



Алгоритм для определения видов ящериц

Состав команды:

Шифон Яков

Скрябин Иван

Кадырова Ольга



Введение

Перед нами стояла задача на основе известных данных выявить критерии для определения вида ящериц рода *Darevskia*. Для этого мы придумали алгоритм, в основе работы которого заложен кластерный анализ.

Шаги работы



01

Анализ имеющихся данных

02

Построение алгоритма

03

Написание кода

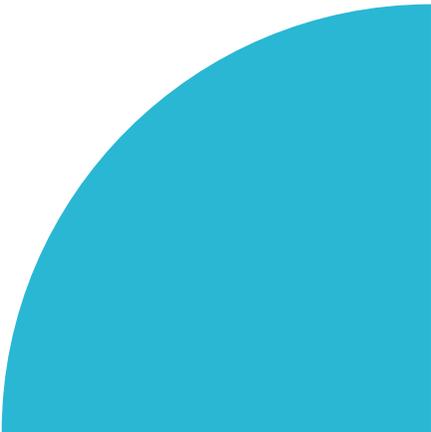
04

Тестирование алгоритма



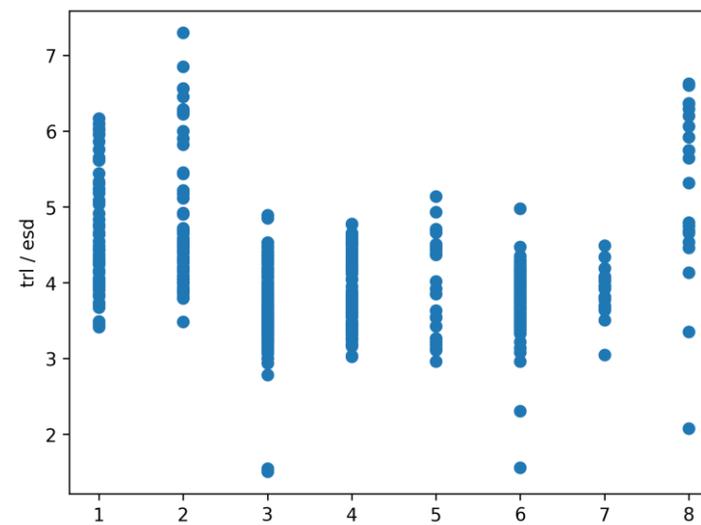
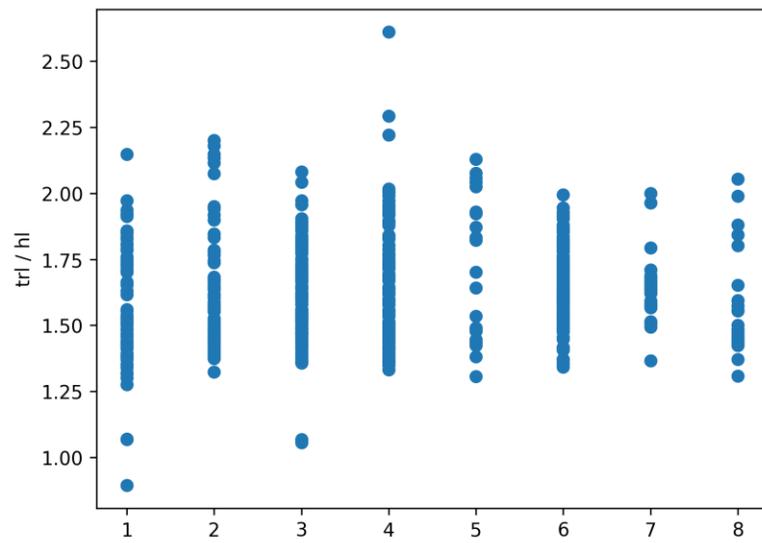
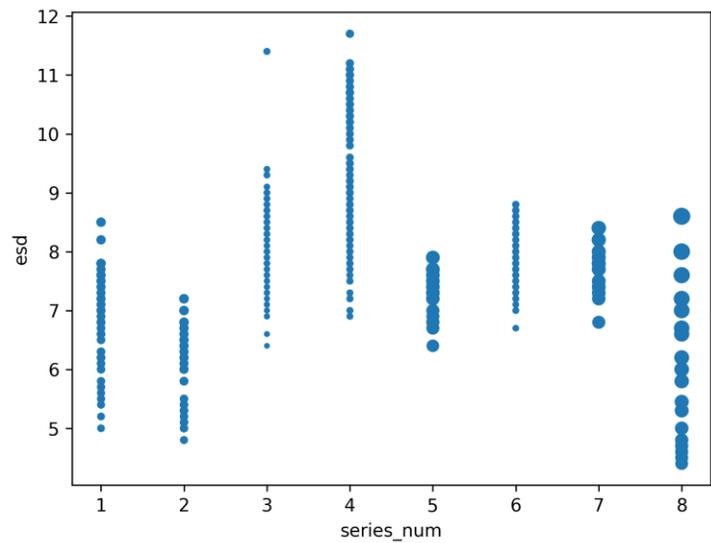
01

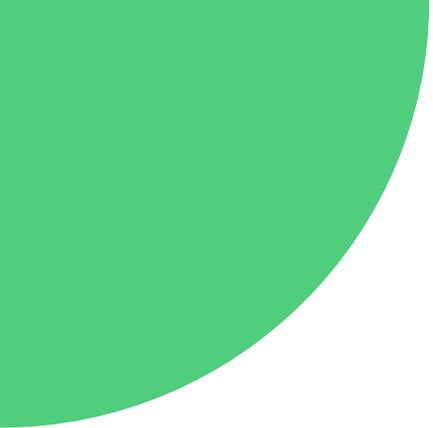
Анализ имеющихся данных



Анализ данных

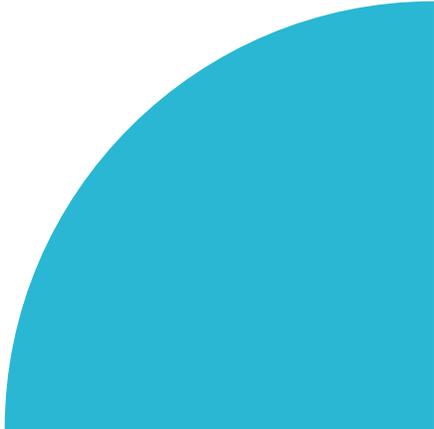
- Построение графиков зависимости особей и их характеристик в том числе их комбинаций
- 5 вид сильно отличается по параметру FPNr от всех остальных
- Применить методы кластерного анализа





02

Построение алгоритма



Идеи



Выявить критерии
для отличия каждого
с каждым



Исключать виды,
которые можем
однозначно отличить



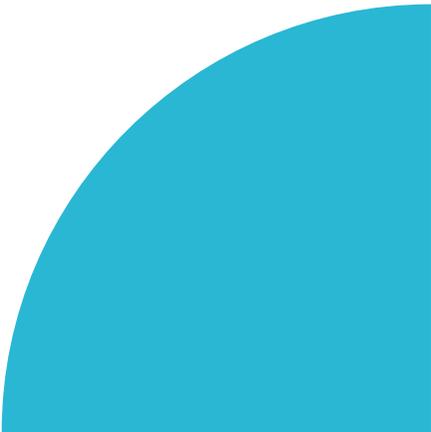
Отличать особей
попарно



Использование
методов
кластерного
анализа



Алгоритм



пространство разбиваем на сетку с шагом 0,1 по каждой оси

Случайным образом
выбираем 8 центроидов

Ищем Евклидово расстояние
для каждой точки до каждого из
центроидов

определяем в группу к тому
центроиду, расстояние до которого
наименьшее

Вычисляем новый центроид как
центр масс кластера

Вычисляем ошибку на основе
реальных данных

Повторяем пока
центроид на предыдущем
шаге не будет совпадать с
положением центроида
на следующем

Тестирование алгоритма

1. Обучаем алгоритм на 80% данных
2. Проверяем его на оставшихся 20%

3. **97%** результативность



Основная задача

- После применения нашего алгоритма для всех пар видов, мы получаем критерии на отличие.
- Далее определяем вид особи, используя критерии, полученные предыдущим алгоритмом.
- 547 особи из 564 определяет верно.
Достоверность нашего определения - 97%

Вывод

Мы разработали алгоритм, который вычисляет критерии для определения каждого вида каждой особи.



Спасибо за внимание!

Мы готовы ответить на ваши вопросы

