Текст защиты дипломной работы

(Слайд 1: Титульный лист)

Добрый день, уважаемые члены аттестационной комиссии! Меня зовут Доготарь Дмитрий, я представляю вашему вниманию свою дипломную работу по теме:

Прогнозирование оттока клиентов телеком-компании и формирование персонализированных рекомендаций

(Слайд 2: Актуальность и цель)

На высококонкурентном рынке телекоммуникаций удержание клиентов — это ключ к успеху. Терять клиентов дорого, а привлекать новых — еще дороже. Поэтому целью моей работы была не просто разработка еще одной модели, которая предсказывает отток, а создание комплексной системы, которая:

* Точно определяет клиентов в группе риска.
* Находит экономически выгодную стратегию их удержания.
* Дает конкретные, персонализированные советы, что именно предложить каждому такому клиенту.

(Слайд 3: План работы)

Для достижения этой цели я прошел через 6 ключевых этапов: от анализа данных и создания новых признаков до обучения моделей, экономического анализа и, наконец, формирования итогового отчета с рекомендациями.

(Слайд 4–6: Анализ данных — что влияет на отток?)

Я начал с исследовательского анализа данных. Главные выводы:

Контракт решает всё: клиенты с помесячным контрактом уходят значительно чаще, чем те, у кого годовой или двухлетний.

Новички в зоне риска: чем меньше клиент с компанией, тем выше вероятность его ухода.

Дополнительные услуги важны: отсутствие техподдержки и онлайн-безопасности - сильные индикаторы будущего оттока.

Также был выявлен дисбаланс классов: около 27% клиентов ушли, что я учёл при обучении моделей.

(Слайд 7: Выбор лучшей модели)

Я протестировал 8 различных моделей — от классической логистической регрессии до современных нейронных сетей, таких как TabNet.

Сначала, с базовыми настройками, лучшие результаты показала TabNet.

Однако после тщательной и автоматизированной настройки гиперпараметров с помощью фреймворка Optuna, лидером стала модель CatBoost, показав ROC AUC 0.8419. Это высокий результат, говорящий о хорошей предсказательной силе. Именно ее я и использовал в дальнейшей работе.

(Слайд 8: Главная идея - от точности к прибыли)

И здесь начинается самое интересное. Высокая точность модели — это хорошо, но бизнесу нужна прибыль. Просто предлагать скидки всем, кто «похож» на уходящего, — прямой путь к убыткам. Мы можем потратить на удержание больше, чем заработаем.

(Слайд 9: Поиск точки максимальной прибыли)

Поэтому я провел экономический анализ. Я задал простые бизнес-допущения: сколько мы тратим на удержание, сколько в среднем принесет удержанный клиент, и какова вероятность, что он примет наше предложение.

Затем я рассчитал чистую прибыль для каждого возможного порога вероятности оттока — от 0 до 100%.

На графике вы видите результат. Максимальная прибыль достигается не при максимальном охвате, а в конкретной точке. Оптимальный порог составил 54,8%. Это значит, что нам нужно нацеливаться только на тех клиентов, чья вероятность ухода, по мнению модели, выше этого значения. Такая стратегия на тестовой выборке принесла бы компании условную прибыль в $24,368.

(Слайд 10: Кого нашли. Что им предложить?)

Итак, мы знаем, кого нужно удерживать. На тестовой выборке это 334 человека. Но что им предложить? Рассылать всем одинаковую скидку — неэффективно.

Я использовал метод k-ближайших соседей (k-NN). Для каждого "рискового" клиента я нашел в данных самого похожего на него, но лояльного клиента — его "двойника".

(Слайд 11: Пример персонализированной рекомендации)

Логика проста: если "двойник" пользуется какой-то услугой, а наш "рисковый" клиент — нет, то именно эту услугу и стоит ему предложить, возможно, со скидкой.

Например, для клиента с ID 260, у которого вероятность оттока 88%, система нашла похожего лояльного клиента, который пользуется телефонной связью. Наш клиент — нет. Рекомендация: "Предложить подключить PhoneService". Это просто, логично и персонализировано. Если же наборы услуг совпадают, базовая рекомендация — предложить более выгодный долгосрочный контракт.

(Слайд 12: Почему клиент уходит? Делаем модель понятной)

Чтобы менеджер по удержанию мог вести диалог с клиентом более предметно, важно понимать причины прогноза. Для этого я использовал инструмент LIME.

На слайде — анализ для клиента с самой высокой вероятностью ухода. LIME наглядно показывает, что главные факторы риска для него — это помесячный контракт, короткий срок пользования услугами и использование оптоволокна, которое, вероятно, дороже. Это дает менеджеру конкретные "болевые точки" для обсуждения.

(Слайд 13: Выводы и результаты)

В результате проделанной работы была создана не просто модель, а полноценная система поддержки принятия решений, которая:

* Точно прогнозирует отток с помощью модели CatBoost.
* Определяет прибыльную стратегию, находя оптимальный сегмент для удержания.
* Дает персонализированные и объяснимые рекомендации, помогая бизнесу принимать взвешенные решения.

(Слайд 14: Дальнейшие шаги)

Проект можно развивать дальше: провести A/B тестирование рекомендаций, чтобы измерить их реальную эффективность, а также обогатить модель новыми данными, например, из логов службы поддержки.

(Слайд 15: Спасибо за внимание!)

На этом мой доклад окончен. Спасибо за внимание! Я готов ответить на ваши вопросы.