

Богданова Т.К., Тужикова И.И.

Москва, НИУ ВШЭ

bogtan@mail.ru, inna.tuzhikova@gmail.com

СТАТИСТИЧЕСКИЕ МЕТОДЫ МОДЕЛИРОВАНИЯ РАСТОРЖЕНИЯ ДОГОВОРОВ СТРАХОВАНИЯ ЖИЗНИ

Страхование жизни является неотъемлемой частью в системе страховой защиты интересов человека. На российском рынке страховых услуг страхование жизни является одним из самых быстрорастущих сегментов. Выявление среди клиентов страховой компании группы риска, состоящей из людей, более других склонных к расторжению договора страхования жизни, позволяет вести целенаправленную работу по их удержанию, что, в конечном итоге, должно привести к уменьшению оттока клиентов и, как следствие, положительно сказаться на финансовых показателях страховой компании.

Страхование жизни: характеристика и особенности. Обществом были разработаны два основных способа защиты своих интересов с целью предотвращения убытков от непредвиденных событий: превентивный и репрессивный [1]. Первый способ – превентивный – предполагает предупреждение реализации негативных рисков. Второй способ – репрессивный – связан с минимизацией неблагоприятных последствий непредвиденного события. В сущности, страхование является особым видом экономической деятельности, связанной с перераспределением риска нанесения ущерба материальным интересам. Данный вид деятельности осуществляется специализированными организациями, обеспечивающими накопление страховых резервов и страховые выплаты при реализации негативных рисков [2].

В рамках страхования выделяется личное страхование – отрасль, обеспечивающая защиту имущественных интересов граждан [3]. В свою очередь, страхование жизни является подотраслью личного страхования. Данная процедура предполагает обязательство страховщика заплатить обозначенную сумму денег в случае смерти страхователя или его дожития до определенного возраста в обмен на уплату страховых премий [4]. Иными словами, страховым случаем в страховании жизни считается смерть до окончания действия договора или продолжающаяся жизнь (дожитие) застрахованного после этого срока. Следует отметить, что страхование жизни является добровольным видом страхования. Данная процедура осуществляется путем оформления договора, который предусматривает обязательство страховщика посредством получения *страховых премий*, уплачиваемых страхователем, выплатить *страховую сумму*, если в течение срока действия договора произойдет предусмотренный страховой случай.

Причины расторжения договоров. Удержание клиента, заключившего договор страхования жизни, является актуальной проблемой для

страховщика. Прекращение выплаты установленной премии и расторжение договора могут быть обусловлены как личными причинами клиента, так и экономической ситуацией в стране. Однако, такой исход отношений между страхователем и страховщиком не выгоден ни одному из субъектов. Во-первых, клиент, расторгнувший договор, получает выкупную сумму, закрепленную в договоре, и некоторый инвестиционный доход. Однако это, в любом случае, будет меньше той суммы, которая была уплачена страхователем. Во-вторых, досрочный разрыв договора страхования жизни заставляет страховую компанию забирать денежную сумму из страхового резерва и, следовательно, лишаться средств для инвестирования.

Следует отметить, что разрыв договора страхования может происходить по ряду причин, однако наиболее сильные негативные последствия и для страховщика, и для страхователя имеют разрыв контракта из-за неуплаты премии или по собственному желанию клиента. Поэтому в данной работе исследуются только два данных типа расторжения договоров. После выделения клиентов, находящихся в группе риска, может быть принят ряд мер, направленных на их удержание. К таким активностям можно отнести смену валюты премии или «финансовые каникулы», представляющие собой отсрочку очередного страхового взноса без расторжения договора.

Информационная база исследования включает 15 688 договоров страхования жизни одной из крупнейших страховых компаний России за период с 2008 по 2015 год. База содержит в себе данные о клиентах: их социо-демографические показатели и характеристики заключенных со страховой компанией договоров. Социально-демографические показатели: пол, возраст, семейное положение, должность, сфера занятости и город заключения договора. Атрибуты договоров: дата начала и окончания сотрудничества, факт и причина расторжения, страховой продукт, валюта договора, периодичность оплаты страховой премии, срок договора, величина премии и некоторые другие денежные атрибуты. Кроме того, на основе количественных переменных, описывающих возраст, срок договора, первоначальную выплачиваемую премию, текущую премию, итоговое количество денег, полученное от клиента, были созданы интервальные категориальные переменные.

Разрабатываемые в рамках данной работы модели вычисляют для каждого клиента страховой компании как вероятность расторжения по собственному желанию, так и вероятность расторжения из-за неуплаты страховой премии. Дифференциация причин расторжения позволит более грамотно разработать политику поведения с такими клиентами.

Строго говоря, предсказание типа расторжения – это задача многомерной классификации, то есть, решающая вопрос о принадлежности к одному из трех и более классов [5]. Это объясняется тем, что зависимая переменная Y принимает три возможных значения:

$$Y = \begin{cases} 0, \text{ договор действует или расторгнут по иной причине} \\ 1, \text{ договор расторгнут по желанию клиента} \\ 2, \text{ договор расторгнут из-за неуплаты} \end{cases}$$

Однако существует ряд способов сведения задачи многомерной классификации к бинарной [6]. В данном случае сведение производится следующим образом. Создаются две отдельные зависимые переменные, каждая из которых идентифицирует одну из двух причин расторжения. Переменная Y_1 является зависимой для предсказания расторжения по желанию, Y_2 – для предсказания расторжения из-за неуплаты:

$$Y_1 = \begin{cases} 1, \text{ расторжение по желанию} \\ 0, \text{ в противном случае} \end{cases}$$

$$Y_2 = \begin{cases} 1, \text{ расторжение из-за неуплаты} \\ 0, \text{ в противном случае} \end{cases}$$

В качестве инструментов для построения моделей использовалось программное обеспечение для статистического анализа IBM SPSS Statistics версии 22 [7] и программное обеспечение для прогнозной аналитики IBM SPSS Modeler версии 18.0 [8].

Для прогнозирования расторжения договора страхования по желанию клиента были сформированы обучающая и тестовая выборки. В обучающей выборке записи о договорах, расторгнутых по собственному желанию клиента, составляют 48.7%. В тестовой выборке этот показатель составляет 6.7%. Для прогнозирования расторжения договоров из-за неуплаты обучающая выборка содержит 51.2% записей о договорах, расторгнутых из-за неуплаты, тестовая - 16.9%. Это приближено к показателю оригинальной информационной базы.

Модели расторжения договоров страхования жизни.

Для решения данной проблемы были построены 4 модели: бинарная логистическая регрессионная модель, регрессия Кокса, модель на основе метода Random Forest, нейросетевая модель. Результаты сравнительного анализа ROC-кривых для всех моделей, построенных на тестовых выборках, для предсказания расторжения договоров страхования жизни по желанию клиента приведены в табл. 1.

Таблица 1.

Сравнение показателей качества классификатора для моделей предсказания расторжения договора по желанию клиента

Модель	TP	TN	Верные предсказания	Точность	F-мера	AUC
Бинарная логистическая модель	68.6%	68.0%	68.0%	0.682	0.684	0.722
Нейросетевая модель	75.4%	68.9%	69.4%	0.708	0.731	0.763
Random Forest	65.1%	80.9%	79.9%	0.773	0.707	0.793
Регрессия Кокса	73.7%	69.2%	69.5%	0.705	0.721	0.779

Из этих моделей в качестве наиболее предпочтительной можно выбрать построенную на основе регрессии Кокса. Точность на обучающей

выборке составила 69.2% (69.2% - доля истинно положительных предсказаний, 69.3% - доля истинно отрицательных предсказаний), на тестовых данных модель показала точность 69.5% (73.7% - доля истинно положительных предсказаний, 69.2% - доля истинно отрицательных предсказаний).

Результаты анализа сравнения ROC-кривых для всех полученных моделей предсказания расторжения договоров страхования жизни из-за неуплаты, построенных на тестовых выборках, приведены в табл.2.

Таблица 2.

Сравнение показателей качества классификатора для моделей предсказания расторжения договора из-за неуплаты

Модель	TP	TN	Верные предсказания	Точность	F-мера	AUC
Бинарная логистическая модель	82.0%	79.6%	80.0%	0.801	0.810	0.865
Нейросетевая модель	77.4%	78.7%	78.5%	0.784	0.779	0.864
Random Forest	92.3%	88.3%	88.9%	0.887	0.905	0.956
Регрессия Кокса	65.4%	40.3%	44.5%	0.523	0.581	0.502

Для модели предсказания расторжения из-за неуплаты на основе модели Random Forest точность на обучающей выборке составила 92.6% (95.8% - доля истинно положительных предсказаний, 89.2% - доля истинно отрицательных предсказаний), на тестовых данных модель показала точность 88.9% (92,3% - доля истинно положительных предсказаний, 88.3% - доля истинно отрицательных предсказаний). Таким образом, получилось, что гораздо точнее предсказывается расторжение договора из-за неуплаты, что соответствует почти всем построенным моделям.

Список использованной литературы:

1. Архипов А. П., Гомелля В. Б., Туленты Д. С. Страхование. Современный курс. – М.: Финансы и Статистика, 2008. – 448 с.
2. Ермасов С. В., Ермасова Н. Б. Страхование: учебник. – М.: Высшее образование, 2008. – 613 с.
3. Закон РФ «Об организации страхового дела в Российской Федерации» № 4015-1 от 27.11.1992 г. в ред. от 08.03.2015 г.
4. Архипов А. П. Страхование: Учебник. – Кнорус, 2012.
5. Bishop C. M. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., 2006. 562 p.
6. Карасиков М. Е., Максимов Ю. В. Поиск эффективных методов снижения размерности при решении задач многоклассовой классификации путем её сведения к решению бинарных задач // Машинное обучение и анализ данных, 2014. Т. 1, № 9.
7. IBM - SPSS Statistics Base [Электронный ресурс].– IBM, 2014. – Режим доступа: <http://www-03.ibm.com/software/products/ru/spss-stats-base>, свободный.
8. IBM - SPSS Modeler [Электронный ресурс].– IBM, 2014. – Режим доступа: <http://www-03.ibm.com/software/products/ru/spss-modeler>, свободный.