**Документация, содержащая информацию, необходимую для эксплуатации экземпляра программного обеспечения**

**Модуль «Предсказание оттока»**

**г. Москва 2025**

# 1. Введение

Модуль «Предсказание оттока» предназначен для прогнозирования вероятности оттока подписчиков онлайн-кинотеатра Premier. Основная цель системы – заблаговременное выявление пользователей с высоким риском оттока для последующей работы с ними через маркетинговые инструменты и повышения бизнес-метрик Заказчика.

# 2. Функционал модуля

Предсказание вероятности оттока на 30 дней (+2 дня на возврат) для всех типов подписок;

Поддержку отдельных моделей для триалистов и пользователей с платной подпиской;

Качество классификации пользователей с ROC-AUC не ниже 0.7;

Автоматическое обновление витрины данных с результатами скоринга ежедневно;

Интерпретируемость предсказаний с возможностью определения ключевых факторов, влияющих на вероятность оттока.

# 3. Начало работы и запуск модуля

Окружение

Модуль «Предсказание оттока» работает в среде Яндекс.Облака и использует следующие сервисы:

Yandex Managed Service for PostgreSQL – для хранения результатов предсказаний;

Проект Data Sphere (Yandex Cloud) – для обучения и предсказаний модели;

S3 bucket (Yandex Cloud) – для хранения промежуточных данных.

**Запуск модуля**

Для работы с модулем администратор должен:

* Войти в Data Sphere (Yandex Cloud)
* Запустить соответствующий Jupyter Notebook

( Воркбенч обновления оттока, Воркбенч с MVP uplift для оттока)

* Проверить соединение с базами данных PostgreSQL и хранилищем S3.
* Запустить пайплайн предсказания.

# **4. Описание алгоритмов и математического обеспечения**

Пайплайн предсказания оттока выполняется в три фазы:

1. Предварительный расчет фичей на Apache Spark, включающий анализ:

Истории просмотров, данных биллинга, показов и кликов (Яндекс.Метрика + Trex) и активности пользователя, сумм его трат и действий на сайте.

1. Обучение и предсказание с помощью временной кросс-валидации моделей градиентного бустинга с бинарным классификатором (отток/не-отток), используя данные за последние 12 месяцев.
2. Обновление и интеграция данных, выгрузка предсказаний в базу PostgreSQL и подготовка отчетов по качеству модели.

**4.1 Потоки данных**

Данные поступают от различных сервисов, включая:

Pass.Media – идентификация пользователей;

Бэкенд-сервисы Premier – информация о подписках и биллинге;

Trex (плеер) – сбор логов событий;

AppMetrica и Яндекс.Метрика – пользовательские события в мобильных и веб-приложениях.

**4.2 Модель данных**

Основные параметры витрины данных о просмотрах:

| **Поле** | **Описание** |
| --- | --- |
| passport\_id | Идентификатор пользователя |
| tv\_id | Идентификатор контента |
| season\_num | Номер сезона |
| episode\_num | Номер эпизода |
| user\_ep\_total\_view\_duaration\_s | Время просмотра эпизода (сек) |
| pdp\_created\_at | Время создания записи |
| actual\_title\_ru | Название контента |
| season\_duration\_s | Длительность сезона (сек) |

**4.4 Интерфейсы взаимодействия**

Модуль передает предсказания через внутренние системы компании в платформу маркетинговых коммуникаций Altcraft, откуда осуществляется взаимодействие с пользователями (email, push, SMS и т.д.).

**5. Ролевая модель**

Система предусматривает три категории пользователей:

Администраторы – сотрудники Предикто, управляющие системой;

Сотрудники Premier – маркетологи, использующие предсказания для кампаний;

Пользователи Premier – получатели персонализированных маркетинговых предложений.

**Администраторы.** Контролируют работу модуля и корректность данных. Запускают пайплайны предсказаний и обновления витрины данных. Настраивают параметры модели и следят за качеством предсказаний. Мониторят работу базы данных и серверов.

**Сотрудники**. Используют дашборды для анализа предсказаний оттока. Планируют маркетинговые кампании на основе полученных данных. Настраивают сегменты пользователей для удержания через Altcraft.

**

*Рис. 1 пример одного из операционных дашбордов для монитроинга состояния базы данных*

**Пользователи**. Получают персонализированные предложения на основе предсказаний оттока. Взаимодействуют с маркетинговыми коммуникациями (push, email, SMS). Реагируют на предложения, что может повлиять на их статус в системе.