

УДК 004.032.26, 005.932.1

**Математическая модель для прогнозирования досрочного расторжения
договоров страхования**

Д.В. Макаревич

Научный руководитель: к.ф.-м.н. М.Е. Семенов

Национальный исследовательский Томский политехнический университет,
Россия, г. Томск, пр. Ленина, 30, 634050

E-mail: dvm48@tpu.ru

Mathematical model for predicting early termination of insurance contracts

D.V. Makarevich

Scientific Supervisor: Ph.D. M.E. Semenov

Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

E-mail: dvm48@tpu.ru

Abstract. *A neural network model is proposed for predicting early termination of insurance contracts at the initiative of the policyholder. Modeling is carried out on the Insurance Churn Prediction dataset and the results are presented.*

Key words: *early termination, insurance contracts.*

Введение

Расторжение договоров страхования по инициативе страхователя – это важный аспект страховой практики, который позволяет застрахованным лицам регулировать свои отношения с страховщиками в случае изменения личных обстоятельств или ухудшения условий, на которых был заключен договор. Страхователь может столкнуться с изменениями в своей жизни, которые могут делать действующий договор страхования нецелесообразным или ненужным.

Целью работы является разработка математической модели и её программная реализация для предсказания обращений за расторжением договора страхования в будущем.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести критический анализ литературы по предметной области.
2. Разработать математическую модель предсказания количества будущих обращений за расторжением договоров страхования.
3. Обучить полученную модель, предварительно подготовив данные.

Актуальность данной работы обусловлена необходимостью предсказания количества будущих обращений за расторжением договоров страхования. Это исследование имеет практическое значение для страховых компаний, поскольку позволяет оптимизировать их стратегию управления рисками и улучшить качество обслуживания клиентов.

Научная новизна данной работы заключается в разработке и внедрении модели нейронной сети для предсказания количества будущих обращений за расторжением договоров страхования.

Признаки (независимые переменные), целевая переменная

Для обучения нейронной модели был использован набор данных (без пропусков и дубликатов), который содержит 7934 записей для 8 анонимизированных признаков, включая числовые и категориальные типы переменных, и целевой переменной labels. Категориальные признаки закодированы числами. Целевой переменной будет являться расторжение договора страхования (labels) в будущем, описание независимых переменных (признаков) приведено в табл. 1. Значения целевой переменной является только 0 и 1, где 0 – расторжения не было, а 1, соответственно, и является самим фактом расторжения.

Описание признаков до нормализации

Наименование признака	Тип признака	Диапазон значений
feature_1	числовой	[-3,081149...33,094776]
feature_2	числовой	[-1,779108...1,825628]
feature_3	числовой	[-1,002478...18,094700]
feature_4	числовой	[-0,569351...19,443647]
feature_5	числовой	[-0,411453...8,127648]
feature_6	числовой	[-0,251940...23,625644]
feature_11	категориальный	[0, 1]
feature_12	категориальный	[0, 1]

Обработка данных и архитектура нейронной сети

На этапе обработки мы провели нормализацию независимых числовых переменных с помощью минимаксного метода:

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min}),$$

где x – исходное значение, $x_{\min}$ ($x_{\max}$) – минимальное (максимальное) значение признака в наборе данных. Далее мы сформировали тренировочную, валидационную и тестовую выборки в пропорции 70/15/15 соответственно. На языке программирования Python с использованием библиотеки Keras мы спроектировали нейронную сеть следующей архитектуры:

1. Слой (8 нейронов на входе, 64 нейронов на выходе, функция активации Relu).
2. Слой (64 нейронов на входе, 32 нейронов на выходе, функция активации Relu).
3. Слой (32 нейронов на входе, 16 нейронов на выходе, функция активации Relu).
4. Слой (16 нейронов на входе, 1 нейрон на выходе, функция активации Sigmoid).

На выходном слое функция Sigmoid была выбрана специально для бинарной классификации, также был выбран стандартный оптимизатор Adam, а в качестве функции потерь была выбрана binary_crossentropy.

Метрики качества

В качестве метрики качества предсказания нейронной модели мы использовали точность (ассурасу), которая на тестовой выборке равна 0,81. На рис. 1 приведена матрица ошибок, из которой видно, что количество истинно предсказанных расторжений равняется 484, правильно предсказанных отсутствий расторжений 476. Ложно предсказанных расторжений 126, а ложных отсутствий расторжений 105.

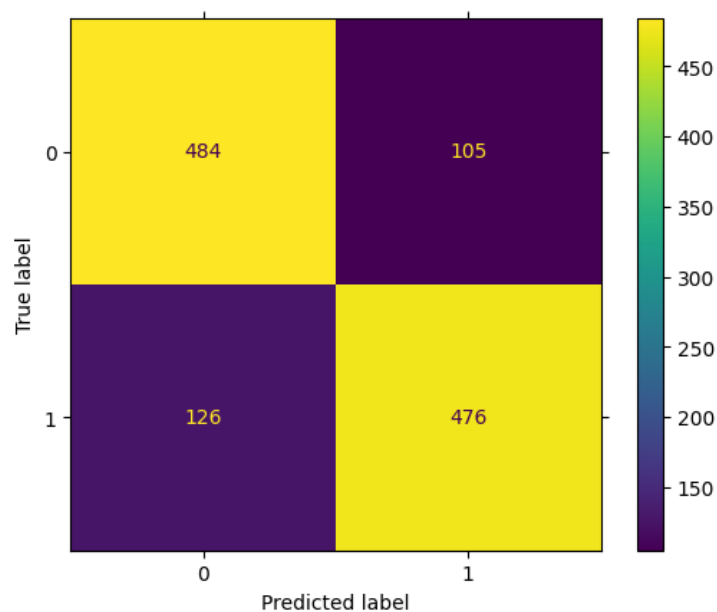


Рис. 1. Матрица ошибок, по осям расположены истинные и предсказанные классы.

Заключение

В результате исследования точность решения задачи бинарной классификации составила 0,81. Для повышения точности требуется более тонкая настройка модели.