

STUDIO GHIBLI

**Neural Network Machine
Learning Model**

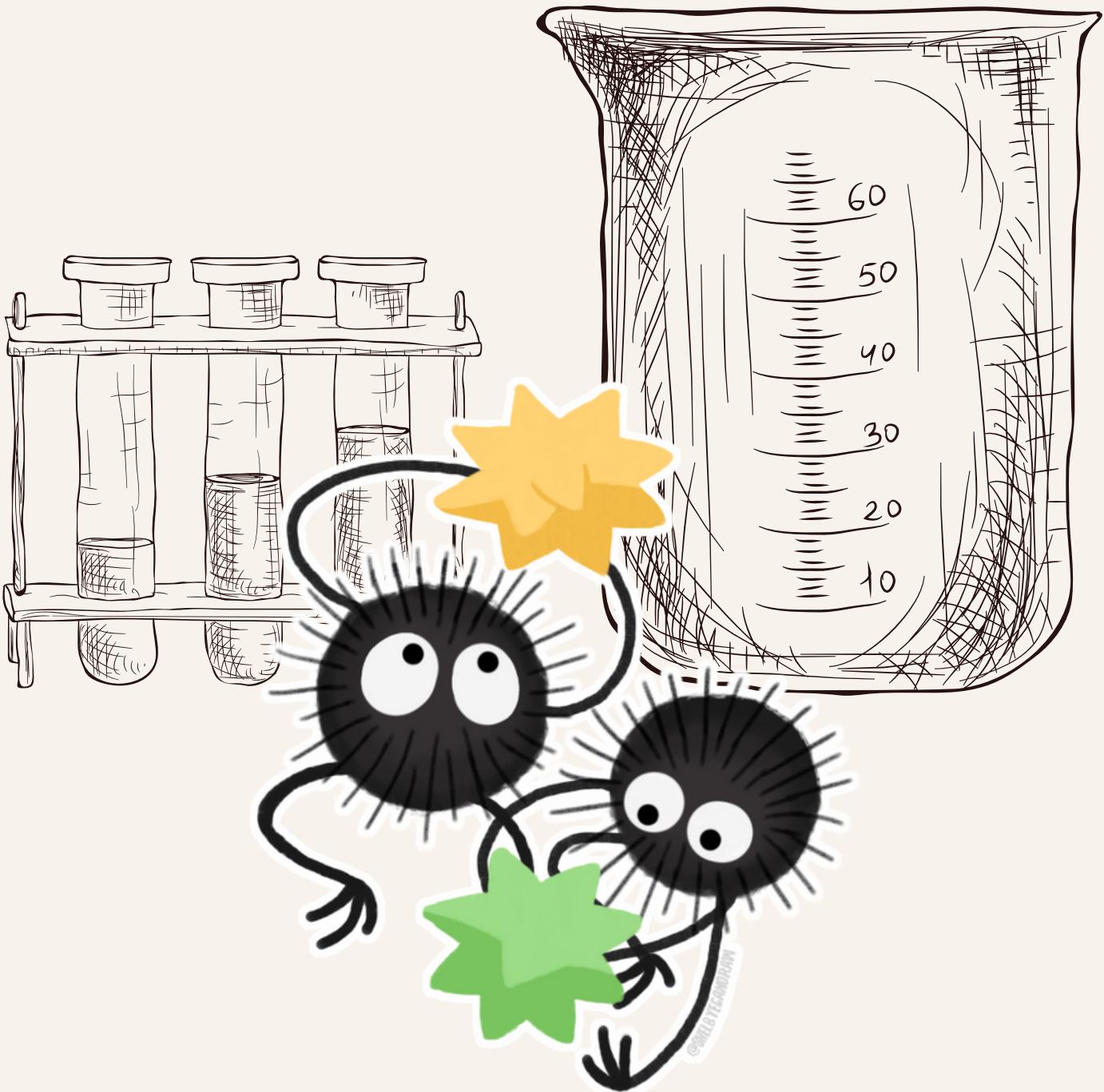
by Thee Xiong



Can a Computer ...

be trained to predict a Studio Ghibli character's species by analyzing their age, height, eye and hair colors and special powers?

This model is based on fictional characters but could easily be modified and applied to predict any other species.

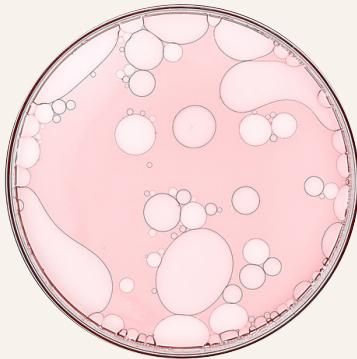


Process



- **Research & Collect Data**

Compiled a csv with character statistics and traits.



- **Preprocess Dataset**

Created DataFrame from csv file, dropped non-beneficial columns , and queried dataset.



- **Build & Test Model**

Neural Network Model was created with 3 hidden layers.



- **Model 2**

Model was tested on dataset that dropped all characters with any NaN values.

Sources for Data Collection

<https://github.com/Laboratoria/CDMX011-data-lovers/blob/master/src/data/ghibli/ghibli.json>

https://disney.fandom.com/wiki/Category:Studio_Ghibli_films

https://ghibli.fandom.com/wiki/Studio_Ghibli

<https://www.kaggle.com/datasets/marshuu/ghibli-characters>





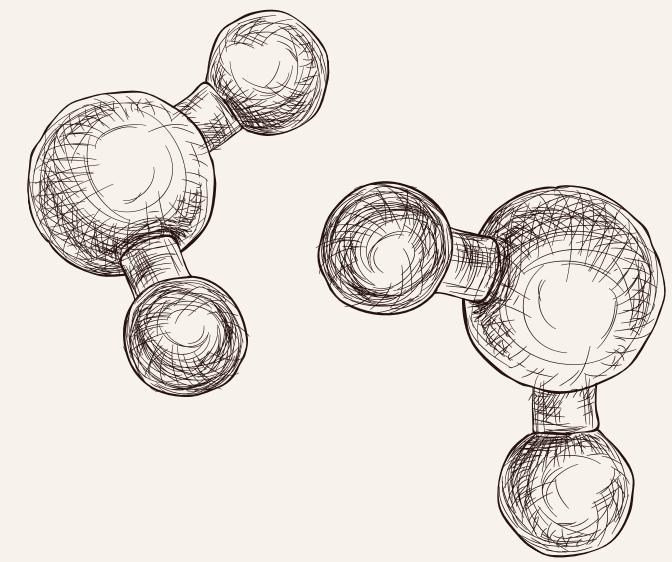
Species Categorizing

01

Human, Witch, Wizard, Spirit, Giant Baby, Cat, Fire Demon, Borrower etc.

02

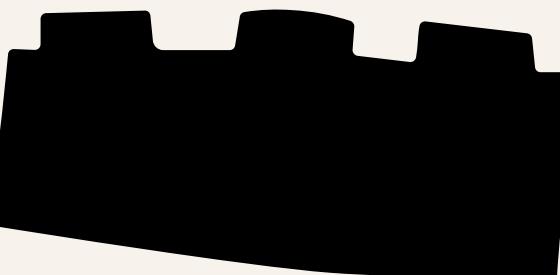
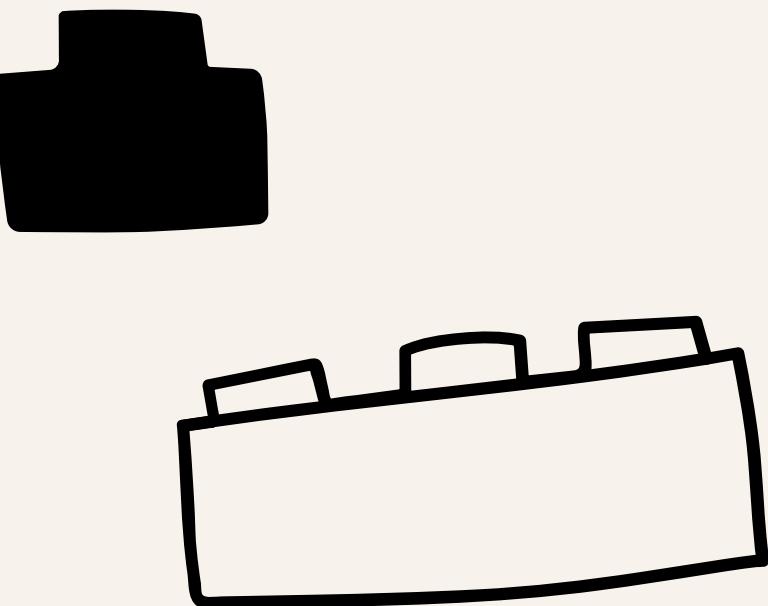
Human/Non-Human



Building the Model

Model 1

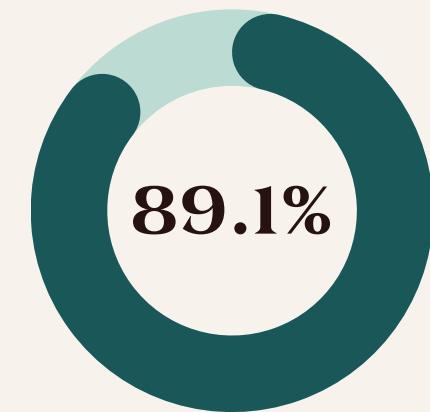
Kept all characters and replaced NaN values with median age and height values.



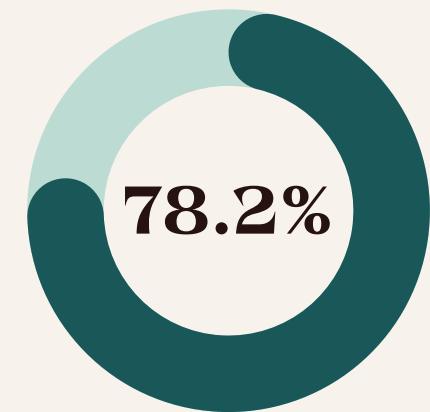
Model 2

Dropped all characters with any NaN values.

Model Results



Model with NaN values
changed to median age and
height values.



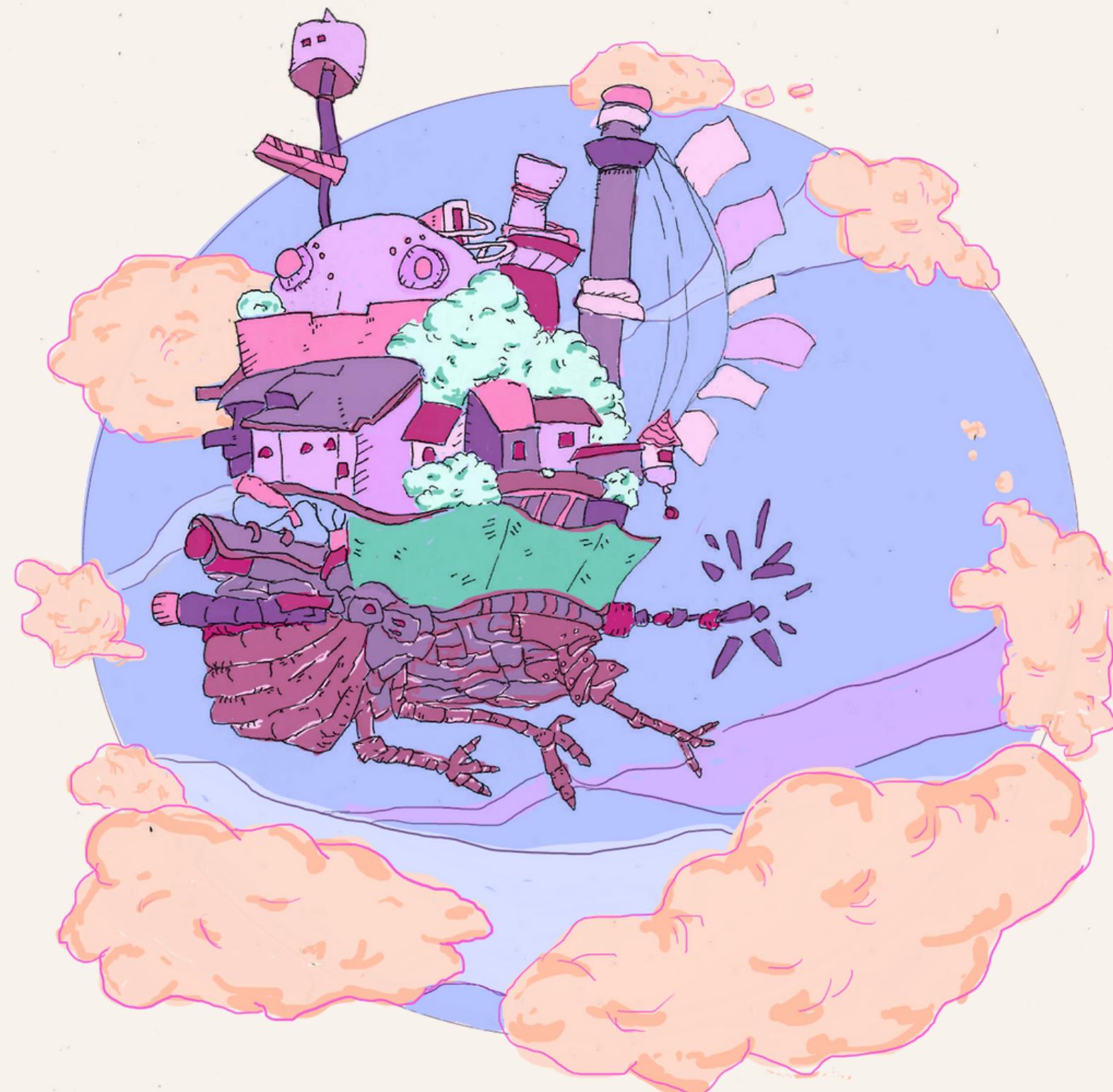
Model with NaN
characters dropped.

Project Links

[Jupyter notebook](#)

[Jupyter notebook 2](#)

[Tableau Dashboard](#)



Conclusions



01

Yes, you can definitely train a computer to categorize species by traits!

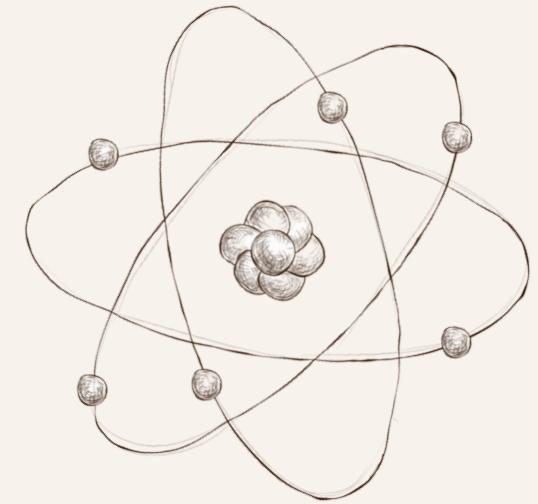
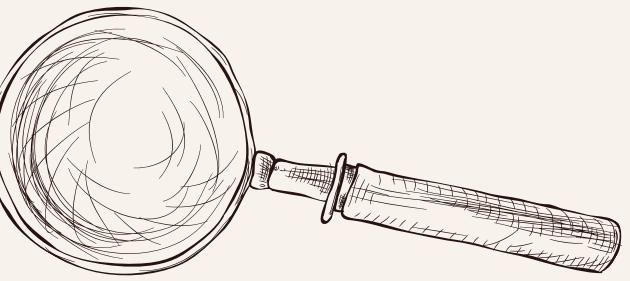
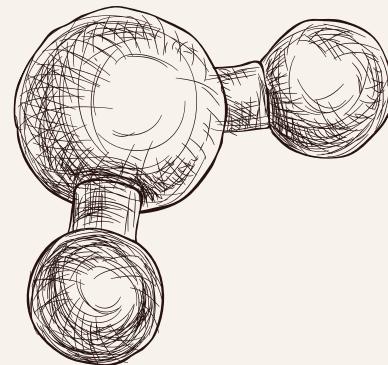
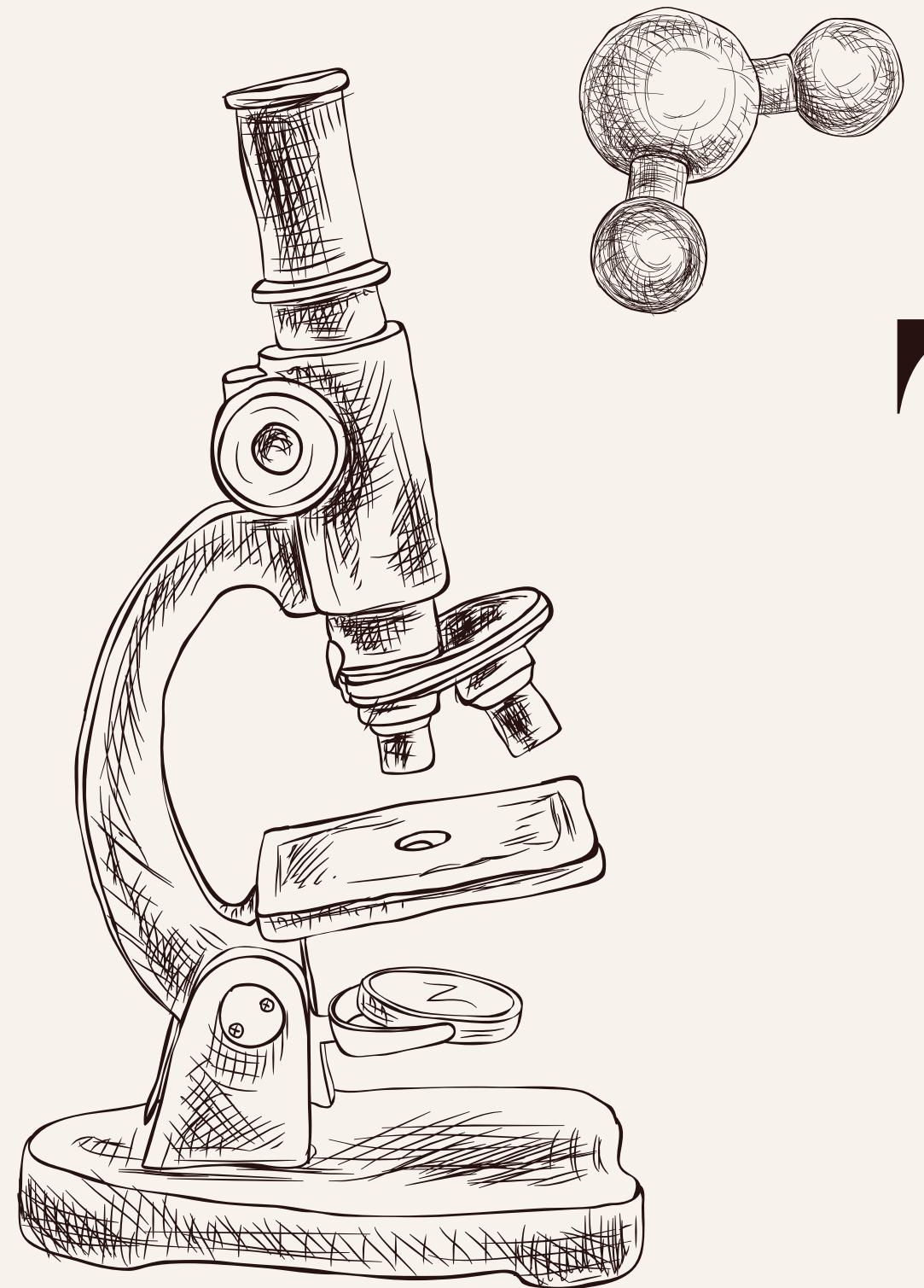
Predicting species can also be used with different types of data.

02

The more diverse your dataset is, the better you can train your model.
You want to avoid overfitting your model.

03

Missing NaN values can decrease your model's success.
Using the median value is one way of keeping data with NaN values.



Thank you!

Do you have any questions?

