

Модель определения абонентов
(msisdn), которые являются
одной и той же Персоной
(человеком).

Данные

[01_Факты.xlsx](#) обучающая выборка: набор пар номеров телефонов (msisdn), каждая из которых с большой долей вероятности принадлежит одной Персоне.

[02_Data.csv](#) Информация о перемещении случайного набора абонентов за период с 2013-05-22 20:00:00 до 2013-05-30 16:29:37

[03_устройства.csv](#) Сокращенный справочник ТАС

[04_event_type.xlsx](#) Справочник типов регистраций с расшифровкой

Описание алгоритма

Алгоритм основан на исследовании о том, что можно практически с 95 % гарантией идентифицировать личность пользователя мобильного устройства, зная только лишь информацию о подключении к базовым станциям , через которые он выходил на связь через определенные промежутки времени (1,5-2 часа).

Jakob, E. B. Context-Aware User Authentication – Supporting Proximity-Based Login in Pervasive Computing / Jakob E. Bardram, Rasmus E. Kjaer, Michael F. Pedersen, – Berlin: UbiComp, LNCS2864. – № 2003. – P. 107-123

Подготовка данных

Для каждого уникального номера телефона (msisdn) агрегируются данные о местоположении базовых станций, к которым подключается абонент с временным окном в 1 час.

Далее набор данных был преобразован в сводную таблицу, где для каждого номера представлена история подключения к базовым станциям за весь период наблюдения.м

На графике нанесены координаты базовых станций к которым подключался каждый уникальный номер за весь период наблюдения. Каждый кластер (пара номеров) из 01_Факты окрашена каждый в свой цвет.

На графике прослеживается корреляция изменений местоположения пары номеров принадлежащей одной персоне.

2013-05-26

2013-05-25

2013-05-24

2013-05-23

-4

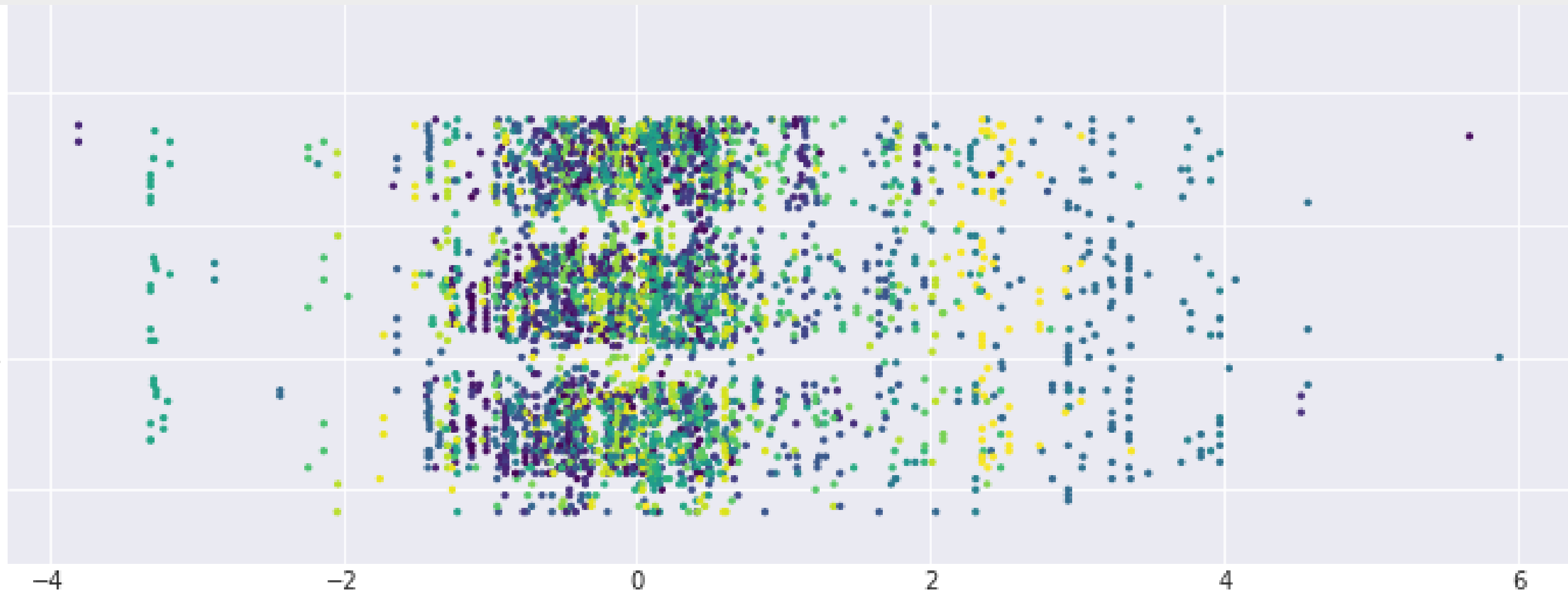
-2

0

2

4

6



Поиск пары номеров

Для поиска пары номеров была применена матрица корреляции между данными всех пар номеров с последующей фильтрации по порогу.