# Adaptive Branching: A Constraint Satisfaction Approach

By Victoria and Leah Zhang

# Overview

Introduction

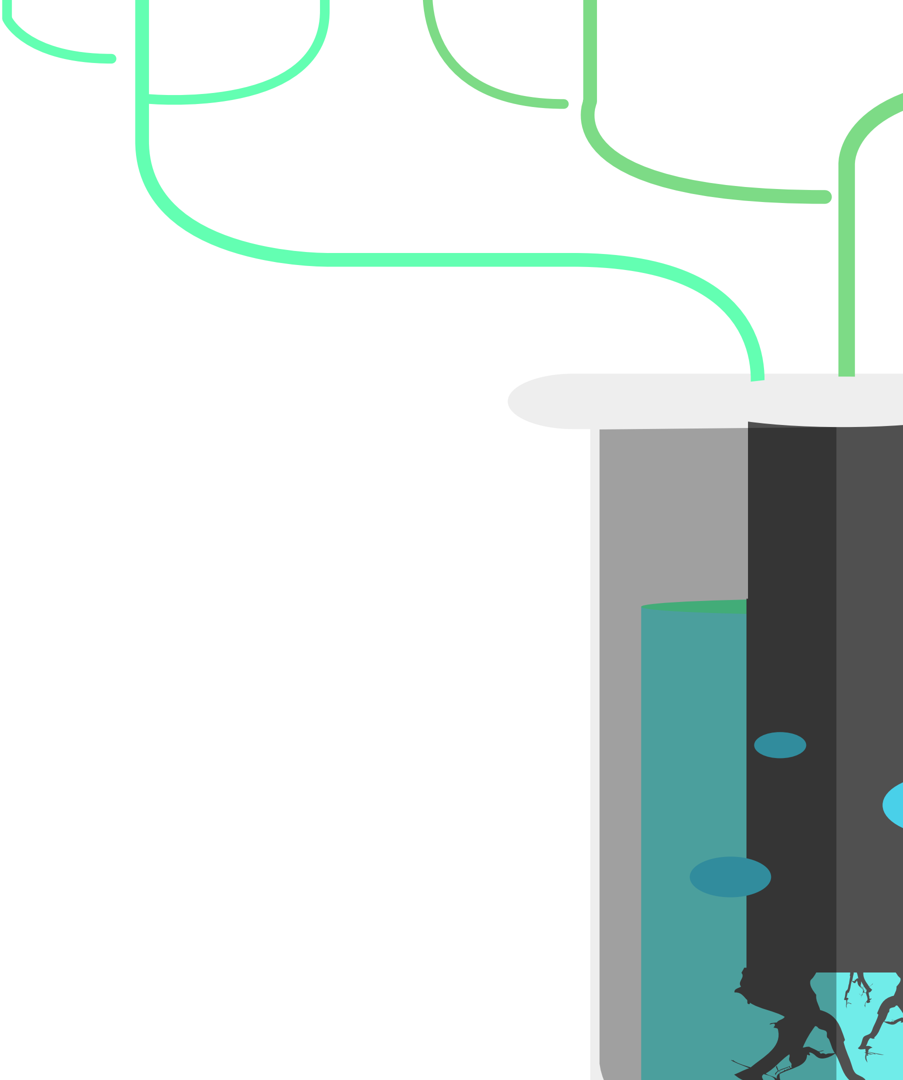
Preprocessing

Methods

Results

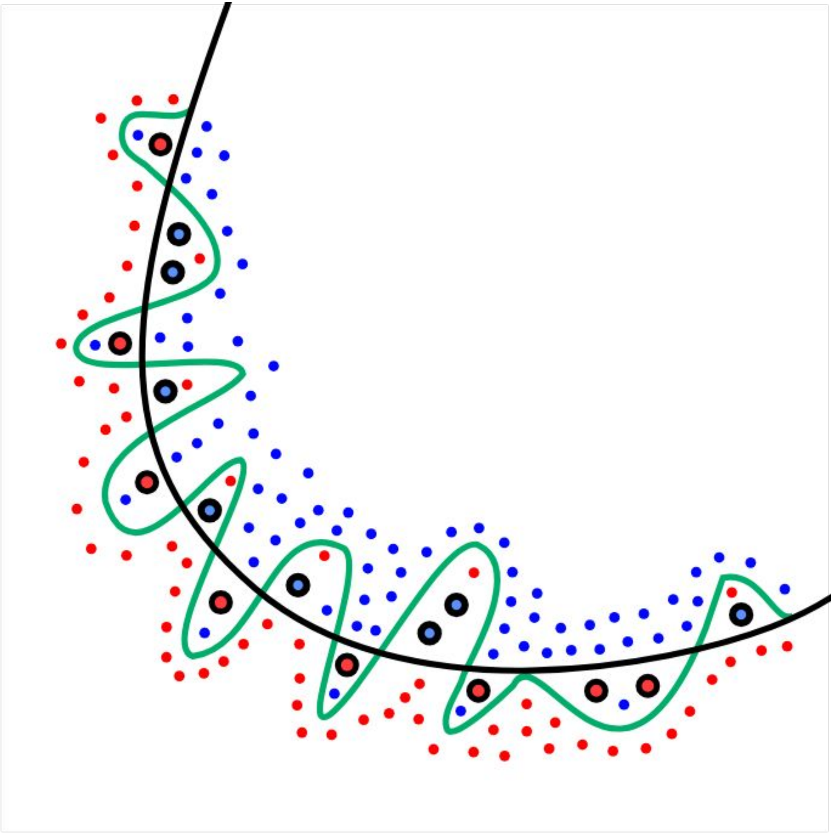
Conclusion

# Introduction



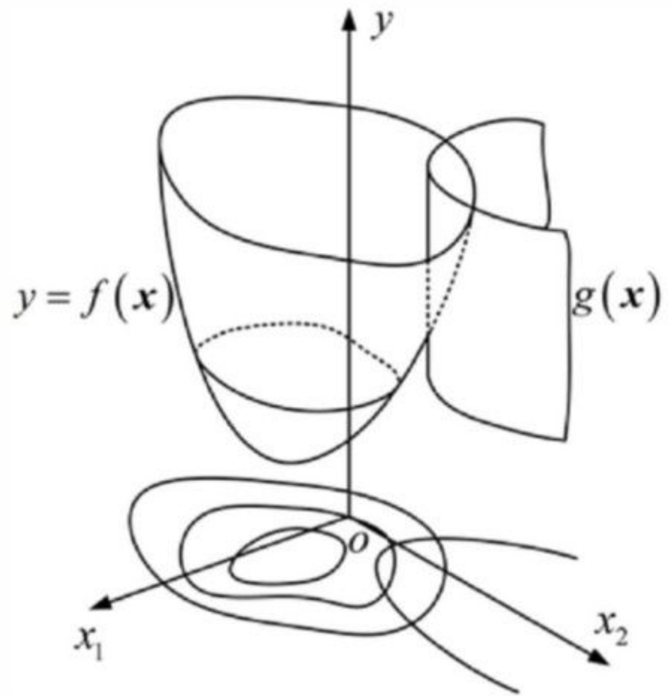
# Our Problem

- Decision Trees struggle with *Overfitting*
- Pre Existing Solutions such as Ensemble Learning and Gradient Boosting lack *Efficiency* and *Interpretability*



# Our Solution

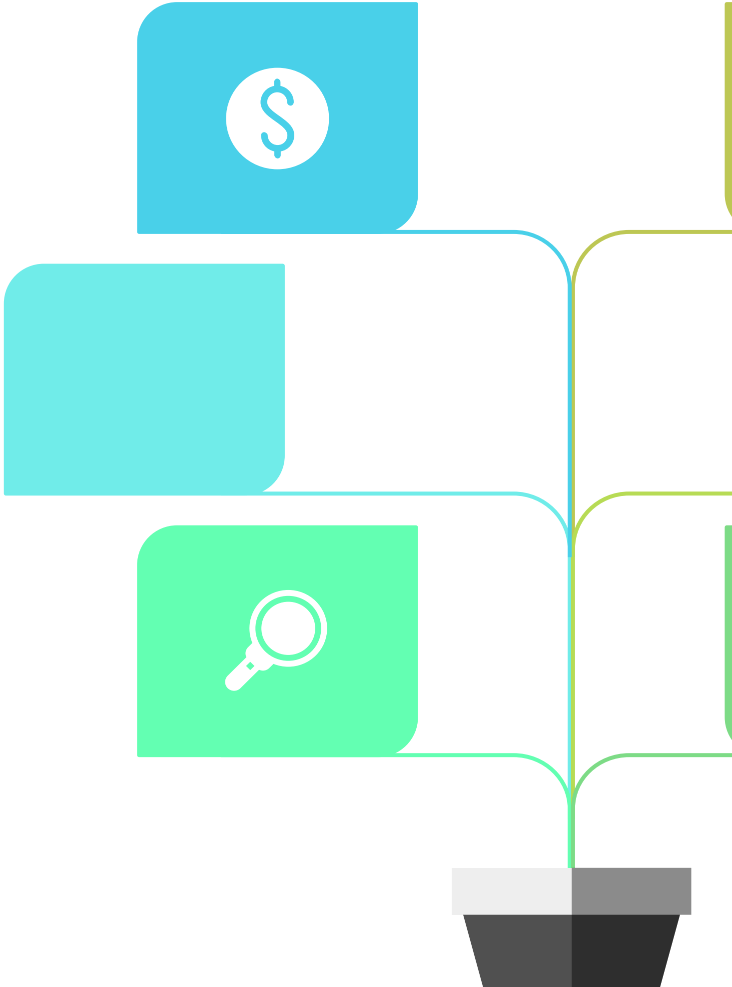
- Traditional Decision Trees algorithms utilize static thresholds for making splits and pruning
- To combat this we use *Dynamic Pruning* and *Constraint Optimization*



## Our Goal

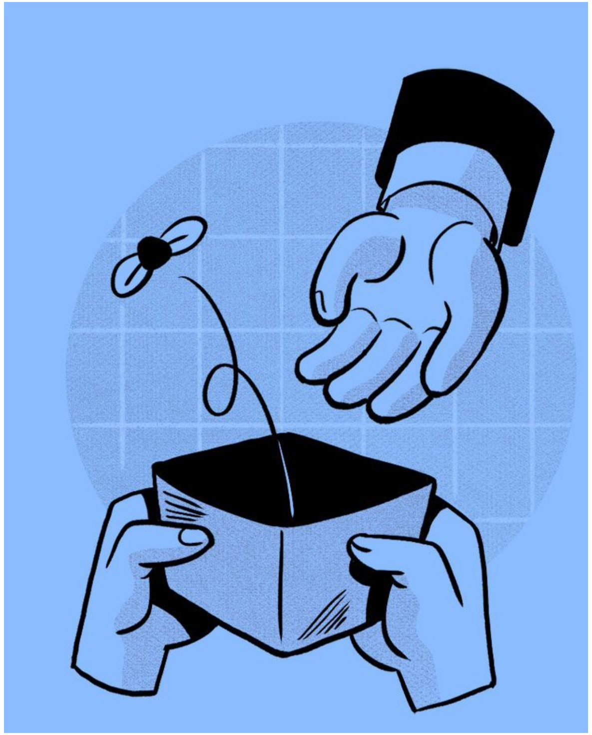
By building upon existing dynamic programming strategies while also emphasizing real-time adaptability, this project aims to construct decision trees without sacrificing either accuracy or efficiency, while also avoiding the lack of interpretability found in state-of-the-art ensemble methods.

# Preprocessing



# Dataset

- Our dataset is on credit risk analysis.
- Class: likelihood of the loan defaulting
- ~32,800 instances
- 11 attributes



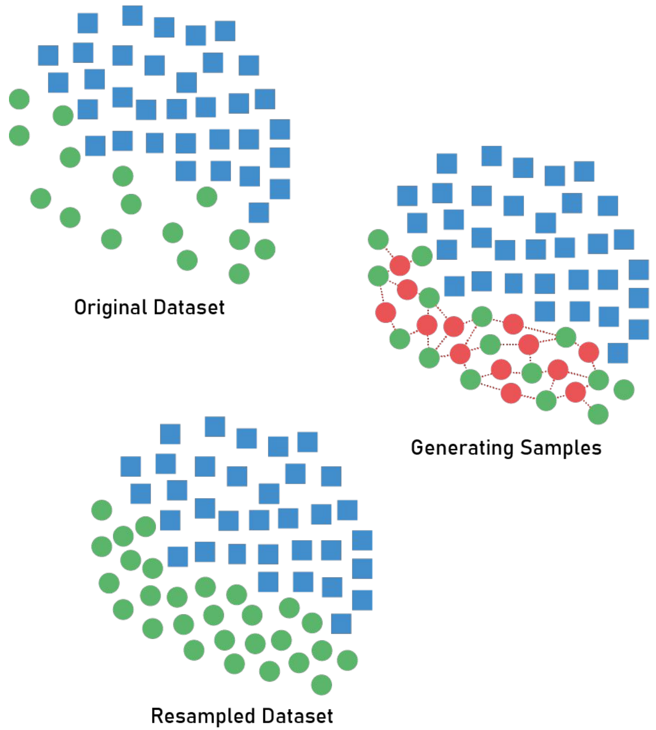
| Attribute | Description |
|---|---|
| ID | Unique identifier for each loan applicant. |
| Age | Age of the loan applicant. |
| Income | Income of the loan applicant. |
| Home | Home ownership status (Own, Mortgage, Rent). |
| Emp_Length | Employment length in years. |
| Intent | Purpose of the loan (e.g., education, home improvement). |
| Amount | Loan amount applied for. |
| Rate | Interest rate on the loan. |
| Status | Loan approval status (Fully Paid, Charged Off, Current). |
| Percent_Income | Loan amount as a percentage of income. |
| Cred_Length | Length of the applicant's credit history. |

# Preprocessing

- irrelevant variables, ‘id’, were removed.
- Missing values found in ‘Emp_length’ and ‘Rate’ were replaced
- Categorical variables, such as ‘Home’ and ‘Intent’ and including the class variable of ‘Default’ were numericized
- Numerical variables of ‘Income,’ ‘Amount,’ and ‘Percent_income’ were normalized through MinMaxScaler().

# Synthetic Minority Oversampling Technique (SMOTE)

- 26.8k values for false, 82% of the total,
- 5745 values for true, 18% of the total.
- SMOTE creates synthetic data.
- Training set initially had 26063 data points with 21468 false and 4596 true
- Now 21468 for both false and true.



# Methods

# Overview

**Baseline**

Using a basic decision trees model

**Random Forest**

A pre existing solution to overfitting for comparison

**Modified**

introduces constraint opt + dynamic pruning

**simulated annealing**

SA to optimize the hyperparamet ers

**Differential equation**

diffeq to optimize hyperparamet ers

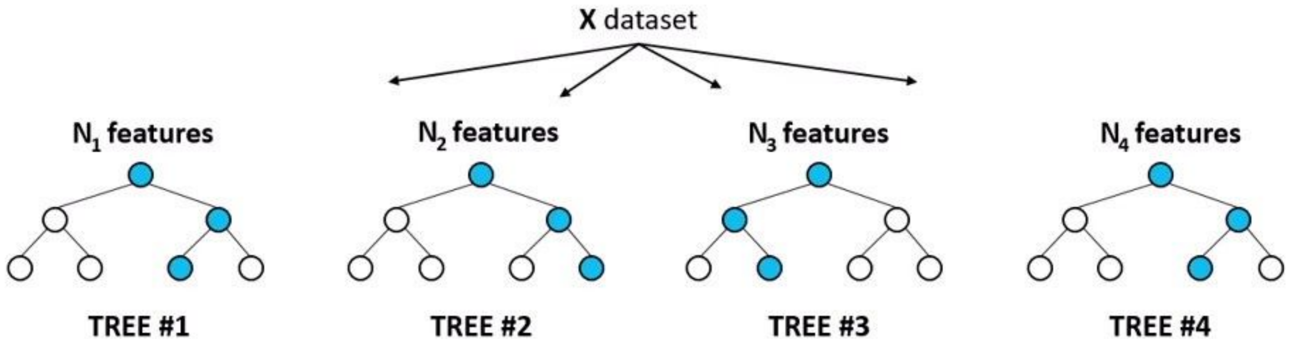
# The Baseline Model

- Standard implementation of a decision tree using the scikit-learn library
- Gini impurity over information gain to reduce computation

$$Gini((D) = 1 - \sum_{i=1}^{n} (p_i)^2$$

# Random Forest

- Ensemble learning is a pre existing solution that addresses overfitting by using multiple decision trees
- Using the 'RandomForestClassifier' model from Scikit-learn. We decided to use 1000 trees and random state 42.



# Modified Constraint Optimization Approach

| Step | Classical Approach | Constraint-Based Approach |
|---|---|---|
| Feature Selection | Gini or entropy based splitting | Penalized gain with dynamic thresholds |
| Pruning | Cost-complexity pruning | Post-pruning using gain penalties |
| Computational Complexity | Greedy exploration of all branches | Prioritized search with reduced search space |

# Feature Selection with Dynamic Information Gain

- Dynamic feature selection process →custom calculation that includes a penalty for small gains.
- G is the standard gain or impurity calculation, P is a penalty factor, and T is a minimum gain threshold.
- Computation to choose features before splitting each node.

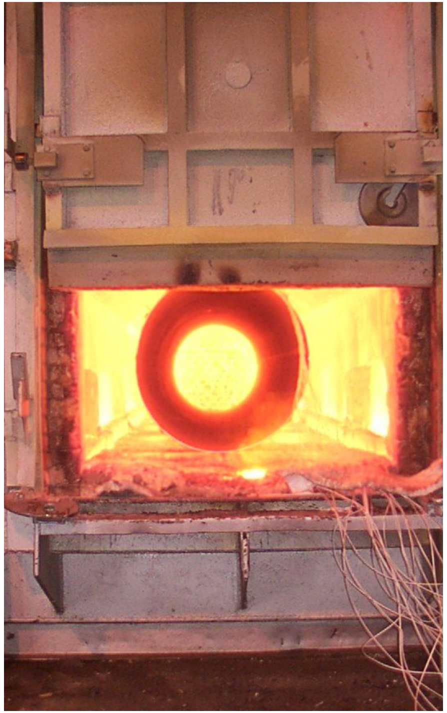
$$G_{adjusted} = G - max(0, P * (T - G))$$

# But how do we decide the best hyperparameters?

spoiler: we do some crazy shenanigans

# Enhancing the Modified Tree with Simulated Annealing

- Inspired by metallurgical annealing ⇒ material is heated and slowly cooled to remove defects.
- “heat” = randomness in the search during the optimization process, which decreases over time to refine the solution.
- SA is a probabilistic algorithm whose strength lies in its ability to escape local minima, allowing it to find near-global optima



# Enhancing the Modified Tree with Simulated Annealing

1. Initialize an initial solution
2. Calculate fitness
3. Generate new solutions by perturbing the current solution slightly
4. Evaluate new fitnesses
5. If the new solution is worse, accept it with probability $P(accept) = exp(\frac{\Delta fitness}{temp.})$
6. Update temperature using a set cooling rate: $T_{new} = T_{old} * cooling\ rate.$
7. Repeat steps 2-5 until a stopping criterion is met (max. number of iterations or a sufficiently low temperature is reached)
8. Return the best solution found after the final iteration.

# Enhancing the Modified Tree with Simulated Annealing

- SA used to optimize the hyperparameters P (penalty factor) and T (minimum gain threshold)
- Fitness for this solution and subsequently-generated solutions was calculated with a function that evaluated the trade-off between accuracy and complexity via a Lagrangian objective
- It fulfills our initial objective

$$Fitness = Accuracy - \lambda * Tree\ Complexity$$

# Alternative Using Differential Equation Optimization

What is a Differential Equation?

- A mathematical equation that relates a function to its derivatives.
- DiffEq a strong tool to model dynamic systems that change continuously.
- In optimization problems, DiffEq can be used to model the evolution of parameters

# Alternative Using Differential Equation Optimization

Our Approach

- Uses similar function to fitness in simulated annealing
- Parameter updates based on gradient
- Stops when either the system reaches the final time point or when the gradients approach zero

$$J(P, T) = Accuracy - \lambda * Number\ of\ Nodes.$$

# The actual equations

$$J(P, T) = Accuracy - \lambda * Number\ of\ Nodes.$$

$$\frac{dP}{dt} = -\frac{\partial J}{\partial P}, \quad \frac{dT}{dt} = -\frac{\partial J}{\partial T}$$

$$\frac{\partial J}{\partial P} \approx \frac{J(P+\epsilon, T) - J(P,T)}{\epsilon}, \quad \frac{\partial J}{\partial T} \approx \frac{J(P,T+\epsilon) - J(P,T)}{\epsilon}$$

# Results

| | Accuracy | F1 Score | Precision | Recall | ROC AUC |
|---|---|---|---|---|---|
| Baseline Tree | 0.80 | 0.49 | 0.45 | 0.54 | 0.70 |
| Random Forest | 0.82 | 0.59 | 0.50 | 0.72 | 0.88 |
| Constraint Optimization | 0.80 | 0.50 | 0.45 | 0.55 | 0.68 |
| Simulated Annealing | 0.81 | 0.52 | 0.48 | 0.58 | 0.78 |
| Differential Equation Optimization | 0.8237 | 0 | 0 | 0 | 0 |

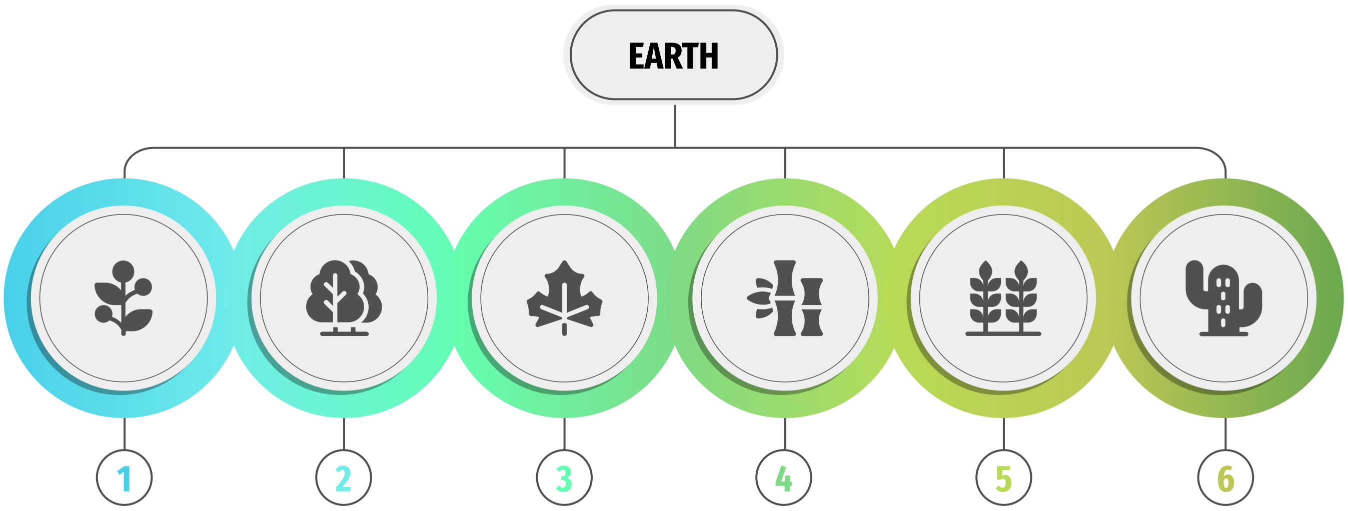
Co

# Conclusion

- Dynamic feature selection is possible
- Future Steps Include
  - Trying a smaller dataset
  - Using real world data prone to noise

# THANK YOU

# Decision Tree Diagrams




**MERCURY**

Mercury is the closest planet

**MARS**

Mars is a cold place, not hot

**JUPITER**

It's a gas giant and the biggest

**VENUS**

Venus is the second planet

**SATURN**

Yes, Saturn is the ringed one

**NEPTUNE**

Neptune is the farthest planet

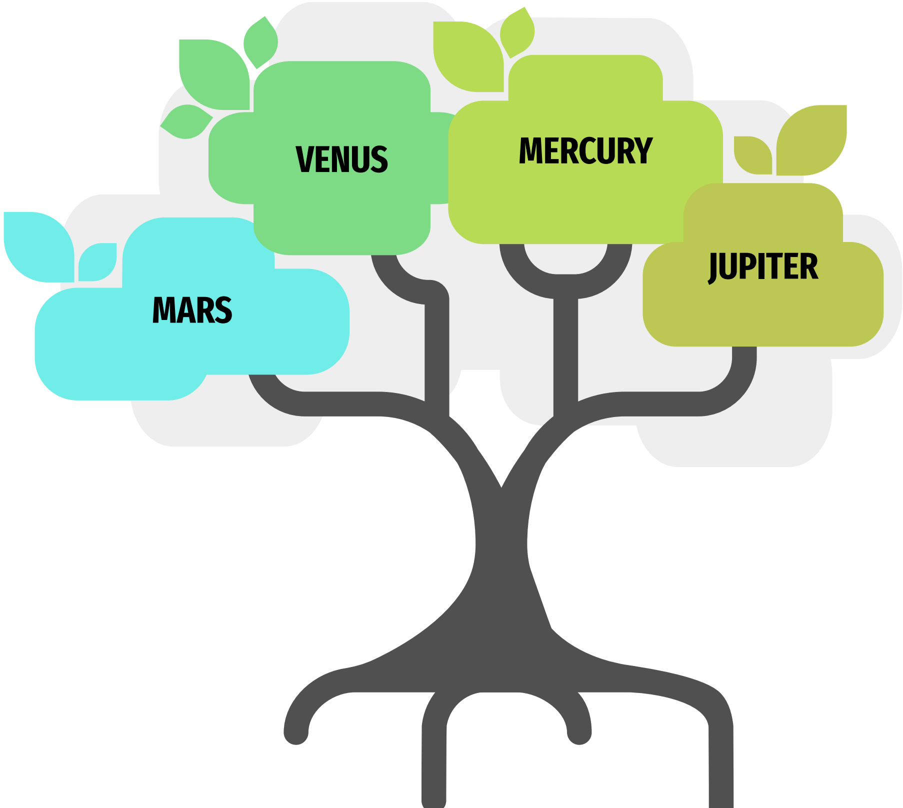
# Decision Tree Diagrams

### MARS

Despite being red, Mars is a cold place

### VENUS

Venus has a very poisonous atmosphere



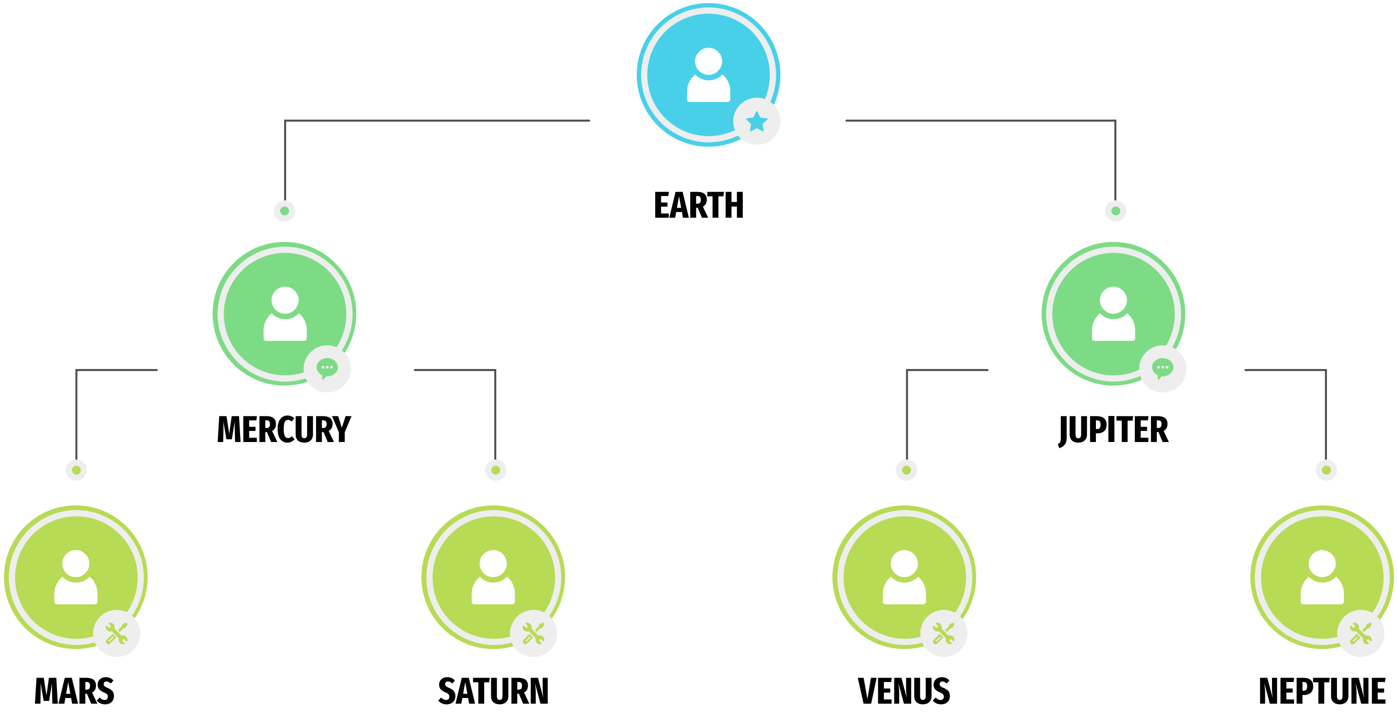

### MERCURY

Mercury is the closest object to the Sun

### JUPITER

Jupiter is a gas giant and the biggest planet

# Decision Tree Diagrams



# Decision Tree Diagrams

### MERCURY

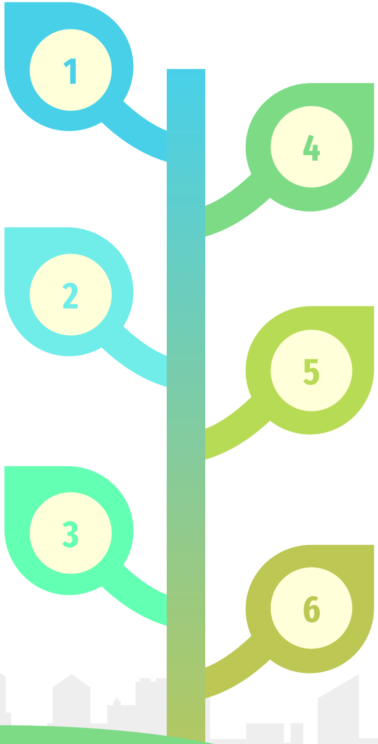
It’s the smallest planet in the Solar System

### VENUS

Venus has a very poisonous atmosphere

### JUPITER

Jupiter is the biggest planet of them all




### MARS

Despite being red, Mars is a cold place

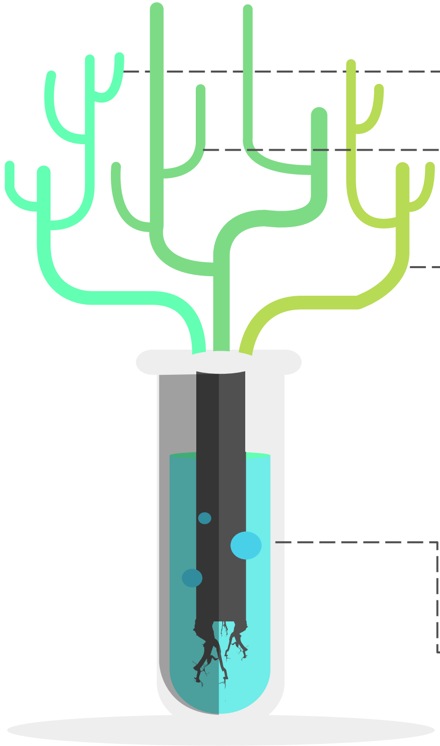
### NEPTUNE

Neptune is the farthest planet from the Sun

### SATURN

Saturn is composed of hydrogen and helium

# Decision Tree Diagrams



**MERCURY** — Mercury is the closest planet to the Sun and the smallest one

**VENUS** — Venus is the second planet from the Sun. It has a beautiful name

**SATURN** — Saturn, the ringed planet in the Solar System

**JUPITER** — Jupiter is the biggest planet in the Solar System

# Decision Tree Diagrams

### MERCURY

Mercury is the closest planet to the Sun and the smallest one

### VENUS

Venus has a beautiful name and is the second planet from the Sun

### JUPITER

Jupiter is a gas giant and the biggest planet in the Solar System

### MARS

Despite being red, Mars is a cold place full of iron oxide dust

### NEPTUNE

Neptune is the farthest planet from the Sun and the fourth-largest one

# Decision Tree Diagrams

### MARS

Despite being red, Mars is cold

### VENUS

Venus has a beautiful name

### NEPTUNE

It's the farthest planet

### SATURN

Yes, Saturn is the ringed one

### JUPITER

Jupiter the biggest planet

### PLUTO

It is now considered a dwarf planet

### MERCURY

Mercury closest planet to the Sun

### EARTH

It's the third planet from the Sun

# Decision Tree Diagrams

**MERCURY**

Mercury is the smallest planet

**VENUS**

Venus has a beautiful name

**JUPITER**

Jupiter is the biggest planet

**MARS**

Mars is a cold place

**NEPTUNE**

It’s the farthest planet

**SATURN**

Yes, Saturn is the ringed one

# Decision Tree Diagrams

**MARS**

Despite being red, Mars is a cold place

**VENUS**

Venus has a very poisonous atmosphere

**MERCURY**

Mercury is the closest object to the Sun

**JUPITER**

Jupiter is a gas giant and the biggest planet

# Decision Tree Diagrams

**MERCURY**

Mercury is the closest planet to the Sun and the smallest one

**VENUS**

Venus has a beautiful name and is the second planet from the Sun

**JUPITER**

Jupiter is a gas giant and the biggest planet in the Solar System

**MARS**

Despite being red, Mars is a cold place full of iron oxide dust

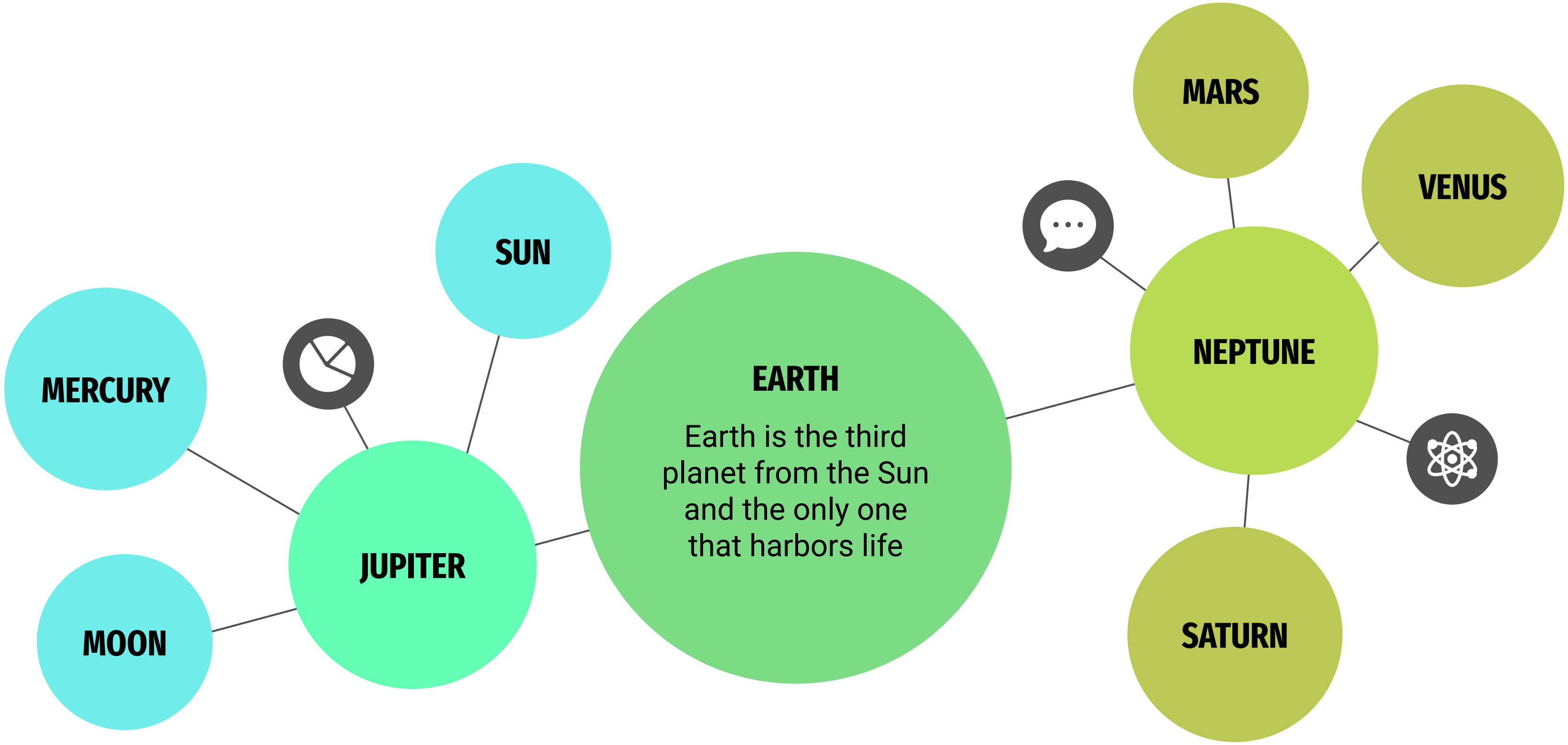
**NEPTUNE**

Neptune is the farthest planet from the Sun and the fourth-largest one

**SATURN**

Planet Saturn is a gas giant composed mostly of hydrogen and helium

# Decision Tree Diagrams



# Decision Tree Diagrams

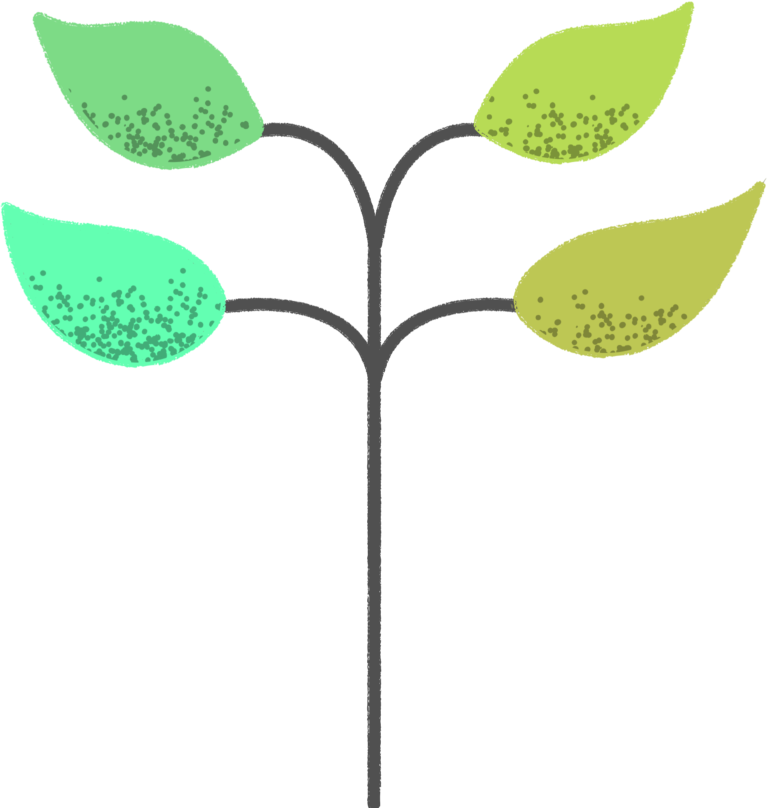
Mercury is the closest planet to the Sun

Venus has a very poisonous atmosphere

Despite being red, Mars is a cold place

Jupiter is a gas giant and the biggest planet

# Decision Tree Diagrams



**MERCURY**

Mercury is the closest planet to the Sun and the smallest one

**VENUS**

Venus has a beautiful name and is the second planet from the Sun

**MARS**

Despite being red, Mars is a cold place full of iron oxide dust

**NEPTUNE**

Neptune is the farthest planet from the Sun and the fourth-largest one

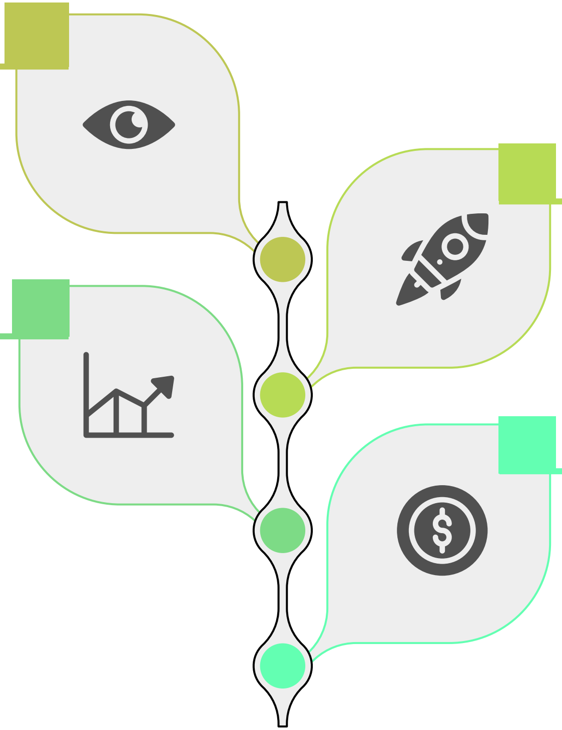
# Decision Tree Diagrams

## MARS

Despite being red, Mars is a cold place

## VENUS

Venus has a very poisonous atmosphere



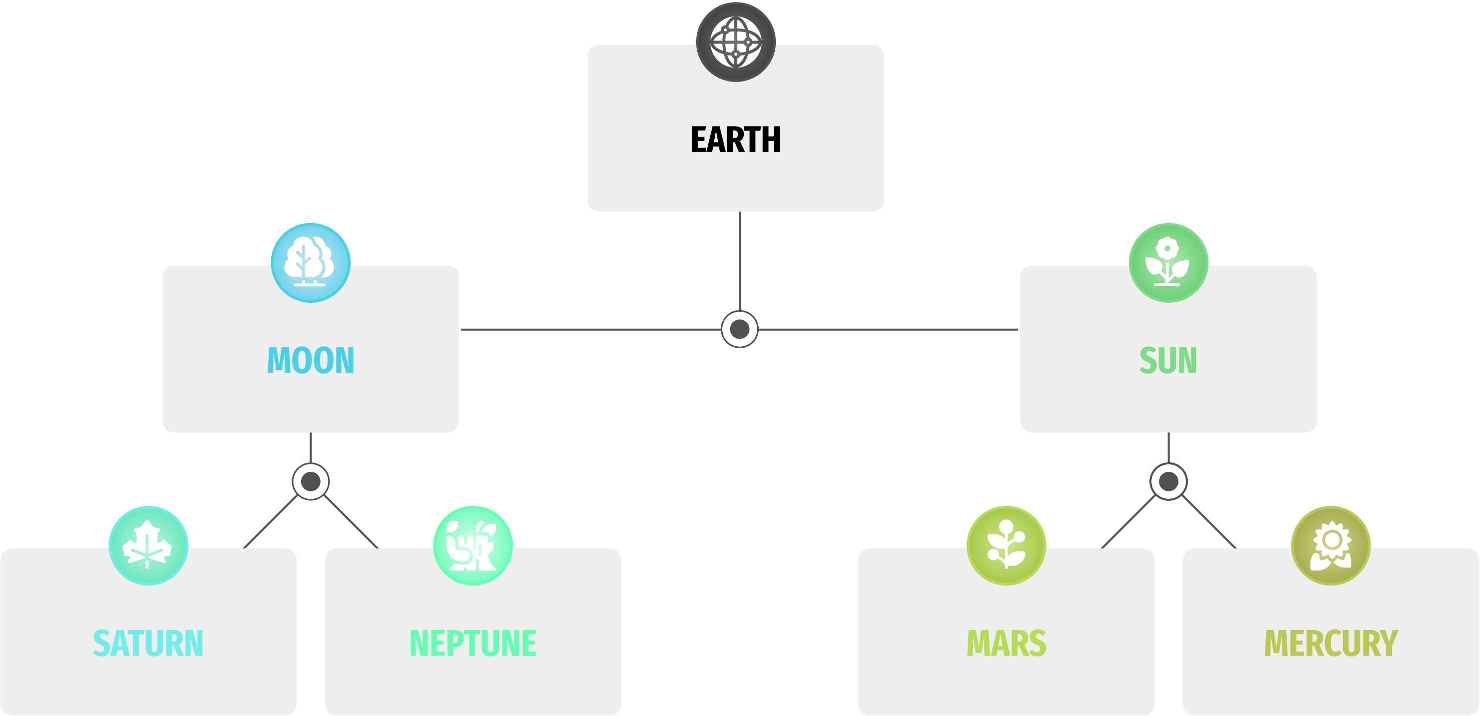
## MERCURY

Mercury is the closest object to the Sun

## JUPITER

Jupiter is a gas giant and the biggest planet

# Decision Tree Diagrams



# Decision Tree Diagrams

**MERCURY**

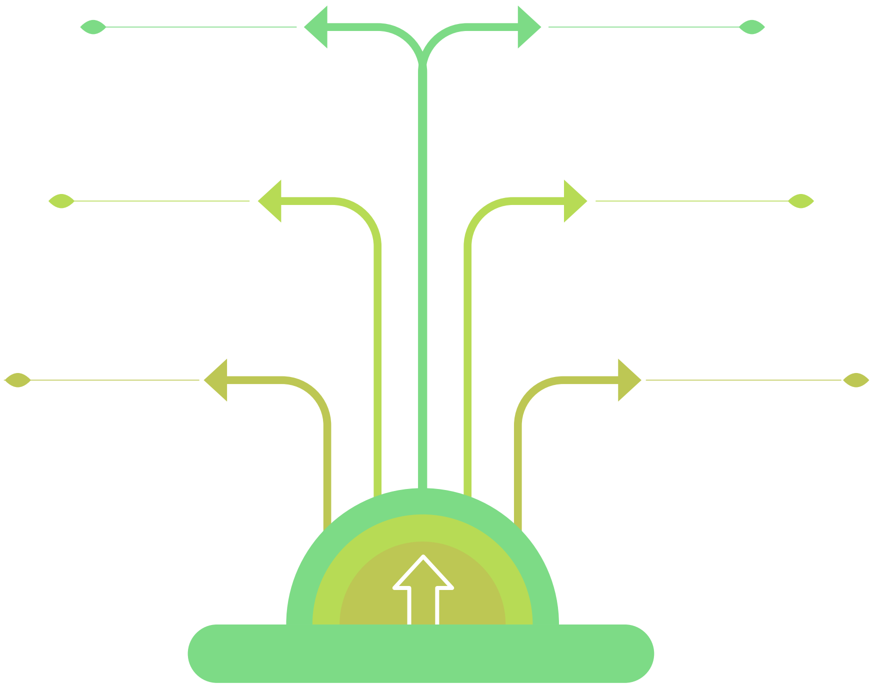
Mercury is the smallest planet

**VENUS**

Venus has a beautiful name

**JUPITER**

Jupiter is  the biggest planet



**MARS**

Mars is a cold place

**NEPTUNE**

It’s the farthest planet

**SATURN**

Yes, Saturn is the ringed one

# Decision Tree Diagrams

### MARS

Despite being red, Mars is a cold place

### MERCURY

Mercury is the closest object to the Sun

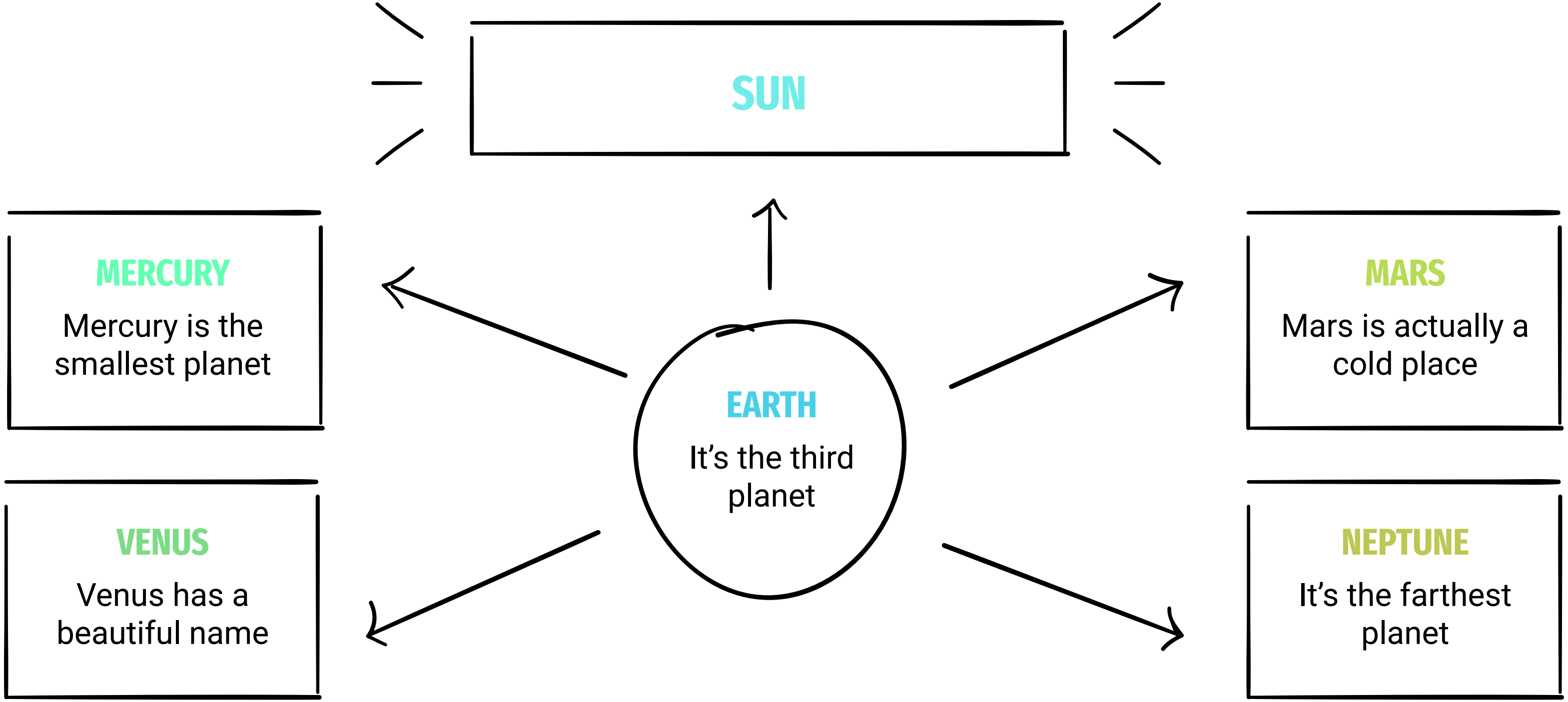
### VENUS

Venus has a very poisonous atmosphere

### JUPITER

Jupiter is a gas giant and the biggest planet

# Decision Tree Diagrams



# Decision Tree Diagrams

**MARS**

Despite being red, Mars is a cold place full of iron oxide dust

**VENUS**

Venus has a beautiful name and is the second planet from the Sun

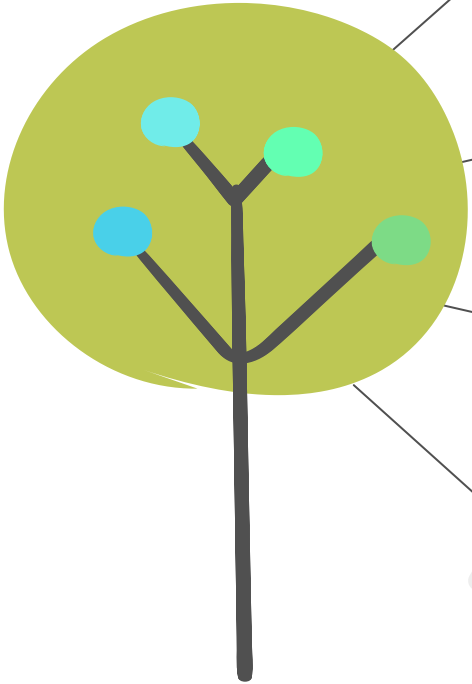
**MERCURY**

Mercury is the closest planet to the Sun and the smallest one

**JUPITER**

Jupiter is a gas giant and the biggest planet in the Solar System

# Decision Tree Diagrams



**MERCURY**

Mercury is the closest planet to the Sun and the smallest one

**VENUS**

Venus is the second planet from the Sun. It has a beautiful name

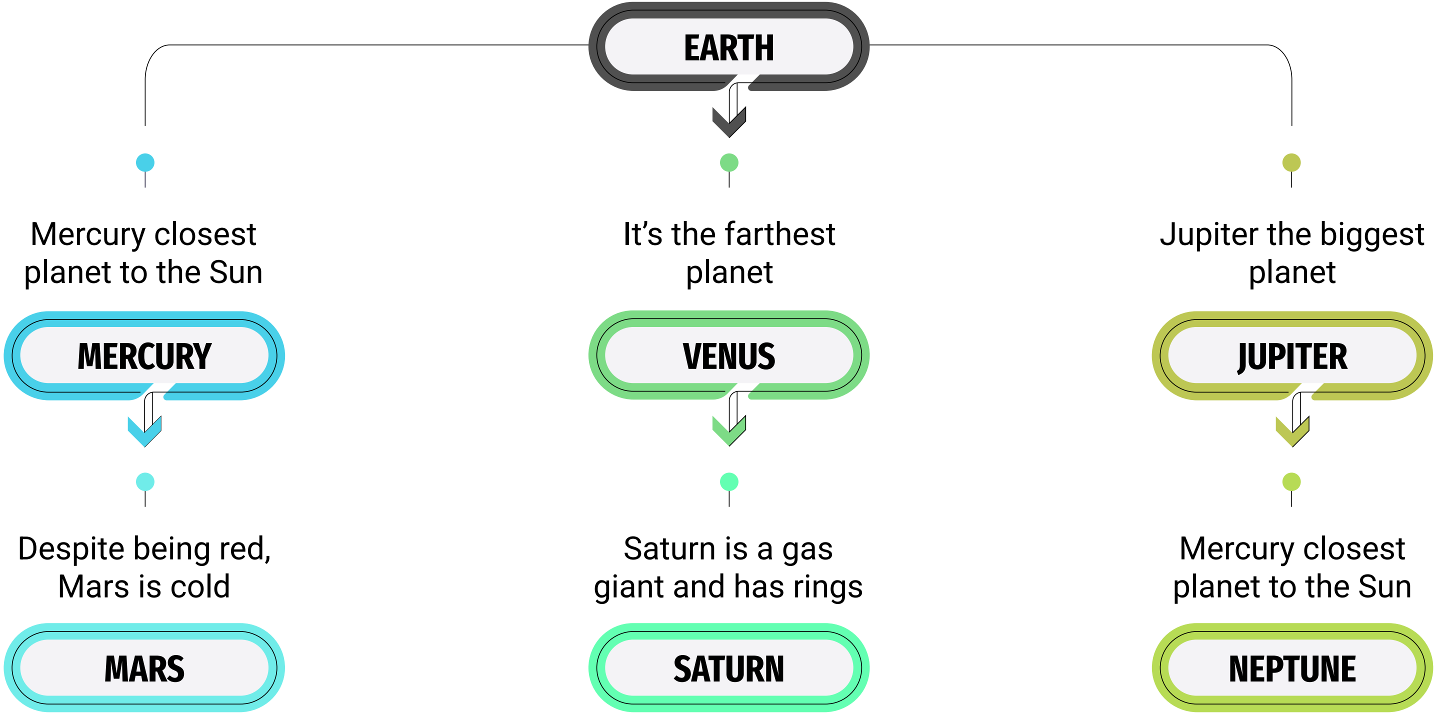
**SATURN**

Saturn is composed of hydrogen and helium

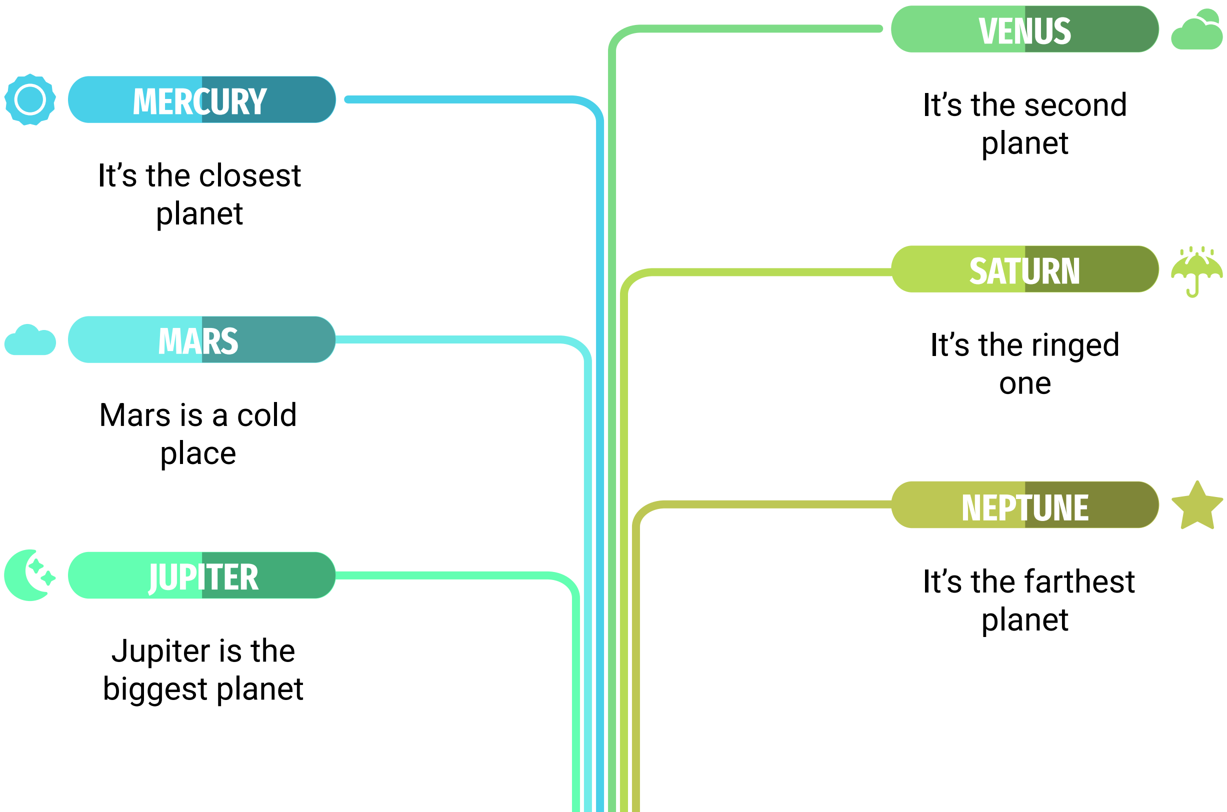
**NEPTUNE**

Neptune is the farthest planet from the Sun

# Decision Tree Diagrams



# Decision Tree Diagrams



# Decision Tree Diagrams

## MERCURY

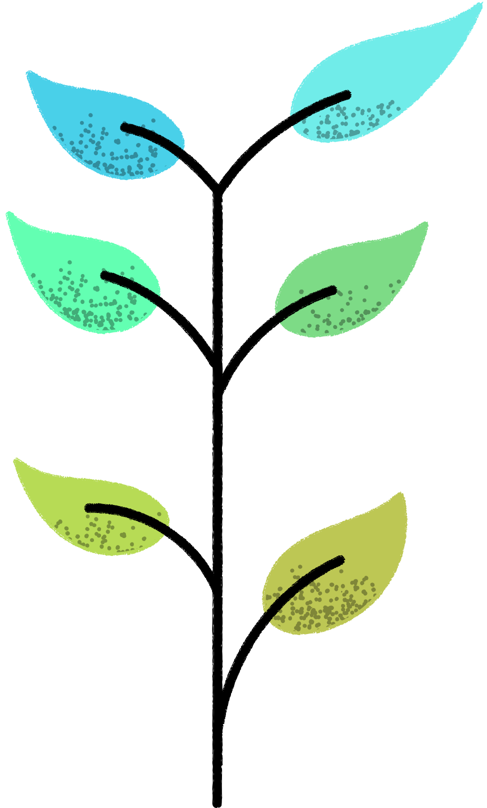
Mercury is the closest planet to the Sun and the smallest one

## VENUS

Venus has a beautiful name and is the second planet from the Sun

## JUPITER

Jupiter is a gas giant and the biggest planet in the Solar System



## MARS

Despite being red, Mars is a cold place full of iron oxide dust

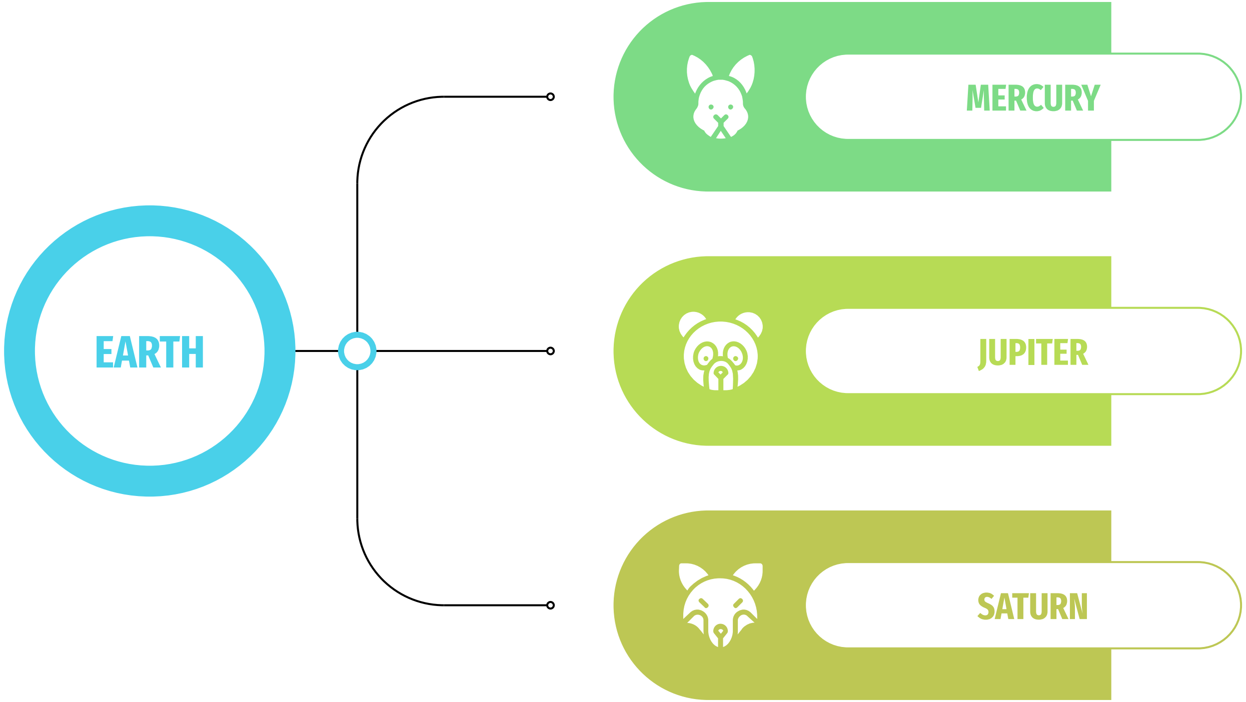
## NEPTUNE

Neptune is the farthest planet from the Sun and the fourth-largest one

## SATURN

Planet Saturn is a gas giant composed mostly of hydrogen and helium

# Decision Tree Diagrams

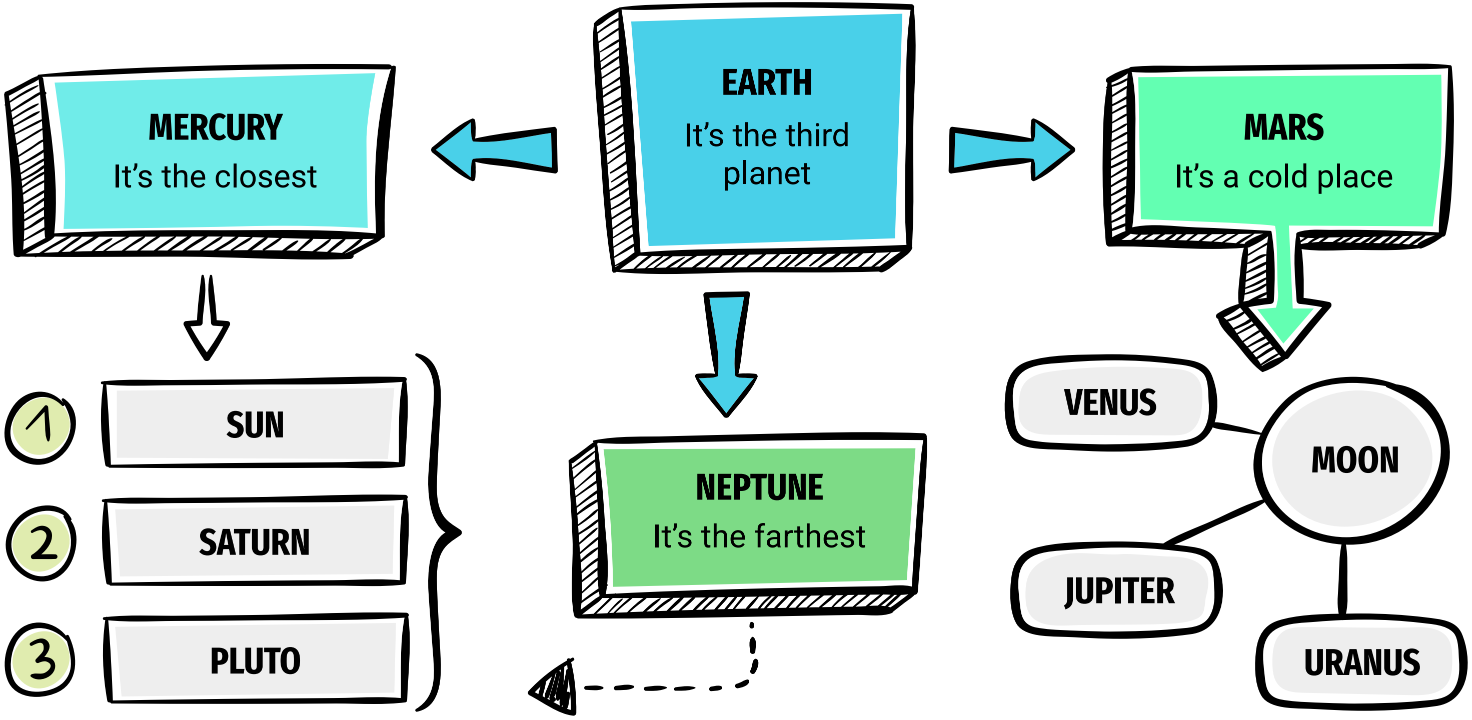



Mercury is the smallest planet in the Solar System

Jupiter is a gas giant and also the biggest planet

Saturn is a gas giant made of hydrogen and helium

# Decision Tree Diagrams



# Instructions for use (free users)

In order to use this template, you must credit **Slidesgo** by keeping the Thanks slide.

**You are allowed to:**

- Modify this template.
- Use it for both personal and commercial purposes.

**You are not allowed to:**

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Delete the “Thanks” or “Credits” slide.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

# Instructions for use (premium users)

In order to use this template, you must be a Premium user on [Slidesgo](https://slidesgo.com).

**You are allowed to:**

- Modify this template.
- Use it for both personal and commercial purposes.
- Hide or delete the “Thanks” slide and the mention to Slidesgo in the credits.
- Share this template in an editable format with people who are not part of your team.

**You are not allowed to:**

- Sublicense, sell or rent this Slidesgo Template (or a modified version of this Slidesgo Template).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

For more information about editing slides, please read our FAQs or visit Slidesgo School:
https://slidesgo.com/faqs and https://slidesgo.com/slidesgo-school

# Infographics

You can add and edit some **infographics** to your presentation to show your data in a visual way.

- Choose your favourite infographic and insert it in your presentation using Ctrl C + Ctrl V or Cmd C + Cmd V in Mac.
- Select one of the parts and **ungroup** it by right-clicking and choosing “Ungroup”.
- **Change the color** by clicking on the paint bucket.
- Then **resize** the element by clicking and dragging one of the square-shaped points of its bounding box (the cursor should look like a double-headed arrow). Remember to hold Shift while dragging to keep the proportions.
- **Group** the elements again by selecting them, right-clicking and choosing “Group”.
- Repeat the steps above with the other parts and when you’re done editing, copy the end result and paste it into your presentation.
- Remember to choose the **“Keep source formatting”** option so that it keeps the design. For more info, please visit **Slidesgo School**.