

Patronizing and Condescending Language Detection in English

Catherine Qingyun Wang
Vicky Xinyi Xiang
Paula Cortés Lemos



What is PCL?

- Perez-Almendros, et al. (2020):
 - “often unconscious but harmful and discriminative”
 - Language that reveals a superior attitude or harmful assumptions (especially about marginalized groups)
 -
- PCL detection
 - Similar to toxicity detection, but not overtly inflammatory

“Don’t worry, I know this is a mistake you usually make, we all make it sometimes, but I am bringing you a solution.”

“Norberto Quisumbing Jr. of the Norkis Group of Companies has a challenge for families who can spare some of what they have: why not adopt poor families and help them break the cycle of poverty?”

Task Description

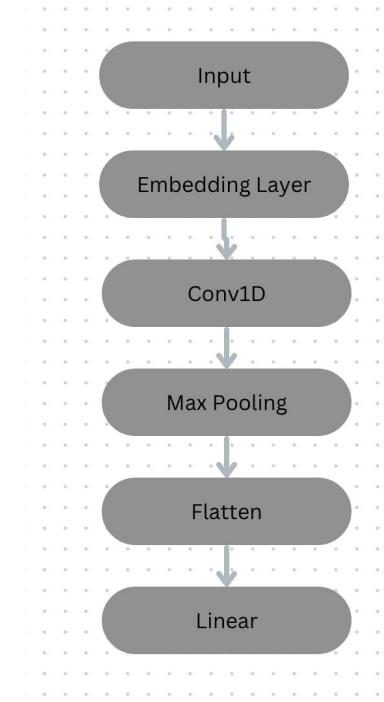
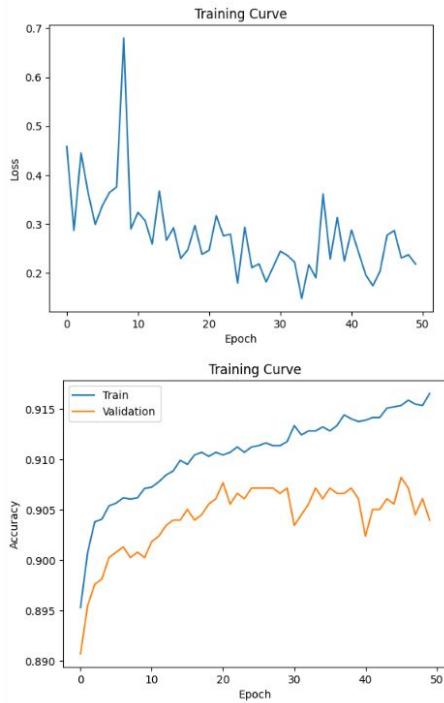
- Binary classification
- Detect presence of patronizing and condescending language in news articles
- Challenges:
 - PCL can be hard for human judges to distinguish, and often subjective
 - PCL is context-dependent, and often contains similar tokens and phrases to objective reporting

"I've also dressed up as a homeless man for a shoot once , but decided to stay like that for a few more hours after experiencing being ignored and feeling invisible. It was an eye opener."

"In our society minorities can seem invisible at times due to language barrier and discrimination."

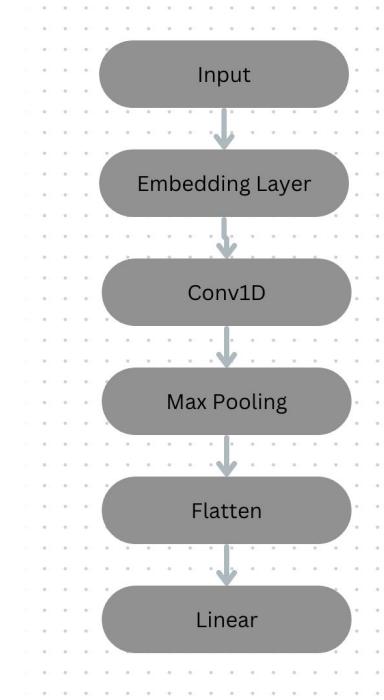
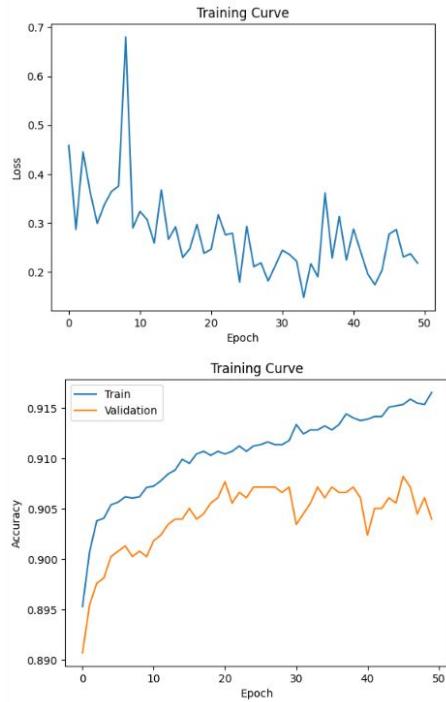
First approach: CNN with static embeddings

- Represent words with GloVe embeddings
- Input sequences represented with average of word embeddings
 - Problem: doesn't capture contextual information



First approach: CNN with static embeddings

- Represent words with GloVe embeddings
- Input sequences represented with average of word embeddings
 - Problem: doesn't capture contextual information



Second approach: BERT model

- Dynamic embeddings and transformer-based models are better at handling context-dependent information
- We choose BERT for its ability to handle bidirectional context (relationships between words) and long-distance dependencies
- Goal: fine-tune a pre-trained BERT model on task-specific data
- Specific model: **BERT-base-uncased**, pre-trained in English with 12 layers, 110M parameters

Preprocessing

- Tokenization
- Remove punctuation, emojis, and stopwords
- Insert start/stop padding tokens (and create attention masks to differentiate padding vs. non-padding)

Tokenized: ['still', 'able', 'walk', ',', 'said', 'jackson', 'therapy', ',', 'travel', 'st', 'ann', 'bay', 'per', 'week', 'proving', 'costly', 'using', 'walker', 'ho', '#bubble', 'around', 'first', ',', 'treatment', 'working', 'wonders', ',', 'due', 'financial', 'difficulties', 'disco', '#nti', '#nu', '#e', 'therapy', 'need', 'financial', 'assistance', ',', 'unable', 'work', 'farm']

Token IDs: [2145, 2583, 3328, 1010, 2056, 4027, 7242, 1010, 3604, 2358, 5754, 3016, 2566, 2733, 13946, 17047, 2478, 5232, 7570, 11362, 2105, 2034, 1010, 3949, 2551, 16278, 1010, 2349, 3361, 8190, 12532, 16778, 11231, 2063, 7242, 2342, 3361, 5375, 1010, 4039, 2147, 3888]



Training

- Optimizer: AdamW
- Linear learning rate scheduling
- 80-20 train/validation split
- Initially trained for 4 epochs (BERT model authors recommend 2-4 for transfer learning)
- After 2 epochs, training loss continuously dropped with no change in validation accuracy; switch to 2 epochs to avoid overfitting



Initial Results

Model	Precision	Recall	F1 Score
Linear NN	0.129	0.038	0.058
Conv1D NN (with undersampling)	0.084	0.343	0.135
BERT (4 epochs)	0.589	0.306	0.403
BERT (2 epochs)	0.565	0.495	0.528

Revised System: Data Augmentation

Random deletion

- Motivation → Chao et al., 2023 -> Random deletion is capable of preserving the semantic characteristics of data, as keywords normally represent just a small fraction of the sentence and are unlikely to be deleted; additionally, the data it creates is more diverse and can thus make the model more generalizable.
- More artificial sentences for the positive class (sentences with PCL).
- 1) iterating through the positive samples, 2) selecting a set of n random words for each sentence, 3) creating n new sentences per sample, each of identical to the original sample except for one deleted word.

Sample sentence: "depicts demonstrations refugees border post , catastrophic living conditions desperate attempt several hundred cross river kilometers camp get macedonia march"

Artificial sentence 1: "depicts demonstrations refugees border post , catastrophic living conditions desperate attempt several hundred cross river kilometres camp get macedonia" [march was deleted]

Initially: 9 new samples (1:1 ratio) → much worse performance.

New run: 3 new samples (2:5 ratio) → better, still under original performance

Revised System: Data Augmentation

Synonym replacement

- Motivation → Vázquez Ramos et al., 2022 -> Implemented synonym replacement for data augmentation and improved their model's performance.
- More artificial sentences for the positive class (sentences with PCL).
- 1) selecting a random word index in each sentence to replace it with a synonym, 2) selecting a random synonym (from list in WordNet, NLTK) for that word, 3) replacing it in the original sample, creating a new sentence. Repeat this process n times.

n value: 3 new samples (2:5 ratio) → better, still under original performance

Sample sentence: "least hungry
traumatised refugees sought refuge
bangladesh since october"

Artificial sentence: "least hungry shock
refugees sought refuge bangladesh since
october" [*traumatised* was replaced by
its synonym *shock*]

Revised System: Data Augmentation

Possible reasons for underperformance

Implementation of data augmentation for class balancing did not improve upon the original model.

- Dev (and test) set itself are imbalanced as well → model learning different base probabilities.
- Complexity of PCL → word deletion and synonym replacement cause loss of nuance that made them PCL.
- Subtle distinctions in synonyms → attenuation of patronizing undertones even if core semantics remain unchanged.

Future → implement more subtle data augmentation techniques (creating less artificial sentences, or use a different method), or foregoing data augmentation altogether.

Results and Discoveries

All runs are with random delete

Leakage?	Learning rate	epsilon	Number of new samples per real sample	Accuracy - epoch 1	Accuracy - epoch 2	Avg training loss - epoch 1	Avg training loss - batch 2
Yes	2e-5	1e-8	0	0.93	0.97	0.33	0.10
No	1e-4	1e-8	9	0.09	0.11	0.71	0.70
No	2e-5	1e-8	3	0.90	0.88	0.58	0.48
No	1e-4	1e-8	3	0.91	0.91	0.61	0.61
No	5e-5	1e-8	3	0.90	0.89	0.56	0.34
No	5e-5	1e-8	4	0.91	0.91	0.65	0.64



Looking Ahead

Structure | Architecture Wise

- Leverage ensemble methods
 - Combined system of BERT, Roberta, Distilbert, Roberta large, Albert
- Potentially, map from existing language sources to other languages and back

Data Wise

- Enhance strategies for class variations among data
- Consider transformation techniques that have not been adopted for this deliverable
 - Used: Synonym Replacement and Random Deletion
 - Unused: Random Insertion and Random Swap

CONTENTS OF THIS TEMPLATE

Here's what you'll find in this Slidesgo template:

1. A slide structure based on a Project Proposal, which you can easily adapt to your needs. For more info on how to edit the template, please visit [Slidesgo School](#) or read our [FAQs](#).
2. An assortment of pictures to use in the presentation can be found in the **alternative resources** slide.
3. A **credits slide**, which you must keep so that proper credits for our design are given.
4. A **resources slide**, where you'll find links to all the elements used in the template.
5. **Instructions for use**.
6. Final slides with:
 - The **fonts and colors** used in the template.
 - More **infographic resources**, whose size and color can be edited.
 - Sets of **customizable icons** of the following themes: general, business, avatar, creative process, education, help & support, medical, nature, performing arts, SEO & marketing, and teamwork.

You can delete this slide when you're done editing the presentation.

TABLE OF CONTENTS

01

ABOUT THE PROJECT

Here you could describe the topic of the section

02

MAJOR REQUIREMENTS

Here you could describe the topic of the section

03

PROJECT GOALS

Here you could describe the topic of the section

04

SNEAK PEEK

Here you could describe the topic of the section

05

PROJECT STAGES

Here you could describe the topic of the section

OUR COMPANY

Mercury is the closest planet to the Sun and the smallest one in the Solar System—it's only a bit larger than our Moon. The planet's name has nothing to do with the liquid metal since it was named after the Roman messenger god, Mercury



WHAT WE ARE WORKING ON

ROAD BRIDGES

Despite being red, Mars is a cold place, not hot. It's full of iron oxide dust, which gives the planet its reddish cast

SUBWAY STATIONS

Yes, Saturn is the ringed one. This planet is a gas giant, and it's composed mostly of hydrogen and helium

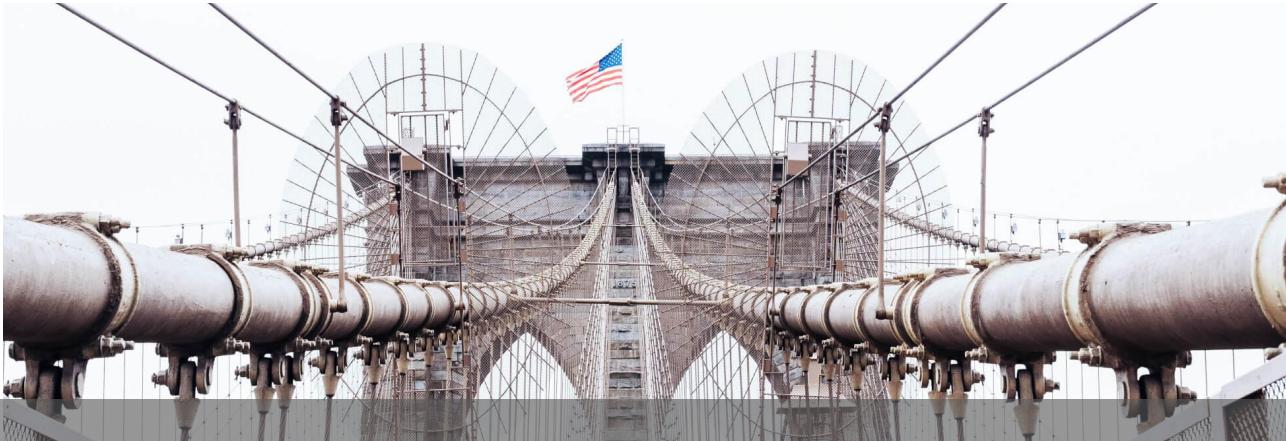
INDUSTRIAL BUILDINGS

Jupiter is a gas giant and the biggest planet in our Solar System. It's the fourth-brightest object in the sky

THERMAL POWER STATIONS

Venus has a beautiful name and is the second planet from the Sun. It's terribly hot, even hotter than Mercury

WHAT WE ARE WORKING ON



GAS PIPELINES

Yes, Saturn is the ringed one. This planet is a gas giant, and it's composed mostly of hydrogen and helium

TUNNEL WORKS

Jupiter is a gas giant and the biggest planet in our Solar System. It's the fourth-brightest object in the sky

WATER TREATMENT PLANTS

Despite being red, Mars is a cold place, not hot. It's full of iron oxide dust, which gives the planet its reddish cast



ABOUT THE PROJECT

Mercury is the closest planet to the Sun and the smallest one in the Solar System—it's only a bit larger than our Moon. The planet's name has nothing to do with the liquid metal since it was named after the Roman messenger god, Mercury



"Here comes the quote. Words full of wisdom that someone important said and can make the reader get inspired."

—SOMEONE FAMOUS

NOW



Mercury is the closest planet to the Sun and the smallest one in our Solar System—it's only a bit larger than our Moon. The planet's name has nothing to do with the liquid metal, since it was named after the Roman messenger god

FUTURE





AWESOME
WORDS

MAJOR REQUIREMENTS



MAJOR REQUIREMENTS



MARS

Despite being red, Mars is a cold place, not hot. It's full of iron oxide dust



JUPITER

It's a gas giant, the biggest planet in our Solar System and the fourth-brightest object in the sky



SATURN

Yes, this is the ringed one. It's a gas giant, composed of hydrogen and helium



MERCURY

Mercury is actually the smallest planet in our Solar System



NEPTUNE

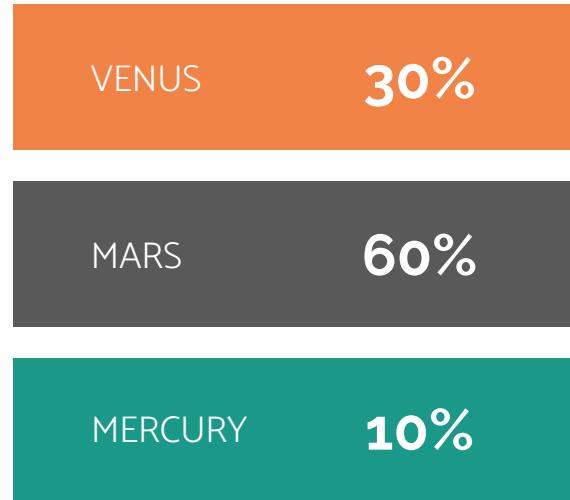
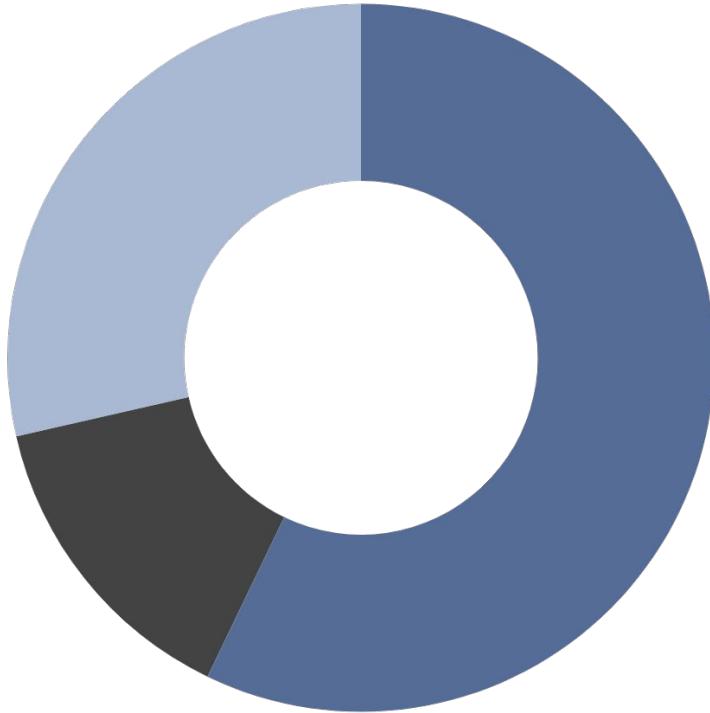
Neptune is the fourth-largest planet in our Solar System and the farthest from the Sun



VENUS

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

BUDGET



If you want to modify this graph, click on it, follow the link, change the data and replace it

OUR NUMBERS



PROJECT GOALS



GOAL 1

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

GOAL 2

Venus has a beautiful name and is the second planet from the Sun. It's terribly hot, even hotter than Mercury

PREDICTED RESULTS

	VENUS	MARS	MERCURY	JUPITER
AREA 1	50	100	20	30
AREA 2	23	34	45	56
AREA 3	234	45	65	56

SNEAK PEEK

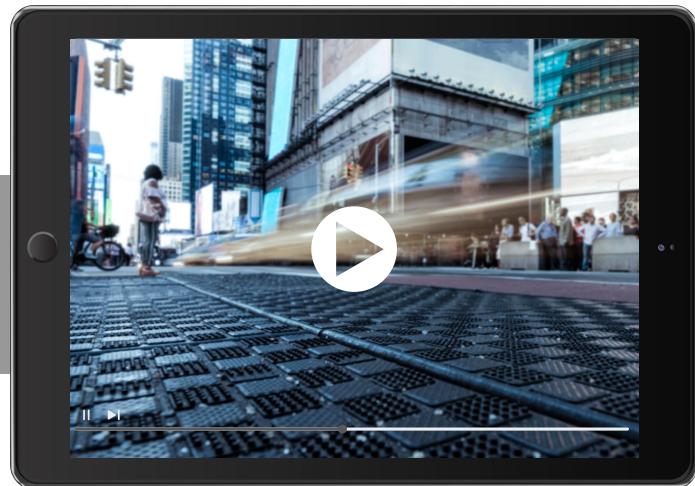


Mercury is the closest planet to the Sun and the smallest one in the Solar System—it's only a bit larger than our Moon. The planet's name has nothing to do with the liquid metal, since it was named after the Roman messenger god, Mercury

SNEAK
PEEK



Insert your
multimedia
content here



PROJECT STAGES

STAGE 1

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

STAGE 2

Mercury is the closest planet to the Sun and the smallest one in the Solar System

STAGE 3

Despite being red, Mars is a cold place, not hot. It's full of iron oxide dust, giving the planet its reddish cast

FUTURE PROJECTS

MERCURY

Mercury is the closest planet to the Sun and the smallest one in the Solar System

MARS

Despite being red, Mars is a cold place, not hot. It's full of iron oxide dust



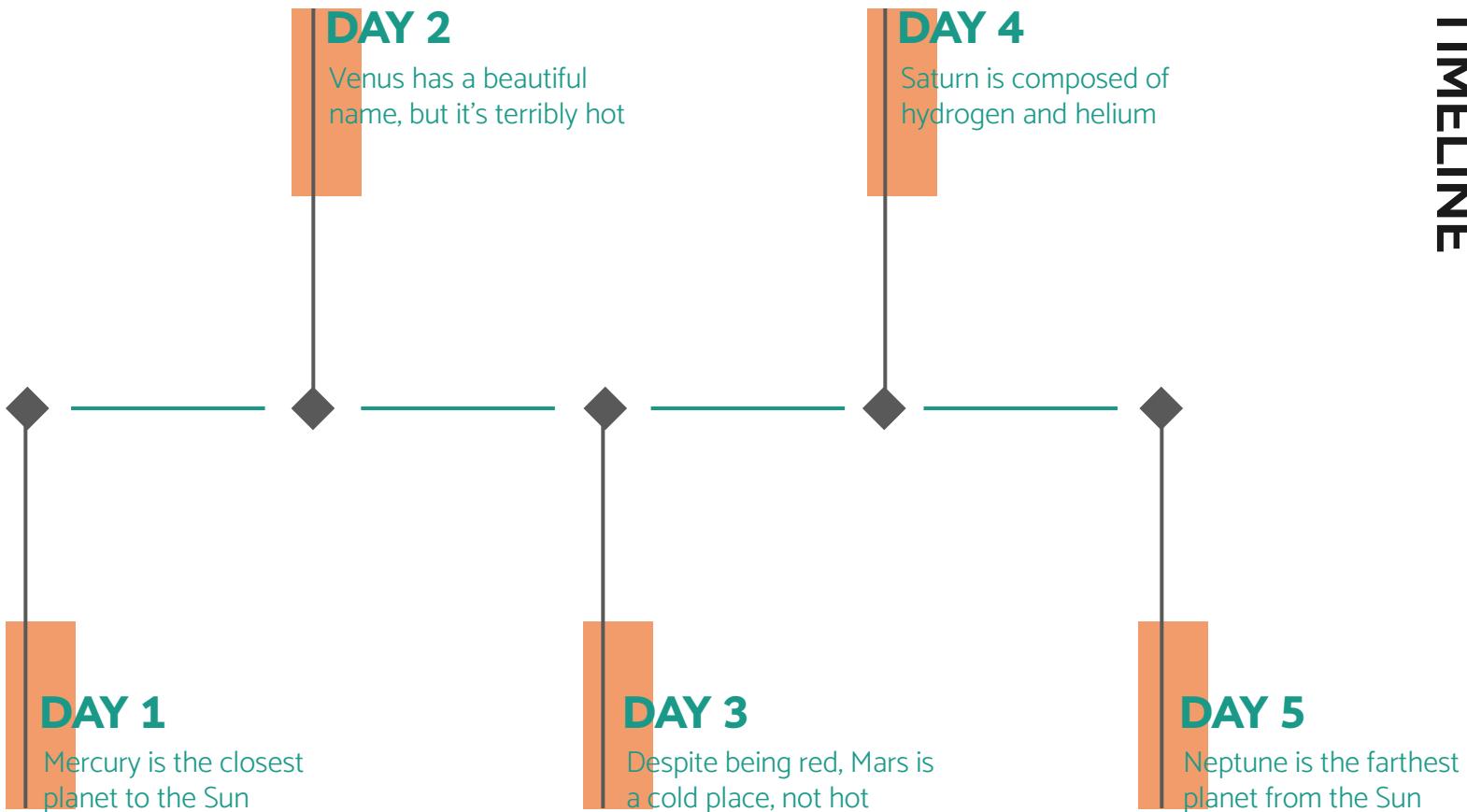
Venus has a beautiful name and is the second planet from the Sun

VENUS

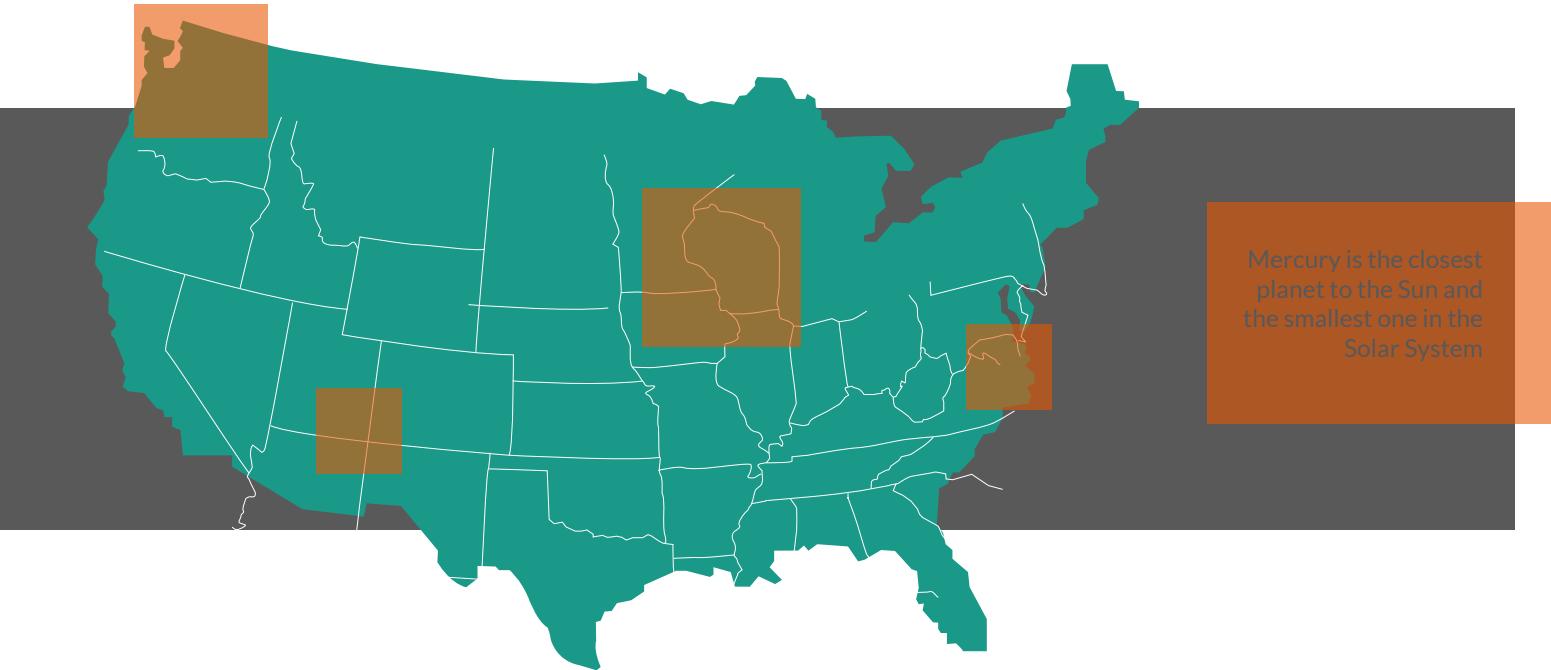
Saturn is the ringed planet. It's a gas giant, composed mostly of hydrogen and helium

SATURN

TIMELINE



OUR LOCATIONS



OUR PARTNERS



MERCURY

Mercury is the closest planet to the Sun and the smallest one in our Solar System—it's only a bit larger than our Moon. The planet's name has nothing to do with the liquid metal, since it was named after the Roman messenger god

VENUS



Venus has a beautiful name and is the second planet from the Sun. It's terribly hot—even hotter than Mercury—and its atmosphere is extremely poisonous. It's the second-brightest natural object in the night sky

OUR TEAM



You can replace
the image on the
screen with your
own



You can replace
the image on the
screen with your
own

THANKS

Does anyone have any questions?
Follow the project updates

addyouremail@freepik.com
+91 620 421 838
yourcompany.com



ALTERNATIVE RESOURCES

Here's an assortment of alternative pictures within the same style of this template.

- Airplane window with sky and sea landscape
- Railway train station long shot
- City skyline at night
- Sunset at beach
- Old bridge over the river near alesund, norway
- Tropical road with desert background
- Mountain range with snow
- Harbor with boats and blue sea
- City roads at night
- Brooklyn bridge and new york

If you want to apply the same filter as the pictures already included in the template, follow these instructions.

In Photoshop:

1. Open the picture in Photoshop.
2. On the Layers panel, click the Create New Adjustment Layer button.
3. Select Photo Filter and click Color.
4. Paste this hex code: #908269
5. Adjust the density to increase or decrease the filter effect.

In Google Slides / PowerPoint:

1. Create a new shape with the same shape as the photo to which you want to apply the filter.
2. Select the color for the filter (#908269) and reduce the opacity until you're happy with the result.
3. If you want to modify the photo, move the filter to a side first.



CREDITS

This is where you give credit to the ones who are part of this pr

- Presentation template by [Slidesgo](#)
- Icons by [Flaticon](#)
- Infographics by [Freepik](#)
- Images created by [Freepik](#)

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.
- 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.

For more information about editing slides, please read our FAQs or visit Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

Instructions for use (premium users)

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

You are allowed to:

- Modify this template.
- Use it for both personal and commercial purposes.
- Share this template with your team and add your name 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 another and separately existing 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>

Fonts & colors used

This presentation has been made using the following fonts:

Livvic

(<https://fonts.google.com/specimen/Livvic>)

Catamaran

(<https://fonts.google.com/specimen/Catamaran>)

#434343

#556d96



Use our editable graphic resources...

You can easily **resize** these resources, keeping the 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. Don't forget to group the resource again when you're done.



...and our set of editable icons

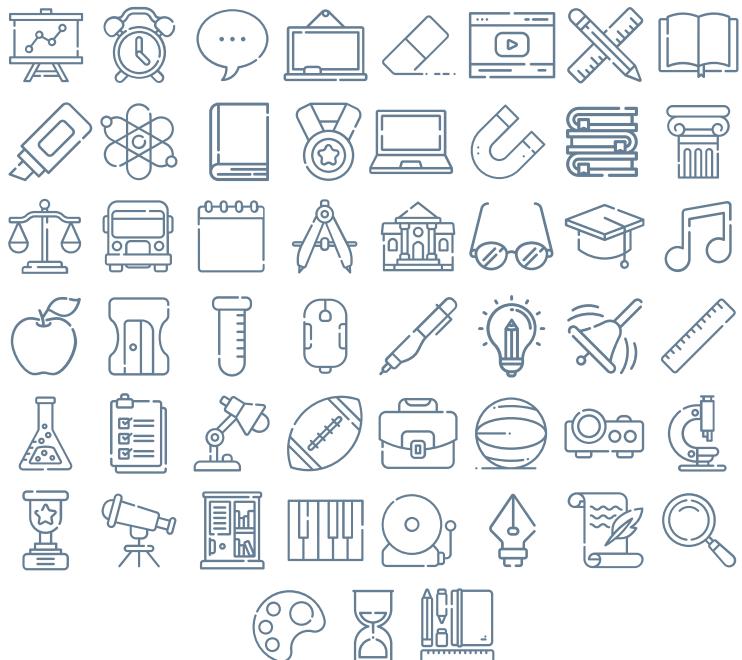
You can resize these icons, keeping the quality.

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



Performing Arts



Nature Icons



SEO & Marketing Icons



