A Global Model-Agnostic XAI method for the Automatic Formation of an Abstract Argumentation Framework and its Objective Evaluation

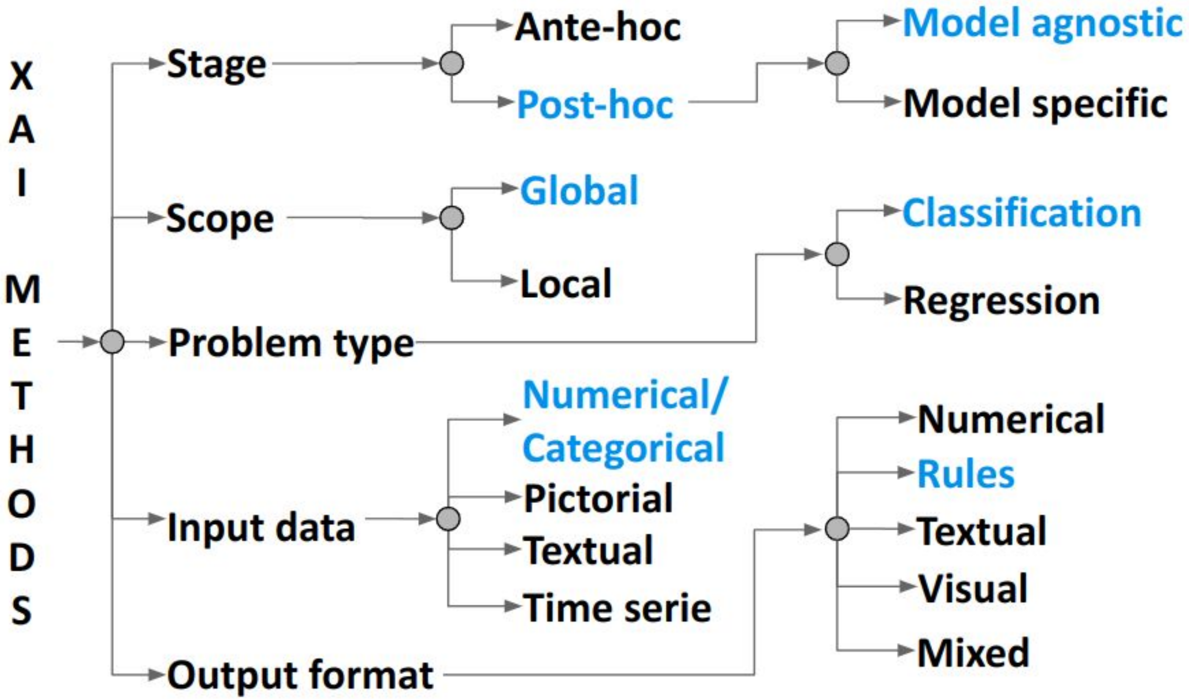
Giulia Vilone, Luca Longo
School of Computer Science, Technological University Dublin

ArgXAI workshop, Cardiff, 12th September 2022

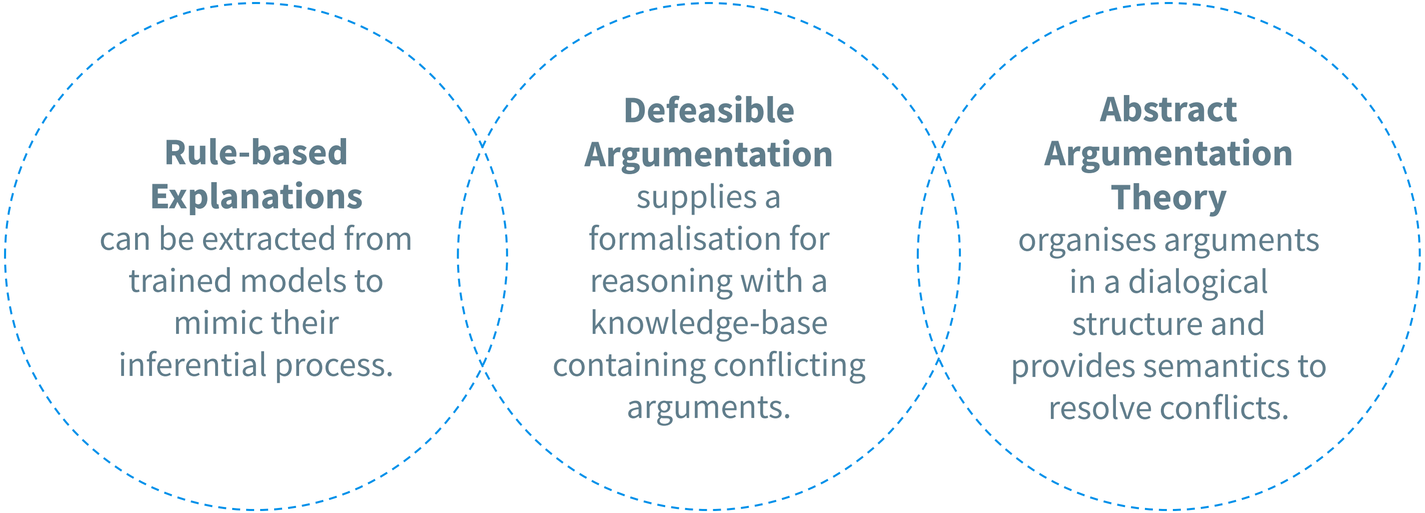
Presentation outline

- Literature review
- State-of-the-art, gap and motivation
- The experiment
- Objective evaluation
- Results
- Conclusions & future work

Literature review



Literature review



Rule-based Explanations can be extracted from trained models to mimic their inferential process.

Defeasible Argumentation supplies a formalisation for reasoning with a knowledge-base containing conflicting arguments.

Abstract Argumentation Theory organises arguments in a dialogical structure and provides semantics to resolve conflicts.

State-of-the-art, gaps and motivations

WEAKNESSES OF XAI METHODS

Generating list of rules just mimicking the inferential process of a model and lacking a richer reasoning process.

ENHANCING EXPLAINABILITY

The existing XAI methods produce rulesets that might not be 1) easily understandable and 2) consistent with existing domain knowledge [2].

NON-MONOTONIC REASONING

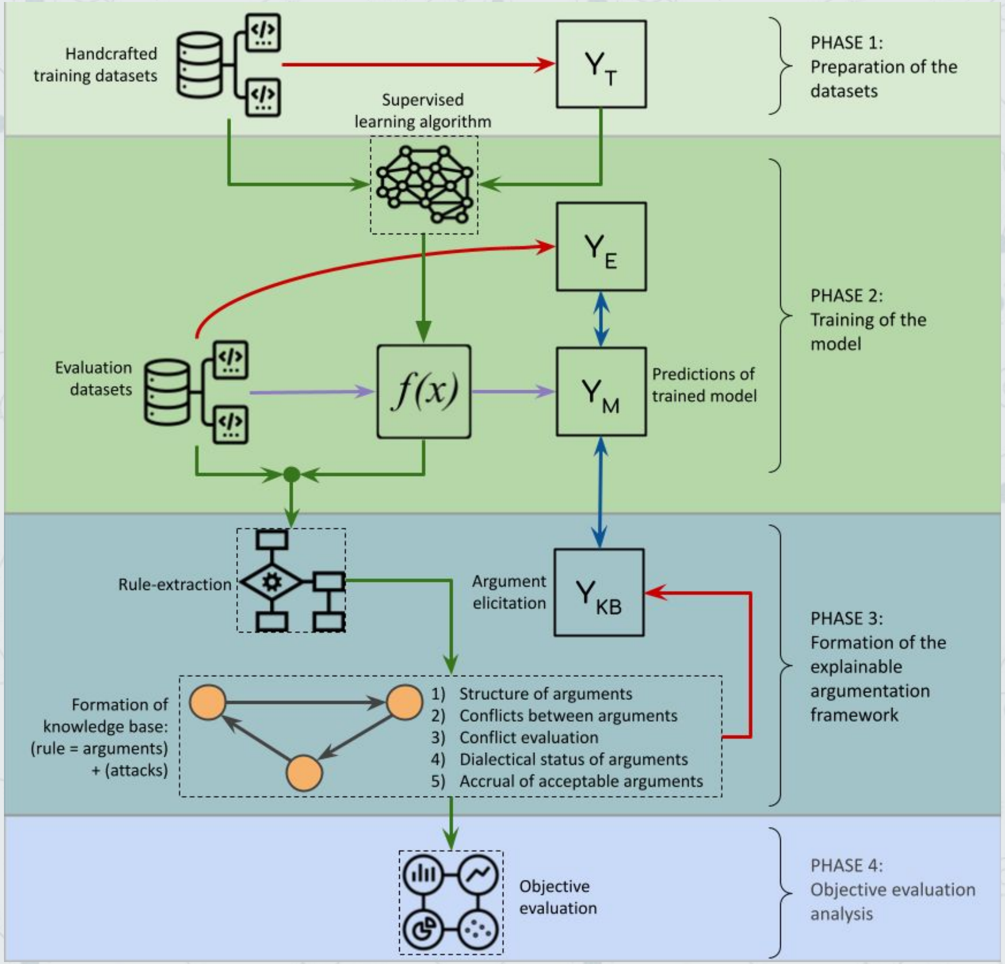
New data can lead to new rules potentially inconsistent with existing rules. No tools to handle these conflicts are provided [3].

ARGUMENTATION THEORY (AT)

AT investigates formal approaches for defeasible reasoning processes. Minimal work exists on the integration of Machine Learning (ML) and AT [4].

The experiment

The process to build the argument-based XAI method in a diagram.



Handcrafted training datasets

Y_T

PHASE 1: Preparation of the datasets

Supervised learning algorithm

Y_E

PHASE 2: Training of the model

Evaluation datasets

$f(x)$

Y_M

Predictions of trained model

Rule-extraction

Argument elicitation

Y_{KB}

Formation of knowledge base: (rule = arguments) + (attacks)

1) Structure of arguments
2) Conflicts between arguments
3) Conflict evaluation
4) Dialectical status of arguments
5) Accrual of acceptable arguments

PHASE 3: Formation of the explainable argumentation framework

Objective evaluation

PHASE 4: Objective evaluation analysis

Phase 1: dataset preparation

Curse of dimensionality

The datasets must have enough samples to train an accurate model.

Machine generated data

Data contained in the datasets must be manually collected or built by domain experts.

Multi-dimensional data

Each dataset must contain a mix of continuous and categorical independent features.

Missing data

None of the selected datasets have missing data, so no action was required.

Multicollinearity

A correlation analysis was carried out to detect and discard highly correlated features.

Unbalanced data

The SMOTE algorithm was applied to the training datasets to up-sample the minority classes.

Phase 1: dataset preparation

	Total number of instances	Number of input features	No. continuous (categorical) features	Number of output classes
Adult	**48,842**	**14**	**6 (8)**	**2**
Avila	**20,867**	**10**	**10 (0)**	**12**
Credit card default	**30,000**	**23**	**20 (3)**	**2**
Hotel bookings	**119,385**	**23**	**16 (7)**	**3**
Online shopper intention	**12,330**	**17**	**14 (3)**	**2**

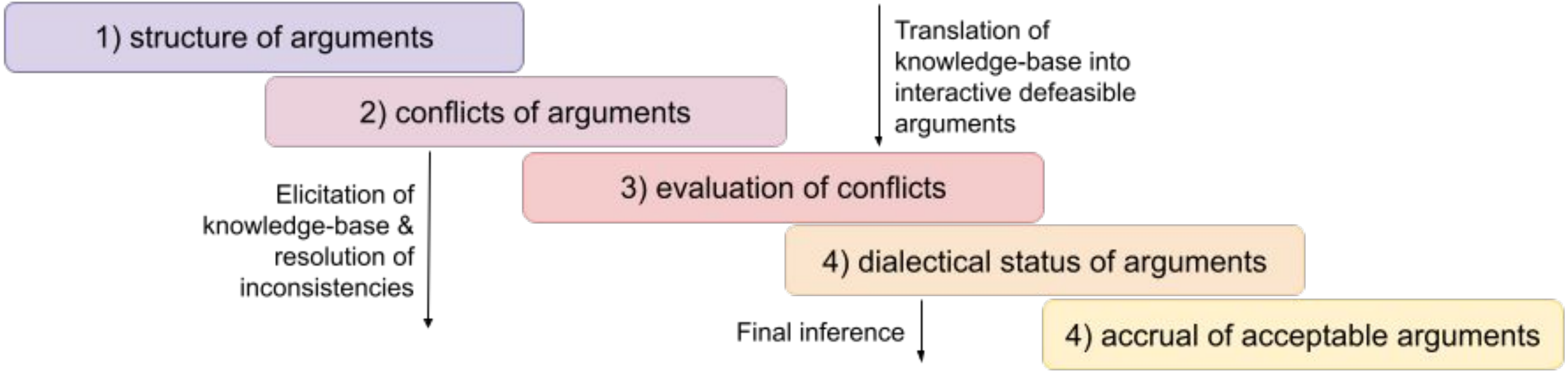
Phase 2: model training

- Feed forward **neural networks**.
- **Grid search** to determine hyperparameters and reach the highest prediction accuracy.
- **Early stopping** of training process after 5 epochs without improvement in validation accuracy to avoid overfitting.

Phase 2: model training

Model hyperparameters	Adult	Avila	Credit card default	Hotel bookings	Online shopper intention
Optimizer	**Adam**	**RMSprop**	**Adamax**	**SGD**	**SGD**
Weight initialisation	**Uniform**	**He-Unif.**	**Normal**	**Lecun-Unif.**	**He-Unif.**
Activation function	**Tanh**	**Relu**	**Softplus**	**Softplus**	**Softmax**
Dropout rate	**0%**	**0%**	**10%**	**0%**	**0%**
Batch size	**128**	**16**	**16**	**8**	**8**
Hidden neurons	**16**	**32**	**32**	**24**	**8**
Accuracy (validation)	**83% (79%)**	**98% (91%)**	**68% (79%)**	**65% (59%)**	**84% (87%)**

Phase 3: argumentation framework



1) structure of arguments

2) conflicts of arguments

3) evaluation of conflicts

4) dialectical status of arguments

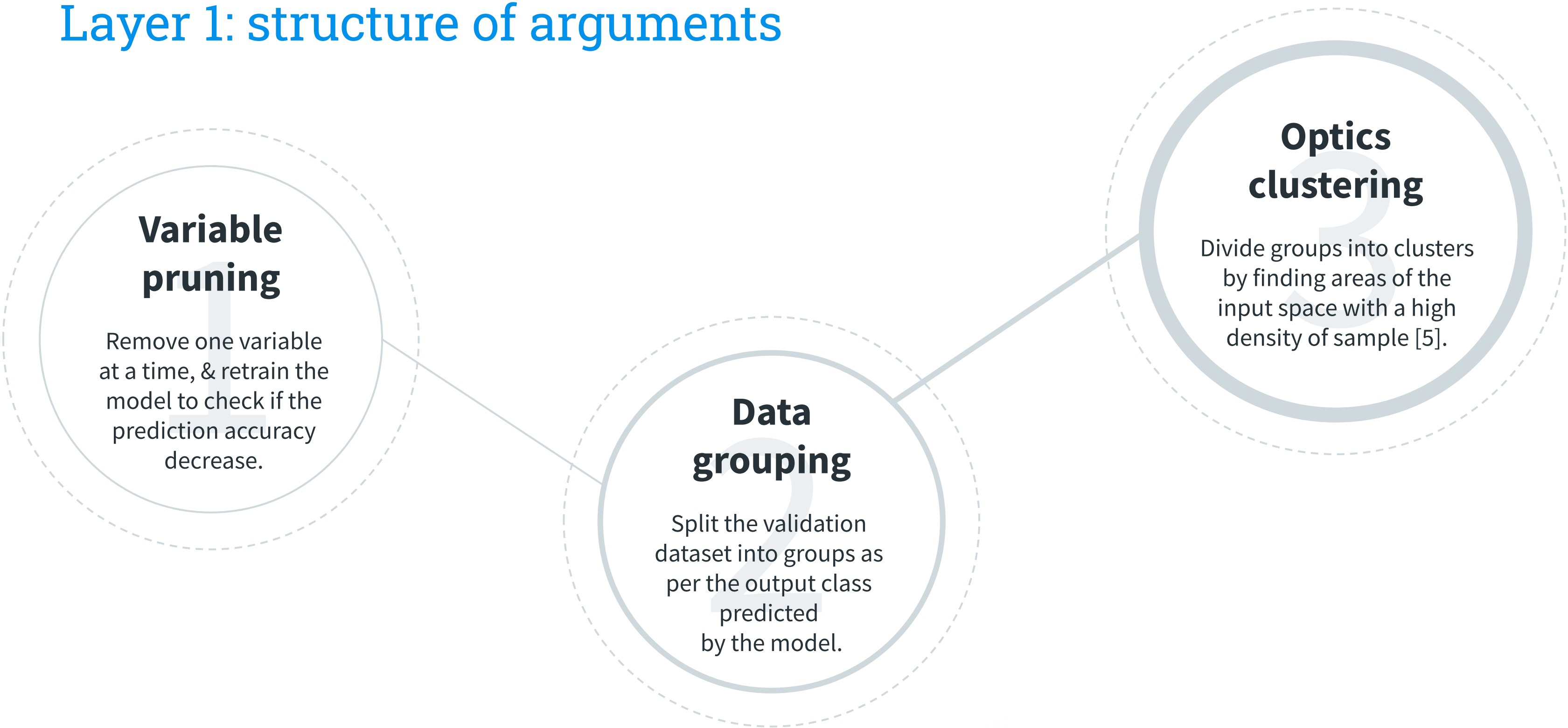
4) accrual of acceptable arguments

Translation of knowledge-base into interactive defeasible arguments

Elicitation of knowledge-base & resolution of inconsistencies

Final inference

Layer 1: structure of arguments



1

Variable pruning

Remove one variable at a time, & retrain the model to check if the prediction accuracy decrease.

2

Data grouping

Split the validation dataset into groups as per the output class predicted by the model.

3

Optics clustering

Divide groups into clusters by finding areas of the input space with a high density of sample [5].

Layer 1: structure of arguments

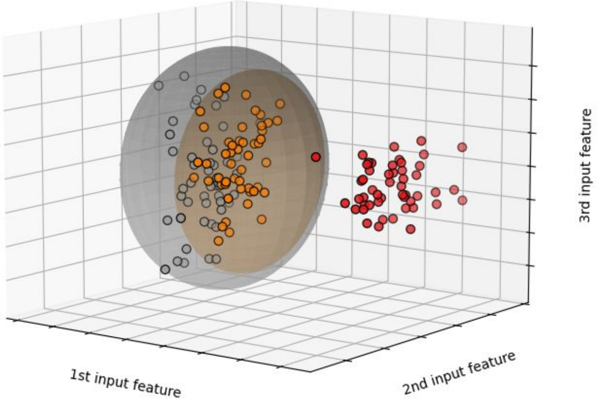
- Each cluster is translated into a rule by determining the min & max values of its samples for each relevant variable.
- The rule's antecedents correspond to these ranges, and the conclusion is the predicted class of the cluster's samples.

$$IF\, m_1 \leq X_1 \leq M_1\, AND \ldots\, AND\, m_N \leq X_N \leq M_N\, THEN\, Class_X$$

Layer 2: attacks between rules

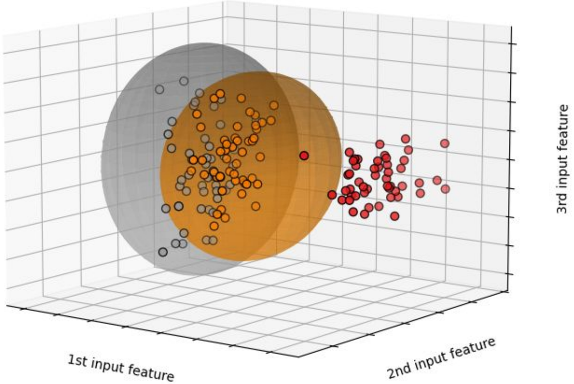
UNDERCUTTING ATTACKS

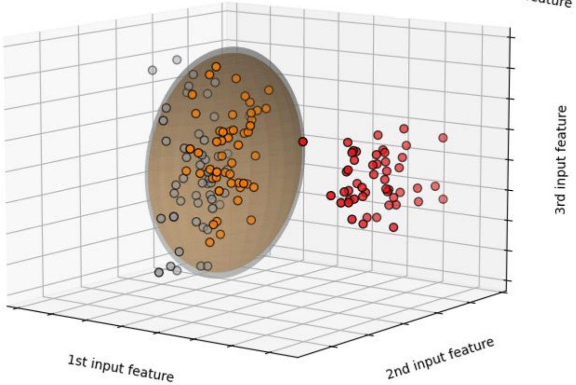
An argument is attacked by arguing that there is a special case that does not allow its application



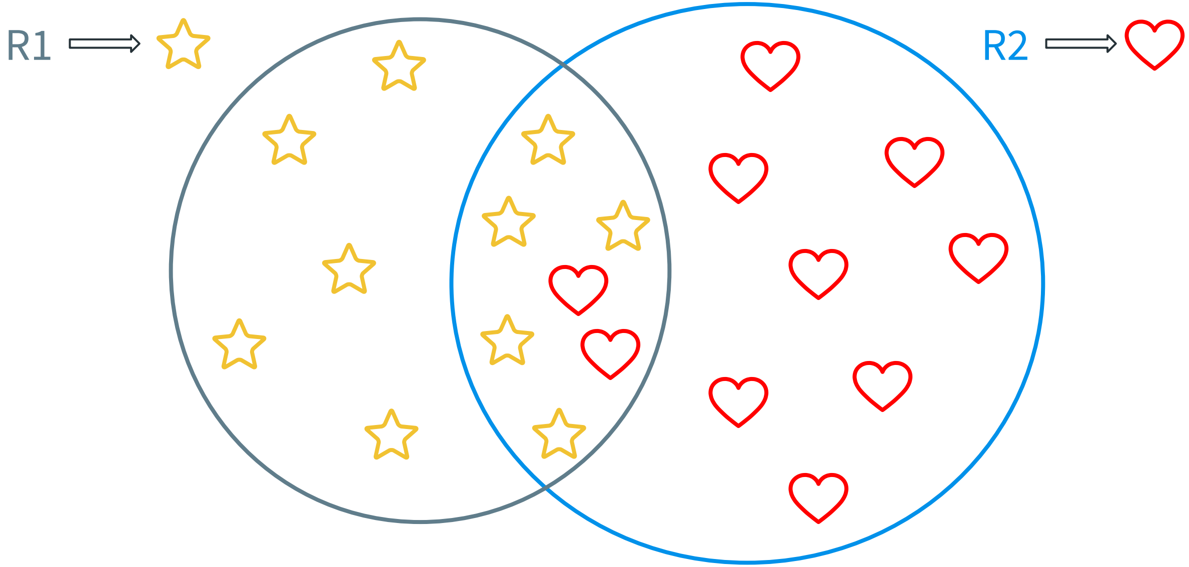
REBUTTING ATTACKS

An argument negates the conclusion of another.

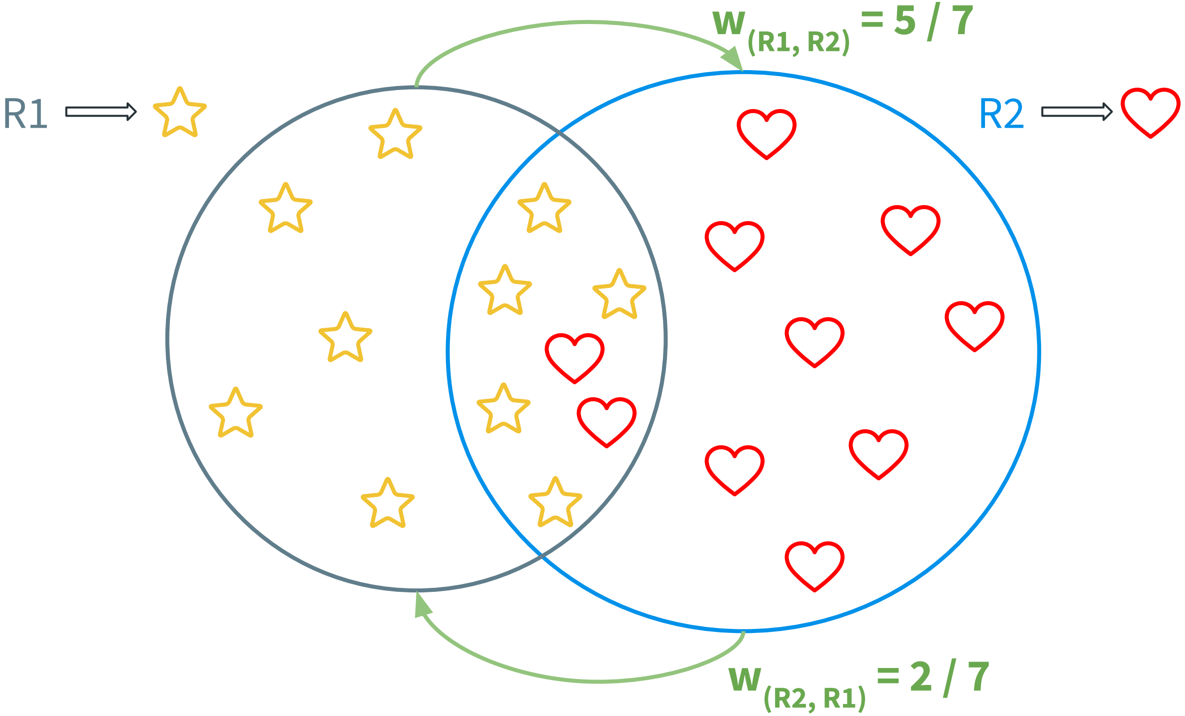




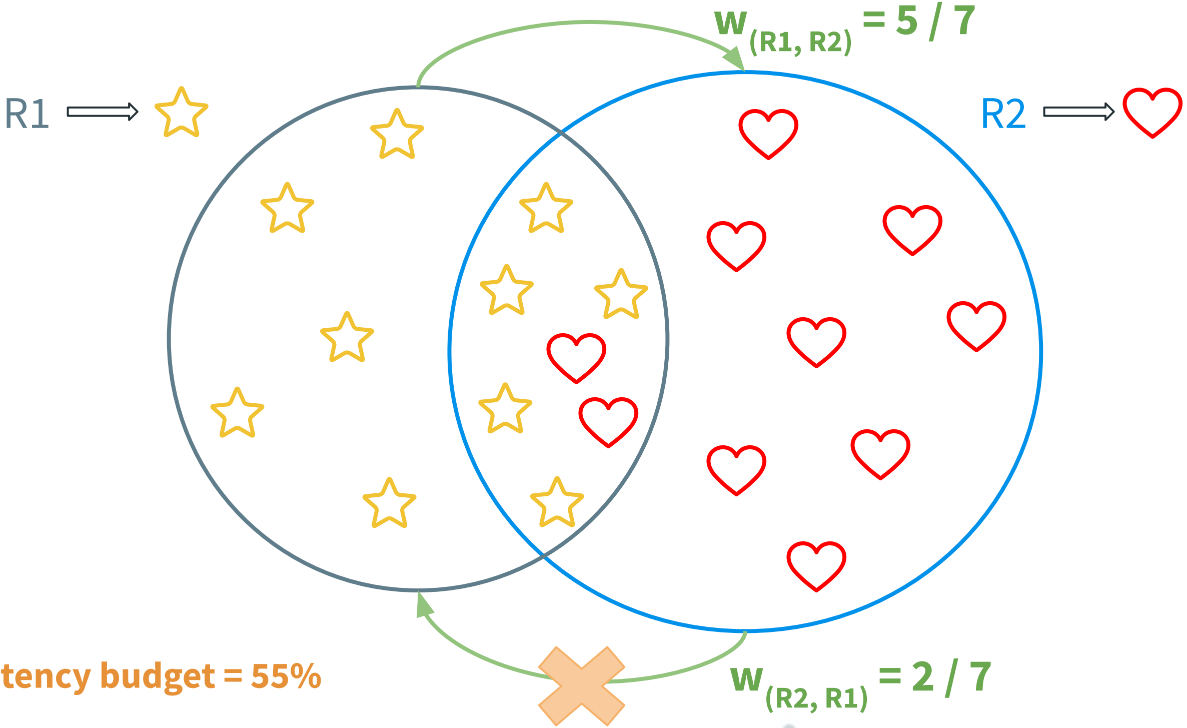
Layer 3: evaluation of attacks



Layer 3: evaluation of attacks



Layer 3: evaluation of attacks



$W_{(R1,\,R2)} = 5\,/\,7$

$W_{(R2,\,R1)} = 2\,/\,7$

Consistency budget = 55%

The other layers

- ## Layer 4: definition of the dialectal status of arguments.

  Ranking-base categoriser semantic - A recursive function that orders a set of active arguments from the most to the list acceptable based on the number of attacks and the ranks of the attacking arguments.

- ## Layer 5: Accrual of acceptable arguments.

  The **highest-ranked argument** was selected and its conclusion was deemed the most rationale. If multiple arguments had the highest rank, they were split into groups according to their conclusion and the group with the highest cardinality was selected.

Phase 4: Objective evaluation

COMPLETENESS % instances covered by the ruleset.

CORRECTNESS % instances correctly classified by rules.

FIDELITY % instances whose predictions of model and rules agree.

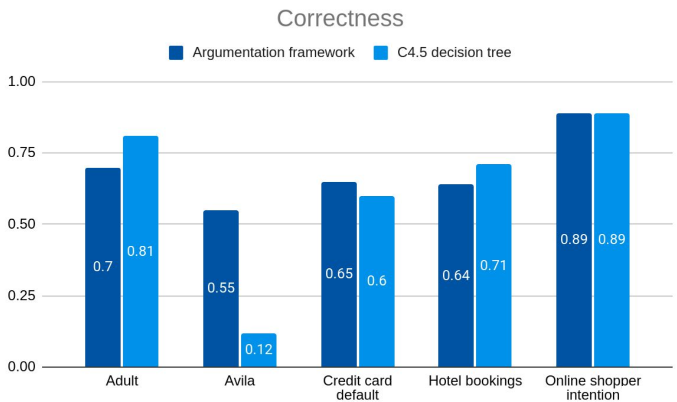
ROBUSTNESS % perturbed instances on which the predictions of model and rules remain unchanged.

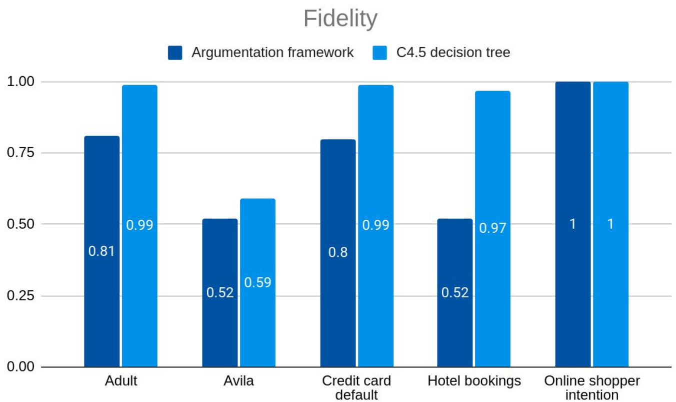
NUMBER OF RULES The cardinality of the ruleset.

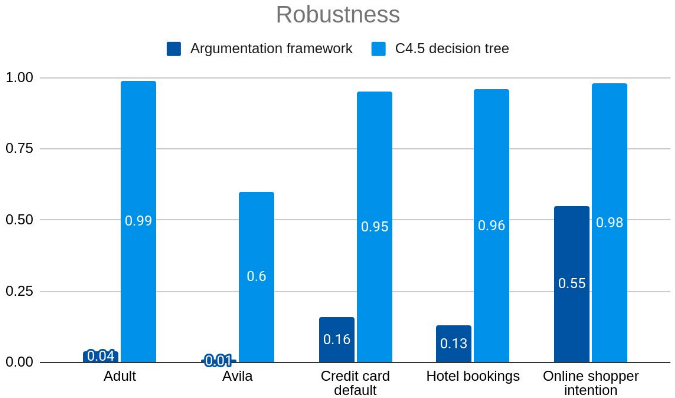
AVG RULE LENGTH The average number of antecedents of the rules.

FRACTION OF CLASSES % of output classes predicted by at least one rule.

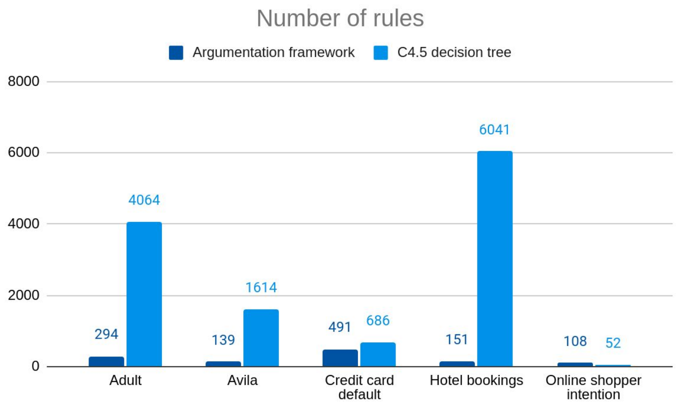
FRACTION OVERLAP The extent of overlap between each pair of rules.



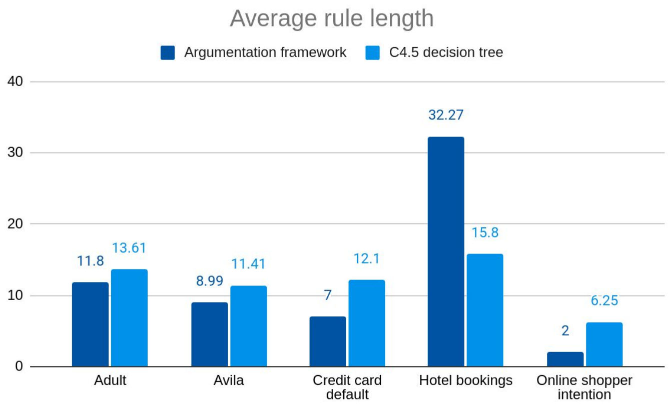




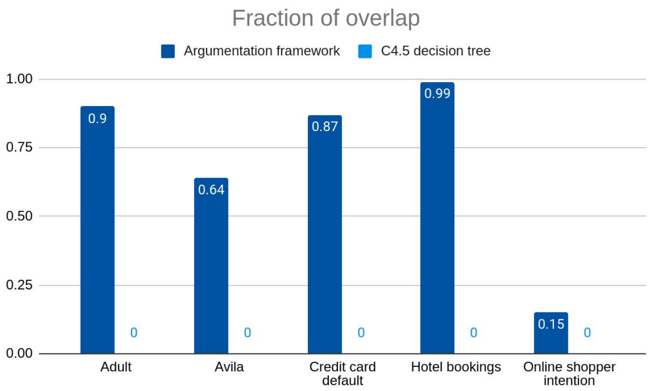
Results of the objective evaluation



Number of rules



Average rule length



Fraction of overlap

Results of the objective evaluation

Conclusions & future work

Objective evaluation

Suggested the presence of a trade-off between ruleset's size and the other metrics: the bigger the ruleset, the higher the score.

Human evaluation

Future work will include a human-centered study to be compared with the outcome of the objective metrics.

Formation of arguments

Fine-tune the inconsistency budget to obtain the optimal set of attacks and arguments, and use semantics designed for weighted argumentation frameworks.

Thanks!

Any questions?

You can find me at:

giulia.vilone@tudublin.ie

www.tudublin.ie

Instructions for use

EDIT IN GOOGLE SLIDES

Click on the button under the presentation preview that says "Use as Google Slides Theme".

You will get a copy of this document on your Google Drive and will be able to edit, add or delete slides.

You have to be signed in to your Google account.

EDIT IN POWERPOINT®

Click on the button under the presentation preview that says "Download as PowerPoint template". You will get a .pptx file that you can edit in PowerPoint.

Remember to download and install the fonts used in this presentation (you'll find the links to the font files needed in the Presentation design slide)

More info on how to use this template at www.slidescarnival.com/help-use-presentation-template

1.

Transition headline

Let’s start with the first set of slides

Quotations are commonly printed as a ***means of inspiration*** *and to invoke philosophical thoughts from the reader.*

This is a slide title

- Here you have a list of items
- And some text
- But remember not to overload your slides with content

Your audience will listen to you or read the content, but won’t do both.

Big concept

Bring the attention of your audience over a key concept using icons or illustrations

You can also split your content

White

Is the color of milk and fresh snow, the color produced by the combination of all the colors of the visible spectrum.

Black

Is the color of ebony and of outer space. It has been the symbolic color of elegance, solemnity and authority.

In two or three columns

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

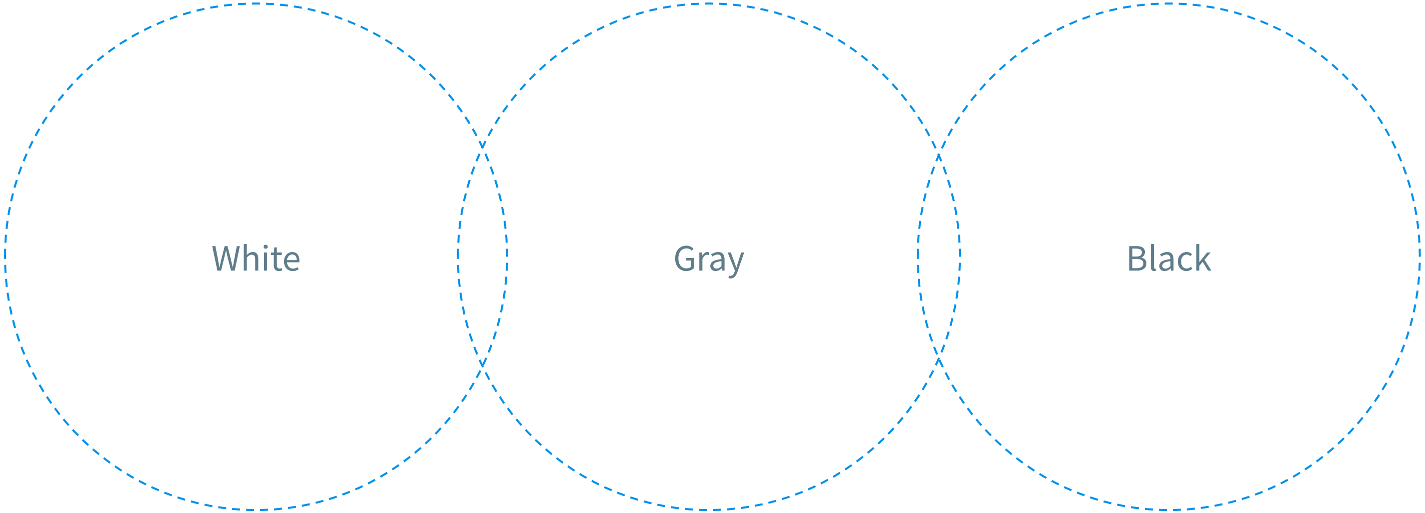
A picture is worth a thousand words

A complex idea can be conveyed with just a single still image, namely making it possible to absorb large amounts of data quickly.

Want big impact?

Use big image.

Use charts to explain your ideas



Or diagrams to explain complex ideas

Example text.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam venenatis nisi at nisl tempor, et luctus diam lobortis. Nulla sit amet metus consequat velit iaculis tempor.

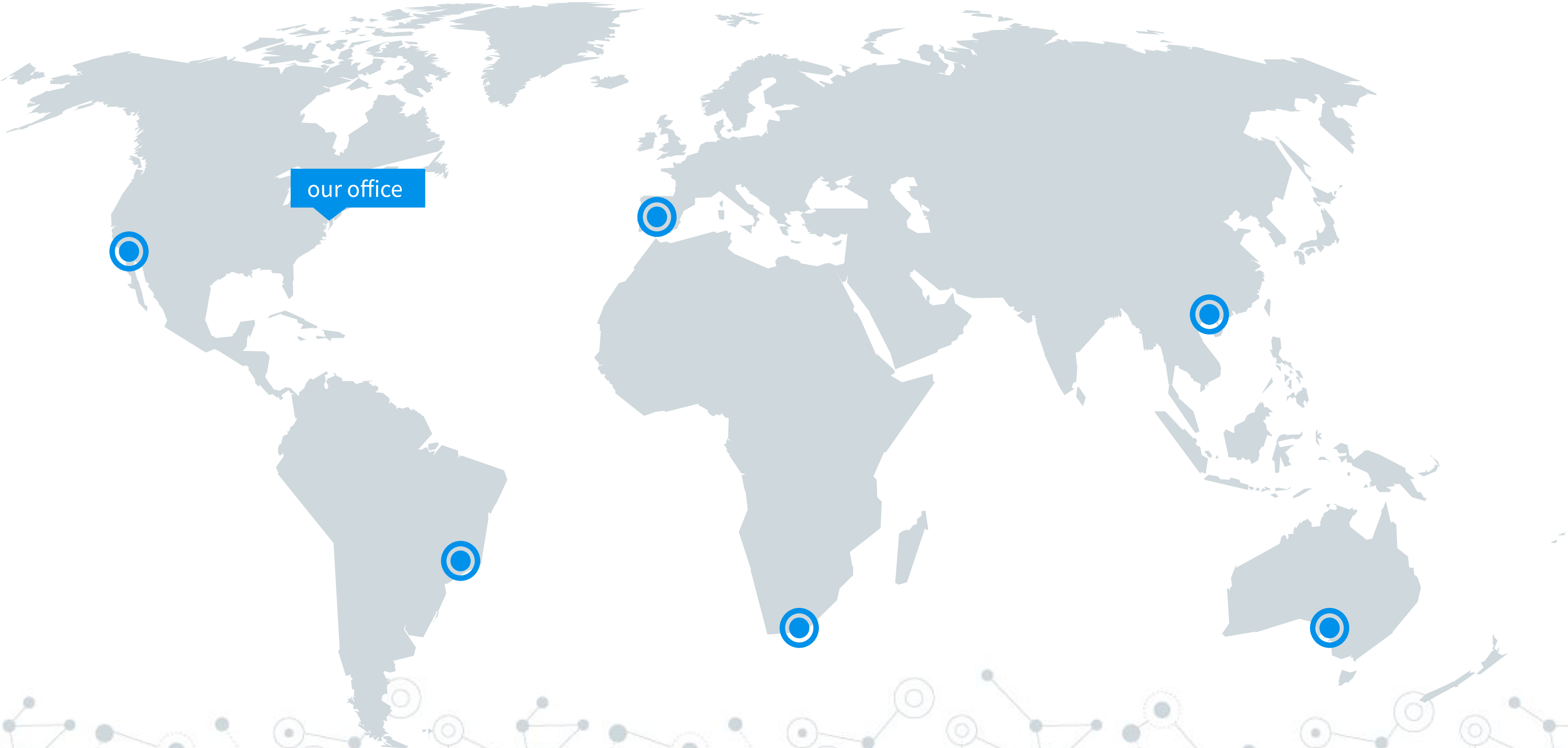
Example text.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam venenatis nisi at nisl tempor, et luctus diam lobortis. Nulla sit amet metus consequat velit iaculis tempor.

And tables to compare data

	A	B	C
Yellow	**10**	**20**	**7**
Blue	**30**	**15**	**10**
Orange	**5**	**24**	**16**

Maps



89,526,124

Whoa! That’s a big number, aren’t you proud?

Presentation design

This presentations uses the following typographies and colors:

- Titles: **Roboto Slab**
- Body copy: **Source Sans Pro**

Download for free at:

https://www.fontsquirrel.com/fonts/roboto-slab

https://www.fontsquirrel.com/fonts/source-sans-pro

You don’t need to keep this slide in your presentation. It’s only here to serve you as a design guide if you need to create new slides or download the fonts to edit the presentation in PowerPoint®

89,526,124$

That’s a lot of money

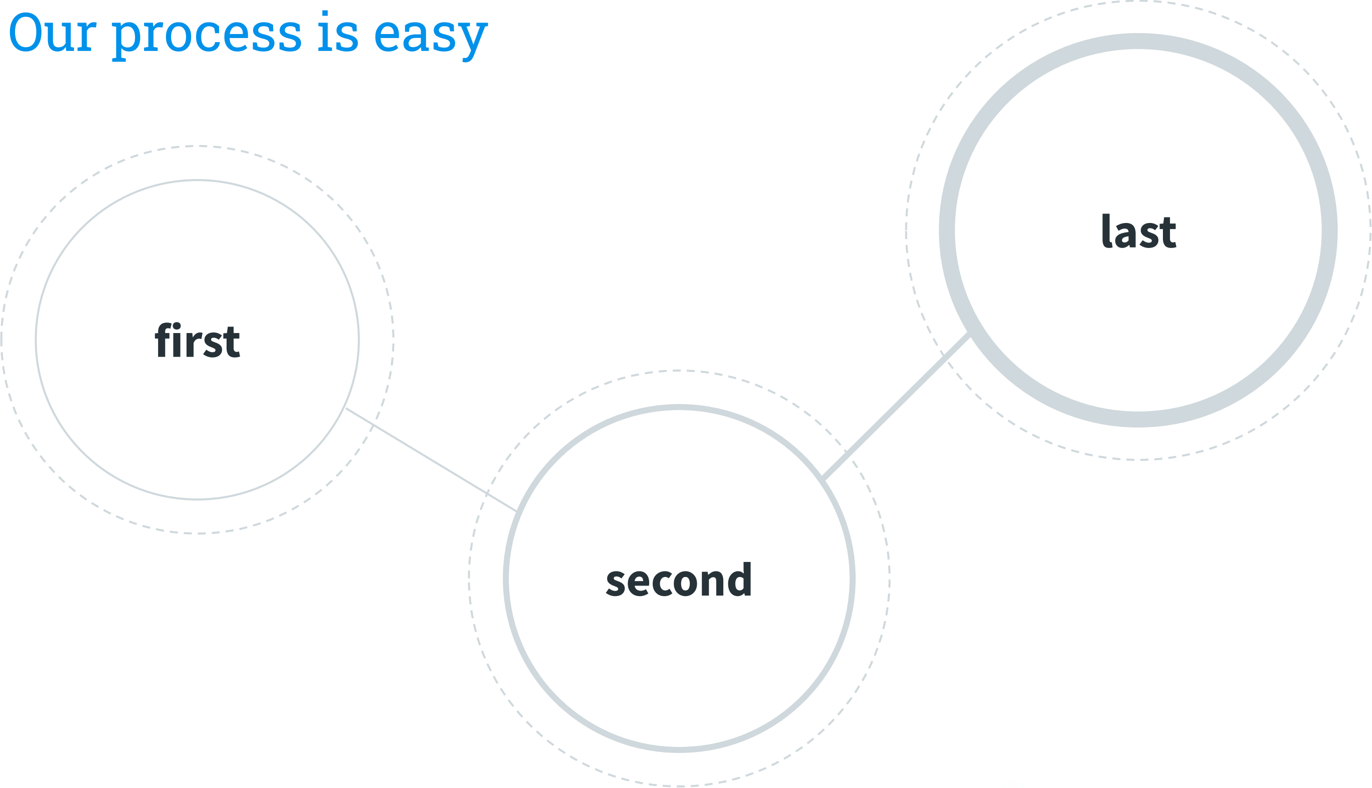
185,244 users

And a lot of users

100%

Total success!

Our process is easy



first

second

last

Let’s review some concepts

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

Yellow

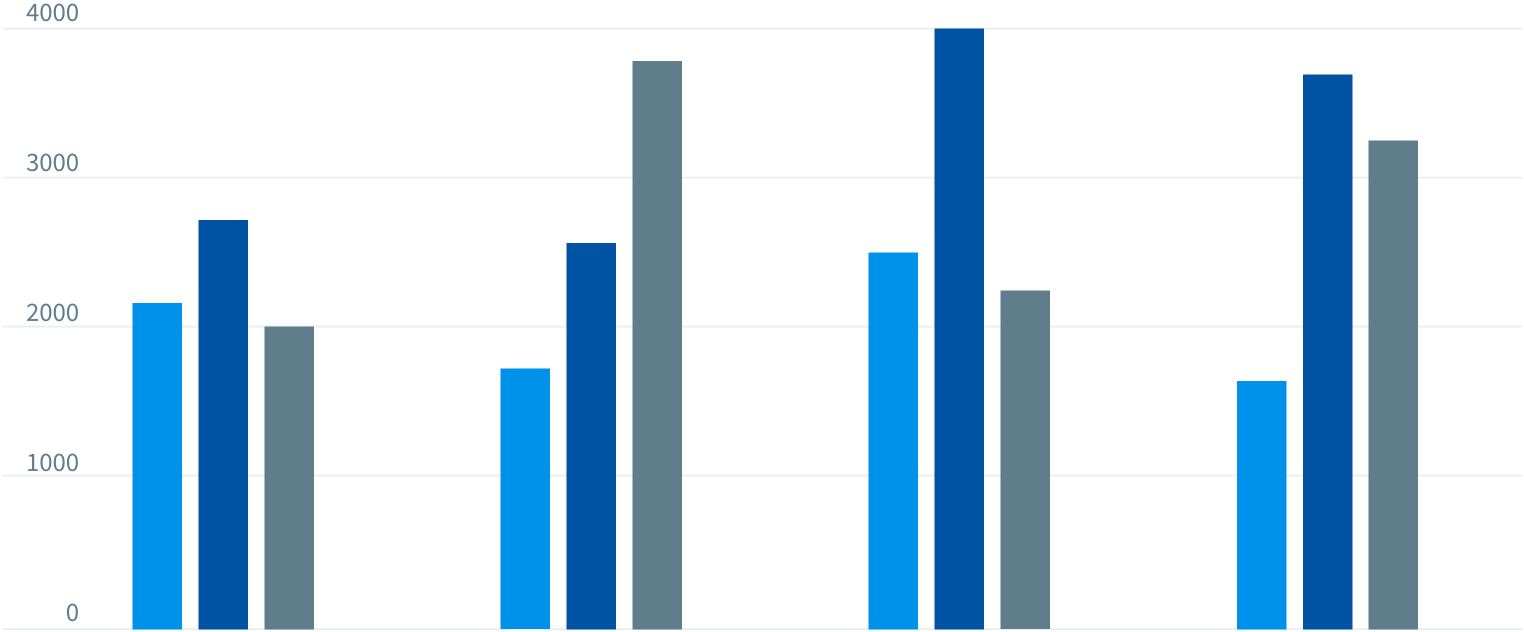
Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

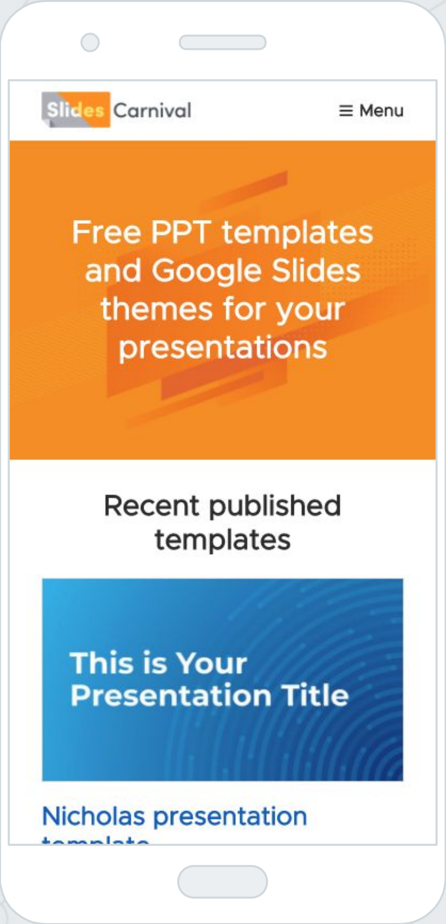
Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.



You can insert graphs from Excel or Google Sheets

Mobile project

Show and explain your web, app or software projects using these gadget templates.



Free PPT templates and Google Slides themes for your presentations

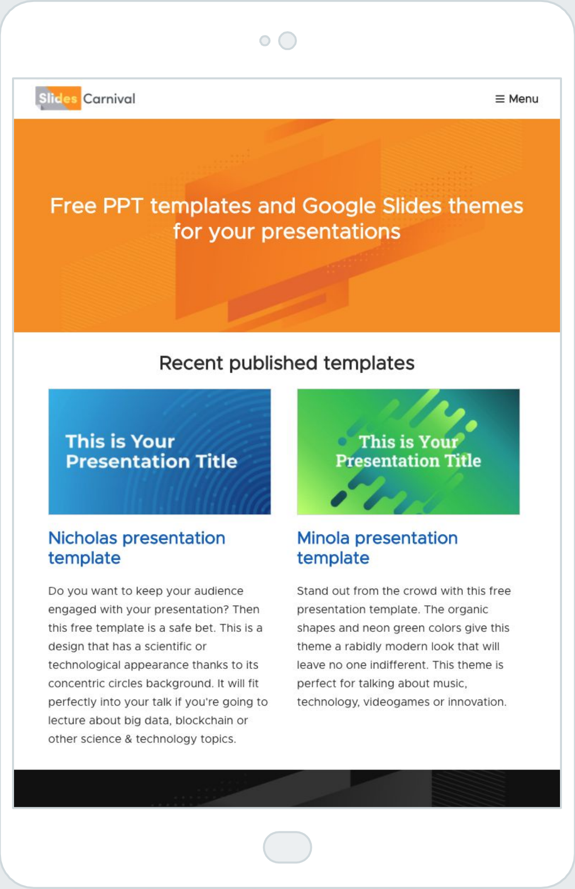
Recent published templates

This is Your Presentation Title

Nicholas presentation template

Tablet project

Show and explain your web, app or software projects using these gadget templates.



Free PPT templates and Google Slides themes for your presentations

Recent published templates

This is Your Presentation Title

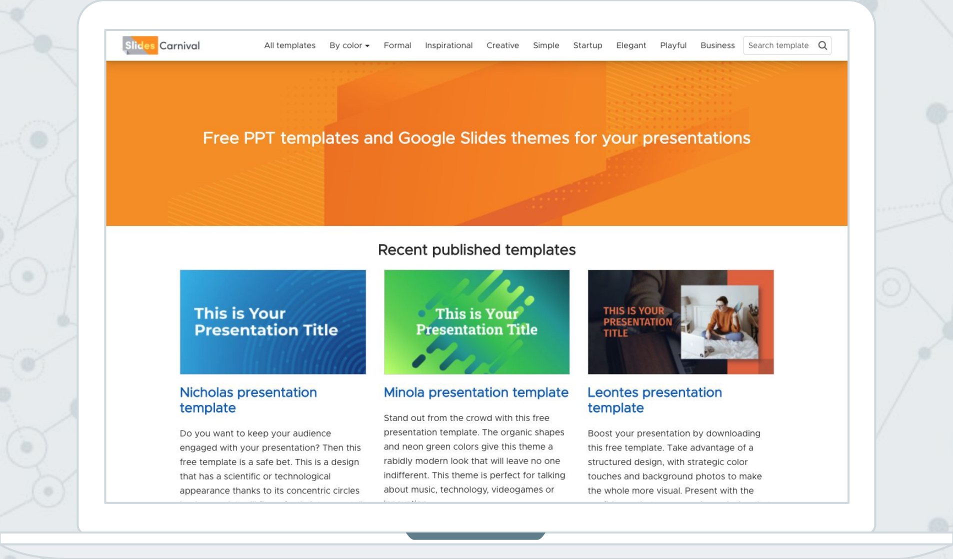
Nicholas presentation template

Do you want to keep your audience engaged with your presentation? Then this free template is a safe bet. This is a design that has a scientific or technological appearance thanks to its concentric circles background. It will fit perfectly into your talk if you're going to lecture about big data, blockchain or other science & technology topics.

This is Your Presentation Title

Minola presentation template

Stand out from the crowd with this free presentation template. The organic shapes and neon green colors give this theme a rabidly modern look that will leave no one indifferent. This theme is perfect for talking about music, technology, videogames or innovation.



Free PPT templates and Google Slides themes for your presentations

Recent published templates

Nicholas presentation template

Do you want to keep your audience engaged with your presentation? Then this free template is a safe bet. This is a design that has a scientific or technological appearance thanks to its concentric circles

Minola presentation template

Stand out from the crowd with this free presentation template. The organic shapes and neon green colors give this theme a rabidly modern look that will leave no one indifferent. This theme is perfect for talking about music, technology, videogames or

Leontes presentation template

Boost your presentation by downloading this free template. Take advantage of a structured design, with strategic color touches and background photos to make the whole more visual. Present with the

Desktop project

Show and explain your web, app or software projects using these gadget templates.

Hello!

I am Jayden Smith

I am here because I love to give presentations.

You can find me at:

@username

Credits

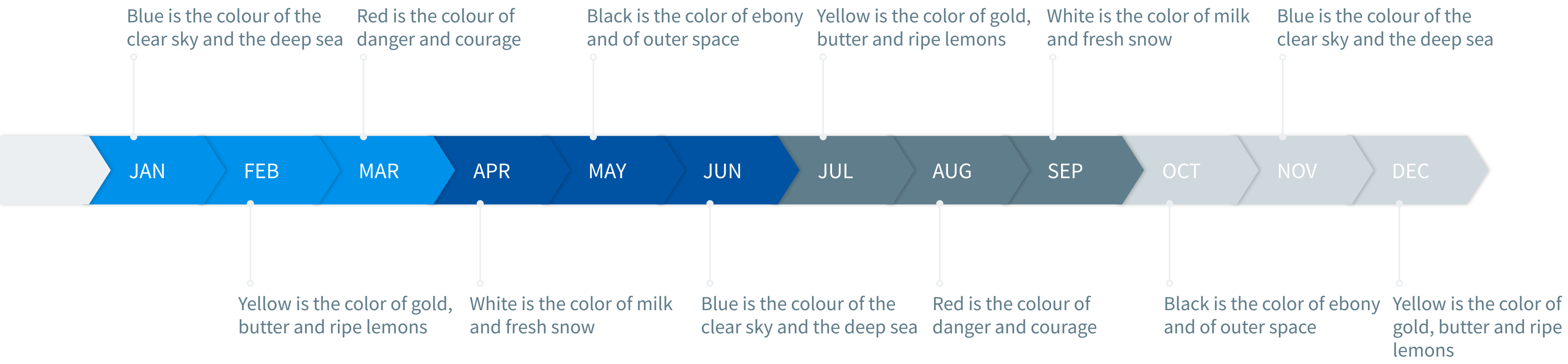
Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash

2.
Extra Resources

For Business Plans, Marketing Plans, Project Proposals, Lessons, etc

Timeline



JAN

Blue is the colour of the clear sky and the deep sea

FEB

Yellow is the color of gold, butter and ripe lemons

MAR

Red is the colour of danger and courage

APR

White is the color of milk and fresh snow

MAY

Black is the color of ebony and of outer space

JUN

Blue is the colour of the clear sky and the deep sea

JUL

Yellow is the color of gold, butter and ripe lemons

AUG

Red is the colour of danger and courage

SEP

White is the color of milk and fresh snow

OCT

Black is the color of ebony and of outer space

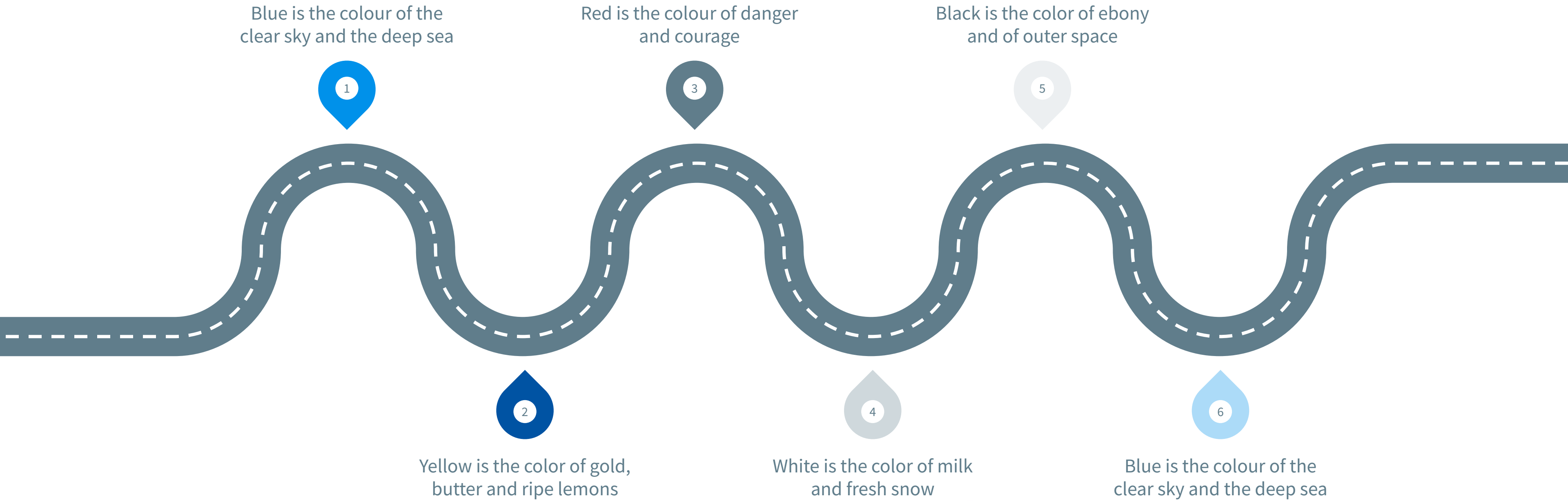
NOV

Blue is the colour of the clear sky and the deep sea

DEC

Yellow is the color of gold, butter and ripe lemons

Roadmap



Blue is the colour of the clear sky and the deep sea

Yellow is the color of gold, butter and ripe lemons

Red is the colour of danger and courage

White is the color of milk and fresh snow

Black is the color of ebony and of outer space

Blue is the colour of the clear sky and the deep sea

Gantt chart

	Week 1							Week 2						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Task 1	🟦	🟦												
Task 2		■	■	■	■	■ ◆								
Task 3		▨	▨	▨										
Task 4							🟦	🟦	🟦	🟦	🟦 ◆			
Task 5						■	■	■	■ ◆					
Task 6											▨	▨	▨	
Task 7													🟦	🟦
Task 8														■

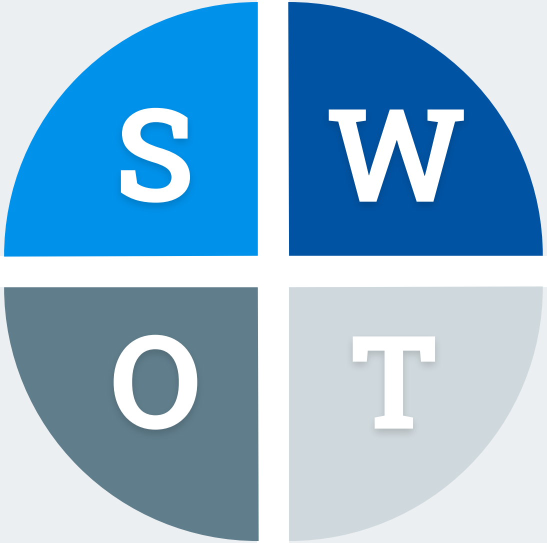
SWOT Analysis

STRENGTHS

Blue is the colour of the clear sky and the deep sea

WEAKNESSES

Yellow is the color of gold, butter and ripe lemons



Black is the color of ebony and of outer space

OPPORTUNITIES

White is the color of milk and fresh snow

THREATS

Business Model Canvas

Key Partners

Insert your content

Key Activities

Insert your content

Key Resources

Insert your content

Value Propositions

Insert your content

Customer Relationships

Insert your content

Channels

Insert your content

Customer Segments

Insert your content

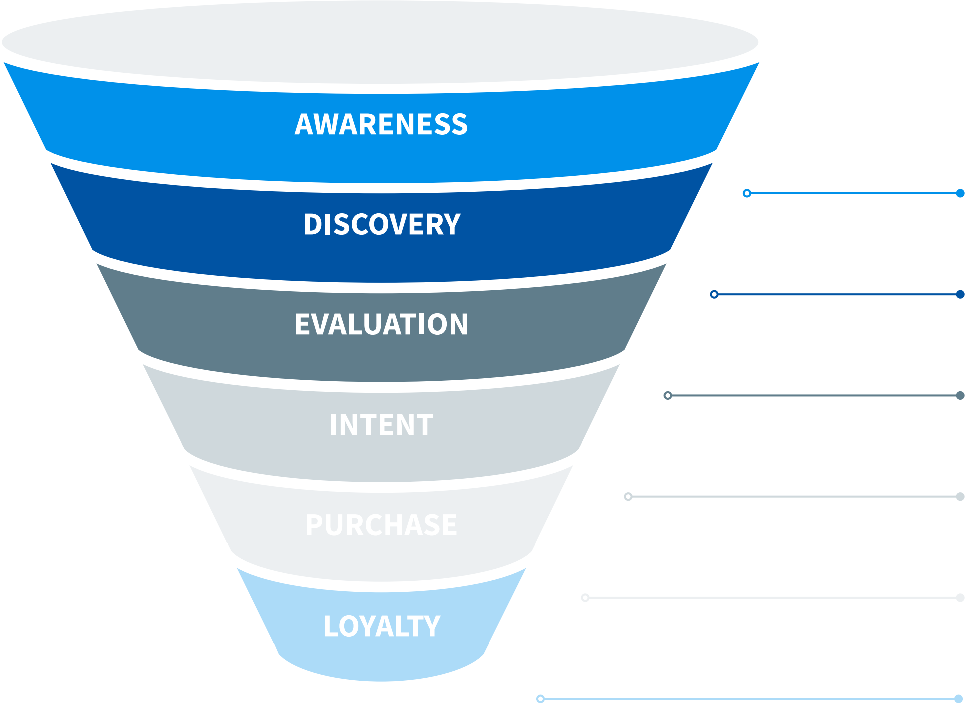
Cost Structure

Insert your content

Revenue Streams

Insert your content

Funnel



Insert your content

Insert your content

Insert your content

Insert your content

Insert your content

Insert your content

Team Presentation

Imani Jackson

JOB TITLE

Blue is the colour of the clear sky and the deep sea

Marcos Galán

JOB TITLE

Blue is the colour of the clear sky and the deep sea

Ixchel Valdía

JOB TITLE

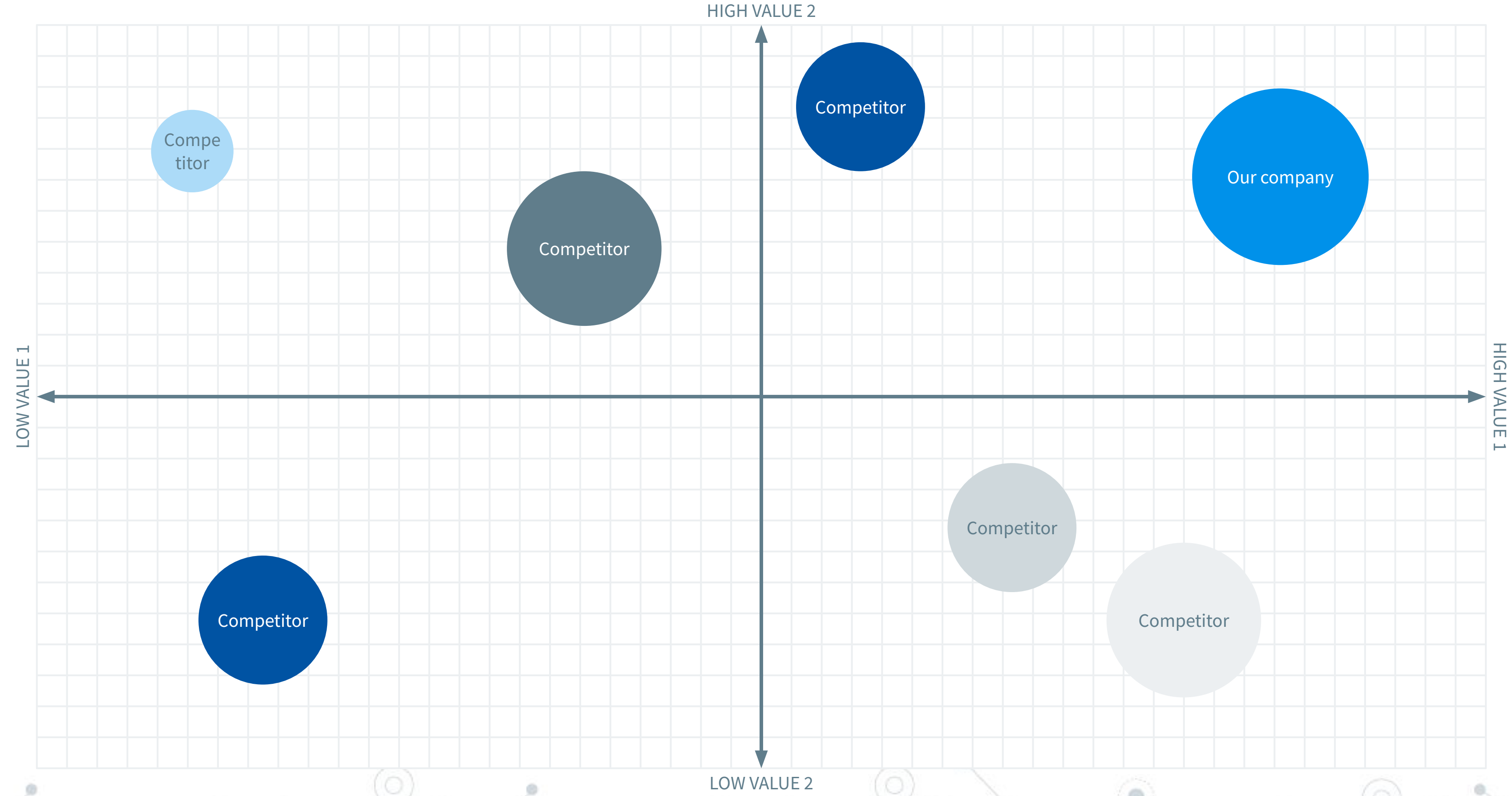
Blue is the colour of the clear sky and the deep sea

Nils Årud

JOB TITLE

Blue is the colour of the clear sky and the deep sea

Competitor Matrix



Weekly Planner

	SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
09:00 - 09:45	Task	Task	Task	Task	Task	Task	Task
10:00 - 10:45	Task	Task	Task	Task	Task	Task	Task
11:00 - 11:45	Task	Task	Task	Task	Task	Task	Task
12:00 - 13:15	✔ Free time	✔ Free time	✔ Free time	✔ Free time	✔ Free time	✔ Free time	✔ Free time
13:30 - 14:15	Task	Task	Task	Task	Task	Task	Task
14:30 - 15:15	Task	Task	Task	Task	Task	Task	Task
15:30 - 16:15	Task	Task	Task	Task	Task	Task	Task

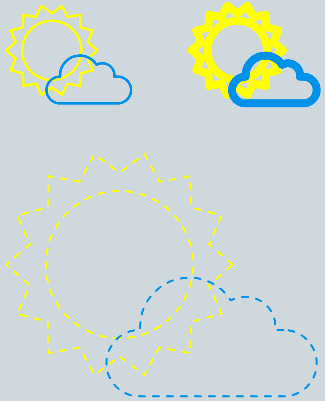
SlidesCarnival icons are editable shapes.

This means that you can:

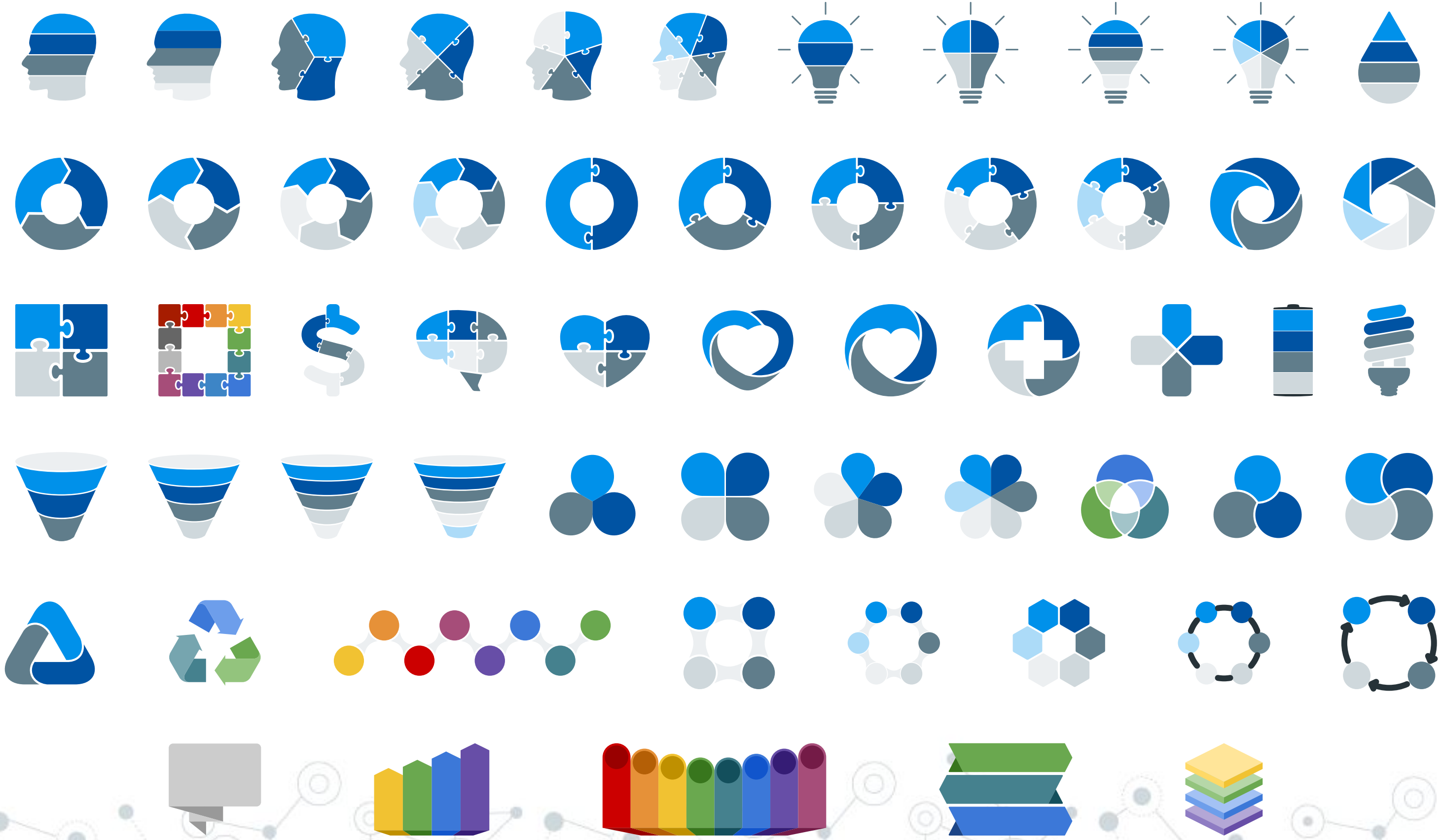
- Resize them without losing quality.
- Change line color, width and style.

Isn’t that nice? :)

Examples:



Diagrams and infographics



You can also use any emoji as an icon!
And of course it resizes without losing quality.
How? Follow Google instructions https://twitter.com/googledocs/status/730087240156643328

and many more...

Slides Carnival

Free templates for all your presentation needs

For PowerPoint and Google Slides

100% free for personal or commercial use

Ready to use, professional and customizable

Blow your audience away with attractive visuals