

# AI FOR GOOD Summer Camp



INDIANA UNIVERSITY  
BLOOMINGTON

# Day 3: NLP



# Webpage link

<https://bit.ly/iuaigood>





# What does NLP stand for?



Natural



Language



Processing





# What is NLP?

When you think of NLP, who or what do you think of?

- NLP is the branch of AI that helps computers detect, understand, interpret, and manipulate **human language** in ways that are context aware.





# But... How does it work?

- Different forms of analysis based on...
  - Data (example: words)
  - Structures (example: grammar)
  - Meaning
- What can this mean?
  - “It’s only 68 degrees in here”



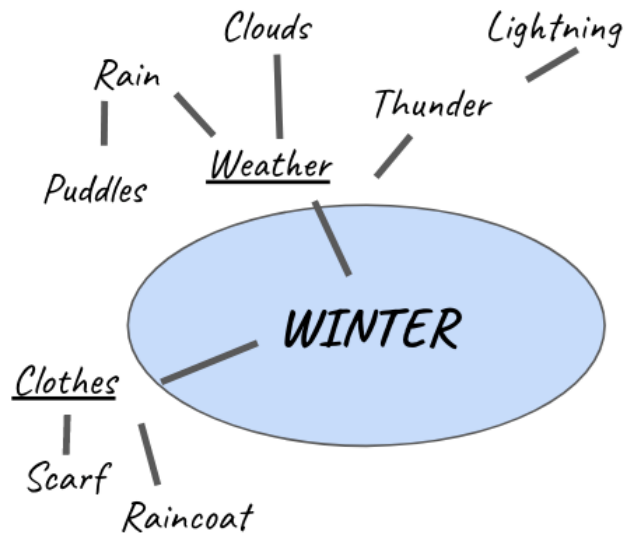


# But... How does it work?

- Language is Data
- Word association
  - the relationship between sentences, words or signs and their meaning

## Winter

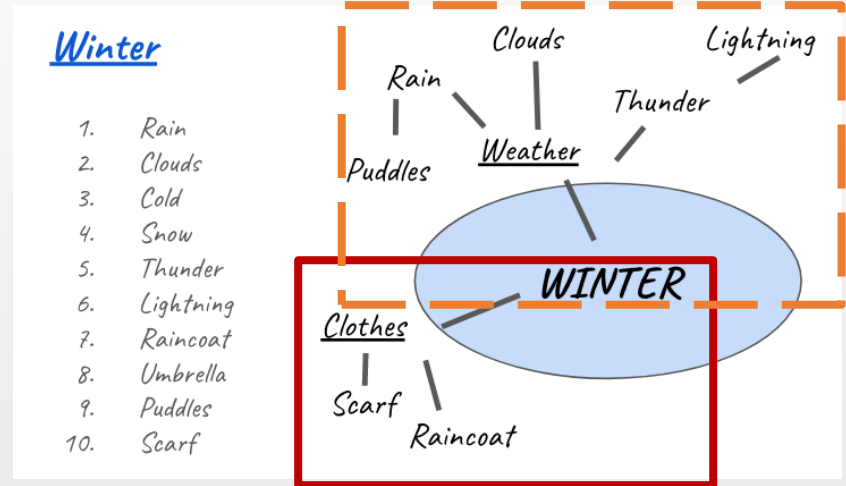
1. Rain
2. Clouds
3. Cold
4. Snow
5. Thunder
6. Lightning
7. Raincoat
8. Umbrella
9. Puddles
10. Scarf





# Let's try it!

- Talk about your target word and the different meanings
- After you receive your word, begin building your word association map
- **Make a cluster:**
  - Draw lines to link words together if they are related
- **Border your clusters:**
  - When you are finished, draw a shape around clusters that are not related to other clusters









# The point...

- Each cluster has a boundary
- Each boundary has been defined by us
- Boundaries can change. *But...how?*
  - Context (restaurant versus school cafeteria)
  - Location (school versus home)
  - Culture (Korea versus US)
  - Trends (1990 versus 2023)
  - Audience (middle schooler versus old timer)
- [Fun fact: this is how algorithms are made!]





**Break Time! :D**





# Going deeper...

[some] techniques to train a model

- Cluster analysis
- Keyword extraction
- Sentiment analysis

is used to

[some] uses of the models we trained

- Make Word Predictions to Generate Text Responses (ChatGPT; autocomplete)
- Determine Meanings of Text (language translation)





# How does autocomplete work for Krista?

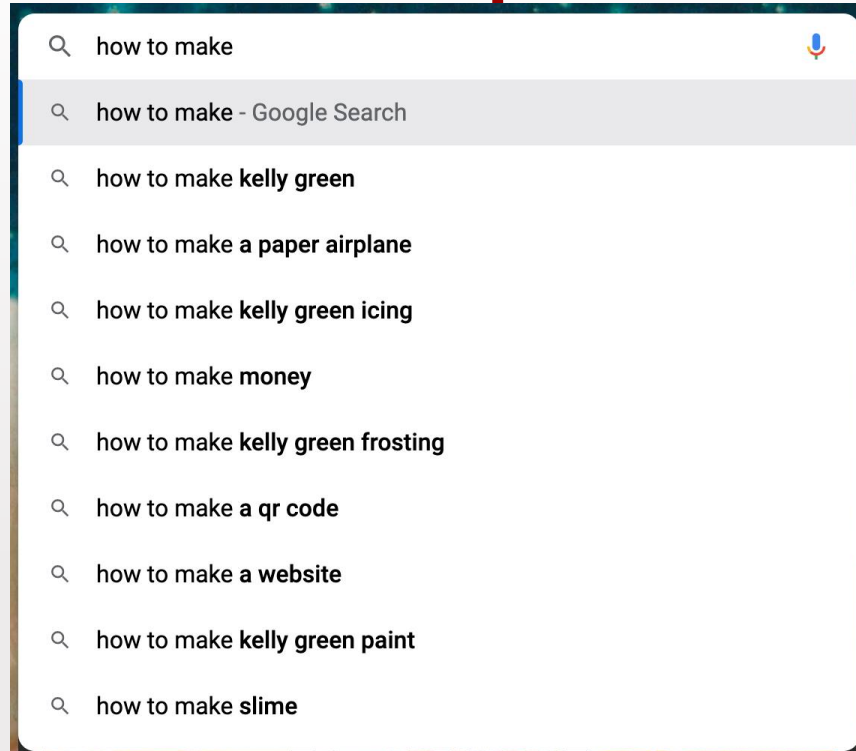
how to make

- how to make **money**
- how to make **friends**
- how to make **water run uphill**
- how to make **a decision**
- how to make **a questionnaire**
- how to make **our ideas clear**
- how to make **telehealth app**
- how to make **strategic alliances work**
- how to make **soap**
- how to make **a good presentation**



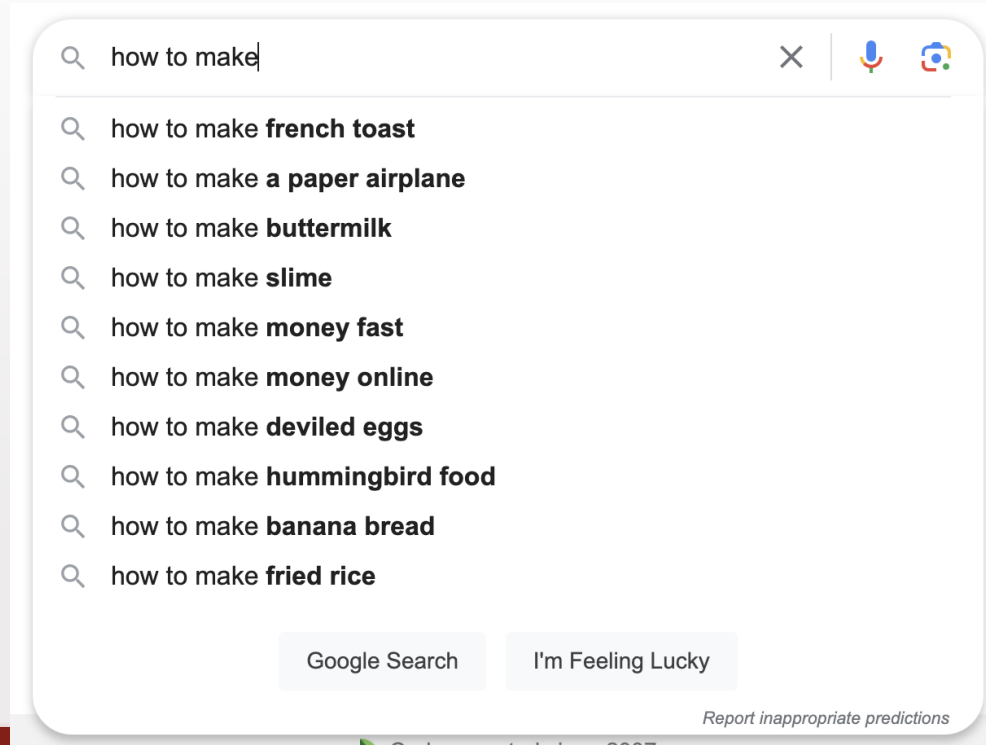


# How does autocomplete work for Sara?





# How does autocomplete work for Haesol?





# What's the algorithm paying attention to?

16

- Google profile
- Social media activity
- Recent searches
- Search history
- Location / zip code
- Date / time
- Other immediate searches in your proximity
- Browser cookies
- Other....?











# Language Translation

Korean ▼ ↔ English ▼

이거 얼마예요? 화장실이 어디예요? ×

igeo eolmayeyo? hwajangsil-i eodiyeyo?

How much is it?  
Where is the restroom?





# What can go wrong?

Korean ▼ ↔ English ▼

개는 엄마다. 개는  
의사야. 개는 변호사  
다. 개는 교수다. 개  
는 간호사다.

×

she is a mother he's  
a doctor he is a  
lawyer he is a  
professor she is a  
nurse

gyaeneun eommada.  
gyaeneun uisaya. gyaeneun  
byeonhosada. gyaeneun  
gyosuda. gyaeneun  
ganhosada.



# Bias in AI systems can happen...





**Break Time! :D**



# What do you think about social media?

Tell our middle school "AI for Good" campers what you think! Your responses are anonymous, and will be used to build data models about public perceptions.

glaze@iu.edu [Switch account](#)



Not shared

\* Indicates required question

How old are you? \*

- 18 years old or less
- Over 18 years old

What is your gender identification? \*

- Male
- Female
- Non-binary
- Other: \_\_\_\_\_

What are some positive things about using social media? \*

Your answer

What are some negative things about using social media? \*

Your answer

allery

Over 18 years old	Female	More news that you may not have access to elsewhere.	Used to show news about people, places, or things.
Over 18 years old	Female	Lost and found fir pets	Time suck, too tied to self-worth, arguments with strangers, selfies... Omg, I hate that them so much, especially if they have their tongue hanging out 🤩, etc etc etc... I'm hoping it all goes "out of style" soon.
Over 18 years old	Male	Staying connected with friends who live far away. Joining and connect	Misinformation that social media companies refuse to block. Targeted adve
Over 18 years old	Female	Keeping up with friends and family that live far away	Very time consuming
Over 18 years old	Male	Connections	Little control
Over 18 years old	Female	Connections to other people and a chance to show others what is in	Comparisons to others. Bullying. Bad actors / people who lie and want to n
Over 18 years old	Female	Sharing, communication, convenient for life	Bullying
Over 18 years old	Male	A lot of information available. A good point of connection with people	A lot of misinformation and disinformation can be disseminated rapidly and
18 years old or less	Male	It can connect people and let u see funny content.	Social media can spread negative things and rumors.
18 years old or less	Male	Social media can teach a lot of things for people and you can find a	There is a lot of mean and creepy people there
18 years old or less	Male	you ca talk with friends and see memories and photos	there is a lot of racism and sexism on it
18 years old or less	Male	you can chat with your friends	cyber bullying
18 years old or less	Female	It allows you to meet new people and keep in touch with your friend	It can become obsessive an toxic.
18 years old or less	Male	vids	drama
18 years old or less	Male	Being able to contact people anywhere that has reception.	Addiction, phishing, anonymous people, sensitive data being leaked.
18 years old or less	Male	Connecting with others around the world.	Trolls on the internet.
18 years old or less	Male	you can find lots of cool stuff, like Emil Reinert, a piano player	could be lots of sus stuff, like profanity
18 years old or less	Female	You can connect with others all around the world.	Some people might see something that might not be appropriate for their
18 years old or less	Male	you can catch up on the news.	you can be told false things.



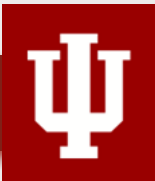
# Keyword Extraction & Sentiment Analysis

What is keyword extraction?

- **Keywords extraction:** finding important words in a text that helps to identify main topic and themes.
- **Sentiment analysis:** NLP can help computers understand how people are feeling.

Negative	Bullying	Community
New	Family	Events
Bad	Addiction	Meet

Sharing	Connected	Contact
Family	Friends	Social
Time	Bullying	New





# Sentiment Analysis

No.	Responses	Emotion (Positive, Negative, Neutral)
1	Use for crimes such as stalking and child abuse, addiction, negative peer pressure	
2	Social media is a negative tool because it can promote negative behaviors such as narrow-minded thinking, bullying, and singular view points. If you are only exposed to certain groups online it is possible to only see information from one perspective. Social media can also be used to skew ideas about people, places, or things.	
3	I feel connected to my community. I can learn what events are happening. However, I am worried if my private information might be shared in the social media. When I share my baby's picture, that means anyone can see and edit the photo?	
4	You can talk to your friends, play games, meet new people	
5	Keeping connected with family & friends.	
6	Comparisons to others. Bullying. Bad actors / people who lie and want to manipulate others. Trolls. Racism.	
7	Extensive exposure to unverified news and addiction for people who aren't able to control its use.	
8	You can socialize and meet people you, otherwise would not have known.	
9	Stalking others profiles, hacking accounts, sharing pictures that should not be.	
	Social media is a positive tool because it allows you to connect with family and friends if you don't see	





# Another word about ethics..



# LET'S EXPLORE NLP TECHNIQUES!

The screenshot displays the NLP4Science web application interface. On the left is a dark sidebar with a user profile for 'Hyejeong Lee' and navigation links for 'Home', 'Workflows', and 'NLP Models'. The main content area features a search bar, a 'Home' header, and a large central illustration of two human heads facing each other, filled with various icons representing AI and data science. To the right, a 'Welcome to NLP4Science' message explains the platform's purpose and includes a green button labeled 'Click Here To Download Demo Dataset!'.

**Home**

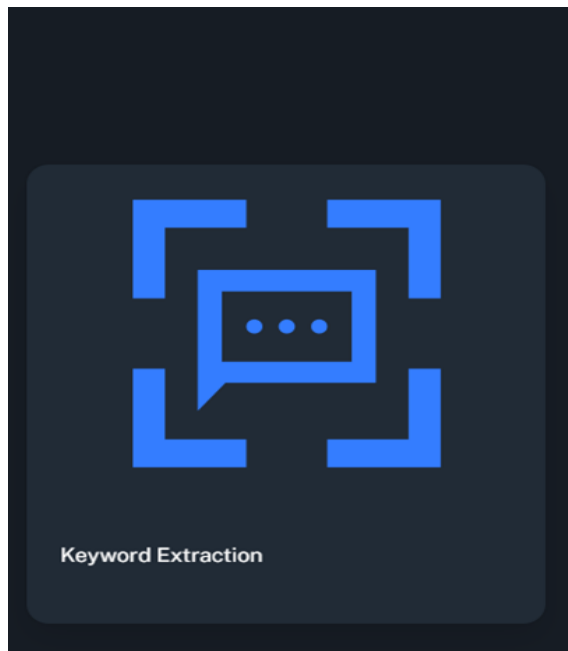
Home

**Welcome to NLP4Science**

NLP4Science is a platform for the development of Natural Language Processing (NLP) tools for the scientific community. NLP4Science makes it easy to create, share, and use NLP tools for scientific text.

[Click Here To Download Demo Dataset!](#)

# LET'S EXPLORE KEYWORD EXTRACTION




<https://app.nlp4science.com/>

[bit.ly/nlp4science](https://bit.ly/nlp4science)



# STEP 1 - REGISTER


**Get started absolutely free.** 

Free forever. No credit card needed.

First name

Last name

Email address

Password  

**Register**

By registering, I agree to Minimal [Terms of Service](#) and [Privacy Policy](#).

# STEP 2 - DOWNLOAD DEMO DATASET

The image shows a web browser window displaying the NLP4Science dashboard. The browser's address bar shows the URL `app.nlp4science.com/dashboard/app`. The dashboard has a dark theme and a sidebar on the left with navigation options: 'Home', 'Workflows', and 'NLP Models'. The 'Home' option is highlighted, and a red arrow points to it from the text 'Home' written in orange below the sidebar. The main content area is titled 'Home' and features a large illustration of two human heads in profile, facing each other, with various icons representing NLP and data science. To the right of the illustration is a 'Welcome to NLP4Science' message. Below the message is a yellow button with the text 'Click Here To Download Demo Dataset!'. A red arrow points from the text 'Click here to download demo datasets' (written in orange below the button) to the button.

app.nlp4science.com/dashboard/app

H Hyejeong Lee

Home

Workflows

NLP Models

Home

Home

Welcome to NLP4Science

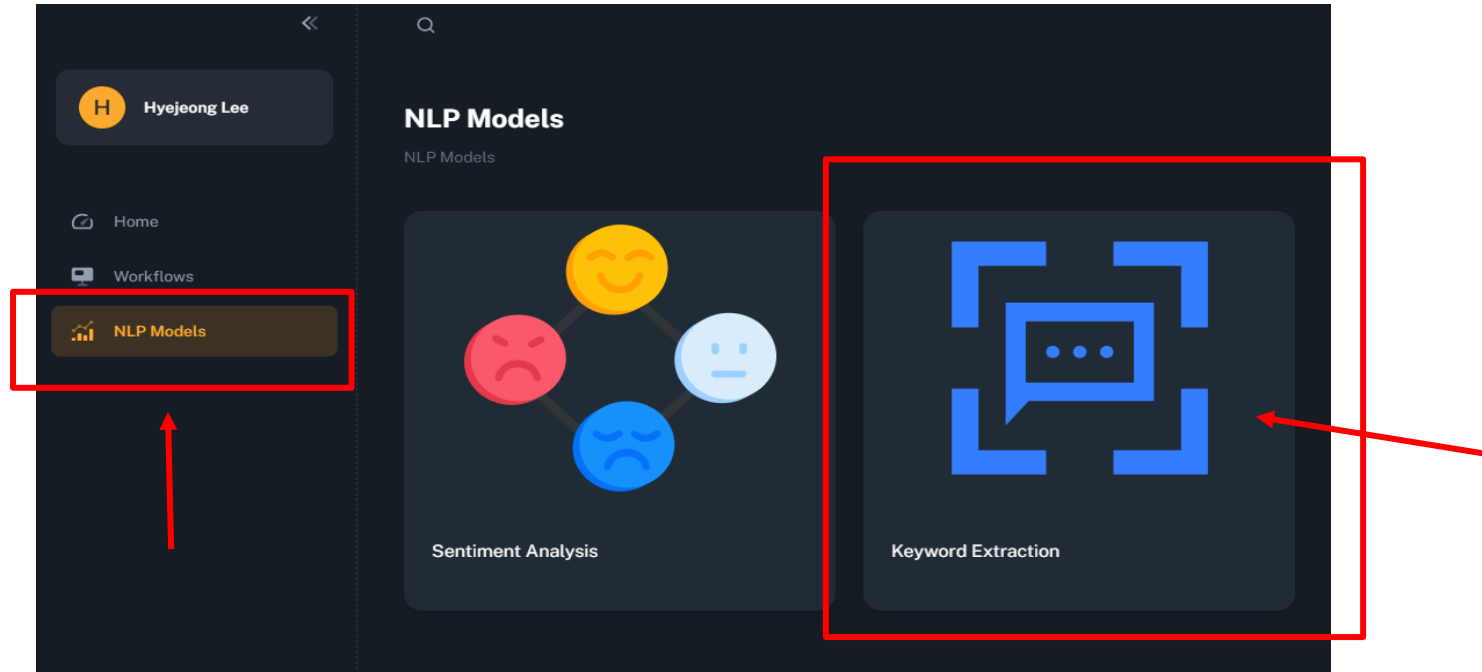
NLP4Science is a platform for the development of Natural Language Processing (NLP) tools for the scientific community. NLP4Science makes it easy to create, share, and use NLP tools for scientific text.

Click Here To Download Demo Dataset!

Click here to download demo datasets



# STEP 3 - KEYWORD EXTRACTION



# STEP 4 - BATCH CLASSIFY

The screenshot shows the AWS SageMaker console interface. On the left is a navigation sidebar with a user profile for 'Hyejeong Lee' and menu items for 'Home', 'Workflows', and 'NLP Models'. The main content area shows a breadcrumb trail: 'NLP Models > Keyword Extraction > Keyword Extraction Keyword'. Below this is a header for 'Keyword Extraction - Keyword' with a description: 'Extract the most relevant keywords from text. These might be aspects (i.e. things mentioned in the text) or opinions (i.e. what is said about aspects)'. A 'Classify' button is visible, with a sub-menu containing 'Batch Classify', which is highlighted with a red box. Below the menu is an 'Upload your data file' section featuring a file upload icon and the text 'Drop or Select file' and 'Drop files here or click [browse](#) through your machine'. A red arrow points from the 'Batch Classify' button to the file upload area. At the bottom right, it specifies 'Allowed \*.csv, \*.txt' and 'max size of 3.1 MB'.

**Click "Batch Classify" and then upload the turtle dataset.**

**Drop or Select file**  
Drop files here or click [browse](#) through your machine

Allowed \*.csv, \*.txt  
max size of 3.1 MB



# STEP 4 - BATCH CLASSIFY

The screenshot shows a 'Batch Classify' interface with a table of items. A red arrow points to the 'Classify' button at the top, and another red arrow points to the checked 'Use this column' checkbox in the table header.

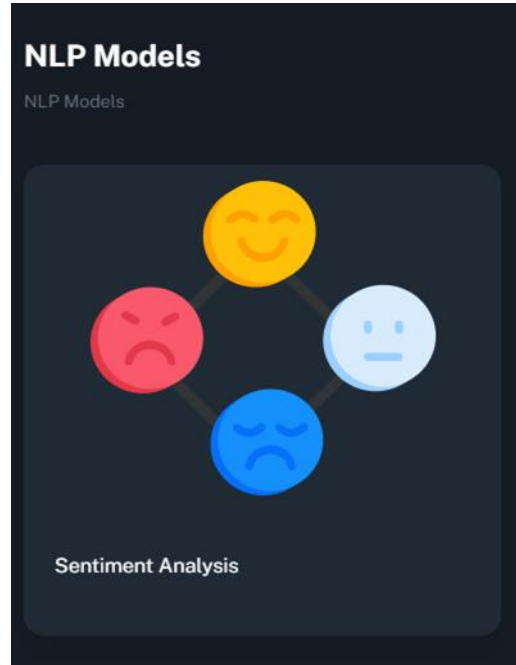
Click "Classify"

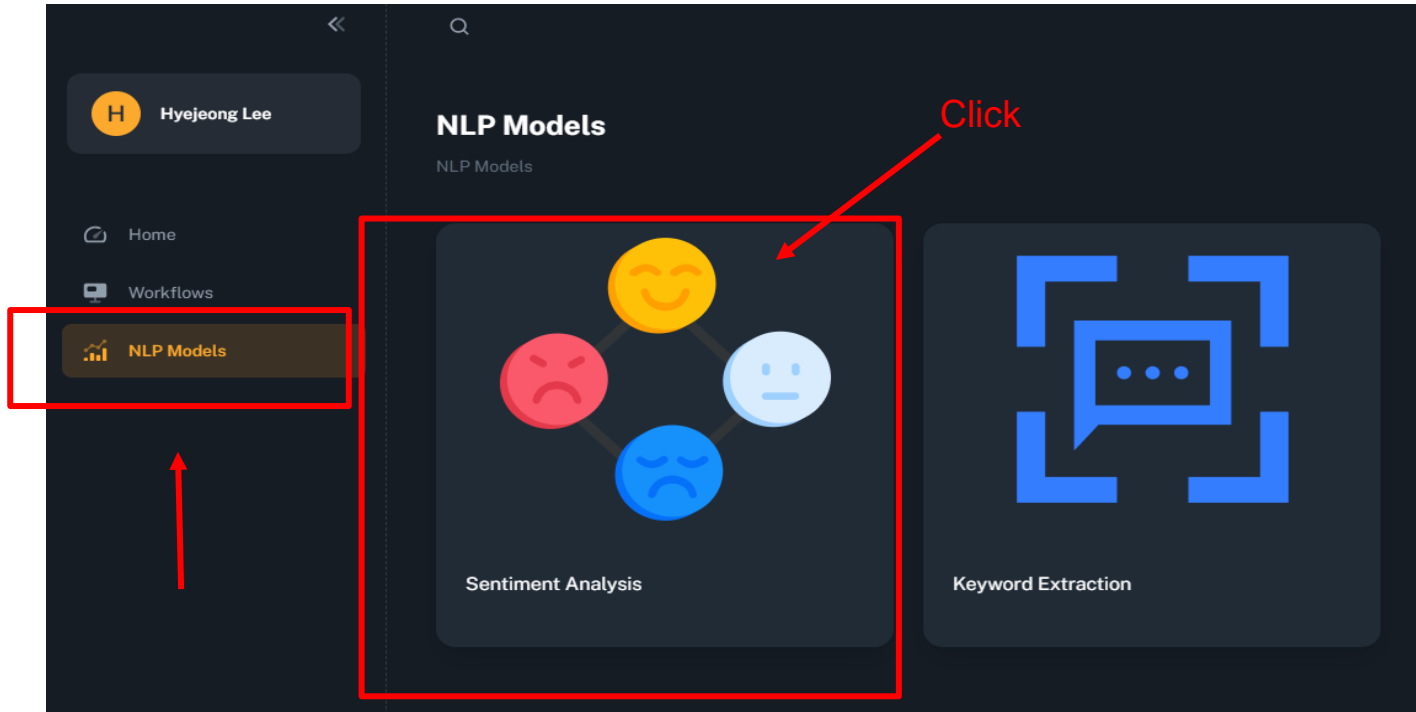
Check "Use this column"

	<input type="checkbox"/> Use this column	<input type="checkbox"/> Use this column	<input type="checkbox"/> Use this column	<input checked="" type="checkbox"/> Use this column
0	Tue Nov 22 10:14:03 +0000 2022	"Not moving heaven and earth to save them is tantamount to burning the... <a href="#">Read More</a>	not moving heaven and earth to save them is tantamount to burning the library before you know what is on the shelves read about doris the turtle and hear from about norfolk reef s amazing biodiversity 🌿👉🌍 <a href="#">Read Less</a>	
1	Tue Nov 22 08:29:31 +0000 2022	@raposatenebrosa @KirbyHonest The reason people fixated on straws was ... <a href="#">Read More</a>	the reason people fixated on straws was because of a video of a turtle with a straw up its nose but straws were developed first and foremost as a hospital aid and there are so many different single use plastics we could target that people don t depend on for basic survival <a href="#">Read Less</a>	
2	Tue Nov 22 04:34:11 +0000 2022	Very few realized the impact of the little video "The turtle and a Str... <a href="#">Read More</a>	very few realized the impact of the little video the turtle and a stra... <a href="#">Read More</a>	
3	Tue Nov 22 03:36:02 +0000 2022	@tedcruz I'm more Conservative than Liberal but I'm an animal lover. Y... <a href="#">Read More</a>	i'm more conservative than liberal but i'm an animal lover you only ne... <a href="#">Read More</a>	



# LET'S EXPLORE SENTIMENT ANALYSIS





**H** Hyejeong Lee

Home

Workflows

**NLP Models**

NLP Models · Sentiment Analysis · Sentiment Analysis Hugging Face

### Sentiment Analysis - Hugging Face

This is a sentiment analysis classifier for texts in English. It works great in customer reviews, but it can be used with any kind of texts. If you want to use this analysis classifier to use, use this one.

Classify **Batch Classify**

**Click "Batch Classify" and then upload the turtle dataset.**

**Upload your data file**

**Drop or Select file**  
Drop files here or click [browse](#) through your machine

Allowed \*.csv, \*.txt  
max size of 3.1 MB



### Sentiment Analysis - Hugging Face

This is a sentiment analysis classifier for texts in English. It works great in customer reviews, but it can be used with any kind of texts. If you are not sure of which sentiment analysis classifier to use, use this one.

Classify **Batch Classify**

Classify

Click "Classify"

Use this column

Use this column

Use this column

Use this column

Check "Use this column"

0

Tue Nov 22 10:14:03 +0000  
2022

"Not moving heaven and earth to save them is tantamount to burning the... [Read More](#)

not moving heaven and earth to save them is tantamount to burning the ... [Read More](#)

1

Tue Nov 22 08:29:31 +0000  
2022

@raposatenebrosa @KirbyHonest The reason people fixated on straws was ... [Read More](#)

the reason people fixated on straws was because of a video of a turtle... [Read More](#)

2

Tue Nov 22 04:34:11 +0000  
2022

Very few realized the impact of the little video "The turtle and a Str... [Read More](#)

very few realized the impact of the little video the turtle and a stra... [Read More](#)

3

Tue Nov 22 03:36:02 +0000  
2022

@tedcruz I'm more Conservative than Liberal but I'm an animal lover. Y... [Read More](#)

I'm more conservative than liberal but I'm an animal lover you only ne... [Read More](#)

4

Tue Nov 22 02:24:56 +0000  
2022

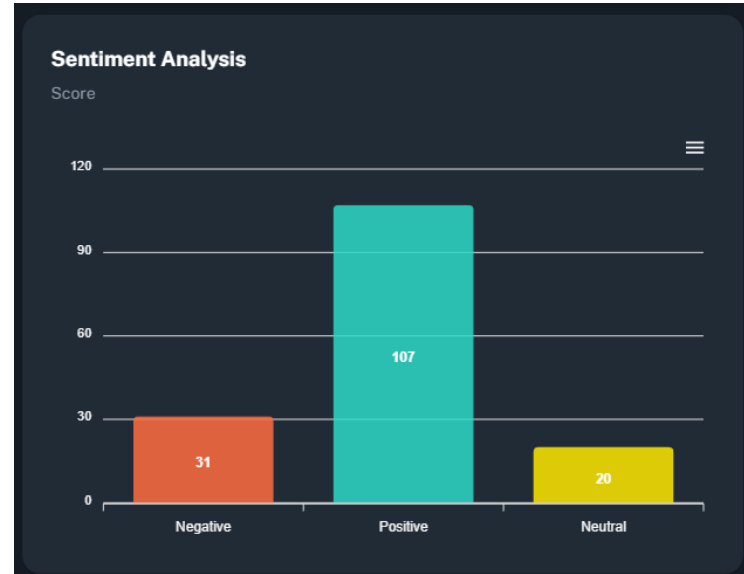
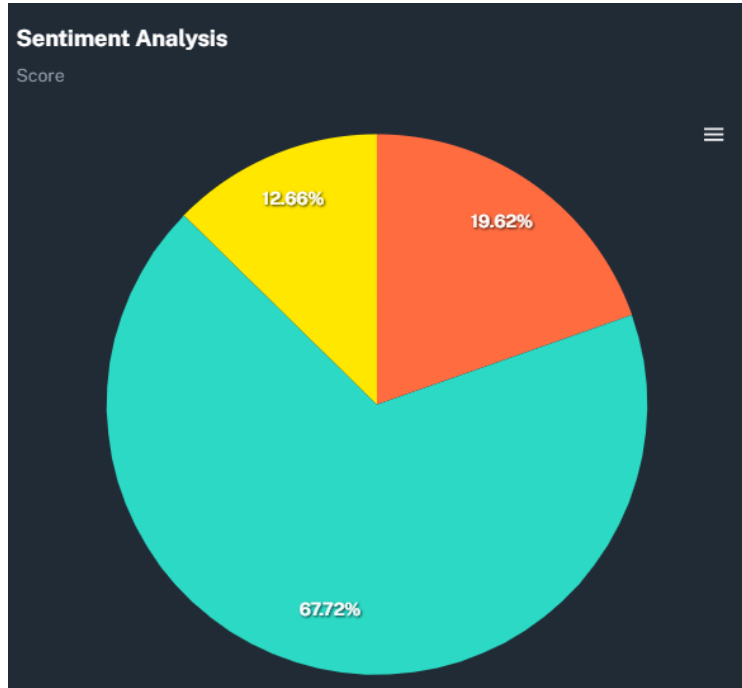
Tokyo Metropolitan Government employees are kicking off the "warm biz"... [Read More](#)

tokyo metropolitan government employees are kicking off the "warm biz"... [Read More](#)

# RESULTS

Text	Score	Confidence
not moving heaven and earth to save them is tantamount to burning the ... <a href="#">Read More</a>	Positive	65%
the reason people fixated on straws was because of a video of a turtle... <a href="#">Read More</a>	Neutral	0%
very few realized the impact of the little video the turtle and a stra... <a href="#">Read More</a>	Positive	36%
i'm more conservative than liberal but i'm an animal lover you only ne... <a href="#">Read More</a>	Positive	86%
tokyo metropolitan government employees are kicking off the "warm biz"... <a href="#">Read More</a>	Positive	77%
give the ocean lover in your life this sea turtle photo t shirt with s... <a href="#">Read More</a>	Positive	79%
have you seen the video of people trying to get a plastic straw out of... <a href="#">Read More</a>	Negative	74%
raph if you count how many times each turtle makes a dive to save some... <a href="#">Read More</a>	Positive	87%

# RESULTS







# NLP or *Not*LP?

- You enter your password into a dialog box by speaking instead of typing
- You ask the virtual assistant (Siri) to find the closest gas station
- When you say, “Alexa, vacuum the living room,” the Roomba in your home does so
- Your phone message app auto-completes words as you are typing
- Your phone uses Face ID recognition to unlock
- You regularly use a dictionary and thesaurus application when you are writing
- You ask the virtual assistant on your phone if you should wear a jacket today, and the response is, “Yes, and don’t forget your umbrella”
- A smart car will avoid hitting an object



keyword extraction  
sentiment analysis  
cluster analysis  
text classification  
algorithmic bias  
data  
human language  
context  
predictions  
meaning  
inference  
autocomplete  
algorithm  
confidence  
text

NLP



**Lunch Time! :D**



The logo features a stylized tree structure where the branches are composed of red circuit board traces, symbolizing the intersection of biology and technology.

# NLP for Science

- <https://app.nlp4science.com/>

