

25 июня 2024 г.

ПРОГНОЗИРОВАНИЕ ВИДА ДВИЖЕНИЯ ЧЕЛОВЕКА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ НА ОСНОВЕ ДАННЫХ С ДАТЧИКОВ СМАРТФОНА

Выполнил:

Студент гр. 5030103/00201

А.А. Балашов

Руководитель:

к.ф.-м.н.

С.А. Щербинин

Консультант:

А.Д. Ершов



АКТУАЛЬНОСТЬ

Актуальность работы заключается в востребованности пользователями получения точных данных о собственной активности.

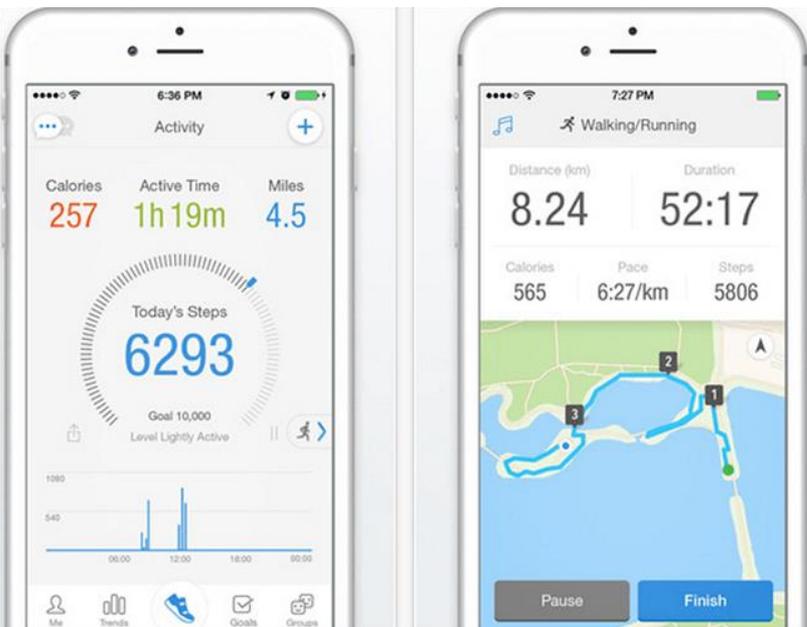
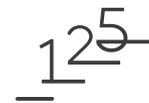


Рис. 1 – Пример интерфейса приложения



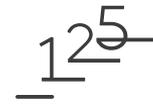
Рис. 2 – График запросов по ключевому слову “шагомер”, мес.

ЦЕЛИ РАБОТЫ



Цель - написание программы, которая считывает данные, полученные с датчиков смартфона, обрабатывает их и строит предиктивную модель, определяющую вид движения человека, предоставляет визуальный анализ активности человека.

ЗАДАЧИ РАБОТЫ



1. Обработка данных, подготовка их к использованию программой;
2. Формулировка задачи в терминах машинного обучения;
3. Реализация алгоритма машинного обучения, определяющего вид активности;
4. Исследование влияния различных параметров на качество и время выполнения программы, настройка параметров для оптимизации показателей;
5. Представление результатов в виде графического анализа активности человека.

ВИДЫ ДВИЖЕНИЯ ЧЕЛОВЕКА

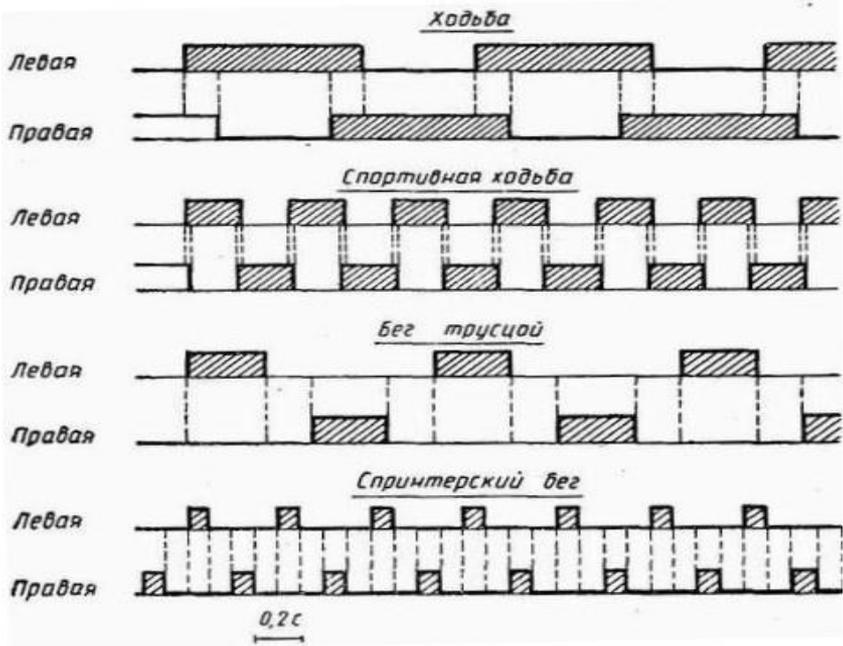


Рис. 3 – Простейшие хронограммы обычной ходьбы, спортивной ходьбы, бега трусцой и спринтерского бега; периоды опоры заштрихованы; вверху – левая нога, внизу – правая

ВХОДНЫЕ ДАННЫЕ

125



	Дата	Время	Имя	Запястье	Активность	
1	2017-6-30	13:51:16.04	viktor	0	0	
2	2017-6-30	13:51:16.25	viktor	0	0	
3	2017-6-30	13:51:16.45	viktor	0	0	
4	2017-6-30	13:51:16.65	viktor	0	0	
5	2017-6-30	13:51:16.85	viktor	0	0	
	Ускорение по оси x	Ускорение по оси y	Ускорение по оси z	Гироскоп по оси x	Гироскоп по оси y	Гироскоп по оси z
1	0.2650	-0.7814	-0.0076	-0.0590	0.0325	-2.9296
2	0.6722	-1.1233	-0.2344	-0.1757	0.0208	0.1269
3	0.4399	-1.4817	0.0722	-0.9105	-0.9105	-2.4367
4	0.3031	-0.8125	0.0888	0.1199	0.1199	-2.9336
5	0.4814	-0.9312	0.0359	0.0527	0.0527	2.4922

Таблица 1 – Входные данные

ПОСТАНОВКА ЗАДАЧИ

Имеется множество объектов X и множество ответов Y .

Существует функциональная зависимость $f: X \rightarrow Y$, но она неизвестна.

Известна лишь совокупность S пар вида (объект, ответ), называемая обучающей выборкой:

$$S = \{(x_i, y_{x_i} = f(x_i)) \in X \times Y \mid i = 1, \dots, l\}.$$

При этом множество ответов Y является конечным, а каждый ответ $y \in Y$ соответствует некоторому классу объектов.

Задача классификации заключается в вычислении для каждого объекта соответствующего ему класса.

В данной работе множество объектов определяется признаками (показаниями датчиков), а классы являются значениями целевой переменной, активности (ходьба, бег).

АНАЛИЗ ДАННЫХ

Графическое представление распределения данных по переменным:

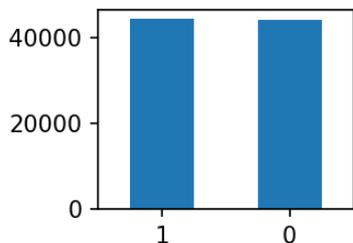


Рис. 4 – Активность

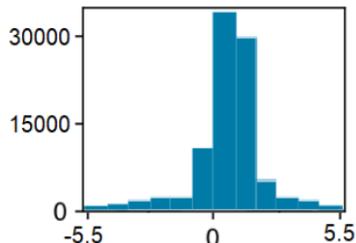


Рис. 5 – Ускорение по x

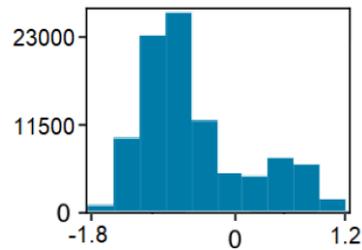


Рис. 6 – Ускорение по y

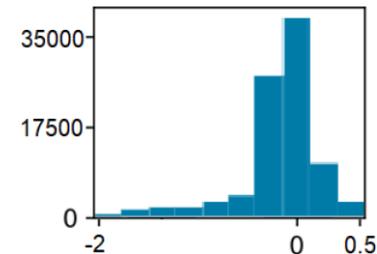


Рис. 7 – Ускорение по z

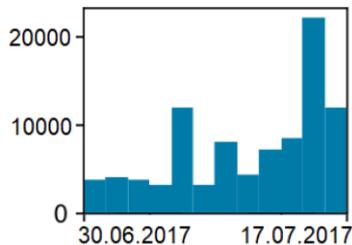


Рис. 8 – Даты

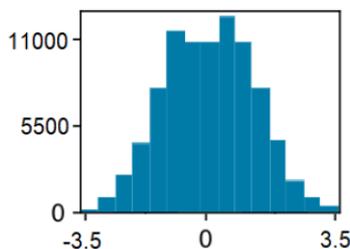


Рис. 9 – Гироскоп по x

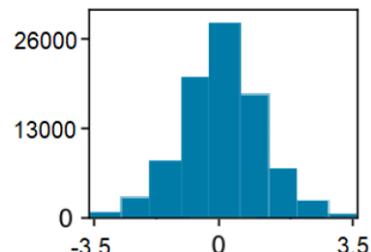


Рис. 10 – Гироскоп по y

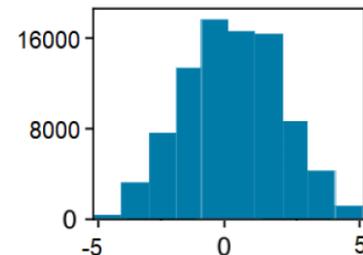


Рис. 11 – Гироскоп по z

ВАЛИДАЦИЯ ДАННЫХ

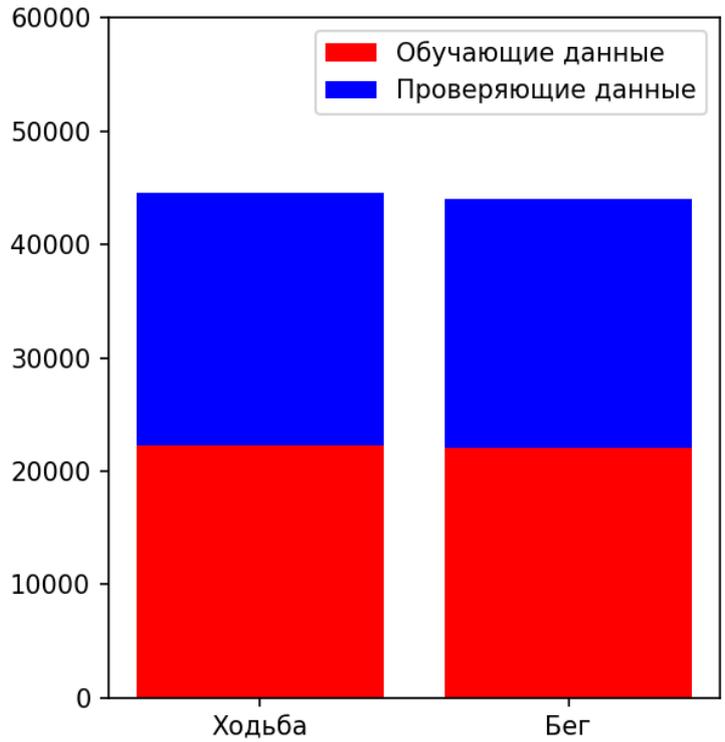


Рис. 12 – Разбиение данных на две группы, красным – распределение значений целевой переменной по обучающей выборке, синим - по проверяющей

ИСПОЛЬЗУЕМЫЕ МЕТОДЫ

1. К-БЛИЖАЙШИХ СОСЕДЕЙ

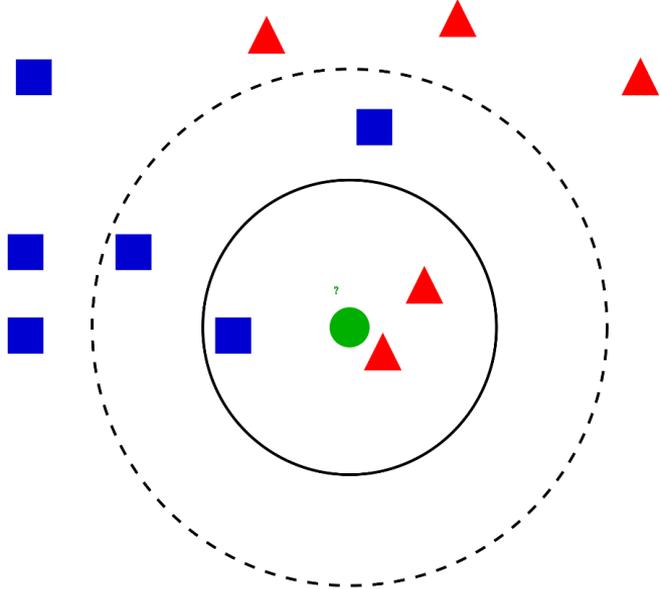


Рис. 13 – Иллюстрация метода К-ближайших соседей

ИСПОЛЬЗУЕМЫЕ МЕТОДЫ

2. МЕТОД СВЕРХСЛУЧАЙНЫХ ДЕРЕВЬЕВ

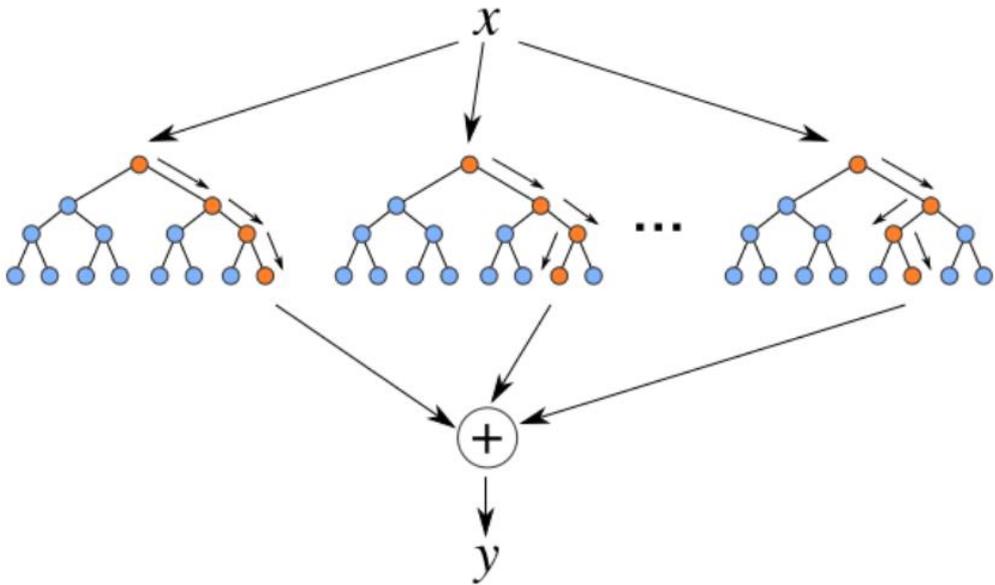
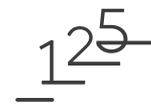


Рис. 14 – Иллюстрация метода сверхслучайных деревьев

ИСПОЛЬЗУЕМЫЕ МЕТОДЫ

3. КЛАССИФИКАТОР БАЙЕСА

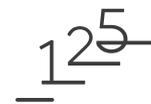


Показания датчиков распределены по Гауссу, поэтому логично использовать гауссовский наивный байесовский классификатор:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Где $P(x_i|y)$ – вероятность отношения признака x_i к классу y , μ_y и σ_y – математическое ожидание и дисперсия x_i .

ОСНОВНЫЕ МЕТРИКИ



Точность позволяет наиболее полно оценить качество модели в случае, когда данные сбалансированы и нет приоритета в качестве оценки одного из классов:

$$\text{Точность} = \frac{\text{Число верных предсказаний}}{\text{Общее число предсказаний}}$$

ОТБОР ПРИЗНАКОВ

125



Зависимость точности от набора признаков

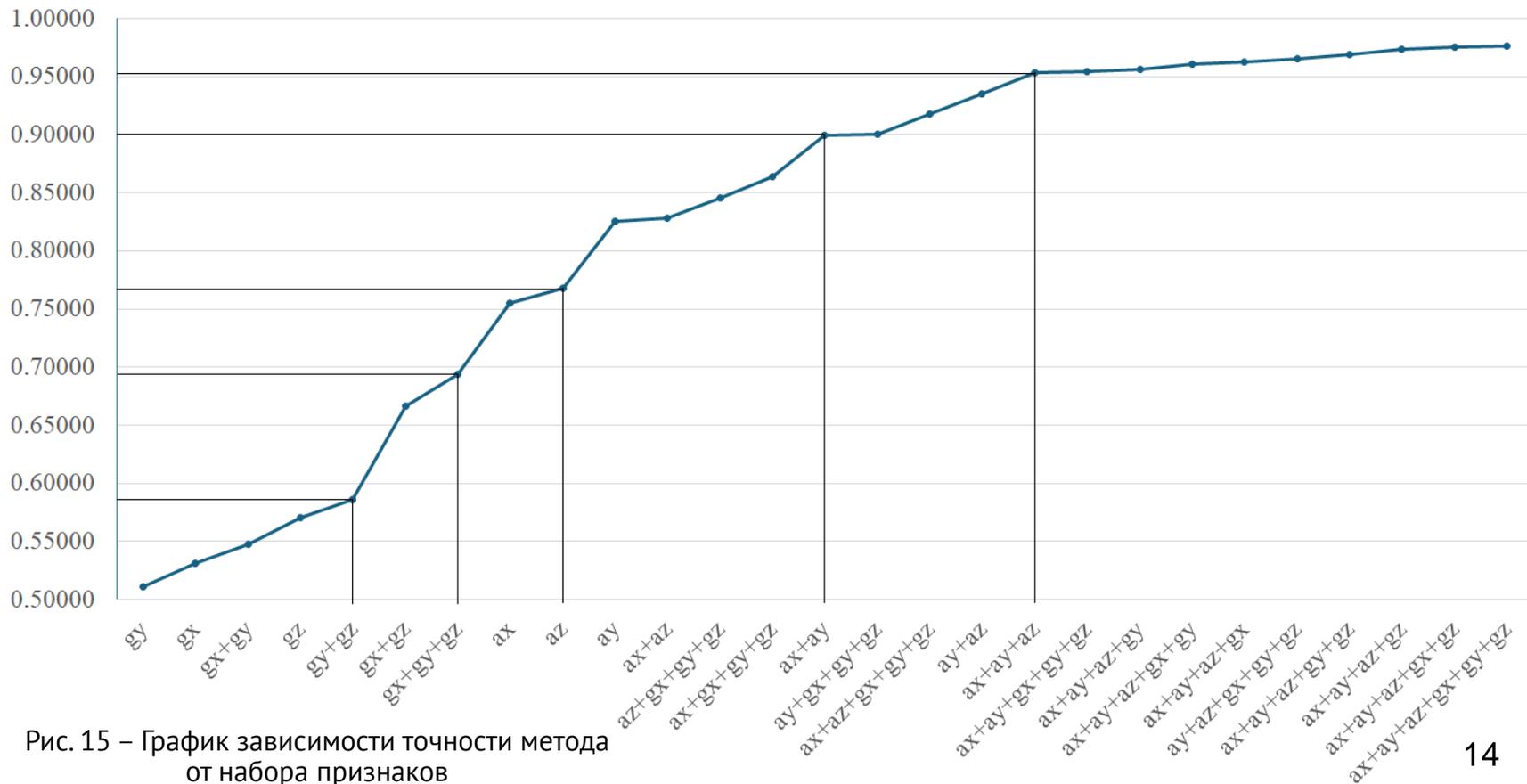


Рис. 15 – График зависимости точности метода от набора признаков

СРАВНЕНИЕ МЕТОДОВ

Метод	Точность	Время, с
К-Ближайших соседей	0.9601	2.7960
Наивный Байес	0.957	0.0164
Сверхслучайных деревьев	0.9538	1.3724

Таблица 2 – Сводная таблица методов,
желтым выделен лучший результат

НАСТРОЙКА МЕТОДА

1. МЕТОД К-БЛИЖАЙШИХ СОСЕДЕЙ

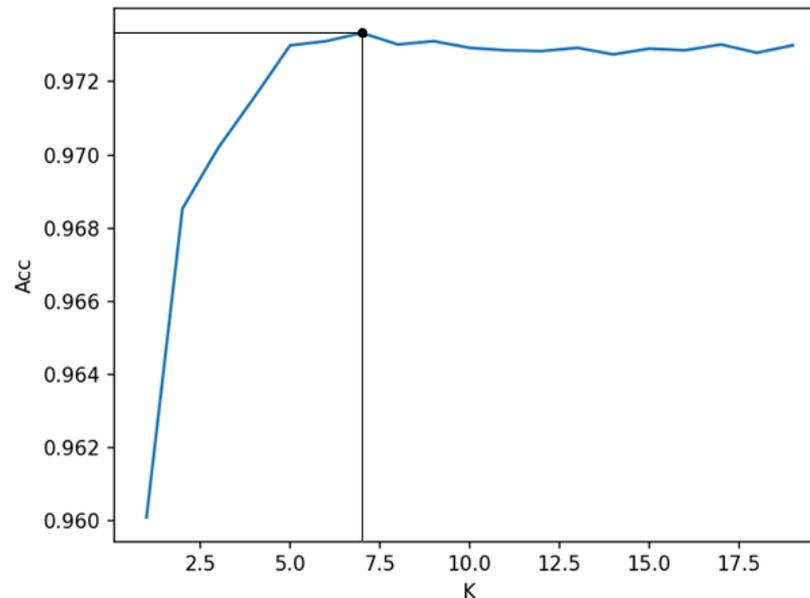


Рис. 16 – График зависимости точности от количества соседей K

НАСТРОЙКА МЕТОДА

2. МЕТОД СВЕРХСЛУЧАЙНЫХ ДЕРЕВЬЕВ

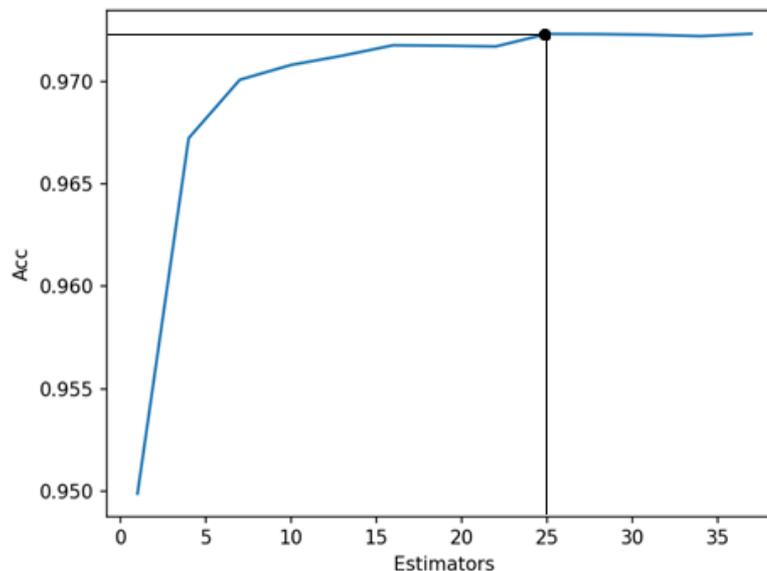


Рис. 17 – График зависимости точности от числа деревьев

НАСТРОЙКА МЕТОДА

2. МЕТОД СВЕРХСЛУЧАЙНЫХ ДЕРЕВЬЕВ

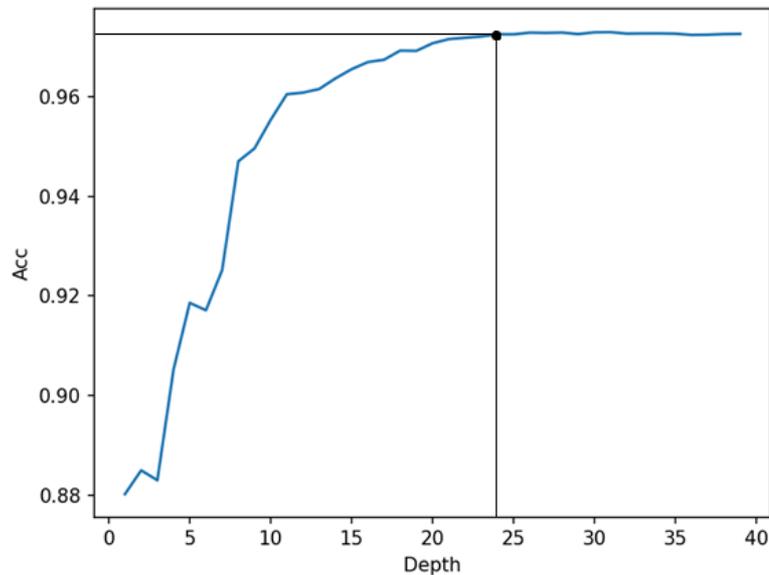
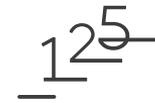


Рис. 18 – График зависимости точности от глубины дерева

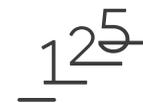
СРАВНЕНИЕ МЕТОДОВ ПОСЛЕ НАСТРОЙКИ



Метод	Точность	Время, с
К-Ближайших соседей	0.9733	1.8856
Сверхслучайных деревьев	0.9725	0.4538
Наивный Байес	0.9570	0.0164

Таблица 3 – Сводная таблица методов после настройки,
желтым выделен лучший результат

АНАЛИЗ РЕЗУЛЬТАТОВ



Метод	Точность
К-ближайших соседей, сдвоенные аномалии удалены	0.99801
К-ближайших соседей, единичные аномалии удалены	0.99589
К-ближайших соседей, без фильтра	0.97334

Таблица 4 – Сводная таблица точности алгоритма
с фильтрами и без

ЭФФЕКТИВНОСТЬ ФИЛЬТРА

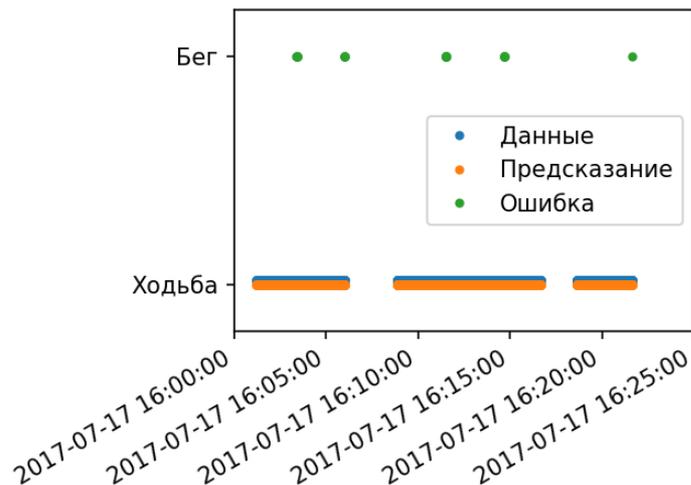


Рисунок 19 – Графическое представление активности человека, масштаб – 25 минут, фильтр на аномалии

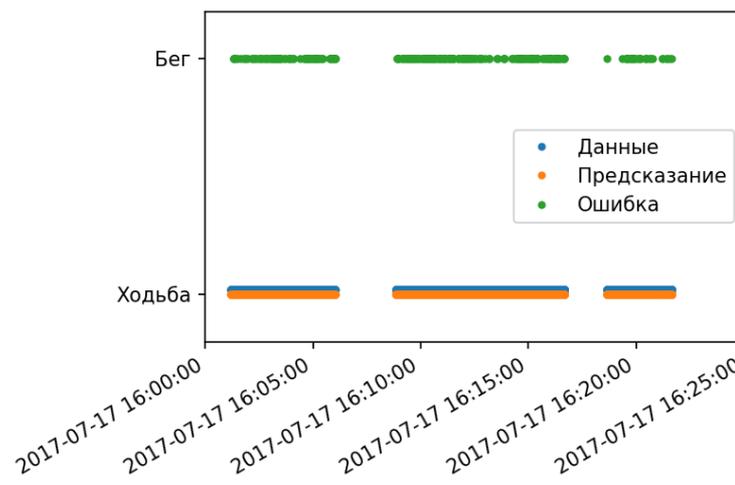


Рисунок 20 – Графическое представление активности человека, масштаб – 25 минут, фильтр отключен

ВИЗУАЛИЗАЦИЯ ОШИБКИ

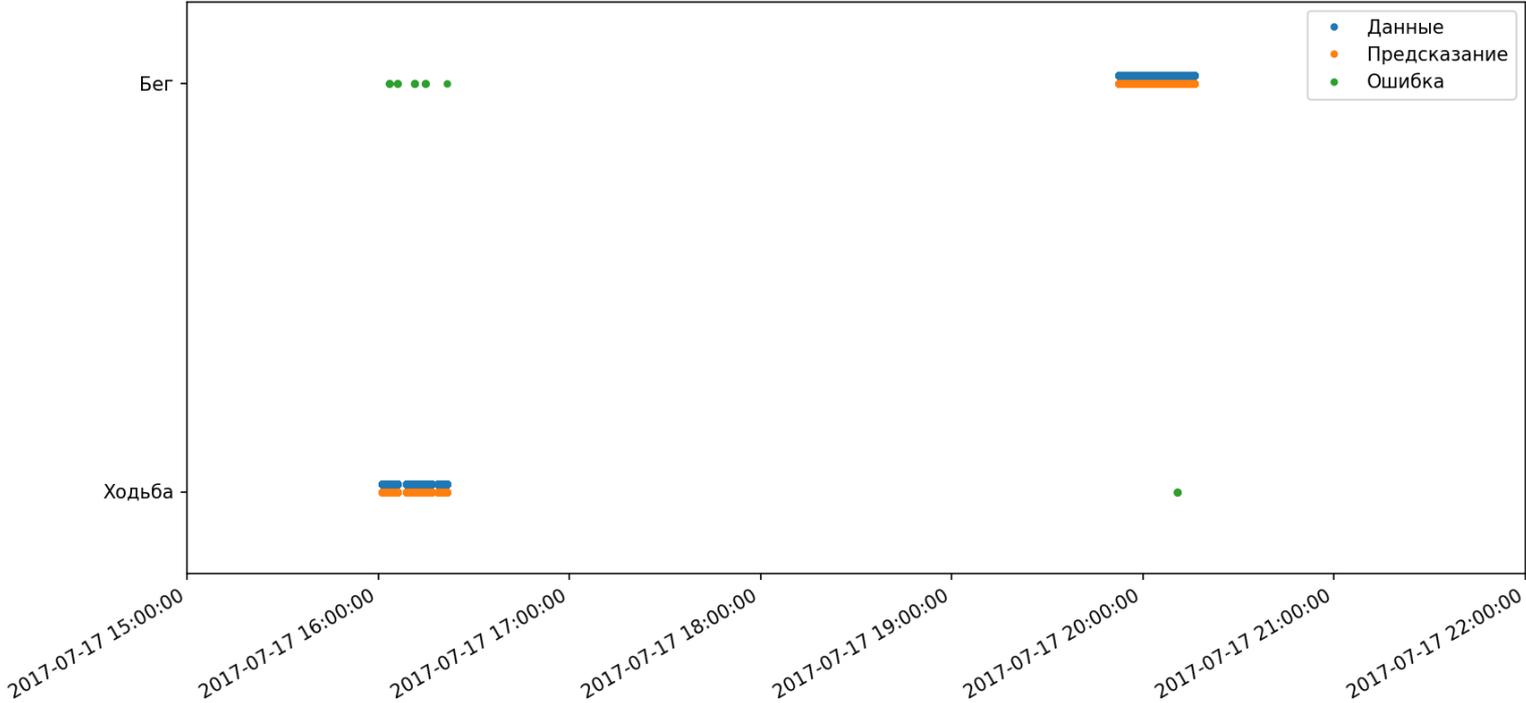


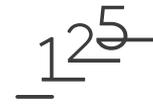
Рис. 21 – Графическое представление активности человека, масштаб – 7 часов, на графике синим показаны исходные данные, оранжевым – правильно предсказанные, зеленым – предсказанные с ошибкой

РЕЗУЛЬТАТЫ В ЦИФРАХ

Точность модели превысила 0.998, то есть за час в среднем будет ошибочно определено только полторы минуты активности.

Время, которое требуется алгоритмам для предсказания 2.5 часов активности, составило 1.9 секунды для самого точного метода и 0.016 секунды для самого быстрого метода.

ЗАКЛЮЧЕНИЕ



В ходе данной работы выполнены следующие задачи:

1. Обработаны данные, подготовлены к использованию программой;
2. Реализован алгоритм машинного обучения, определяющий значение переменной "активность";
3. Исследовано влияние различных параметров на качество и время выполнения программы, произведена настройка параметров для оптимизации показателей
4. Проверена и подтверждена гипотеза о ложности единичных и сдвоенных аномалий;
5. Представлены результаты в виде графического анализа активности человека.

СПАСИБО ЗА ВНИМАНИЕ!

