<имя функции>

- Назначение функции
- Используемый метод
- Параметры входа
- Параметры выхода

1. Статистические методы

- **check_duplicate_events**
  • Поиск дубликатов событий с одинаковыми временными метками и типами событий.
  • Подсчет точных совпадений по ключевым полям
  • ts, event, counter_id, randPAS_session_id, ip.
  • Словарь с дубликатами или сообщением об их отсутствии.

- **check_session_start**
  • Проверка корректности начала сессий.
  • Валидация начальных значений порядковых номеров
  • randPAS_session_id, page_view_order_number, event_order_number, ip.
  • Словарь с ошибками или подтверждением корректности.

- **check_order_relation**
  • Проверка соответствия между номерами событий и просмотров страниц.
  • Сравнение порядковых номеров событий и просмотров
  • randPAS_session_id, event_order_number, page_view_order_number, ip, ts.
  • Словарь с ошибками или подтверждением корректности.

- **check_numbering_sequence**
  • Проверка последовательности нумерации событий.
  • Анализ непрерывности последовательности номеров
  • randPAS_session_id, event_order_number, ip, ts, event.
  • Словарь с ошибками или подтверждением корректности.

- **detect_location_changes**
  • Анализ смены местоположения пользователей.
  • Трекинг изменений геоданных во времени

  - randPAS_user_passport_id, ts, ip, geo_city_id.
  - DataFrame с информацией о сменах местоположения.

- **zscore_detector**
  - Обнаружение аномалий методом Z-Score.
  - Стандартизация и пороговая фильтрация
  - Одномерный массив значений.
  - Бинарный массив.

- **iqr_detector**
  - Обнаружение аномалий методом IQR.
  - Межквартильный размах
  - Одномерный массив значений.
  - Бинарный массив.

- **modified_zscore_detector**
  - Обнаружение аномалий устойчивым методом Z-Score.
  - Медианные абсолютные отклонения
  - Одномерный массив значений.
  - Бинарный массив.

- **percentile_detector**
  - Обнаружение аномалий по перцентилям.
  - Перцентильный анализ
  - Одномерный массив значений.
  - Бинарный массив.

- **count_device_types**
  - Подсчет распределения типов устройств.
  - Группировка и агрегация
  - ua_is_mobile, ua_is_tablet, ua_is_pc.
  - Series с количеством по типам устройств.

2.Методы машинного обучения

- **detect_anomalous_users**
  • Поиск аномальных пользователей по времени на страницах.
  • Isolation Forest
  • randPAS_user_passport_id, url, ts.
  • DataFrame с аномальными пользователями.

- **isolation_forest_detector**
  • Обнаружение аномалий через Isolation Forest.
  • Алгоритм Isolation Forest
  • Одномерный массив значений.
  • Бинарный массив.

- **lof_detector**
  • Обнаружение аномалий через LOF.
  • Local Outlier Factor
  • Одномерный массив значений.
  • Бинарный массив (1 — аномалия).

3.Анализ временных рядов

- **analyze_city_activity**
  • Выявление аномальной активности по городам.
  • Z-score + скользящие окна
  • geo_city_id, ts.
  • Кортеж из DataFrame с аномалиями и общей активностью.

- **detect_anomalous_time_windows**
  • Поиск аномальных временных окон активности.
  • Анализ временных интервалов
  • randPAS_user_passport_id, ts.
  • DataFrame с аномальными окнами.

- **detect_anomalous_device_shares**
  • Анализ аномальных изменений в распределении устройств.
  • Перцентильный анализ долей
  • ts, ua_device_family.
  • DataFrame с аномалиями.

- **detect_anomalous_page_views**
  • Выявление аномальных изменений в просмотрах страниц.
  • Анализ изменений во времени
  • ts, url, randPAS_user_passport_id.
  • DataFrame с аномалиями.

4.Сетевой анализ

- **detect_suspicious_ips**
  • Поиск IP с аномальным числом пользователей.
  • Анализ распределения пользователей по IP
  • ip, randPAS_user_passport_id, ts.
  • DataFrame с подозрительными IP.

- **detect_user_activity_spikes**
  • Обнаружение пользователей с аномально высокой активностью.
  • Анализ временных интервалов между событиями
  • randPAS_user_passport_id, ts, ip.
  • DataFrame с подозрительными пользователями.

5.Ансамблевые методы

- **majority_anomaly_vote**
  • Объединение результатов разных методов.
  • Мажоритарное голосование
  • Произвольное число бинарных массивов.
  • Бинарный массив.

6.Визуализация

- **plot_anomalies_comparison**
  • Визуализация сравнения методов обнаружения аномалий.
  • Совмещенное отображение результатов

  - Исходные данные и бинарные массивы аномалий.
  - График.

- **plot_location_change_distribution**
  - Визуализация распределения изменений местоположения.
  - Логарифмическая гистограмма
  - location_changes_df, threshold.
  - График распределения.

- **plot_top_anomalous_cities**
  - Визуализация топ аномальных городов.
  - Временные ряды с маркерами
  - city_activity_df, top_n.
  - График временных рядов.

- **plot_top_suspicious_ips**
  - Визуализация подозрительных IP-адресов.
  - Столбчатая диаграмма
  - suspicious_ips_df, top_n.
  - Столбчатая диаграмма.

- **plot_ip_bubble_chart**
  - Визуализация анализа IP в формате bubble chart.
  - Пузырьковая диаграмма
  - suspicious_ips_df.
  - Пузырьковая диаграмма.

- **plot_anomalous_time_windows**
  - Визуализация аномальных временных окон.
  - Стековая диаграмма с маркерами
  - anomalous_df, full_stats_df.
  - График временных рядов с выделением аномалий.

- **plot_anomalous_page_views**
  • Визуализация аномальных просмотров страниц.
  • Scatter plot с размером точек
  • anomalies_df.
  • Точечная диаграмма.

- **plot_anomalous_users**
  • Визуализация аномальных пользователей.
  • Scatter plot с цветовой кодировкой
  • user_df.
  • Точечная диаграмма.

- **plot_device_share_changes**
  • Визуализация изменения долей устройств.
  • Линейный график с маркерами
  • device_shares, anomalous_windows.
  • Линейный график.

- **plot_user_activity_spikes**
  • Визуализация спайков активности пользователей.
  • Комбинированный bar+line chart
  • suspicious_users, top_n, exclude_top_outlier.
  • Комбинированный график.