

Low Code для разработки AI-моделей в страховании

Июнь 2023 г.



Сергей Алешкин
Head of Data Science
АО «СОГАЗ»

СОГАЗ

Лицензии Банка России СР №1208, СИ №1208, ОС №1208-02,
ОС №1208-03, ОС №1208-04, ОС №1208-05, ПИС №1208, АО «СОГАЗ».

Рынок AI в России: большой потенциал

Использование AI в бизнес-процессах российских компаний



61%
рынка AI

анализ данных с помощью AI

33%
рынка AI

NLP-обработка естественного языка

Источники:

- Альманах «Искусственный интеллект». Состояние отрасли в России и мире, ЦНТИ на базе МФТИ
- IDC Worldwide Artificial Intelligence Spending Guide

AI на российском страховом рынке

30% компаний
применяют AI
в бизнес-процессах

25% клиентов
готовы полностью
перейти на цифровое
страхование

Основные бизнес-процессы, в которых используется AI:

- проведение скоринга при предстраховой проверке
- аналитическое выявление мошеннических убытков
- прямые продажи страховых продуктов
- проведение кампаний целевого маркетинга
- **кросс-продажи продуктов и услуг существующим клиентам**

Источники:

- Показатели цифровизации страхового рынка в Российской Федерации, Финансовый университет
- Исследование BCC и AC&M Consulting

AI – ключевой элемент цифровой трансформации:

Корпоративное хранилище данных,
отвечающее современным принципам работы с **BigData**

Использование моделей машинного обучения (ML) в бизнес-процессах

(персонализация продаж, оптимизация товарных запасов, оптимизация графика работы сотрудников в точках продаж, предотвращение потерь из-за нечестных сотрудников и т. п.)

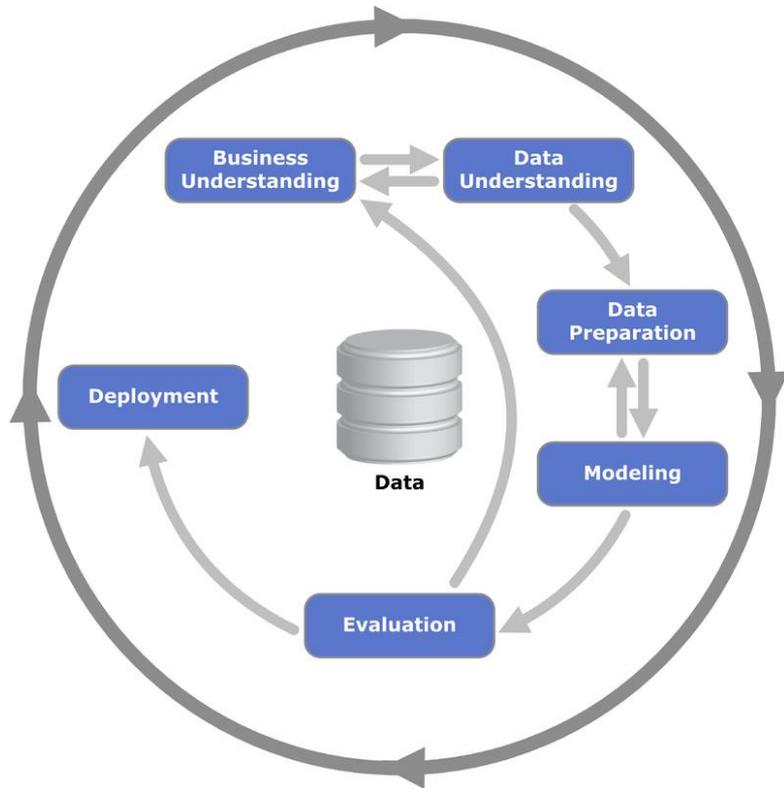
Обучение сотрудников
компании работе с современными цифровыми технологиями

Интерактивные цифровые отчеты на базе существующих систем бизнес-аналитики (Power BI, ClickView и т. д.).

Digital marketing для продвижения товаров и услуг (сквозная аналитика, модели атрибуции конверсий, автоматизация управления контекстной рекламой, call tracking и т. д.)

Цифровые сервисы
удаленный доступ к услугам и продуктам компании для клиентов

Межотраслевой стандарт исследования данных



- Постановка бизнес-целей (Business Understanding)
- Предварительное изучение данных (Data Understanding)
- Подготовка данных (Data Preparation)
- Построение модели (Modeling)
- Оценка качества модели (Evaluation)
- Внедрение (Deployment)

Основные типы ML-задач в страховании

Обучение «без учителя»:

- Кластеризация
- Заполнение пропущенных значений
- Сокращение размерности
- Визуализация данных

Практические задачи:

Предстраховой скоринг: выделение наиболее убыточного сегмента из множества всех клиентов.

Целевой маркетинг: разбиение множества всех клиентов на кластеры для выявления типичных представителей основных клиентских групп.

Кросс-продажи: выявление сочетаний товаров, часто встречающихся вместе в покупках клиентов разных групп.

Обучение «с учителем»:

Построение алгоритма на основе анализа множества объектов с аналогичными параметрами и известными ответами для этих объектов, в т. ч.:

- Задачи классификации (принадлежность к классу)
- Задачи регрессии (действительные числа)

Практические задачи:

Тарификация: оценка прогнозируемого убытка по полису на основе анализа имеющегося портфеля.

Предстраховой скоринг: выявление потенциальных мошенников.

Кросс-продажи: выявление клиентов с высоким уровнем спроса на выбранный тип продукта.

Преимущества Low Code-библиотек для ML

Скорость разработки моделей:

Небольшой объем кода для написания моделей позволяет значительно снизить сроки разработки.

Мультиплатформенность:

Могут быть развернуты в разных программных средах. Совместимы с разными операционными системами и будут работать на разных устройствах.

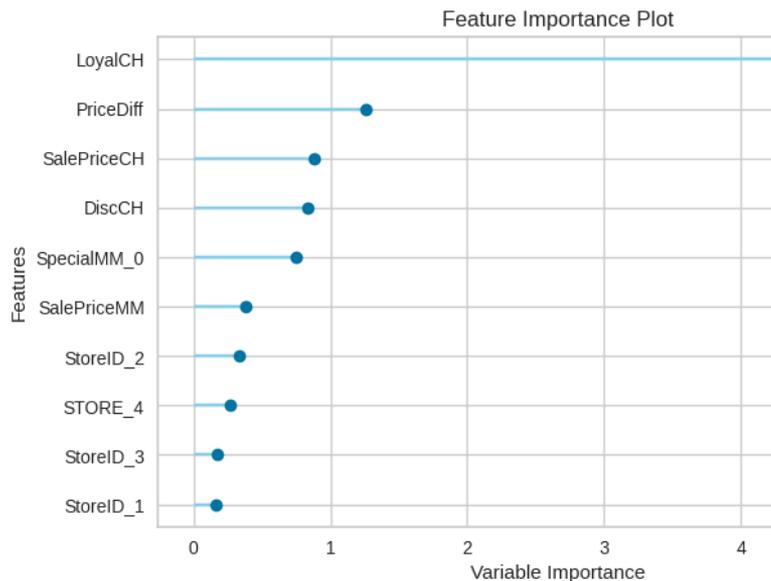
Снижение требований к знанию программирования:

Работа с данными с помощью визуальных интерфейсов. Чтобы готовить эффективные модели, специалистам достаточно знать всего несколько базовых библиотек (TensorFlow, PyTorch и т. д.).



Примеры работы Low Code-библиотек (на основе Русарет)

Оценка значимости признаков



Выбор оптимальной ML-модели

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lr	Logistic Regression	0.8343	0.8979	0.7559	0.8026	0.7758	0.6446	0.6479	0.273
lda	Linear Discriminant Analysis	0.8316	0.9012	0.7701	0.7879	0.7761	0.6412	0.6443	0.014
ridge	Ridge Classifier	0.8289	0.0000	0.7594	0.7885	0.7700	0.6340	0.6379	0.012
gbc	Gradient Boosting Classifier	0.8210	0.9028	0.7459	0.7807	0.7594	0.6172	0.6210	0.096
ada	Ada Boost Classifier	0.8102	0.8835	0.7245	0.7661	0.7432	0.5929	0.5950	0.083
lightgbm	Light Gradient Boosting Machine	0.8102	0.8859	0.7389	0.7613	0.7469	0.5956	0.5988	0.081
rf	Random Forest Classifier	0.7982	0.8769	0.7244	0.7416	0.7305	0.5695	0.5719	0.451
et	Extra Trees Classifier	0.7915	0.8513	0.7070	0.7355	0.7190	0.5536	0.5558	0.452
dt	Decision Tree Classifier	0.7889	0.7833	0.7283	0.7250	0.7244	0.5538	0.5561	0.014
nb	Naive Bayes	0.7554	0.8386	0.7527	0.6587	0.6991	0.4951	0.5020	0.013
knn	K Neighbors Classifier	0.6991	0.7575	0.5725	0.6061	0.5849	0.3509	0.3532	0.115
svm	SVM - Linear Kernel	0.5083	0.0000	0.5414	0.2630	0.3266	0.0356	0.0389	0.015
qda	Quadratic Discriminant Analysis	0.4303	0.5144	0.8565	0.3961	0.5124	0.0274	0.0371	0.015

CPU times: user 3.9 s, sys: 366 ms, total: 4.26 s
Wall time: 19.7 s

Примеры востребованных Low Code-библиотек

Русcaret



H₂O AutoML



Auto-ViML



AI в продажах: изменение поведения потребителя...

80%

КЛИЕНТОВ

готовы платить за
удобство и
качество сервиса, а
не за бренд

63%

КЛИЕНТОВ

готовы делиться
персональной
информацией для
лучшего сервиса

32%

КЛИЕНТОВ

готовы уйти от
любимого бренда
после первой же
ошибки

76%

КЛИЕНТОВ

ожидают
персонального
предложения от
компаний

Источники:

- Experience is everything: Here's how to get it right. PwC
- Исследование потребительского поведения в 2022, «Ромир»

... меняет подход к работе с клиентской базой

Применение AI на каждом этапе сотрудничества с клиентом



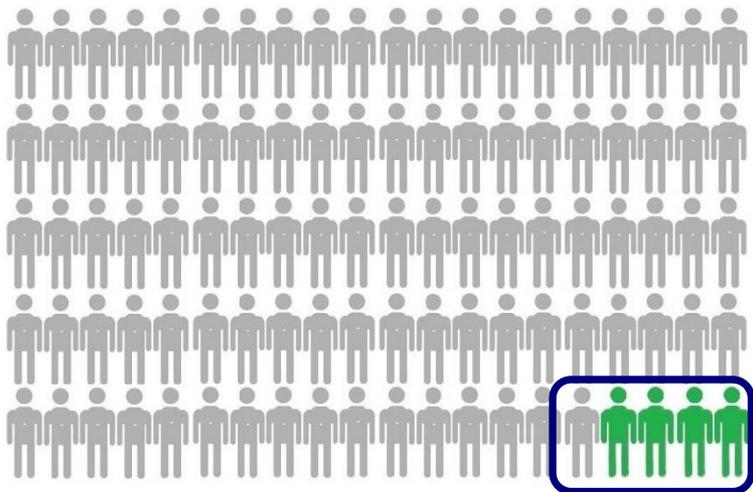
Не нужно забивать гвозди микроскопом

Применение AI в продажах для различных продуктов и сегментов клиентов

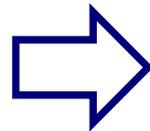
	Низкомаржинальные продукты	Высокомаржинальные продукты
Мало данных о клиентах	<p>Онлайн-продажи с помощью голосовых роботов с относительно простыми скриптами</p> 	<p>Предварительные продажи с помощью голосовых роботов с дальнейшим переключением на опытных продавцов</p> 
Данных о клиентах достаточно	<p>Отбор клиентов с помощью AI-модели и дальнейшая коммуникация с помощью голосовых роботов со сложными скриптами (похожи на человека), оформление онлайн</p> 	<p>Отбор клиентов с помощью AI-модели и дальнейшая обработка опытными продавцами</p> 

Применение AI в кросс-продажах клиентам в ОСАГО

4 человека из 100
рассмотрят покупку полиса
добровольного страхования



* согласно существующей статистике



Клиенты, отобранные AI:

40%

клиентов готовы купить продукты
добровольного страхования в
ближайшее время

98%

клиентов, не заинтересованных в
покупке, исключаются
алгоритмом

**в 30
раз**

снижаются трудозатраты сети
на кросс-продажи

Отбор клиентов производится с использованием ансамбля моделей актуализируемого на основе A/B тестирования. В актуальной версии используются модели Logistic Regression и XGboost.

СПАСИБО ЗА ВНИМАНИЕ!

