Random Forest Classifier

Here is where your presentation begins

Table of Contents

Introduction

One of the most popular algorithms to implement due to its simplicity and general accuracy is Random Forest.

Random Forest is an ensemble of decision trees, in which each tree is comprised of data drawn from a sample of a data set with a replacement called a bootstrap sample.

Combines predictions from individual trees to make final decisions

Developed by Leo Breiman and Adele Cutler in 2001

Decision Tree Review

A decision tree is a supervised machine learning algorithm that makes decisions by following a tree-like structure of conditions:

Structure: Consists of nodes (decision points), branches (outcomes), and leaves (final decisions)

Decision Process: Starts at the root node and follows paths based on feature values until reaching a leaf node

Splitting Criteria: Uses metrics like Gini impurity, entropy, or information gain to determine optimal splits

Interpretability: Highly interpretable - can follow the logic path for any prediction

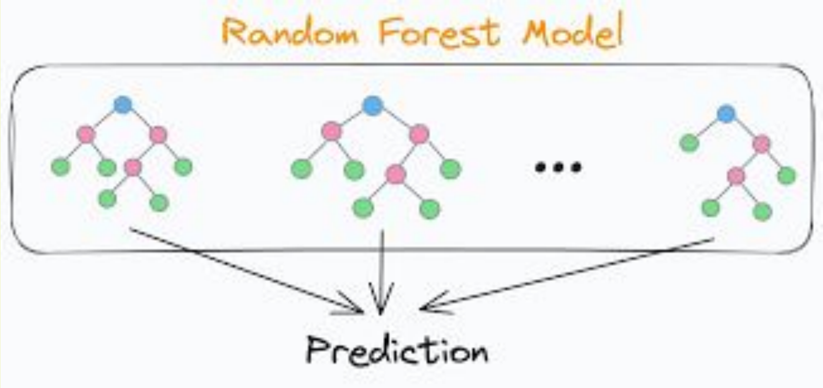
Visualization: Can be visualized as a flowchart, making it accessible to non-technical stakeholders

Decision trees work by recursively partitioning the data space, asking questions like "Is feature X > threshold Y?" at each node. The goal is to create segments where the samples are as homogeneous as possible regarding the target variable.

In other words, they work great with the data used to create them but are not flexible when it comes to classifying new samples.

Random Forest Builds Upon Decision Trees

While one decision tree might make errors due to the specific way it was trained, combining many trees with different perspectives helps cancel out individual mistakes and produces more robust predictions.
Random Forests essentially sacrifice some of the interpretability of decision trees to gain significant improvements in predictive power and stability.



Supervised or Unsupervised

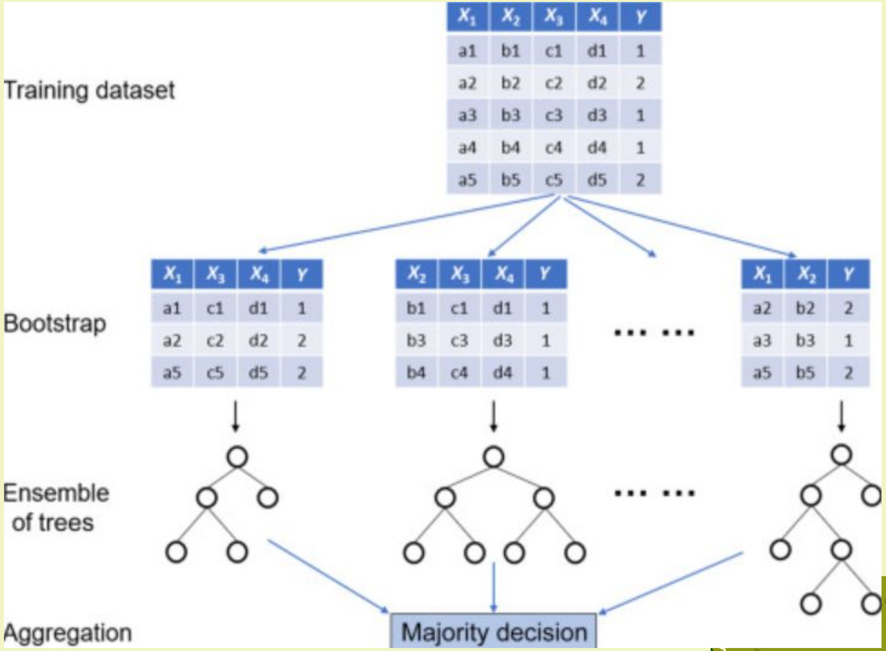
Random Forest is primarily supervised learning, using labeled training data to learn patterns and make predictions on new, unseen data.

Unsupervised methods exist that leverage the concept of Random Forest such as using the proximity matrix to perform clustering

Random Forest

Random forest can be broken down into 4 stages

1. Bootstrapping (Bagging Technique)
2. Feature Randomness (Feature Bagging)
3. Tree Construction
4. Voting for Classification



Bootstrapping

The algorithm randomly selects subsets of the training data (with replacement) to train individual decision trees.

Each tree is trained on a slightly different dataset.

Creates multiple subsets of training data through random sampling with replacement
Each subset trains a different decision tree

Feature Randomness

When splitting a node in a decision tree, only a random subset of features is considered rather than all available features.

This reduces correlation between the trees and improves generalization.

At each split, only considers a random subset of features
Typically $\sqrt{p}$ features (where p is total number of features)

Tree Construction

Each decision tree in the forest is trained independently on its bootstrapped dataset.

Trees are grown to full depth (or stopped based on pre-defined conditions like max_depth, min_samples_split).

Each tree learns different patterns due to different training samples and feature selections meaning trees are decorrelated due to different training samples and feature subsets

Aggregation/Voting

Once all trees are trained, they make independent predictions on new data.

The final prediction is determined by majority voting, meaning the class that receives the most votes from individual trees is chosen.

Key Parameters

n_estimators: Number of trees in the forest
max_features: Maximum number of features considered for splitting
max_depth: Maximum depth of trees
min_samples_split: Minimum samples required to split a node
min_samples_leaf: Minimum samples required at a leaf node
bootstrap: Whether to use bootstrap samples

Advantages vs Disadvantages

Robust to overfitting compared to individual decision trees
Handles high-dimensional data well
Automatically ranks feature importance
Works well with both categorical and numerical data
Handles missing values effectively
No assumption about data distribution
Parallel processing capability

Less interpretable than single decision trees
Computationally intensive for large datasets
Slower prediction than simpler models
May overfit on noisy datasets
Biased in favor of features with more levels (categorical variables)

Real World Applications

- Healthcare: Disease prediction, diagnostic assistance
- Finance: Credit scoring, fraud detection
- Marketing: Customer segmentation, churn prediction
- E-commerce: Product recommendations
- Environmental Science: Climate modeling, species distribution
- Manufacturing: Quality control, predictive maintenance
- Cybersecurity: Threat detection, anomaly identification

Random forest vs Gradient Boosting

- **Ensemble approach**: Builds trees sequentially (boosting)
- **Tree building**: Each new tree focuses on correcting errors made by previous trees
- **Error correction**: Explicitly minimizes errors through gradient descent on loss function
- **Training process**: Usually uses shallow trees (weak learners)
- **Overfitting resistance**: More prone to overfitting, requires careful tuning
- **Speed**: Slower to train due to sequential nature
- **Tuning difficulty**: More hyperparameters to tune, requiring more careful optimization

Random Forest: When you need a robust model with minimal tuning, have very high dimensional data, or when training speed is important
Gradient Boosting: When you need the highest possible predictive performance and have time for proper hyperparameter tuning

Random Forest vs KNN

- Random Forest creates a model during training, while KNN is "lazy" and only computes during prediction
- Random Forest naturally handles feature importance, while KNN treats all features equally
- Random Forest generally performs better on high-dimensional data
- KNN is conceptually simpler but can be computationally expensive for large datasets

Which one works better depends on your specific use case, dataset characteristics, and computational constraints.

Common Questions

How many trees should I use in my Random Forest?
This is often determined experimentally, but typically ranges from 100-500 trees
More trees generally lead to better performance up to a point of diminishing returns
Computational resources become a limitation with very large numbers of trees

How does Random Forest handle imbalanced data?
By default, not very well - tends to favor majority class
Solutions include using class weights, undersampling/oversampling techniques, or setting sample_weight parameter

When should I choose Random Forest over other algorithms?
When interpretability is important but single decision trees are too unstable
With mixed data types (categorical and numerical features)
When you have a moderate amount of features but don't want to perform extensive feature selection

Main Take Away

- Random Forest is a versatile ensemble learning method
- Combines multiple decision trees with randomization
- Offers excellent performance with minimal tuning
- Provides feature importance rankings
- Works well across various domains and data types
- Balances accuracy and overfitting

Example

https://colab.research.google.com/drive/1ew7BgbgbHiHZ9JAcuDhuSy6F2p-Y9qyZ?usp=sharing

The end.

Our academic areas

Mars

Despite being red, Mars is a cold planet

Venus

Venus is the second planet from the Sun

Neptune

Neptune is the farthest planet from the Sun

Mercury

Mercury is the closest planet to the Sun

Saturn

Saturn is a gas giant with several rings

Jupiter

Jupiter is the biggest planet of them all

Our mission and vision

Mission

Mercury is the closest planet to the Sun and the smallest one in the Solar System—it's only a bit larger than the Moon

Vision

Venus has a beautiful name and is the second planet from the Sun. It's hot and has a very poisonous atmosphere

Our values

Mars

Mars is actually a very cold place

Neptune

Neptune is very far from the Sun

Jupiter

Jupiter is the biggest planet of them all

Saturn

Saturn is a gas giant and has several rings

Our success

Mercury

Mercury is the closest planet to the Sun and the smallest of them all

75%

Venus

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

50%

Mars

Despite being red, Mars is actually a cold place. It’s full of iron oxide dust

25%

98,300,000

Big numbers catch your audience’s attention

Awesome words

A picture always reinforces the concept

Images reveal large amounts of data, so remember: use an image instead of a long text. Your audience will appreciate it

Statistics

Venus

Venus has a beautiful name

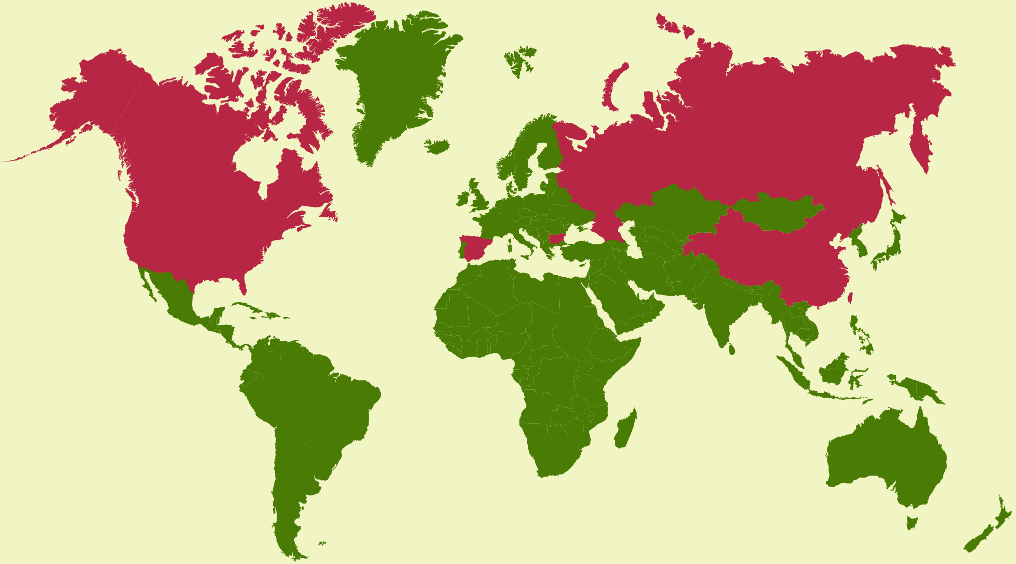
Mercury

It's the closest planet to the Sun

Follow the link in the graph to modify its data and then paste the new one here. **For more info, click here**

Protected areas in world

Number of protected areas	
Equal to or greater than 20	
Equal to or less than 19	



Mockups

You can replace the images on the screen with your own work. Just right-click on it and select “Replace images”

Our goals

Objectives	Description
Academic success	Mercury is the closest planet to the Sun
	Mars is actually a very cold place
Community	Venus is the second planet from the Sun
	Jupiter is the biggest planet of them all
Diversity	Saturn is composed of hydrogen and helium
	Neptune is the farthest planet from the Sun

Student progress

01

Earth is the only planet known to harbor life

02

Saturn is a gas giant and has several rings

03

Despite being red, Mars is actually a cold planet

04

Mercury is the closest planet to the Sun

05

Neptune is very far away from Earth

06

Jupiter is the biggest planet of them all

Enrollment process

01

Venus is the second planet from the Sun

02

Despite being red, Mars is actually a cold place

03

Jupiter is the biggest planet of them all

04

Saturn is composed of hydrogen and helium

05

Neptune is the farthest planet from the Sun

Special reminders

Do you know what helps you make your point clear?
Lists like this one:

- They're simple
- You can organize your ideas clearly
- You'll never forget to buy milk!

And the most important thing: the audience won't miss the point of your presentation

Thanks!

Do you have any questions?

youremail@freepik.com

+34 654 321 432

yourwebsite.com

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution

Alternative resources

Here’s an assortment of alternative resources whose style fits that of this template:

- Flat jungle background

Resources

Did you like the resources in this template? Get them on these websites:

Vectors:

- Hand drawn flat design enchanted forest illustration

Photos:

- Close up on cute bird in the woods
- Gecko
- Front view smiley man working
- Medium shot woman holding book

Instructions for use

If you have a free account, in order to use this template, you must credit Slidesgo by keeping the Thanks slide. Please refer to the next slide to read the instructions for premium users.

As a Free user, you are allowed to:

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

You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

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

Instructions for use (premium users)

As a Premium user, you can use this template without attributing [Slidesgo] or keeping the "[Thanks]" slide.

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 our blog:
https://slidesgo.com/faqs and https://slidesgo.com/slidesgo-school

Fonts & colors used

This presentation has been made using the following fonts:

Barlow
https://fonts.google.com/specimen/Barlow

Lato
https://fonts.google.com/specimen/Lato

#155800

#f1f5c4

#a1704e

#909919

#864326

#bd906a

#b72644

#4a7c05

Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out **how it works**.

Pana

Amico

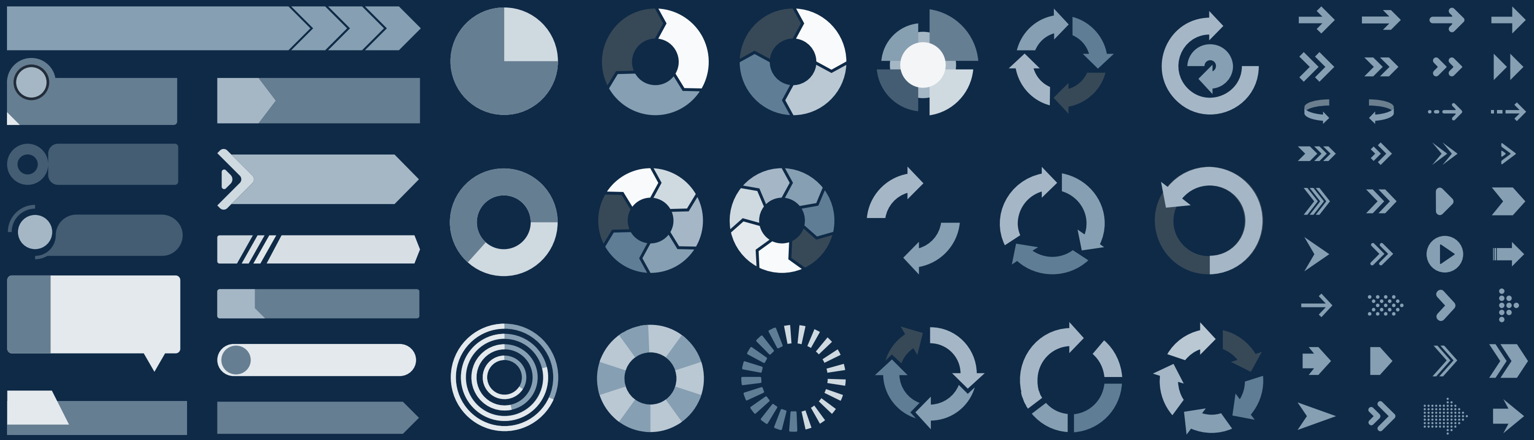
Bro

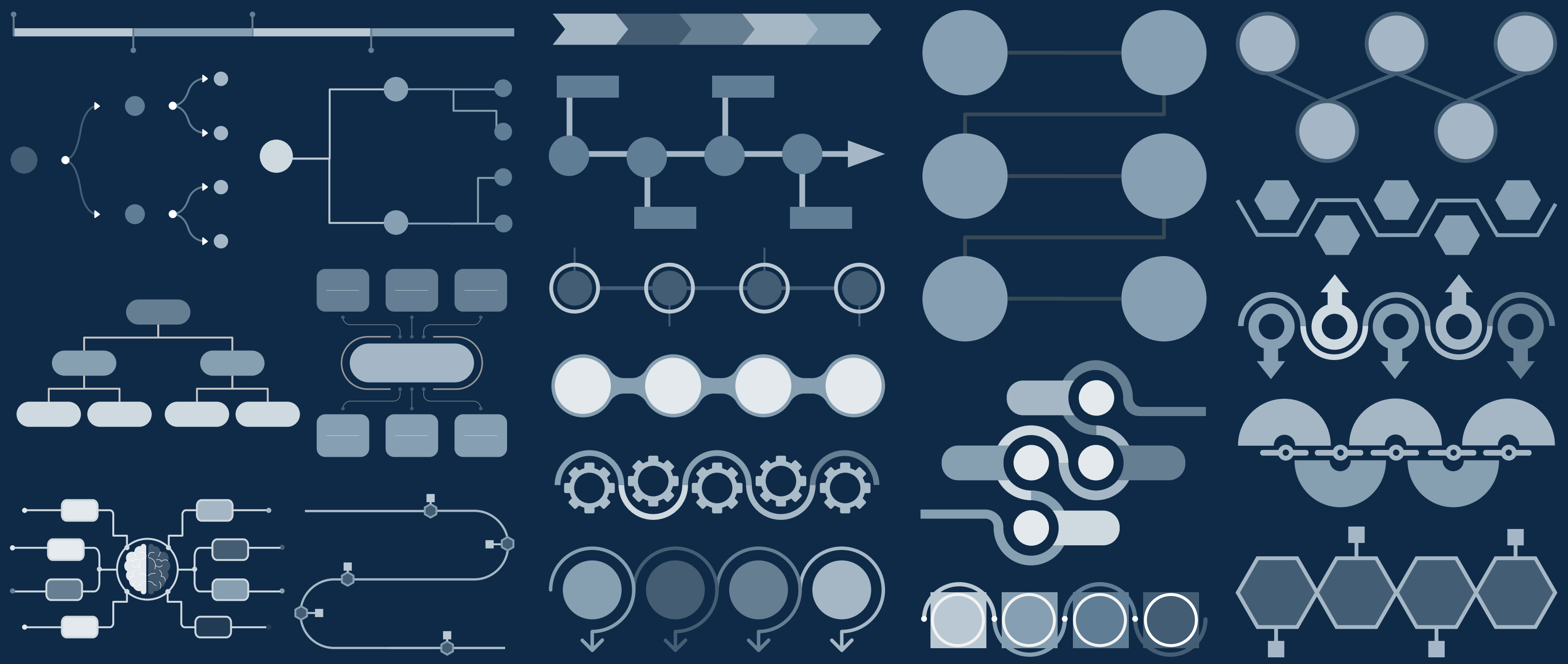
Rafiki

Cuate

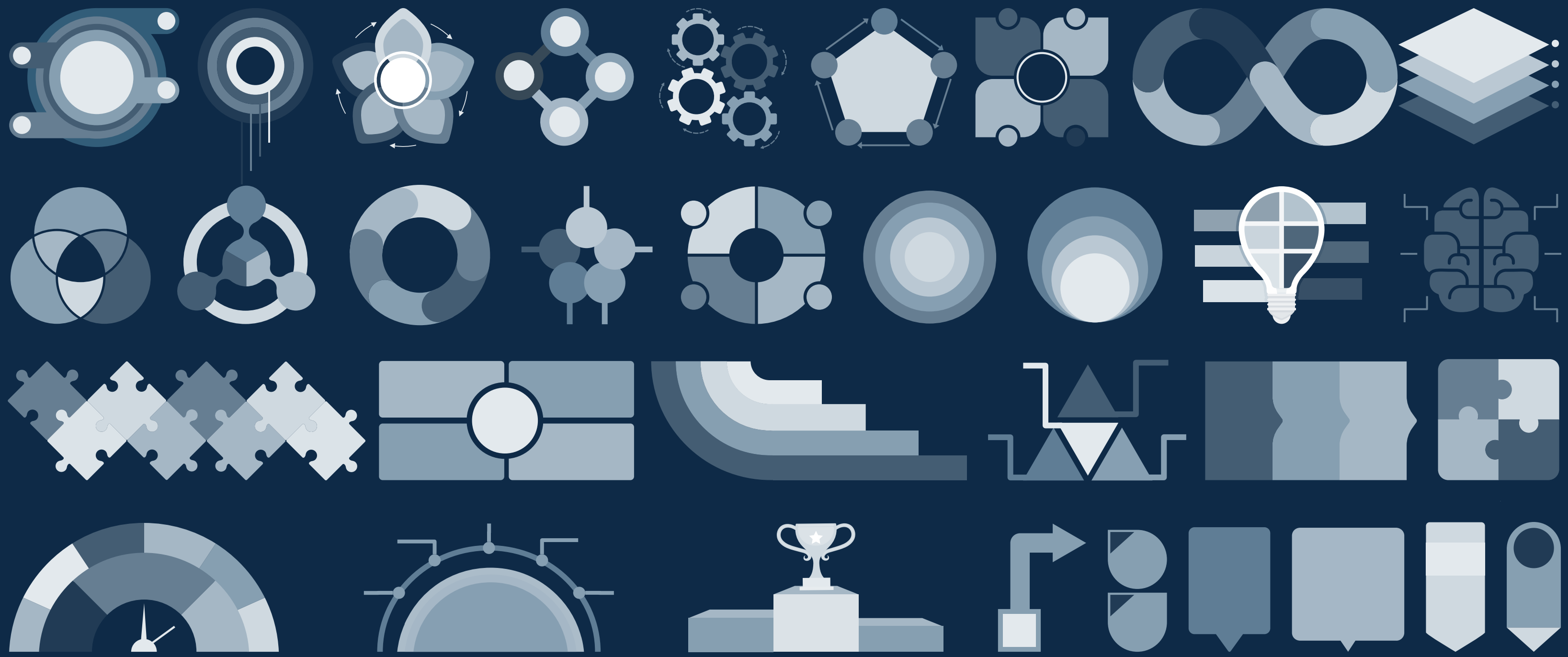
Use our editable graphic resources...

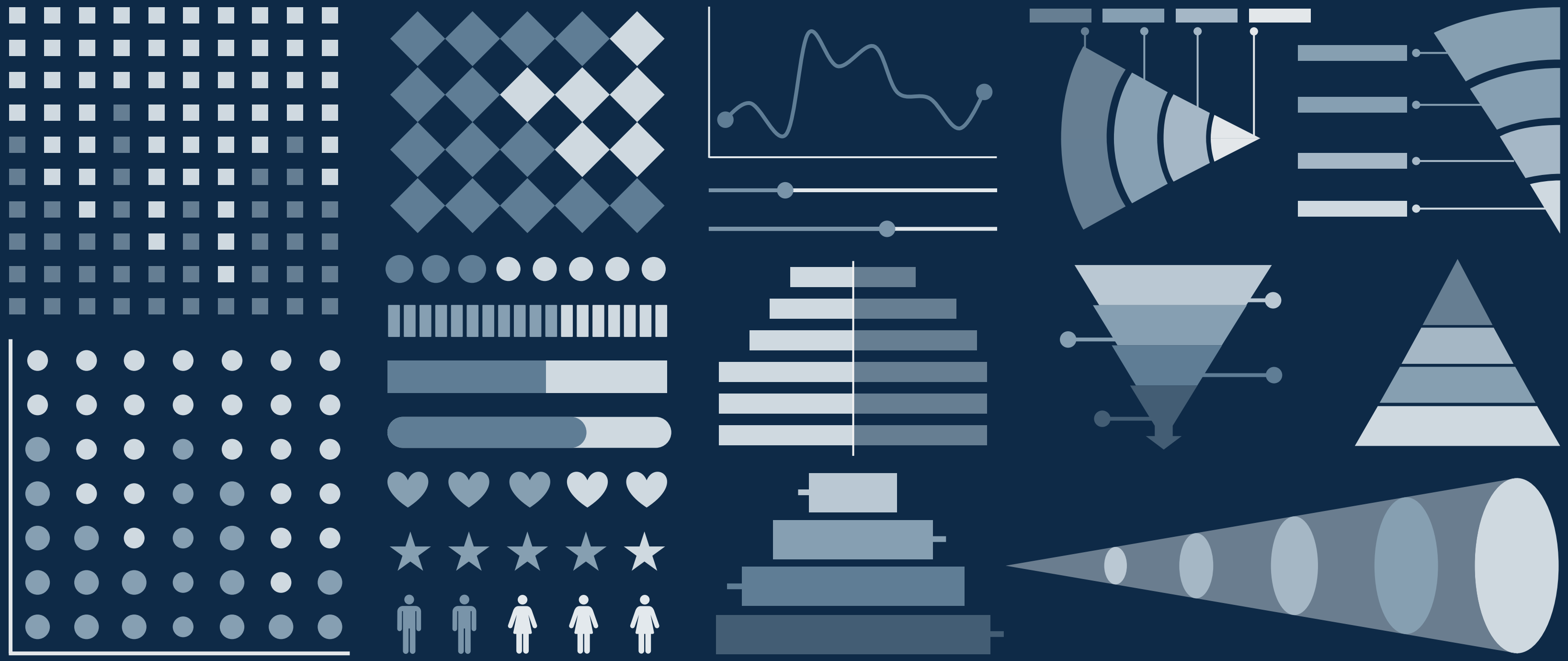
You can easily **resize** these resources without losing quality. To **change the color**, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more **infographics** on Slidesgo.






JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
PHASE 1
Task 1
Task 2
PHASE 2
Task 1
Task 2
JANUARY
FEBRUARY
MARCH
APRIL
PHASE 1
Task 1
Task 2





...and our sets of editable icons

You can **resize** these icons without losing quality.
You can **change the stroke and fill color**; just select the icon and click on the **paint bucket/pen**.
In Google Slides, you can also use **Flaticon’s extension**, allowing you to customize and add even more icons.

Educational Icons

Medical Icons

Business Icons

Teamwork Icons

Help & Support Icons

Avatar Icons

Creative Process Icons

Performing Arts Icons

Nature Icons

SEO & Marketing Icons