $p_i \geq t$, где t - порог классификации. Преимущество логистической регрессии состоит в том, что знак коэффициента показывает направление влияния признака, а величина $\exp(\beta_j)$ характеризует изменение отношения шансов при увеличении признака на одну единицу при фиксированных остальных факторах.

Таблица 1

Оцененные параметры логистической модели

Фактор	Коэффициент β	$\exp(\beta)$	Влияние на риск
Срок использования, мес.	-0,051	0,950	снижает
Сессии в месяц	-0,033	0,968	снижает
Обращения в поддержку	0,464	1,591	повышает
Задержка оплаты	0,948	2,581	повышает
Абонентская плата	0,018	1,019	повышает

Полученные коэффициенты имеют содержательную бизнес-интерпретацию. Увеличение срока использования на один месяц связано со снижением отношения шансов оттока примерно на 5,0%, а дополнительная сессия - примерно на 3,2%. Одно дополнительное обращение в поддержку увеличивает отношение шансов примерно в 1,59 раза. Наиболее сильным фактором стала задержка оплаты: при ее наличии отношение шансов возрастает приблизительно в 2,58 раза. Эти результаты позволяют

использовать модель не только для ранжирования клиентов, но и для выявления проблемных элементов клиентского пути.

Качество модели оценивалось на тестовой выборке, Площадь под ROC-кривой составила 0,757, что указывает на способность модели различать клиентов с разным уровнем риска независимо от конкретного порога [2], При стандартном пороге $t = 0,50$ доля правильных классификаций составила 0,796, точность положительного класса - 0,600, полнота - 0,294, Высокая по сравнению с полнотой точность означает, что модель выбирает относительно небольшую, но более концентрированную группу риска. Для удерживающей кампании это не всегда является оптимальным, поскольку часть потенциально уходящих клиентов остается вне воздействия.

Для выбора порога предлагается учитывать ожидаемый экономический эффект. Пусть q - вероятность успешного удержания действительно уходящего клиента, M - маржинальный доход от его сохранения, C - затраты на одно воздействие, $NTP(t)$ и $NFP(t)$ - количество истинно и ложно положительных прогнозов при пороге t . Тогда оценка эффекта определяется выражением:

$$E(t) = NTP(t) \cdot q \cdot M - [NTP(t) + NFP(t)] \cdot C.$$

В расчетном примере приняты $q = 0,35$, $M = 60$ условных единиц и $C = 6$ условных единиц. Такие параметры не претендуют на универсальность: в реальной информационной системе они должны формироваться из данных CRM, истории кампаний и управленческого учета. Их использование позволяет показать, как статистический прогноз преобразуется в оптимизационную задачу бизнес-информатики.

Таблица 2

Результаты выбора порога классификации

Порог t	Точность	Полнота	Выбрано клиентов	Эффект, усл. ед.
0,25	0,385	0,608	161	336
0,35	0,472	0,490	106	414
0,50	0,600	0,294	50	330
0,65	0,714	0,147	21	189

Сценарный расчет показывает, что максимальный эффект достигается при $t = 0,35$ и составляет 414 условных единиц. При этом для воздействия выбираются 106 клиентов, точность составляет 0,472, а полнота - 0,490. Стандартный порог 0,50 дает более высокую точность 0,600, однако ожидаемый эффект снижается до 330 условных единиц из-за пропуска значительной части клиентов, фактически склонных к оттоку. Следовательно, максимизация статистической метрики и максимизация управленческого результата являются различными задачами.

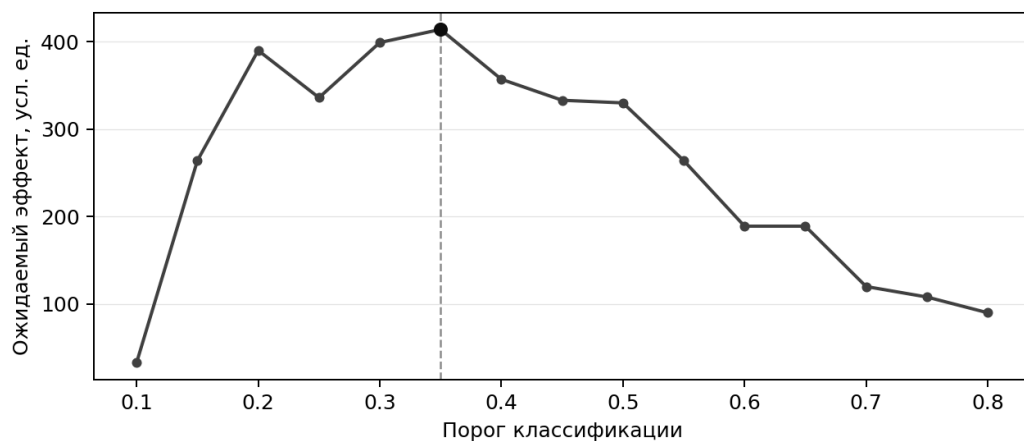


Рис. 1. Зависимость ожидаемого экономического эффекта от порога классификации

График подтверждает наличие внутреннего максимума. При слишком низком пороге кампания охватывает много клиентов, однако затраты на ложные срабатывания снижают итоговый результат. При чрезмерно высоком пороге число воздействий уменьшается, но одновременно растут потери от невыявленного оттока. Поэтому порог должен рассматриваться как управляемый параметр информационной системы, а не как фиксированная техническая константа.

Практическая реализация модели может быть выполнена в виде аналитического модуля CRM-системы. На вход поступают актуальные клиентские признаки, после чего рассчитывается вероятность оттока, формируется ранжированный список и применяется порог, соответствующий текущим ограничениям бюджета. Для контроля качества необходимо хранить версии модели, параметры кампании и фактические результаты удержания.

Такой контур обеспечивает обратную связь между математическим прогнозированием и управлением бизнес-процессом.

Ограничением исследования является использование имитационной выборки и линейной формы зависимости в логит-пространстве. В дальнейшем модель может быть дополнена нелинейными алгоритмами, временными признаками, анализом последовательностей действий и оценкой неопределенности. Однако даже базовая логистическая регрессия обладает важным преимуществом: она прозрачна для аналитика и руководителя, позволяет проверить направление влияния факторов и облегчает обоснование управленческого решения [4].

Таким образом, разработанная модель объединяет методы прикладной математики и задачи бизнес-информатики. Она оценивает вероятность оттока, объясняет влияние факторов и позволяет выбирать порог воздействия по экономическому критерию. Расчетный пример показывает, что оптимальный для бизнеса порог может быть ниже стандартного значения 0,5. Предложенный подход может служить основой для интеллектуальной системы поддержки решений в сфере управления клиентским опытом.

Литература:

1. Hosmer D. W., Lemeshow S., Sturdivant R. X. Applied Logistic Regression. 3rd ed. - Hoboken : John Wiley & Sons, 2013. - 528 p. - DOI: 10.1002/9781118548387.
2. Fawcett T. An introduction to ROC analysis // Pattern Recognition Letters. - 2006. - Vol. 27, No. 8. - P. 861-874. - DOI: 10.1016/j.patrec.2005.10.010.
3. Provost F., Fawcett T. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. - Sebastopol : O'Reilly Media, 2013.
4. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning: with Applications in R. 2nd ed. - New York : Springer, 2021. - DOI: 10.1007/978-1-0716-1418-1.