

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ПОВТОРНЫХ ОБРАЩЕНИЙ И ПРЕДОТВРАЩЕНИЯ ПОВТОРНЫХ ВЫЕЗДОВ В ТЕЛЕКОМ-СЕРВИСЕ

О.А. Авилова¹, А.С. Ягодкин¹, А.В. Кладов¹

¹ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

Аннотация. Повторные обращения и повторные выезды в телекоммуникационном сервисе увеличивают операционные затраты и ухудшают соблюдение соглашений об уровне обслуживания (SLA), поэтому актуальна разработка предиктивных моделей риска рецидива и их внедрение в контур принятия решений. В работе рассматривается постановка задачи прогнозирования повторных событий на горизонтах 7, 14 и 30 дней, методика подготовки телеком-данных из CRM, OSS, диагностики CPE/ONT и Workforce Management, а также прототип информационной системы, которая приоритизирует инциденты и выдает рекомендации для снижения доли повторных выездов. Исходной базой выступает датасет SLA_Telecom_71430, содержащий 71 430 связей телекоммуникационных событий за 182 дня наблюдений с GPON/XGS-PON сетей, включающий 68 телеком-специфичных признаков и три целевые переменные, отражающие риск нарушения SLA на разных временных горизонтах. Применение методов машинного обучения с учетом дисбаланса классов и экономической асимметрии ошибок позволяет строить модели для превентивного управления качеством обслуживания.

Ключевые слова: телекоммуникации, SLA, повторные обращения, повторные выезды, машинное обучение, предиктивная аналитика, несбалансированные данные, MLOps, CRM, OSS, Workforce.

APPLYING MACHINE LEARNING TO PREDICT REPEAT CALLS AND PREVENT REPEAT CALLS IN TELECOM SERVICE

O.A. Avilova¹, A.S. Yagodkin¹, A.V. Kladov¹

¹Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. Repeat calls and repeat calls in telecommunications services increase operating costs and impair service level agreement (SLA) compliance. Therefore, the development of predictive

models for recurrence risk and their integration into decision-making is relevant. This paper discusses the problem of predicting repeat events over 7-, 14-, and 30-day horizons, a methodology for preparing telecom data from CRM, OSS, CPE/ONT diagnostics, and Workforce Management, and a prototype information system that prioritizes incidents and provides recommendations for reducing the rate of repeat calls. The initial database is the SLA_Telecom_71430 dataset, containing 71,430 telecommunications event associations over 182 days of observations from GPON/XGS-PON networks. It includes 68 telecom-specific features and three target variables reflecting the risk of SLA violations at different time horizons. The use of machine learning methods, taking into account class imbalance and economic asymmetry of errors, enables the construction of models for proactive service quality management.

Keywords: telecommunications, SLA, repeat calls, repeat visits, machine learning, predictive analytics, imbalanced data, MLOps, CRM, OSS, Workforce.

Традиционные реактивные подходы к мониторингу, основанные на анализе логов и пороговых алертах, демонстрируют высокий уровень ложных срабатываний и недостаточную способность к раннему выявлению проблем. Предиктивная аналитика на основе машинного обучения открывает возможность перехода к превентивному управлению: прогнозирование риска повторного события за несколько дней позволяет усилить диагностику на ранних этапах, улучшить качество закрытия обращений и рационально распределить ресурсы полевой службы. Однако реализация таких систем в телекоммуникационном домене сталкивается с рядом специфических вызовов, включая крайнюю несбалансированность данных, необходимость интеграции гетерогенных источников информации, учет экономической асимметрии ошибок и обеспечение эксплуатационной готовности через практики MLOps.

Целью настоящей работы является разработка методологии прогнозирования повторных обращений в телекоммуникационном сервисе на основе интеграции данных из множественных источников и применения специализированных техник машинного обучения для несбалансированных данных с последующим проектированием информационной системы предотвращения повторных выездов. Научная новизна заключается в комплексном подходе, объединяющем телеком-специфичное формирование признаков, методы cost-sensitive обучения и MLOps-архитектуру для промышленного применения.

Постановка задачи и источники данных

Процесс обработки обращений в телекоммуникационной компании представляет собой многоуровневую систему, где контактный центр выполняет пер-

вичную регистрацию инцидентов в CRM-системах с целевым показателем разрешения с первого контакта выше 70%. При невозможности решения на первом уровне происходит эскалация во вторую линию технической поддержки, где инженеры проводят углубленную диагностику с использованием систем операционной поддержки, анализируя метрики сетевого трафика, ошибки канального уровня, телеметрию оптических терминалов и клиентского оборудования. При необходимости физического вмешательства формируется задание для полевой службы через системы управления рабочей силой, где планирование выездов оптимизируется с учетом геолокации, навыков техников и временных окон SLA.

Повторное обращение трактуется как рецидивирующий тикет по той же или близкой проблеме в заданном временном окне после закрытия первичного обращения, что напрямую отражается в росте нагрузки на все уровни поддержки и повторных работах полевой службы. Основные причины повторных событий включают операционные факторы (недостаточная квалификация персонала, низкий уровень разрешения с первого контакта, формальное закрытие без устранения первопричины), технические факторы (неполный ремонт на выезде, скрытые дефекты оборудования, деградация оптоволокну, перегрев клиентских устройств), организационные факторы (потеря истории тикета при эскалации, несогласованные временные окна SLA между уровнями) и клиентские факторы (несоблюдение инструкций, завышенные ожидания).

Экономическое влияние повторных событий существенно: стоимость одного повторного выезда составляет 3-10 тысяч рублей, при коэффициенте переделок выше 15% совокупные затраты полевой службы возрастают на 25-40% годовых. Нарушения SLA влекут штрафы до 5-10% от месячной платы при доступности 99.72-99.9% и до 15% при 98.9-99.45%. Три и более рецидива удваивают вероятность оттока клиента с базового уровня 3% до 6-8% ежемесячно, особенно для корпоративных заказчиков.

Данные для прогнозирования формируются из четырех основных источников. CRM и системы тикетинга фиксируют клиентские взаимодействия, включая временные метки создания и закрытия обращений, категории проблем, уровни приоритета, количество эскалаций, текстовые описания инцидентов и оценки удовлетворенности клиентов. OSS и сетевой мониторинг обеспечивают телеметрию инфраструктуры с высокой дискретизацией, включая метрики трафика, ошибки передачи данных на канальном уровне, задержки, потери пакетов, утилизацию процессоров и памяти сетевого оборудования. Диагностика клиентского оборудования CPE/ONT предоставляет параметры оптических терминалов

(принимаемая и передаваемая оптическая мощность, аварийные сигналы потери сигнала, температура, пропускная способность портов, утилизация беспроводных интерфейсов). Данные полевой службы документируют физические вмешательства через записи о выездах, включающие идентификаторы техников, GPS-траектории, выполненные действия, использованные материалы, результаты пост-ремонтной диагностики и признаки переделок.

Основной базой исследования выступает датасет SLA_Telecom_71430, включающий 71 430 связанных телекоммуникационных событий, собранных с оптических сетей GPON/XGS-PON в период с 1 января по 1 июля 2025 года (182 дня наблюдений). Датасет содержит 68 телеком-специфичных признаков, разделенных на семь категорий, и три целевые переменные, прогнозирующие нарушения SLA на горизонтах 7, 14 и 30 дней соответственно.

Методология прогнозирования

Формирование признаков учитывает специфику телекоммуникационного домена и включает несколько категорий переменных. Признаки оптического доступа отражают состояние и динамику оптических терминалов: оптический бюджет линии, рассчитываемый как разность между мощностью передатчика на станции и принимаемой мощностью терминала с учетом потерь в разветвителях; показатель деградации как превышение фактического затухания над техническим порогом; тренд изменения принимаемой мощности, вычисляемый через линейную регрессию за последние 24 часа; плотность аварийных сигналов потери сигнала; утилизация беспроводных интерфейсов как произведение загрузки канала и числа подключенных устройств; температурные аномалии через Z-оценку отклонения от среднего.

Сетевые признаки из систем операционной поддержки включают интенсивность ошибок канального уровня (циклические избыточные коды, ошибки выравнивания кадров) за различные временные окна, всплески трафика через 95-й перцентиль минутных измерений, показатели задержки через 95-й перцентиль времени кругового обхода, запас пропускной способности как дополнение среднего трафика до максимальной емкости линии. Признаки истории обращений и полевой службы фиксируют временной лаг до предыдущего рецидива, частоту эскалаций как отношение числа эскалаций к общему числу тикетов за 30 дней, коэффициент переделок для клиента как отношение повторных выездов к общему числу за 90 дней, кросс-признаки типа числа тикетов, зарегистрированных

после аварийных сигналов оптических терминалов. Текстовые описания инцидентов обрабатываются методами TF-IDF для извлечения частотных терминов, моделями BERT для получения семантических векторов с последующим снижением размерности через анализ главных компонент, латентным размещением Дирихле для выделения тематических кластеров симптомов.

Селекция показателей эффективности рекурсивным исключением с кросс-проверкой, что позволяет отобрать подмножество наиболее информативных функций по критерию взаимной информации с независимой переменной. Итоговый набор включает в себя признаки информативности выше порога, обеспечение баланса между полным описанием и вычислительной эффективностью.

Моделирование предполагает сравнение базовых решений и продвинутых алгоритмов с учетом специфики несбалансированных данных. Базовая модель строится на основе логистической регрессии с балансировкой весов классов, что дает точку отсчета для оценки улучшений. Ансамблевые методы на основе градиентного бустинга применяются с параметром масштабирования позитивного класса, ограничением границ деревьев, контролируемой скоростью обучения, ранней остановкой обучения по валидационной выборке и пользовательской уменьшенной потерей, наблюдательной стоимостью ошибок. Рекуррентные нейронные сети с механизмами внимания, используемые для непрерывной обработки временных рядов с двумя слоями, регулярной обработки посредством выпадения и взвешенной задержки бинарной кросс-энтропии.

Гиперпараметры оптимизируются методом древневидного структурированного Парзеневского измерения посредством нескольких сотен итераций с максимизацией площади под кривой точности-полнота при валидационной выборке. Балансировка классов прогрессивной синтетической генерацией включает меньшинства с учетом категориальных признаков, что позволяет достичь более равномерного соотношения классов в обучающей выборке. Калибровка вероятностей применяет метод изотонической регрессии при валидационной выборке для улучшения соответствия предсказанных вероятностей реальной частоте событий.

Метрики качества для несбалансированных данных включают площадь под кривой точности-полнота, которая более информативна при минимальном дисбалансе классов по сравнению с площадью под ROC-кривой; полноту в топ-K предсказаний, исследовательскую долю происходящих событий среди K

наиболее рискованных обращений; F1-меру как гармоническое, среднее и полноты; Экономический лифт как отношение разности ожидаемых затратных ситуаций и предложенной модели к ожидаемым затратам случайной стратегии.

Информационная система и результаты

Проектирование информационной системы предотвращения повторных выездов включает интеграцию с корпоративными ограничениями управления поддержкой с клиентами, поддержку, управление рабочей силой и хранение данных через сервис машинного обучения, механизмы приоритизации и выдачи рекомендаций. Система предотвращения в существующем процессе обслуживания: при регистрации или обработке обращения применяется скорейшее, после чего обеспечивается действие – глубокая диагностика оптической линии, контроль качества закрытия с верификацией, ограничение первопричины, назначение более опытного метода риска с соответствующей сертификацией, дополнительное тестирование оборудования или отложенная проверка через несколько дней после закрытия.

Экспериментальная оценка по выбору методов превосходит базовые решения: квадрат под кривой точность-полнота дает результаты на 60-65% выше результатов модели, полнота в топ-100 предсказаний 90%, что означает выявление большинства проблемных случаев среди наиболее рискованных применений, экономический эффект представляет собой более двукратное улучшение со случайной стратегией. Интерпретация факторного риска с помощью аддитивных объяснений показывает, что ключевыми предикторами происходит деградация оптического сигнала выше порога, малый временной лаг до начала рецидива, высокий уровень групповых ошибок канального уровня, перегрузка беспроводных интерфейсов и отрицательный тренд, принимаемой оптической динамикой, что согласуется с доменными инженерами и повышением доверия к моделям.

Пилотное внедрение в режиме А/В-тестирования с определением трафика на контрольной и экспериментальной группы показывает снижение среднего времени восстановления на 50-55% для применений, тщательного применения прогнозной приоритизации, за счет более раннего порога и ограничения первопричина. Экономический эффект происходит за счет снижения операционных затрат на 25-30% в квартальном исчислении благодаря сокращению повторных выездов и повышению качества закрытия заявок с первого раза, что в денежном выражении вы на этапе составляет несколько миллионов долларов экономии при

инвестициях в развитие и интеграцию системы, что приводит к доходности инвестиций на уровне 200-250% в годовом горизонте.

Предложенная методология прогнозирования повторных обращений в телекоммуникационных службах на основе гетерогенных данных из систем CRM, OSS, диагностики клиентского оборудования и полевой службы с применением методов машинного обучения, адаптированных для несбалансированных данных и экономической асимметрии ошибок, обеспечивает эффективность в задачах превентивного управления качественным обслуживанием. Ключевыми факторами успеха являются телекомспецифические признаки, отражающие физику оптического доступа, динамику сетевой метрики и историю обращений; учет ошибок в стоимости посредством обучения с учетом затрат и снижения порогов принятия решений; интерпретируемость моделей для прогнозирования конкретных действий; MLOps-практики для обеспечения эксплуатационной помощи и контроля качества во времени.

Разработанная информационная система предотвращает повторные выезды, закрывает типичный межсетевой контакт между аналитикой и внедрением через сквозную интеграцию цепочек от данных к моделям, от моделей к рекомендациям, от рекомендаций к действиям полевой службы и от действий к измеримому эффекту в виде снижения времени восстановления и затрат. Практическая готовность обеспечить пилотное управление, продемонстрировав улучшение операционных показателей и экономическую эффективность решений.

Направления дальнейшего развития включают федеративное обучение между территориально распределенными кластерами оборудования для учета региональной специфики без централизации значимых данных; реализация моделей на пограничных вычислительных узлах клиентского оборудования для локального обнаружения аномалий в режиме реального времени; интеграцию больших языковых моделей для углубленного анализа текстовых описаний инцидентов и автоматической генерации рекомендаций по диагностике и устранению неисправностей; Расширение горизонта прогнозирования и детализации направлено на поддержку более широкого спектра управленческих последствий от оперативного реагирования, а также направлено на изменение адаптивности.

Список литературы

1. Суглобова К.Ю. (2020) Экономический эффект от цифровизации отраслей реального сектора экономики в России. Требования к сетям связи нового поколения. REDS: Телекоммуникационные устройства и системы, 10 (4), 30-38.

2. Воспроизводство экономического потенциала региона в условиях цифровой экономики. Морозов И.В. Казань, 2023.
3. Проблемы безопасности интернета вещей. Верещагина Е.А., Капещкий И.О., Ярмонов А.С. Москва, 2021.
4. Информационные системы и технологии в АПК. Бабкина А.В., Быстренина И.Е., Горбачев М.И., Греченева А.В., Ермолаева О.С., Журавлев М.В., Зейлигер А.М., Ивашова О.Н., Красовская Л.В., Лемешко Т.Б., Лосев А.Н., Мелентьев А.Н., Моторин О.А., Никаноров М.С., Пучкова О.С., Пчелинцева С.В., Солошенко А.Д., Степанцевич М.Н., Щедрина Е.А., Худякова Е.В. и др. Москва, 2023.
5. Методы, модели и алгоритмы в системах безопасности : машинное обучение, робототехника, страхование, риски, контроль. Топольский Н.Г., Вилисов В.Я. Москва, 2021.

References

1. Suglobova K. Yu. (2020) The Economic Effect of Digitalization of the Real Economy in Russia. Requirements for Next-Generation Communication Networks. REDS: Telecommunication Devices and Systems, 10 (4), 30-38.
2. Reproduction of a Region's Economic Potential in the Context of the Digital Economy. Morozov I. V. Kazan, 2023.
3. Internet of Things Security Issues. Vereshchagina E. A., Kapetsky I. O., Yarmonov A. S. Moscow, 2021.
4. Information Systems and Technologies in the Agro-Industrial Complex. Babkina A.V., Bystrenina I.E., Gorbachev M.I., Grecheneva A.V., Ermolaeva O.S., Zhuravlev M.V., Zeiliger A.M., Ivashova O.N., Krasovskaya L.V., Lemeshko T.B., Losev A.N., Melentyev A.N., Motorin O.A., Nikanorov M.S., Puchkova O.S., Pchelintseva S.V., Soloshenkov A.D., Stepantsevich M.N., Shchedrina E.A., Khudyakova E.V. and others. Moscow, 2023.
5. Methods, models and algorithms in security systems: machine learning, robotics, insurance, risks, control. Topolsky N.G., Vilisov V.Ya. Moscow, 2021.