AI for Children

Artificial Intelligence Curriculum for Elementary and Secondary Schools

Card

# Dataset

01 Machine Learning Deck

https://kurikulum.aidetem.cz/cards

Created by: Radek Špáta & Eva Nečasová
Expert guarantors: Tomáš Mlynář, Pavel Kordík
Methodological consultants: Peťa Dovhunová, Ondra Brém
Language proofreading: not yet done
Last update: 01/2025
Version: 06

These teaching materials were translated using ChatGPT.
Please note possible imperfections in the expressions or wording.

Form for comments

Teaching Material for the AI Curriculum for Elementary and Secondary Schools
Computer Science at Secondary Schools - Cards, Machine Learning Deck

# Dataset

### Basic information about the dataset

Think of a dataset as an album filled with different pictures. Each picture comes with a caption that describes what's in it – like "cat in the sun," "dog in the snow," or "family on the beach." When we want to teach an AI system to recognize what's in an image, we let it "look through" this album – in the form of a dataset. The system "reads" our captions and searches the images for recurring visual patterns – for instance, snow tends to be white, and so on. The more images and captions the album includes, the better the system gets at identifying different things in pictures.

### Advanced definition

A dataset typically refers to a large collection of data used to train, test, and validate machine learning systems (also called machine learning models). These data can come in many forms – voice recordings, music, videos, images, numerical files, and more. The choice of dataset largely determines what the model will be capable of, how it will behave, and what kinds of tasks it can solve. Ideally, a dataset should contain high-quality and task-relevant data to ensure that the model performs reliably. In practice, however, developers sometimes prioritize quantity over quality.

### Machine Learning Deck

Dataset
Supervised Learning
Bias
Reinforcement Learning

Lesson presentation in PDF

Editable presentation in Canva

Note: For the sake of simplicity, this teaching material refers to datasets only in the context of images and supervised learning.
Note 2: Gender equality is a key value for AI for Children, but to keep our teaching materials concise, we use masculine grammatical forms.

Classroom Activity

# Alien Detective Agency

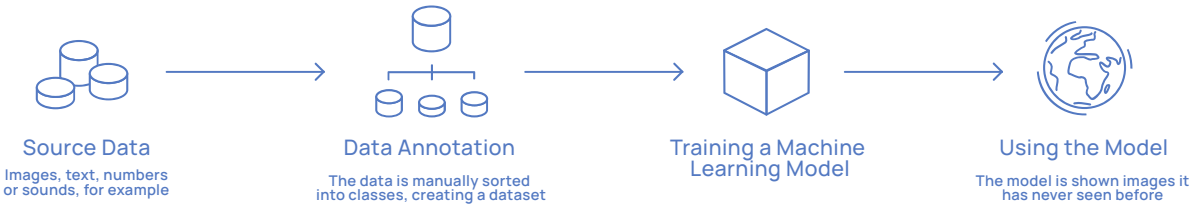
45 min

## Activity description

In this activity, students take on the role of an alien detective agency. Their mission is to figure out which alien family a mysterious child belongs to – a small foundling who has suddenly appeared in their village. You'll give students (or project on the screen) a set of portraits showing members of two alien families: the Fluffs and the Earls. The students' task is to correctly sort the family members and describe their visual features. Based on those observations, they'll try to determine which family the foundling comes from. There's no wrong answer – but the clues suggest that the child might, thanks to shared visual traits, be the offspring of members from both families.

## How the detective agency connects to the dataset

A dataset is a collection of data used to train, validate and test artificial intelligence systems. This data can take many forms such as images, audio, video or tables. In our case, the dataset is represented by a series of generated images of aliens. To train a system, also known as a machine learning model, the data must first be labeled. This process is called annotation.
Here, we annotate the data by sorting the aliens into groups called classes and giving these classes names – the Fluff family and the Earl family. As the model trains, it looks for similarities within the data. These are patterns that humans might describe as ears, whiskers, fur or colors. Based on these patterns, the model learns to classify new images it has not seen before, deciding which class – or alien family – they most likely belong to: the Fluffs or the Earls.




## Lesson Overview

**Recommended Age, Lesson Length**
Children aged 11–15, 45 minutes.

**Building Blocks**
Dataset.

**What Are the Students Learning?**
A dataset is a collection of data prepared for training artificial intelligence systems.

**Why Are They Learning This?**
Goal of the Machine Learning Card Deck:
To critically evaluate how AI systems make decisions.

**How Do We Know They Have Learned It?**
They explain the term dataset and describe the process of how it is created.

**Tools**
Teacher: Sets of alien family cards, worksheets, presentation equipment, presentation slides.

**Digital Competence**
Communication and Collaboration.

**Bloom's Taxonomy**
Remembering: Students become familiar with the term dataset and understand its meaning.
Analysing: They identify key attributes within different alien families and determine which features played a major role in classifying the foundling.

**Five Big Ideas**
1-B-II Processing (Feature Extraction).
2-A-II Representation (Symbolic representations).
3-A-II Nature of Learning (Finding patterns in data).

# Engage

10 min

Discuss

Presentation slide 03

**How do you find your way around a supermarket?**
Students will likely respond that it's easier to navigate because items are grouped into departments – like food, electronics, personal care products or clothing. The discussion should help students realize that people naturally sort things into categories based on different principles – such as shared features, how items are used (food, tools, clothing, toiletries, etc.), and many other possible criteria. Students can also come up with their own ideas for how things can be organized – not just in supermarkets, but in places like app stores or online shops as well.

A key insight should be that we describe things, and this allows us to categorize, filter, search for similarities and discover connections.

# Understand

25 min

Activity 1 (25 min)

In this activity, students will work with worksheets. We recommend cutting them up before the lesson and handing them out step by step. If you prefer not to print the worksheets, you can rely on the presentation alone, along with printed tables from page 08 of this teaching material.

Worksheet 1 upper part (5 min)

Presentation slide 05 and 06

**Give each group the top part of worksheet 1 with members of the Fluff and Earl families.**
If you don't want to print the cards, you can project them from the presentation (slide 05). Students discuss how to sort the characters, looking for differences and shared visual traits. They decide on their own how to group them. Note: Students often start by arranging the characters by age, which is fine. Later, they should ideally sort them into the two families (see slide 06).

The Fluff Family

Earl family

7 min

Middle part Worksheet 1

In the middle section of worksheet 1, students will find a **table**. After sorting the characters into two families, their task is to write down at least three shared traits for each family of aliens. These might include ears, fur, color or other visual features.

| The Fluff Family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____

| The Earl family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____

3 min

Worksheet 1 bottom part

Presentation slide 07

**Next, show students the portrait of the small foundling.** If they're not working with printed worksheets, project the portrait from the presentation (slide 07). Based on the attributes they selected earlier, students now score the foundling in the table. For each attribute that matches between the foundling and the family, they give one point in the "Foundling" column. If there's no match, they write zero.

Then they add up all the points, and based on the higher score, they decide which family the foundling belongs to.

Present & discuss

**Students present their results to the rest of the class.** Which attributes did they select for each family? Did they see these traits as clearly shared across the family members? Would they choose different attributes if they could do it again? Was the foundling assigned to the right family?

10 min

Analyze & explain

Presentation slide 08 and 09

Have students read slides 08 and 09 of the presentation and then discuss together. They can summarize what they did during the lesson and how it connects to the idea of a dataset. There's no need for perfectly correct terminology – what matters is understanding this simple principle: When training an AI program, we need to show it source data and describe that data by sorting it into categories. That's how a dataset is created. A dataset is data prepared specifically for training an AI system.

Tesla autopilot

**Play the short video (0:31) for the students.**
Address: www.youtube.com/watch?v=fKXztwtXaGo In the video, students can see how a self-driving car's program perceives its surroundings. It highlights recognized objects by drawing bounding boxes around them. To make this possible, the program had to be trained on many examples of different objects – such as pedestrians, cyclists, traffic signs and more. Humans created a dataset by sorting these objects into categories. The AI system then used that dataset during training to look for patterns and similarities among the examples.

# Reflect

10 min

Describe & evaluate

**Let's go back to the topic of the lesson and share:**
**What did you learn about datasets and their classification?**
**Why is it important to define categories and attributes correctly?**

**AI programs can automatically decide what's in an image. But they need people to provide the right data. Describe how and what people need to prepare for an AI system.**
For AI systems to be able to decide what something is, they need data prepared by people—images, sounds, etc.—that are accurately labeled and sorted into categories. These organized sets of data are called datasets.

**Can you think of some examples where AI could be used similarly to self-driving cars? How hard would it be to create a dataset for that?**
Possible answers: Face recognition at airports for security reasons or unlocking phones, automated quality control on factory lines (when the system is trained to spot defective products)...

**Sometimes the world is very complex. For AI programs to recognize certain objects, they have to be trained on a truly massive amount of data. How many images do you think were needed to teach a self-driving car to correctly recognize a pedestrian or a crosswalk?**
It took millions of images.

# Follow-up lesson

**In this lesson, students worked with worksheets to understand the theoretical concept of a dataset.**
In the follow-up lesson, they will continue with the theme of the alien detective agency, but this time, they'll let an AI program decide where the little found creature belongs.

Discuss in groups what makes the individual aliens different and what they have in common.
Then, based on your judgment, sort them into two families – the Fluffs and the Earls.


| The Fluff Family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

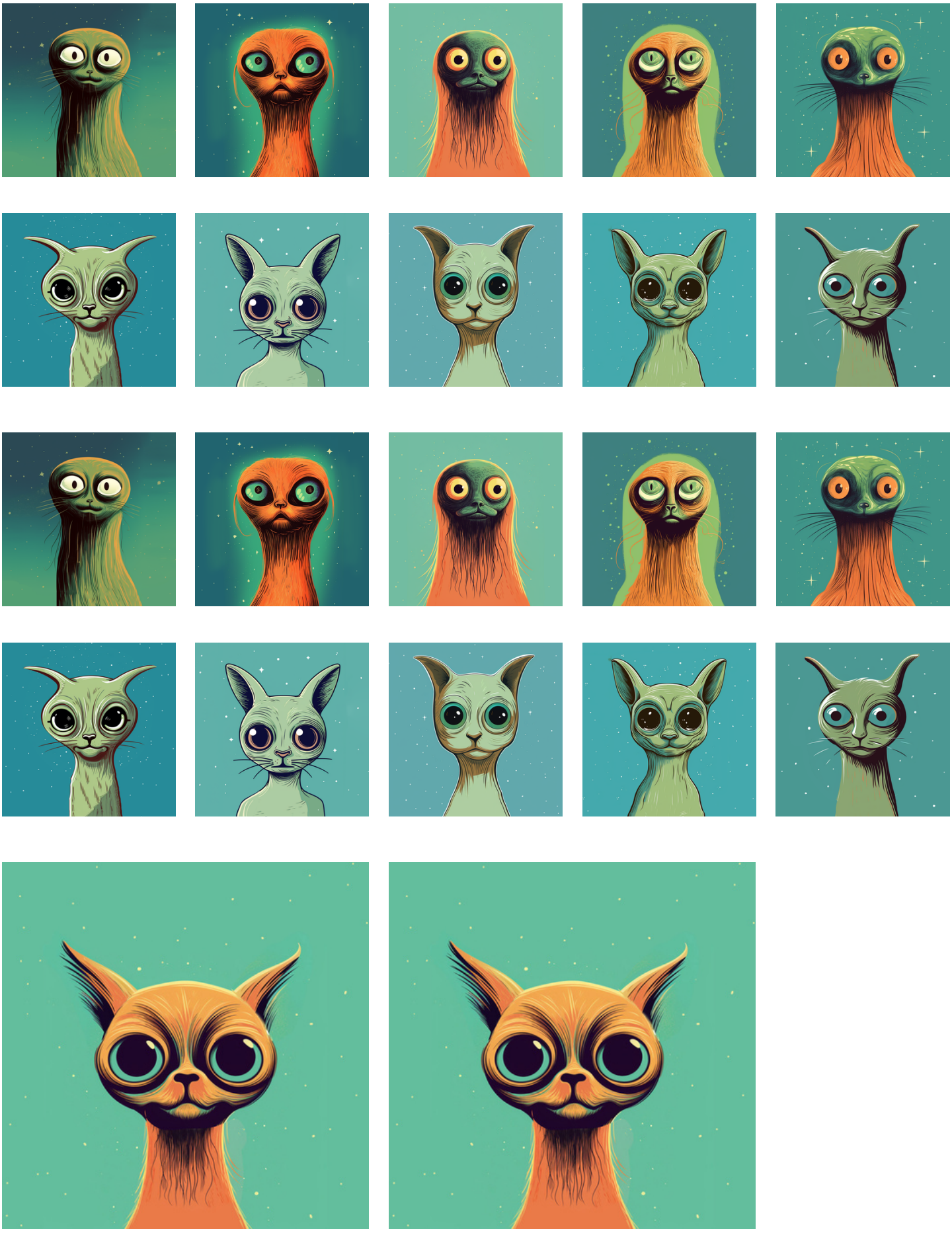
Total score _____

| The Earl family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____

Take a look at the portrait of the little foundling.

Based on the traits you chose in the previous part of the activity, score the foundling's image in the table. If the foundling shares a trait with a family, write one point in the "Foundling" column. If the family and the foundling have nothing in common, write zero. Then, add up all the points. The family with the highest total will be the one the foundling belongs to. And if it ends in a tie...? 😜



# Dataset



| The Fluff Family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____

| The Earl family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____


| The Fluff Family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____

| The Earl family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____



| The Fluff Family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____

| The Earl family – shared attributes | Foundling |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |

Total score _____