**Документация, содержащая описание функциональных характеристик модуля программного обеспечения**

**Модуль «Предсказание оттока»**

**г. Москва 2025**

**Содержание**

[1. Предназначение ПО 3](#_heading=h.s2xpnm9x5zoo)

[2. Описание функциональных характеристик 3](#_heading=h.7ep7zxhbxr7i)

[3. Архитектура модуля «Предсказание оттока» 3](#_heading=h.fuxsqccpbuer)

[4. Схема работы 4](#_heading=h.3idki31in34i)

[5. Используемые языки программирования и фреймворки 6](#_heading=h.x6192k3qe3ls)

# Предназначение ПО

Модуль "Предсказание оттока" предназначен для заблаговременного определения вероятности оттока подписчиков. Это позволяет эффективно выстраивать коммуникацию с пользователями с целью их удержания и повышения бизнес-метрик Заказчика.

# Описание функциональных характеристик

Модуль предоставляет вероятность оттока по каждому пользователю (скоринг пользователей).

* В основе продукта - ML-модели предсказания оттока.
* Прогнозирование вероятности оттока в течение 30 дней + 2 дня на "возврат".
* Отдельные модели для триалистов и пользователей с платной подпиской.
* ROC-AUC моделей не ниже 0.7.
* Финальный результат - витрина с вероятностями оттока, обновляемая ежедневно.
* Инструмент мониторинга качества модели.
* Интерпретация результатов с возможностью определения ключевых атрибутов.
* Используемые данные:
	+ Биллинг
	+ Статистика пользования сервисом
	+ Поведение пользователей
	+ Ключевые характеристики пользователей
* Информация для рекомендаций оптимальных каналов взаимодействия с пользователем и ценностных предложений.

# Архитектура модуля «Предсказание оттока»



*Рис. 1 Диаграмма концептуальной архитектуры модуля*

Функциональность: предсказание оттока, интеграция с маркетинговыми инструментами.

Технологический стек: Yandex.Облако (Data Sphere, Dataproc), Python, PostgreSQL

Перечень сервисов, явно или неявно являющихся источником данных для модуля «Предсказание оттока»:

Pass.Media - cервис для регистрации, авторизации и аутентификации пользователей Премьер;

Бэкенд сервисы - различные бэкенд сервисы, обеспечивающие работу кинотеатра;

Сервис подписок - cервис для управления подписками пользователей, содержащий информацию об оплатах подписки;

Трекс (Плеер) - cбор логов (событий) плеера в режиме реального времени

AppMetrica - cервис для сбора пользовательских событий в мобильных приложениях;

Яндекс.Метрика - cервис для сбора пользовательских событий в web приложении.

# Схема работы

Источники данных:

| Яндекс Метрика | События пользователя в web приложении  |
| --- | --- |
| АппМетрика | События пользователя в мобильных приложениях  |
| Trex | События пользователя в приложениях для smart TV  |
| Player/Trex | Логи плеера (Трекс) |
| Бэкенд сервисы/Каталог контента | Метаданные контента  |
| Бэкенд сервисы/Сервис подписок | Информация об оплате подписок (биллинг) |

Ролевая модель

| Администраторы  | разработка и эксплуатация модуля. |
| --- | --- |
| Сотрудники партнеров  | работа с маркетинговыми кампаниями через Altcraft |
| Пользователи  | получение персонализированной маркетинговой коммуникации |

Все данные попадают в **Платформу данных** которая собирает информацию из всех источников. Сохраняет и упорядочивает её. Создаёт отчёты и прогнозы.

Технологии: Python, ClickHouse, S3, DataSphere, Dataproc, Dagster.

На основе собранных данных модель машинного обучения предсказывает, кто может отказаться от подписки

Использует разные факторы такие как просмотры, платежи, активность. Строит прогноз для каждого пользователя. Помогает выбрать лучший способ удержания (например, скидки или персональные рекомендации).

Технологии: DataSphere, Dataproc, PostgreSQL, Python.

Предсказания передаются в Altcraft – систему для маркетинга.

Если пользователь рискует уйти, ему можно отправить push-уведомление или письмо с персональным предложением.

**Идентификация, авторизация и аутентификация пользователей**

Администраторы - Сотрудники, отвечающие за разработку и эксплуатацию продукта. Используют инструменты разработки ПО и консоль Яндекс. Облака Идентификация, авторизация и аутентификация таких сотрудников происходит на основании аккаунтов Яндекс

Сотрудники (и/или подрядчики, парнтеры и т.д.) не взаимодействуют напрямую с модулем «Предсказание оттока». Они имеют возможность работать с результатами работы модуля через платформу маркетинговой коммуникации Altcraft, данные в которую передаются из модуля «Предсказание оттока». Сама платформа Altcraft не является частью модуля «Предсказание оттока»

Пользователи. Персонализированный скоринг вероятности оттока получают зарегистрированные пользователи при наличии достаточного количества информации о их действиях. Идентификация зарегистрированных пользователей для предсказания оттока происходит на основании ID. Дальнейшее маркетинговое взаимодействие с пользователями (на основе результатов работы модуля) происходит через инструменты платформы Altcraft

Пайплайн скоринга клиентов осуществляется в 3 фазы:

Фаза предрасчета фичей (Apache Spark):

* Биллинг
* История просмотров
* Взаимодействие с web-сервисом (Яндекс.Метрика, Trex)

Фаза переобучения модели (градиентный бустинг, временная кросс-валидация, учет сезонных факторов).

Фаза обновления данных (выгрузка вероятностей оттока в PostgreSQL, мониторинг качества предсказаний).

**

*Рис. 2 Потоки данных от источников до сервиса API рекомендаций организованы в соответствии со следующей диаграммой DFD*

# Используемые языки программирования и фреймворки

| Языки программирования | Python |
| --- | --- |
| Фреймворки и библиотеки | ML-библиотеки: CatBoost, Scikit-learn, XGBoost |
| Обработка данных | Pandas, NumPy |
| Хранилища данных | PostgreSQL, S3 (Yandex.Cloud) |
| Оркестрация | Apache Airflow, Dagster |
| Облачные сервисы | Yandex DataSphere, Dataproc |

1. Эксплуатация системы

Администрирование компонентов

PostgreSQL - Регулярные обновления и мониторинг производительности.Настройка резервного копирования и восстановления. Оптимизация запросов и индексов для улучшения скорости обработки данных.

Data Sphere - Управление вычислительными ресурсами. Мониторинг выполнения задач машинного обучения. Настройка окружения для ML-разработки.

S3 бакет - Управление правами доступа. Контроль версий и управление архивами данных. Мониторинг использования и оптимизация хранилища.

Внешние интеграции

Altcraft - Передача данных о предсказанном оттоке для настройки маркетинговых кампаний. Обратная связь о результатах взаимодействия с пользователями.

PostgreSQL - Интеграция с аналитическими сервисами для построения BI-отчетов. Использование SQL-запросов для получения агрегированной информации о пользователях.

Yandex.Cloud - Автоматизированное масштабирование инфраструктуры. Настройка логирования и мониторинга через встроенные инструменты Yandex.Cloud.