# **# 3 Предсказание оттока**

**Пункт 1 [Решение задачи предсказания оттока игрока]**

А. Сбор и подготовка данных

* Сбор данных: необходимо собрать данные о каждом игроке, включая уникальный ID, время прохождения каждого шага туториала и шаг, на котором игрок прекращает игру.
* Предобработка: очистка данных, обработка пропусков, кодирование категориальных переменных и нормализация данных.

Б. Исследовательский анализ данных (EDA)

Анализ поведения игроков: оценка распределения оттока по шагам туториала, выявление шаблонов и аномалий.

В. Выбор признаков

Признаки: использование времени прохождения каждого шага, общего времени в игре, предыдущих действий игрока и других доступных метрик.

Г. Разработка модели

* Выбор модели: можно использовать модели классификации, такие как логистическая регрессия, случайный лес, градиентный бустинг или нейронные сети.
* Обучение модели: разделение данных на обучающую и тестовую выборки, обучение модели на обучающей выборке.

Д. Валидация и Тюнинг Модели

* Кросс-валидация: применение кросс-валидации для оценки стабильности модели.
* Тюнинг гиперпараметров: использование методов, таких как Grid Search или Random Search, для оптимизации гиперпараметров.

**Пункт 2 [Оценка качества полученной модели]**

А. Метрики качества

* Точность (Accuracy): общая доля правильных предсказаний.
* Точность (Precision) и полнота (Recall): важны для баланса между избежанием ложных предсказаний оттока и упущением реальных случаев.
* F1-мера: гармоническое среднее между точностью и полнотой.
* ROC-AUC: площадь под кривой рабочих характеристик приемника.
* Confusion Matrix: предоставляет полное представление о производительности модели.

Б. Выбор Оптимального Метода Оценки

* F1-мера и ROC-AUC считаются наиболее подходящими, так как они предоставляют сбалансированный взгляд на производительность модели в условиях несбалансированных классов (большинство игроков, скорее всего, не прекратят играть на каждом шаге).
* Confusion Matrix поможет понять специфику ошибок модели (ложноположительные и ложноотрицательные результаты).

Заключение

1. Подход к решению задачи предсказания отвала игрока включает в себя тщательный анализ данных, выбор подходящих моделей и метрик для оценки их эффективности.
2. Особое внимание следует уделить балансу между различными типами ошибок предсказания, чтобы максимизировать общую пользу от модели в реальных условиях эксплуатации игры.

**Пункт 3 [Дизайн эксперимента и методологию по проверке гипотезы геймдизайнеров]**

Дизайн Эксперимента

*1. Формирование Групп*

* Контрольная группа: игроки, проходящие стандартный туториал.
* Экспериментальная группа: игроки, для которых туториал изменяется на основе предсказаний модели.

*2. Рандомизация*

Произвольное распределение новых игроков между контрольной и экспериментальной группами для обеспечения сопоставимости групп.

*3. Временной горизонт*

Определение периода эксперимента, достаточного для сбора значимых данных (например, несколько недель или месяцев).

*4. Трекинг метрик*

* Основная метрика: процент игроков, завершивших туториал.
* Вторичные метрики: время, проведенное в игре, уровень вовлеченности, конверсия в платящих пользователей и т.д.

Методология Проверки Гипотезы

*1. Предварительный анализ данных*

Проверка данных на наличие аномалий и предварительный анализ для убеждения в корректности сбора данных.

*2. Анализ Результатов*

* Сравнение процентов завершения туториала между контрольной и экспериментальной группами.
* Использование статистических тестов, например, t-теста, для определения наличия статистически значимых различий.

*3. Дополнительный анализ*

* Анализ вторичных метрик для понимания влияния изменений на общее поведение игроков.
* Оценка потенциального влияния изменений на долгосрочную удержание пользователей и монетизацию.

*4. Оценка результатов*

* Анализ полученных результатов для определения эффективности внесенных изменений.
* При положительных результатах – рассмотрение возможности постоянного внедрения изменений в туториал.
* При отрицательных или неоднозначных результатах – дальнейший анализ и возможная корректировка подхода.

Заключение

1. Эксперимент должен быть спланирован и проведен с учетом строгих статистических методов для обеспечения точности и надежности результатов.
2. Оценка эффективности изменений в туториале на основе данных, собранных в ходе эксперимента, позволит сделать обоснованное решение относительно дальнейшего использования предложенного подхода в игре.