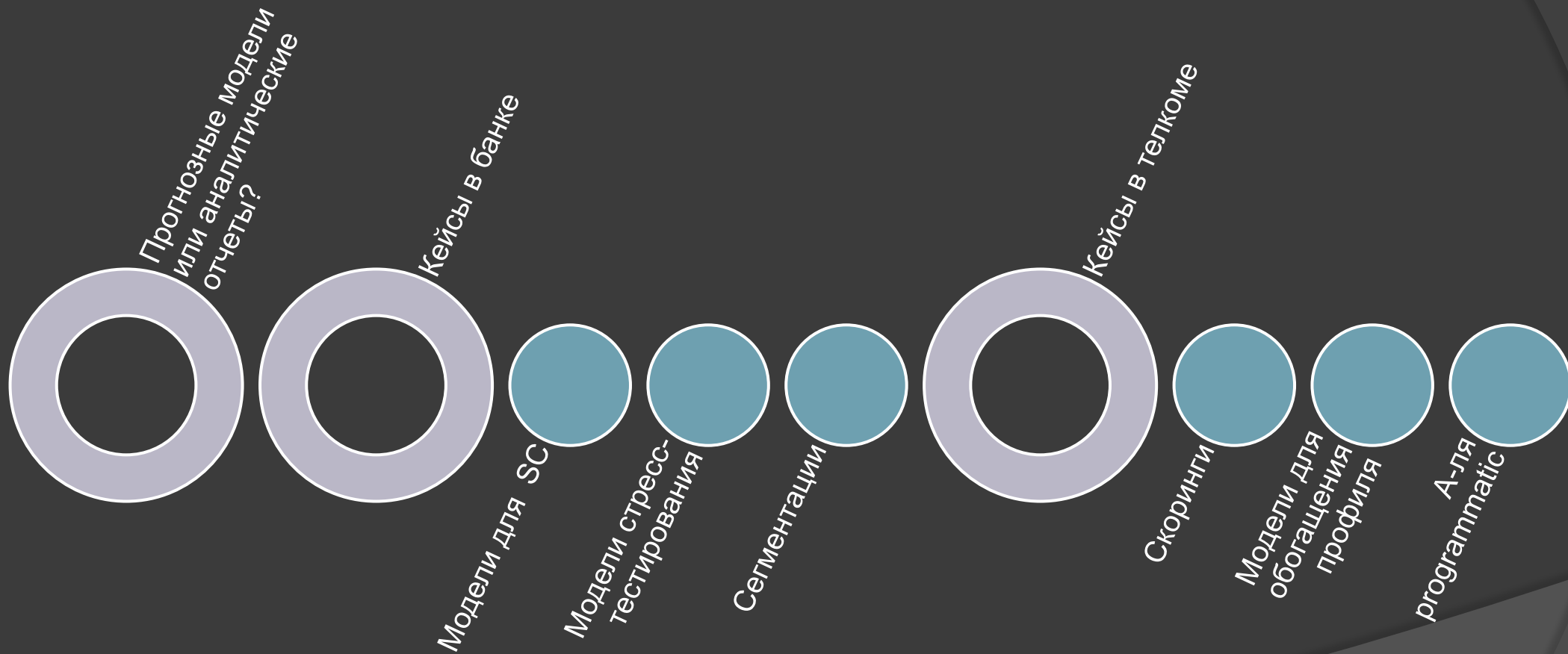


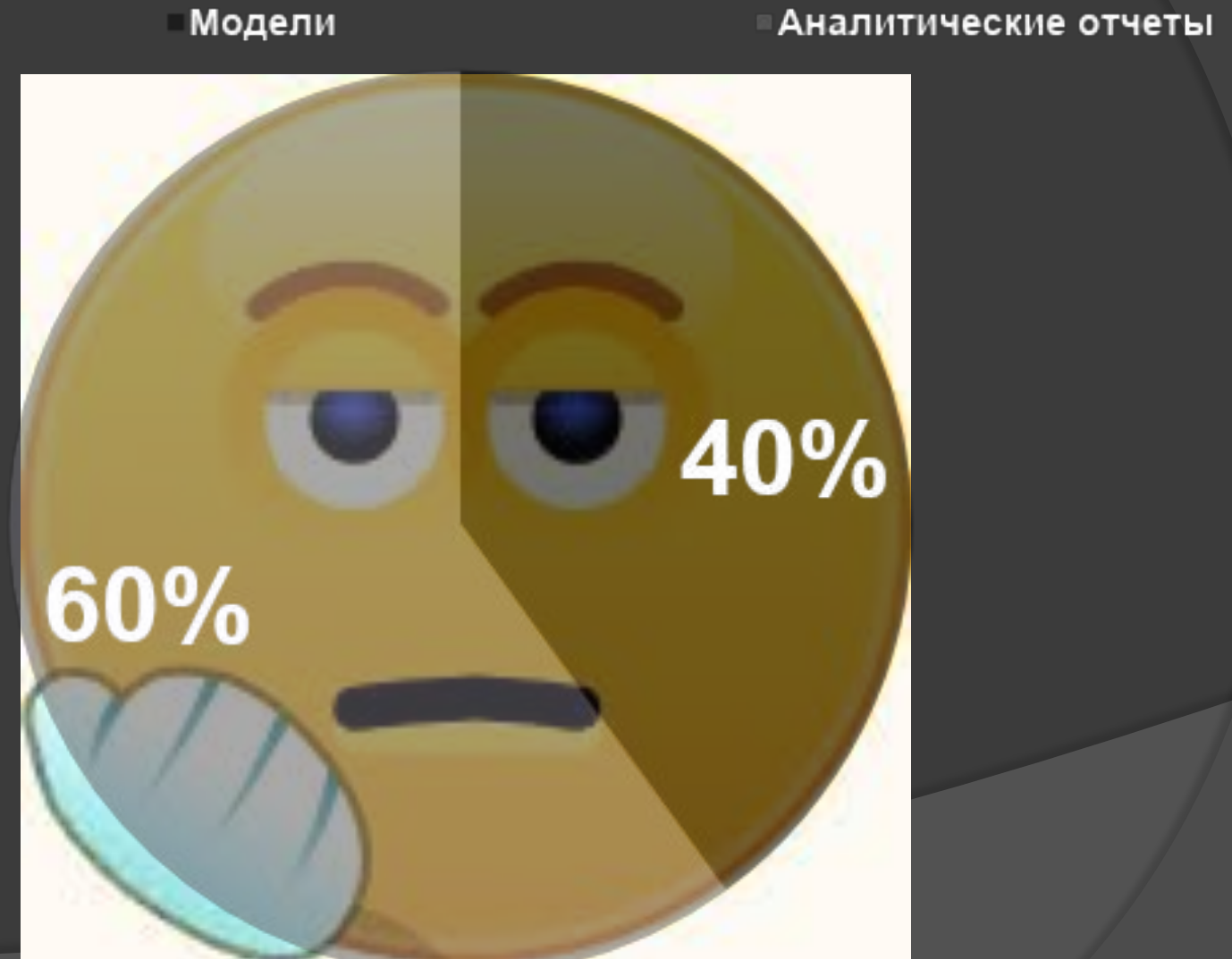
АНАЛИЗ ДАННЫХ В БАНКЕ И ТЕЛЕКОМЕ

О чем поговорим?



Прогнозные модели или аналитические отчеты?

Мой опыт...



Банки...

Наша первая модель – оптимизация затрат в Soft Collection

Проблематика:

Количество клиентов выходящих на просрочку растет с каждым месяцем.

Колл центр расширять никто не хочет в виду высоких затрат на содержание, мотивацию, обучение и прочее затраты.

Нужен новый метод который позволит при неизменных затратах сохранять уровень по сборам

Как действует бизнес «ДО»:

Звонит с первого дня просрочки (далее регулярные звонки и СМС)

Предложение по пилоту:

Построить модель по выявлению «случайных» просрочников

Что сделали:

Построили регрессионную модель в SAS Guide с таргетом: хорошие -до 7 дней, средние –до 30, злостные – остальные)

Как оценивать результат:

В тестовых группах тестировали как отсутствие коммуникации с хорошими, так и более дешевый канал коммуникации.

Результат:

Пилот успешный, экономия состоялась за счет более грамотной политики управления каналами коммуникации

Модель стресс-тестирования

Проблематика:

В рамках МСФО 9, необходимо построить модель стресс-тестирования: влияние макропоказателей на кредитный портфель банка

Компании или её регулирующие органы предпочитают стресс-тестирование, когда хотят посмотреть на то, как ведут себя финансовые инструменты в случае определенной стрессовой ситуации, например:

- Что случится, если фондовые рынки обрушатся более чем на $X\%$ в этом году?
- Что произойдет, если ВВП падает на $Z\%$ в течение данного года?
- Что произойдет, если процентные ставки вырастут, по крайней мере на $Y\%$?
- Что делать, если половина инструментов в портфеле будет расторгнута через пять лет?
- Что произойдет, если цена на нефть вырастет на 200% ?

Что использовали:

Инструменты по работе с временными рядами в IBM SPSS Modeller

Про что важно не забыть:

Не стационарные временные ряды необходимо приводить в стационарный вид

Иначе корреляция будет между чем угодно

При рассмотрении влияния переменных с лагом на целевую, необходимо убедиться в адекватности данного влияния и отсутствию сменны знака на разных уровнях лага.

Результат:

Не получилось построить модели с коэффициентом детерминации более 50% , за исключением одного продукта. Что, как нам показалось и логично

Сегментации

Для чего обычно использовала:

- Для понимания своих клиентов (соц демо, P&L, поведение)
- Для формирования программ лояльности (но тут не совсем банк, точнее ломбард был)
- Для поиска вдохновения в разработке новых продуктов
- Для проведения маркетинговых исследований
- Прелюдия перед моделированием

Какие проблемы бывали:

Не достаточно данных для полной картинки. Особенно в очень узких задачах, когда например необходимо выяснить, кто в вашей базе клиентов – курит, водит автомобиль и т.д.

Решение:

Провести опрос силами СС по своей базе задавая интересующие вопросы – получить обучающую выборку.
Далее дело техники 😊

Кейсы телекома

Телеком скоринг - Аналог кредитного скоринга

Для чего нужен и как используется?

Используется банками и микро кредитными организациями как

- дополнительный предиктор в локальной скоринговой карте
- в качестве прескоринга для таргетных/триггерных рассылок/звонков
- как фильтр для телко- лидогенерации

Какая ценность для банков?

- Увеличение коэффициента Gini (по опыту от 3 баллов до 8) – точность прогноза
- Экономия на неэффективных звонках/СМС
- Экономия на не эффективных лидах

Модели для обогащения профиля

О чем это?

У телеком компаний недостаточно регистрационных данных для профилирования клиентов по соц. Демографическому портрету (пол, возраст, семейное положение, наличие детей и т.д.)

Для чего это нужно?

- Для более точных СВМ моделей,
- Анализа клиентов,
- Сегментации и идентификации их потребностей клиентов

Как решали вопрос?

- Сбор обучающей выборки с помощью СС
- Прогонка всей базы по данной модели

Что делали мы?

1. Построили модель по полу:
Модель отдельно по тем по кому есть 8 марта и 23 февраля и отдельно по тем по кому нет. Качество первой сильно лучше : ~80% и ~60% precision соответственно
2. Модель по возрасту:
 - Выборка оказалась сильно меньше чем по полу
 - Строили отдельно для 5 возрастных групп, трех и двух. Чем меньше сегментов не выше качество каждой в отдельностиМодель по семейному статусу была на очереди, но долго собиралась обучающая выборка...

Что еще можно?

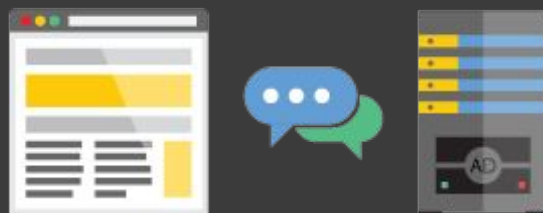
Сегменты по интересам на базе DPI, звонковой активности: любитель сериалов, автомобилей, фитнеса и т.д.

А-ля programmatic

1 User visits a website



2 Website communicates with an ad exchange/SSP



3 SSP offers user information to DSPs



4 Advertisers place bids through a DSP



5 Highest bid wins the ad



Опыт коллег в России про создание собственной рекомендательной системы, которая предлагает каждому сегменту показать наиболее релевантный оффер

Рекламодатели дают информацию по предложению которое будет таргетироваться

Далее описание оффера преобразуют в вектор (модель «word to vec»), и таргетируют на ту аудиторию (look-a-like), которая ранее отвлекалась на самый похожий оффер. И со временем получая ОС модель подстраивается под свою аудиторию

Такой таргетинг повышает отклики мин на 5% и сокращает отток мин на 15%*

*- по данным коллег из России