TECH SEO CONNECT

How to incorporate ML into your SEO day-to-day

A bit of ML theory, and a quickfire of ideas to implement ML in your daily routine (with templates ✨)

Lazarina Stoy.

Founder of MLforSEO, Marketing Consultant

October 17 - 18 2024

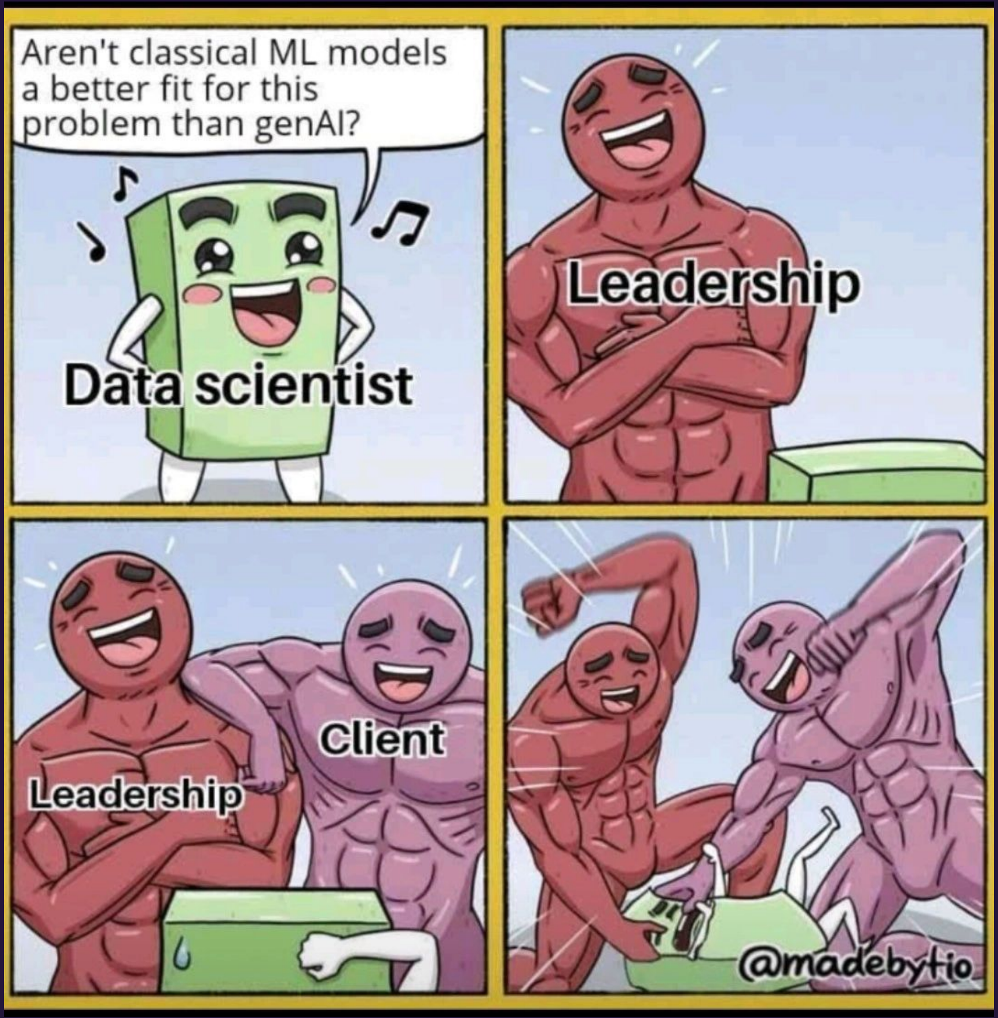
CAM Raleigh 409 W. Martin Street Raleigh, NC 27603

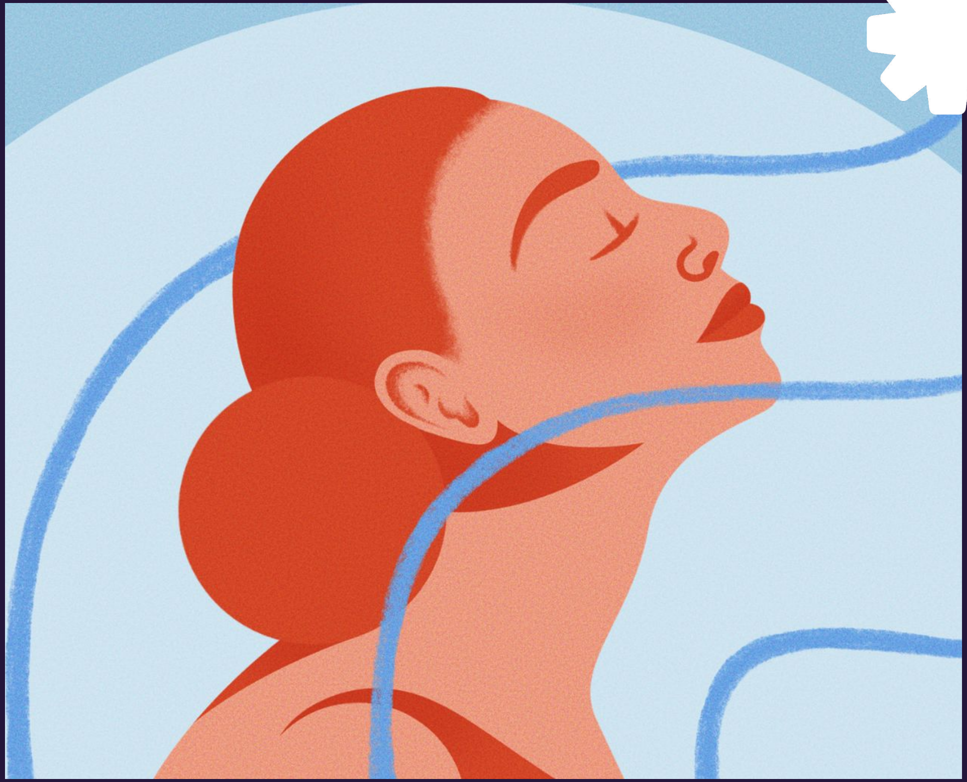
We'll cover...

01 - What you *really* need to start

02 - Where you can implement ML APIs straight away, instantly

03 - What you need to grow 🌱


Aren't classical ML models a better fit for this problem than genAI?
Data scientist
Leadership
Leadership
Client
@madebytio



01. What you *really* need to start


WHERE
LIMITED
BELIEFS
COME
FROM
IMPOSTER SYNDROME
NEGATIVE PAST EXPERIENCE
DEEPER PSYCHOLOGICAL REASONS
FEAR OF FAILURE
planio

Here's what you *really* need

To know

- **When** to search for ML
- **What** model to use
- How to **find** suitable ML tools
- What you can **achieve** in a short time-frame
- How to drive **value** via ML

Let's start with the basics.

For each potential project consider **three aspects**.

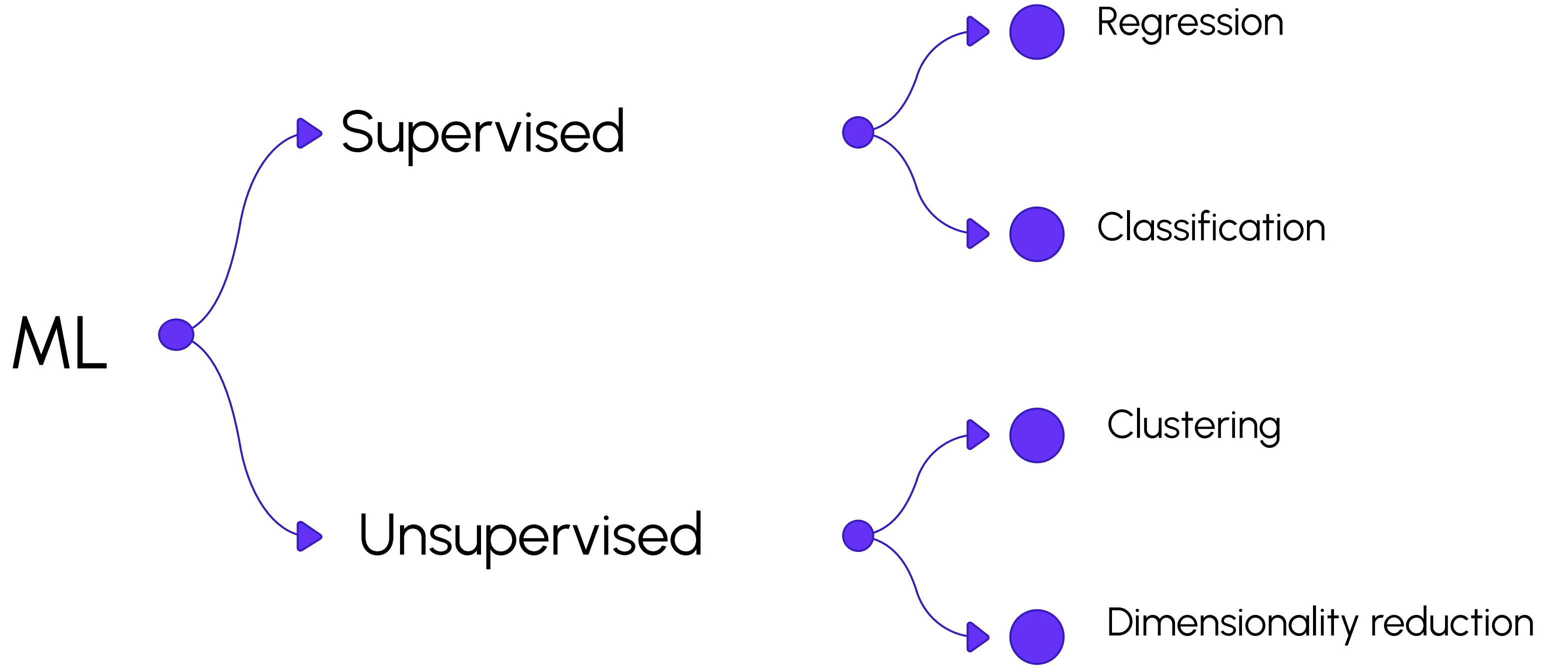
Characteristics of

task

data

solution

Task characteristics


ML
Supervised
Regression
Classification
Unsupervised
Clustering
Dimensionality reduction

ML

- You have labelled data to validate results
  - Make predictions
  - Split into groups, based on existing classes
- You don't have a way to validate results
  - Find patterns and group based on similarity
  - Simplify or transform your data

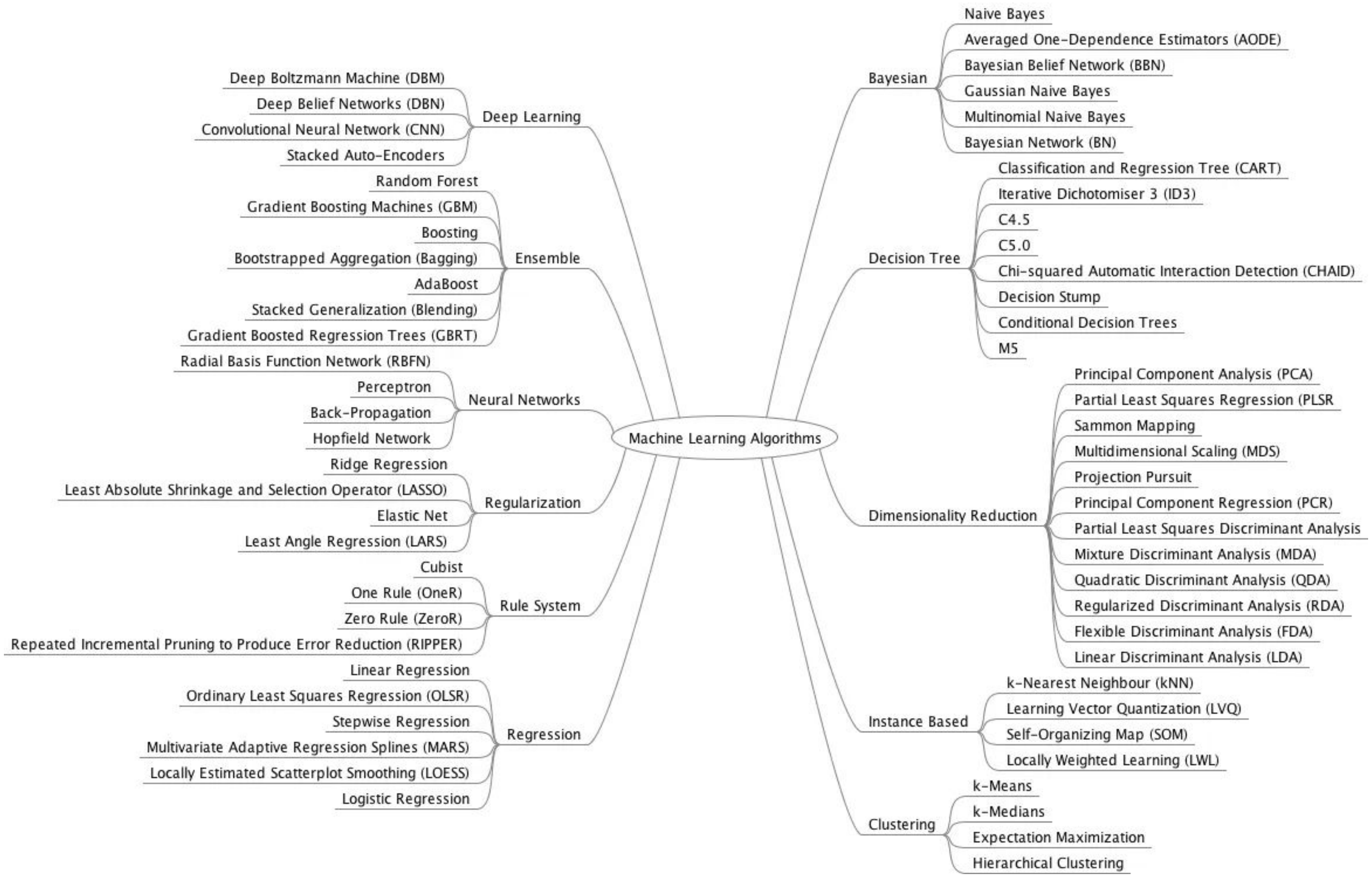
Needless to say, this is a very simplified view.

Top Machine Learning Algorithms

datacamp

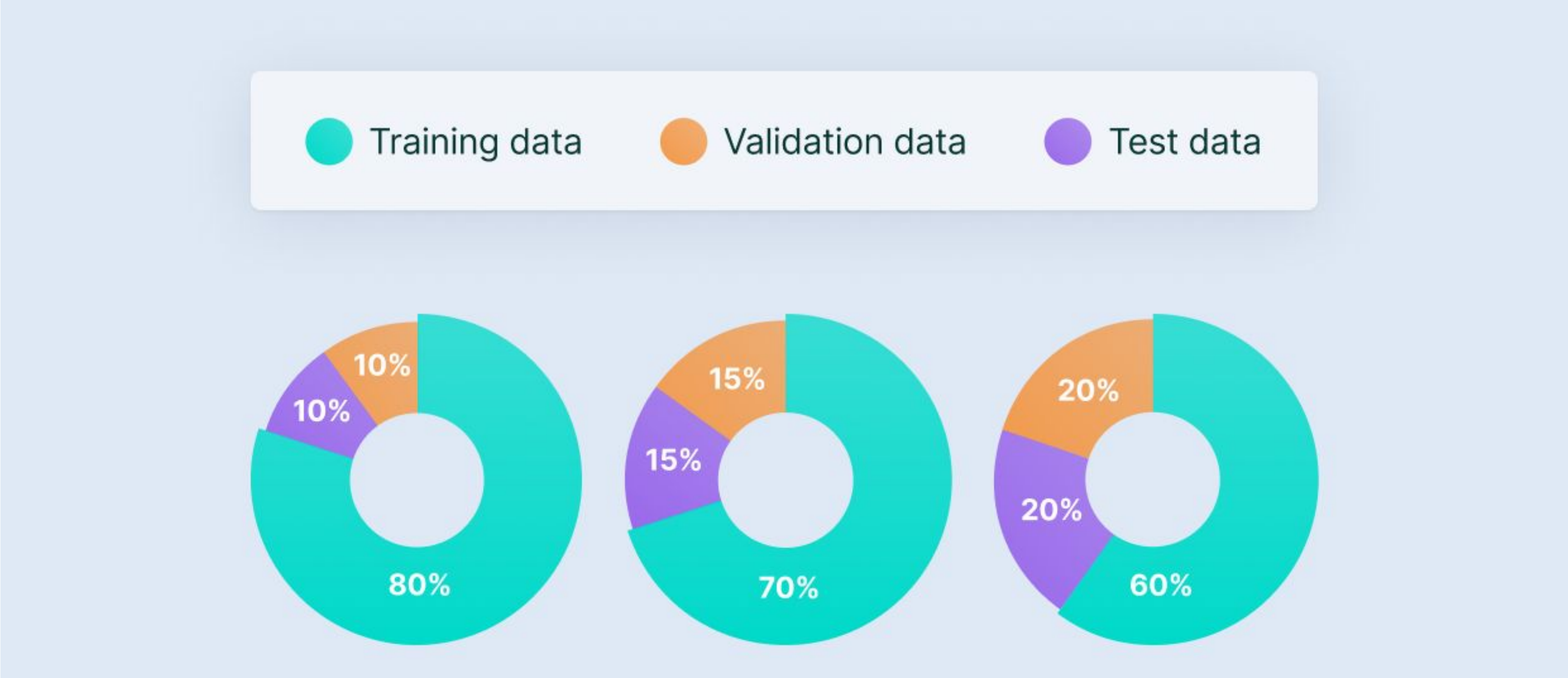
		ALGORITHM	DESCRIPTION	APPLICATIONS	ADVANTAGES	DISADVANTAGES
Supervised Learning	Linear Models	**Linear Regression**	A simple algorithm that models a linear relationship between inputs and a continuous numerical output variable	USE CASES 1. Stock price prediction 2. Predicting housing prices 3. Predicting customer lifetime value	1. Explainable method 2. Interpretable results by its output coefficients 3. Faster to train than other machine learning models	1. Assumes linearity between inputs and output 2. Sensitive to outliers 3. Can underfit with small, high-dimensional data
Supervised Learning	Linear Models	**Logistic Regression**	A simple algorithm that models a linear relationship between inputs and a categorical output (1 or 0)	USE CASES 1. Credit risk score prediction 2. Customer churn prediction	1. Interpretable and explainable 2. Less prone to overfitting when using regularization 3. Applicable for multi-class predictions	1. Assumes linearity between inputs and outputs 2. Can overfit with small, high-dimensional data
Supervised Learning	Linear Models	**Ridge Regression**	Part of the regression family — it penalizes features that have low predictive outcomes by shrinking their coefficients closer to zero. Can be used for classification or regression	USE CASES 1. Predictive maintenance for automobiles 2. Sales revenue prediction	1. Less prone to overfitting 2. Best suited where data suffer from multicollinearity 3. Explainable & interpretable	1. All the predictors are kept in the final model 2. Doesn't perform feature selection
Supervised Learning	Linear Models	**Lasso Regression**	Part of the regression family — it penalizes features that have low predictive outcomes by shrinking their coefficients to zero. Can be used for classification or regression	USE CASES 1. Predicting housing prices 2. Predicting clinical outcomes based on health data	1. Less prone to overfitting 2. Can handle high-dimensional data 3. No need for feature selection	1. Can lead to poor interpretability as it can keep highly correlated variables
Supervised Learning	Tree-Based Models	**Decision Tree**	Decision Tree models make decision rules on the features to produce predictions. It can be used for classification or regression	USE CASES 1. Customer churn prediction 2. Credit score modeling 3. Disease prediction	1. Explainable and interpretable 2. Can handle missing values	1. Prone to overfitting 2. Sensitive to outliers
Supervised Learning	Tree-Based Models	**Random Forests**	An ensemble learning method that combines the output of multiple decision trees	USE CASES 1. Credit score modeling 2. Predicting housing prices	1. Reduces overfitting 2. Higher accuracy compared to other models	1. Training complexity can be high 2. Not very interpretable
Supervised Learning	Tree-Based Models	**Gradient Boosting Regression**	Gradient Boosting Regression employs boosting to make predictive models from an ensemble of weak predictive learners	USE CASES 1. Predicting car emissions 2. Predicting ride hailing fare amount	1. Better accuracy compared to other regression models 2. It can handle multicollinearity 3. It can handle non-linear relationships	1. Sensitive to outliers and can therefore cause overfitting 2. Computationally expensive and has high complexity
Supervised Learning	Tree-Based Models	**XGBoost**	Gradient Boosting algorithm that is efficient & flexible. Can be used for both classification and regression tasks	USE CASES 1. Churn prediction 2. Claims processing in insurance	1. Provides accurate results 2. Captures non linear relationships	1. Hyperparameter tuning can be complex 2. Does not perform well on sparse datasets
Supervised Learning	Tree-Based Models	**LightGBM Regressor**	A gradient boosting framework that is designed to be more efficient than other implementations	USE CASES 1. Predicting flight time for airlines 2. Predicting cholesterol levels based on health data	1. Can handle large amounts of data 2. Computational efficient & fast training speed 3. Low memory usage	1. Can overfit due to leaf-wise splitting and high sensitivity 2. Hyperparameter tuning can be complex
Unsupervised Learning	Clustering	**K-Means**	K-Means is the most widely used clustering approach—it determines K clusters based on euclidean distances	USE CASES 1. Customer segmentation 2. Recommendation systems	1. Scales to large datasets 2. Simple to implement and interpret 3. Results in tight clusters	1. Requires the expected number of clusters from the beginning 2. Has troubles with varying cluster sizes and densities
Unsupervised Learning	Clustering	**Hierarchical Clustering**	A "bottom-up" approach where each data point is treated as its own cluster—and then the closest two clusters are merged together iteratively	USE CASES 1. Fraud detection 2. Document clustering based on similarity	1. There is no need to specify the number of clusters 2. The resulting dendrogram is informative	1. Doesn't always result in the best clustering 2. Not suitable for large datasets due to high complexity
Unsupervised Learning	Clustering	**Gaussian Mixture Models**	A probabilistic model for modeling normally distributed clusters within a dataset	USE CASES 1. Customer segmentation 2. Recommendation systems	1. Computes a probability for an observation belonging to a cluster 2. Can identify overlapping clusters 3. More accurate results compared to K-means	1. Requires complex tuning 2. Requires setting the number of expected mixture components or clusters
Unsupervised Learning	Association	**Apriori algorithm**	Rule based approach that identifies the most frequent itemset in a given dataset where prior knowledge of frequent itemset properties is used	USE CASES 1. Product placements 2. Recommendation engines 3. Promotion optimization	1. Results are intuitive and Interpretable 2. Exhaustive approach as it finds all rules based on the confidence and support	1. Generates many uninteresting itemsets 2. Computationally and memory intensive. 3. Results in many overlapping item sets

MLFORSEO
Machine Learning for SEO


Machine Learning Algorithms
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering

Also...

	Self-Train	A machine learning model you train from scratch, with your own data.
NEWBIE	**Pre-Train**	A machine learning model that a third-party has trained.
	Fine-Tune	A machine learning model that a third-party has trained, that you retrain, improve, adapt, fine-tune with your own data. You train the model further on a more specific dataset.


Training data
Validation data
Test data
10%
10%
80%
15%
15%
70%
20%
20%
60%

Data characteristics

Is your input data...

Textual? **Numeric?** **Image-based?** **Time series?**

Solution characteristics

Is this task mission critical?

✅

Don't rely on AI.

(seriously)

❌

Do you need
consistent results,
every time?

✅

Avoid unsupervised ML.
Avoid generative AI.
Avoid deep learning.

(yeah, really)

❌

Do you need results to be easy to understand?

✅

Skip deep learning.

(yes, that includes LLMs)

Do you need to explain them to stakeholders?

❌

Is it okay that simply **on average** the output outperforms existing methods?

✅

Okay, then.

Take a look at ML options.

Assess usefulness of ML using multiple factors:

- insights
- complexity
- accuracy
- scalability
- assets (data in = data out)
- resources
- bottom line

How to find what you need whenever you have an idea

Keep **queries** specific to data, task, and solution.

Writing meta descriptions

Data Characteristics	Textual	Page content
Task characteristics	Unsupervised	• It is transformational (Page content to Page Summary in less than 160 characters) • It can also be generative (write them from scratch)
Solution characteristics	Mission critical?	No
	Different results OK?	Yes
	Explanation of process needed?	Not really.
	Outperform current methods?	Yes, much faster to get to a good enough result and satisfy a hygiene condition.

Title / H1 Optimisations

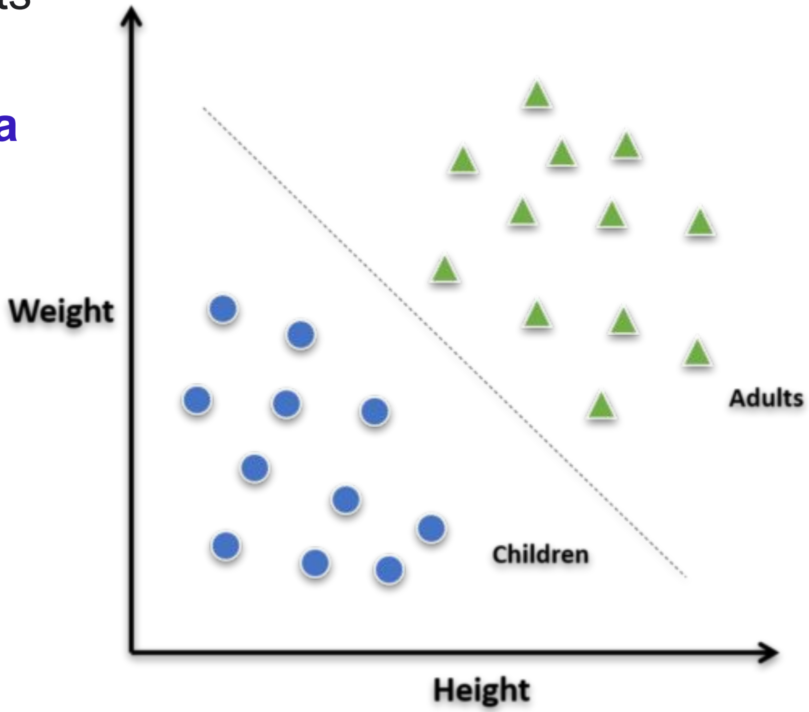
Data Characteristics	Textual	Page content
Task characteristics	Unsupervised	• It is transformational (Page content to Page Summary in less than 60 characters) • It can also be generative (write them from scratch)
Solution characteristics	Mission critical?	Could be, depending on the industry
	Different results OK?	Could be critical for certain industries.
	Explanation of process needed?	Sometimes.
	Outperform current methods?	Yes, much faster to get to multiple first drafts, to pass onto an editor.

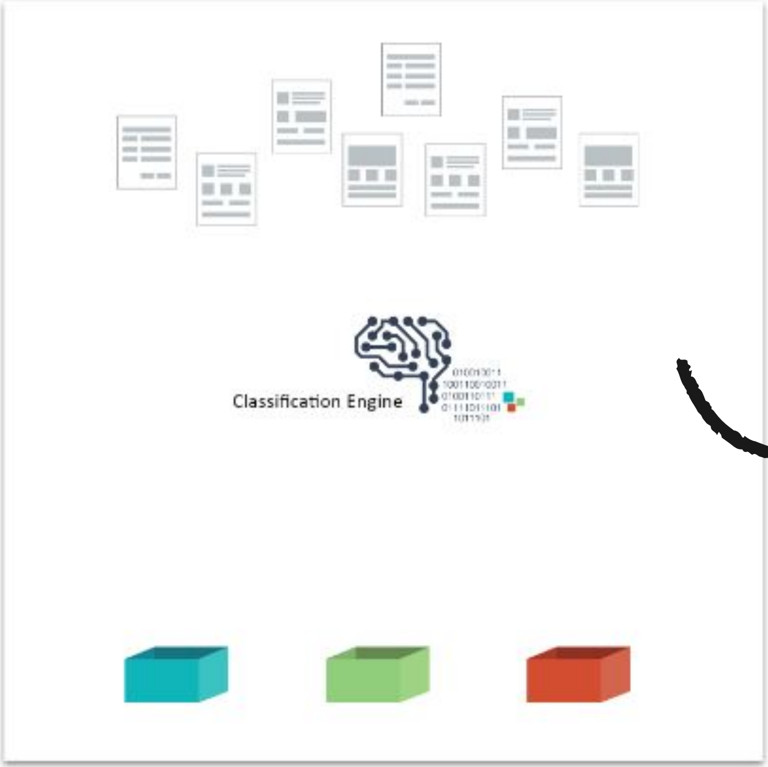
Image captioning/ Alt tag generation		
Data Characteristics	Image	Image library
Task characteristics	Unsupervised	Using a Pre-trained model image recognition model Generative AI / Image recognition
Solution characteristics	Mission critical?	No
	Different results OK?	Yes
	Explanation of process needed?	Not really
	Outperform current methods?	Yes, much faster.

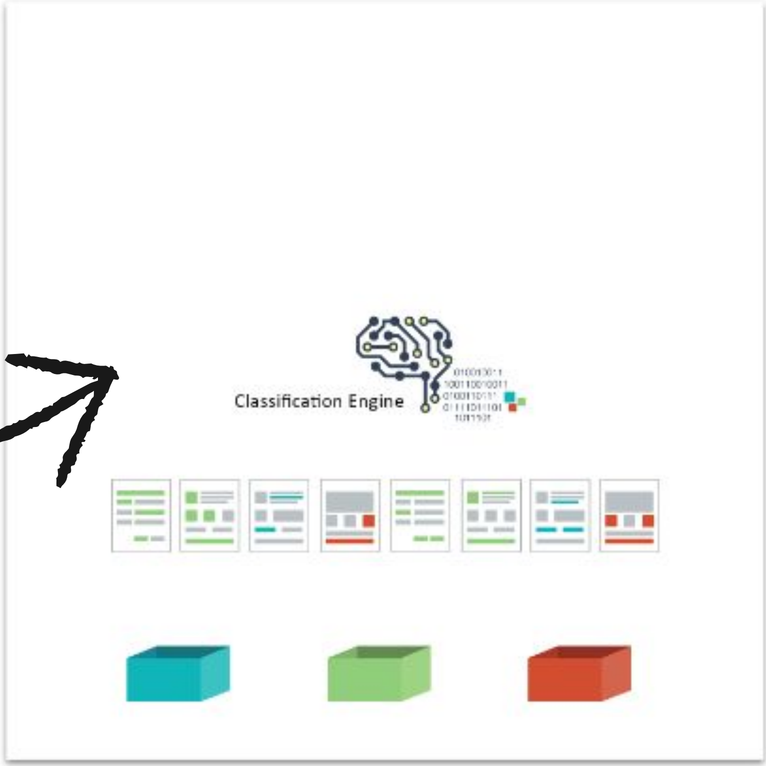
02. What to do for immediate value


Entertainment
News
Education

Classification sorts data into specific categories **using a labeled dataset**.




Classification Engine


Classification Engine

Quick check-in

- Classification is a **supervised** machine learning approach.
- It involves sorting **data** (documents, pages, keywords) into **pre-labeled** categories
- Applications include - content audit, competitor audit

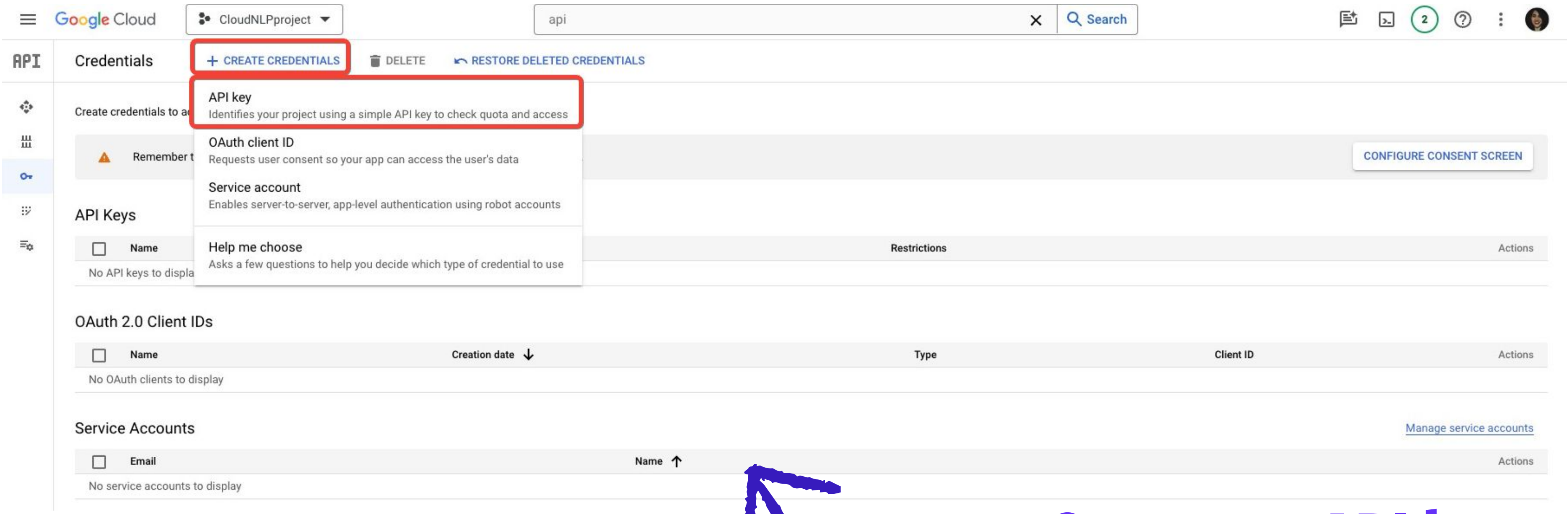
With Google's Natural Language API, you can **classify** documents in **1,300+** predefined categories

Text Classification

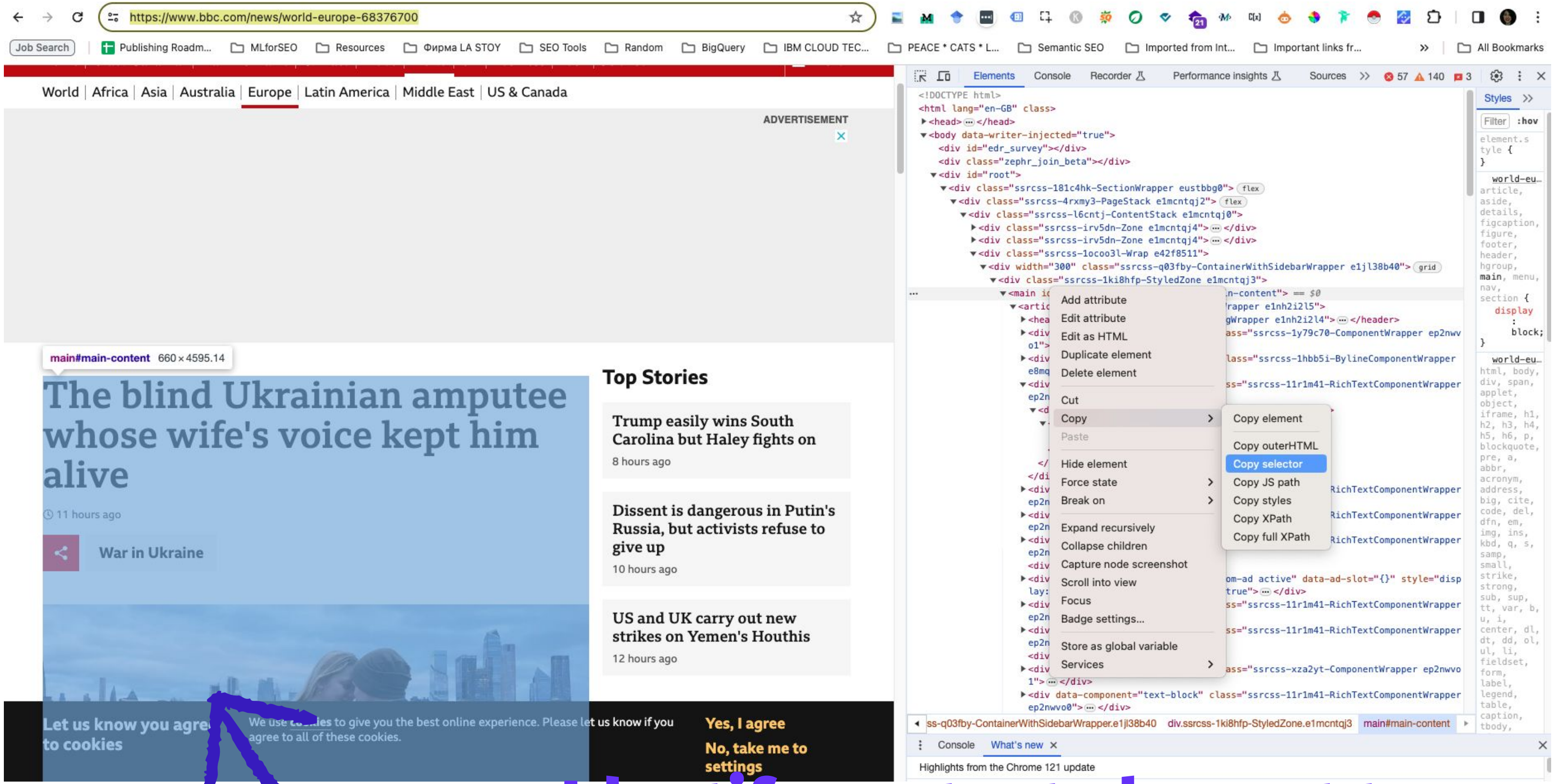
How to do Text Classification with Google's Natural Language API in Google Sheets (Apps Script)

Lazarina Stoy. · Mar 27, 2024

Process will take no more than 20 minutes



Create an API key



Identify content element to scrape

MLforSEO
Machine Learning for SEO

@lazarinastoy | #techseoconnect

TECH SEO CONNECT


Custom Extraction
Extract selected elements of internal HTML pages. See our User Guide for examples. The results can be seen in the Custom Extraction Tab.
Content BBC
CSSPath
#main-content > article
Extract Text
Attribute (optional)
Delete All
Add
OK
Scrape with 🐸

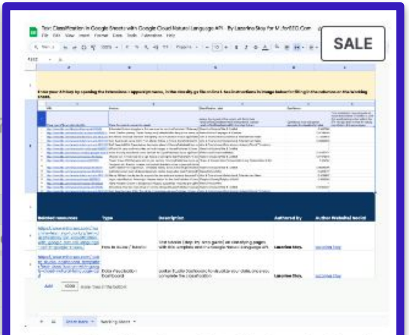
Text Classification in Google Sheets with Google Cloud Natural Language API - By Lazarina Stoy f...

File Edit View Insert Format Data Tools Extensions Help

2:999 | fx https://www.bbc.com/news/world-europe-68376700

	A	B	C	
1	URL	Content	Classification Label	Confi
2	https://www.bbc.com/news/world-europe-68376700	The blind Ukrainian amputee whose wife's voice kept him alivePublished11 h		
3	https://www.bbc.com/news/world-europe-68255490	Exhausted Ukraine struggles to find new men for front linePublished12 Febru		
4	https://www.bbc.com/news/world-us-canada-68395414	South Carolina primary: Donald Trump easily defeats Nikki Haley in her hom		
5	https://www.bbc.com/news/entertainment-arts-68362810	Kim Petras on sexual liberation and fighting TikTokPublished10 hours agoSh		
6	https://www.bbc.com/news/entertainment-arts-68395354	SAG Awards red carpet 2024: From Margot Robbie to Emma StonePublishe		
7	https://www.bbc.com/news/entertainment-arts-68395355	SAG Awards 2024: Oppenheimer dominates ahead of OscarsPublished8 ho		
8	https://www.bbc.com/news/world-middle-east-68395173	US and UK carry out fresh strikes on Houthi targets in YemenPublished12 h		
9	https://www.bbc.com/news/uk-scotland-glasgow-west-67980670	Inside the long-abandoned tunnel beneath the ClydePublished2 hours agoS		
10	https://www.bbc.com/news/world-europe-68322527	Ukraine war: Is Avdiivka's fall a sign Russia is turning the tide?Published17 F		
11	https://www.bbc.com/news/uk-wales-68210255	Travel: How a £525 bet gave birth to your morning commutePublished		
12	https://www.bbc.com/news/world-europe-68384341	Two years into Russia's invasion, exhausted Ukrainians refuse to give upPub Anthill resident"I'm no politician," confesses Valeriy, a man in his 80s perche		
13	https://www.bbc.com/news/world-europe-68393412	Authorities return body of Alexei Navalny to mother 8 days after deathPublis		
14	https://www.bbc.com/news/entertainment-arts-68391330	Wendy Williams thanks fans for support after dementia and aphasia diagnosi		
15	https://www.bbc.com/news/world-asia-68378651	Japan naked festival: Women join Hadaka Matsuri for first timePublished10		
16	https://www.bbc.com/news/world-europe-68395030	Alexei Navalny: Dissent is dangerous in Russia, but activists refuse to give u		
17	https://www.bbc.com/news/world-europe-68359252	Rosenberg: How two years of war in Ukraine changed RussiaPublished3 day		
18	https://www.bbc.com/news/entertainment-arts-68395352	SAG Award winners 2024: The full list of nominees and winsPublished13 ho		
19	https://www.bbc.com/news/entertainment-arts-68362811	Stray Kids: How K-Pop took over the global charts in 2023Published3 days a		
20	https://www.bbc.com/news/entertainment-arts-68317736	Gareth Edwards: The Creator director on shaking up Hollywood's visual effe		
21	https://www.bbc.com/news/newsbeat-68382142	Chuckie: 1Xtra presenter feels R&B has special year aheadPublished1 day a		
22	https://www.bbc.com/news/entertainment-arts-68338730	Alia Bhatt: The young Bollywood star taking on HollywoodPublished2 days ag		

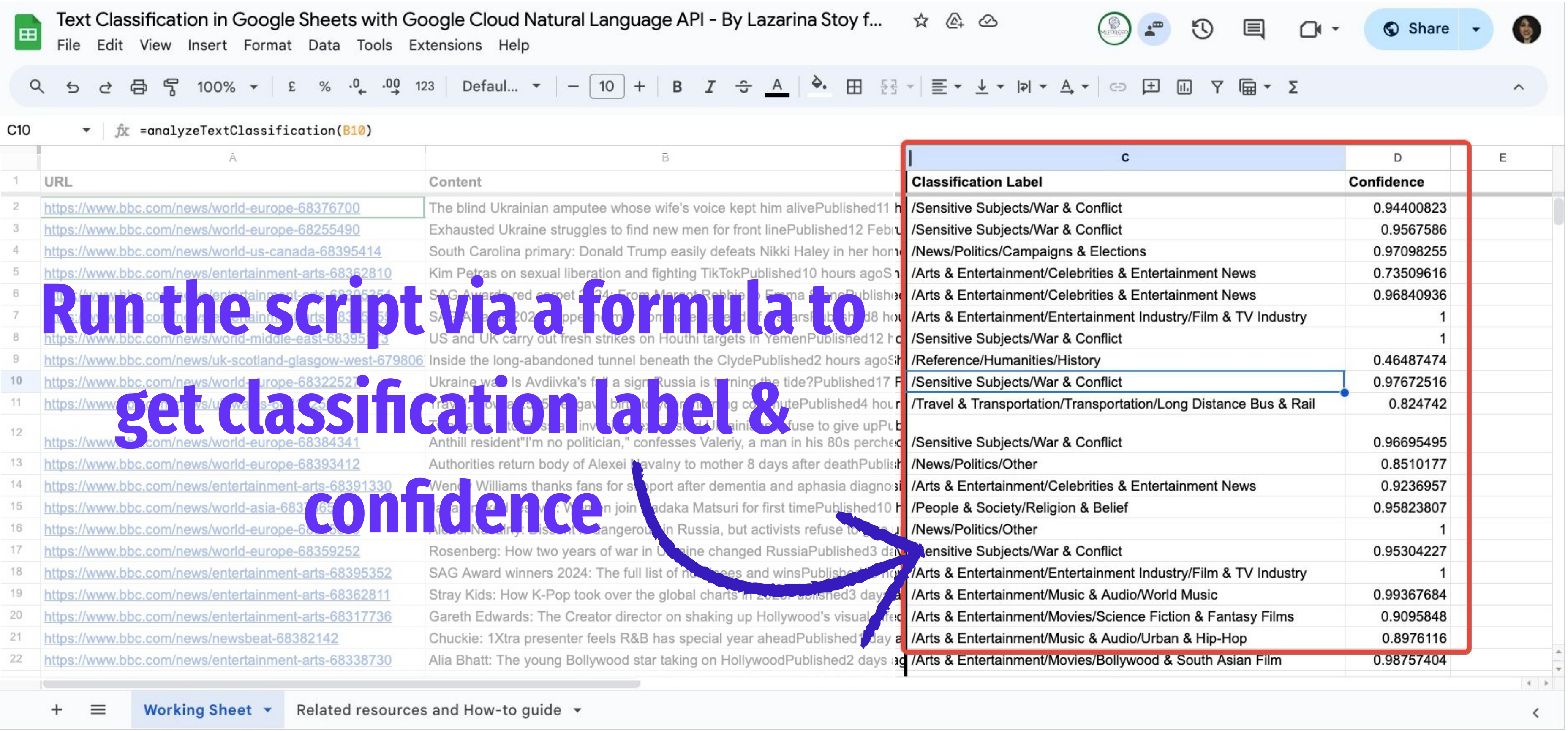
Working Sheet | Related resources and How-to guide | Count: 254

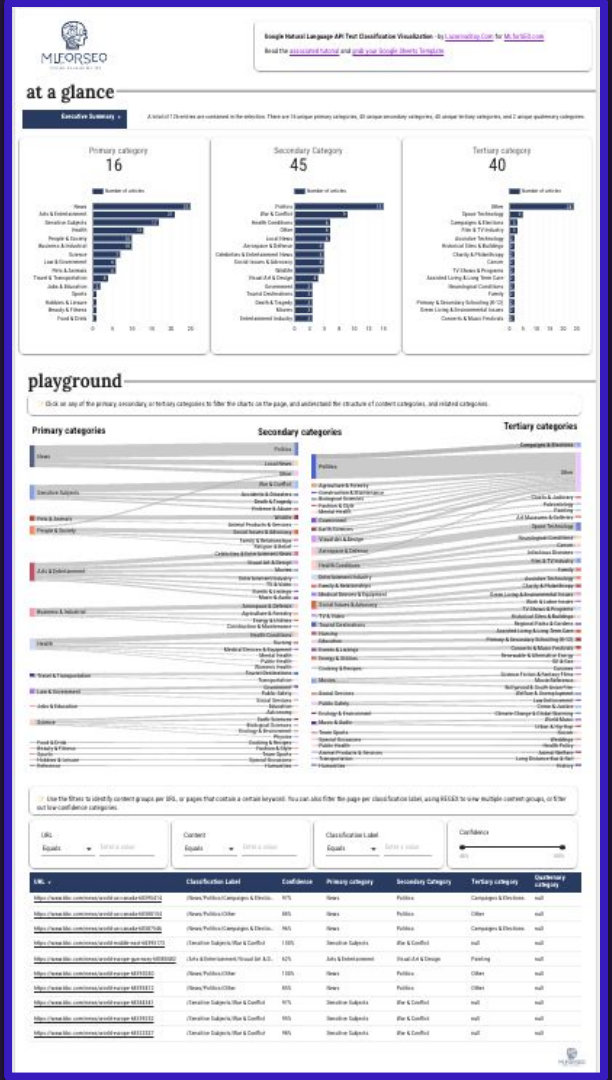


Text Classification with with Google Cloud Natural Language API (Google Sheets Template with Apps Script)

Enter your URLs and content

Run the script via a formula to get classification label & confidence

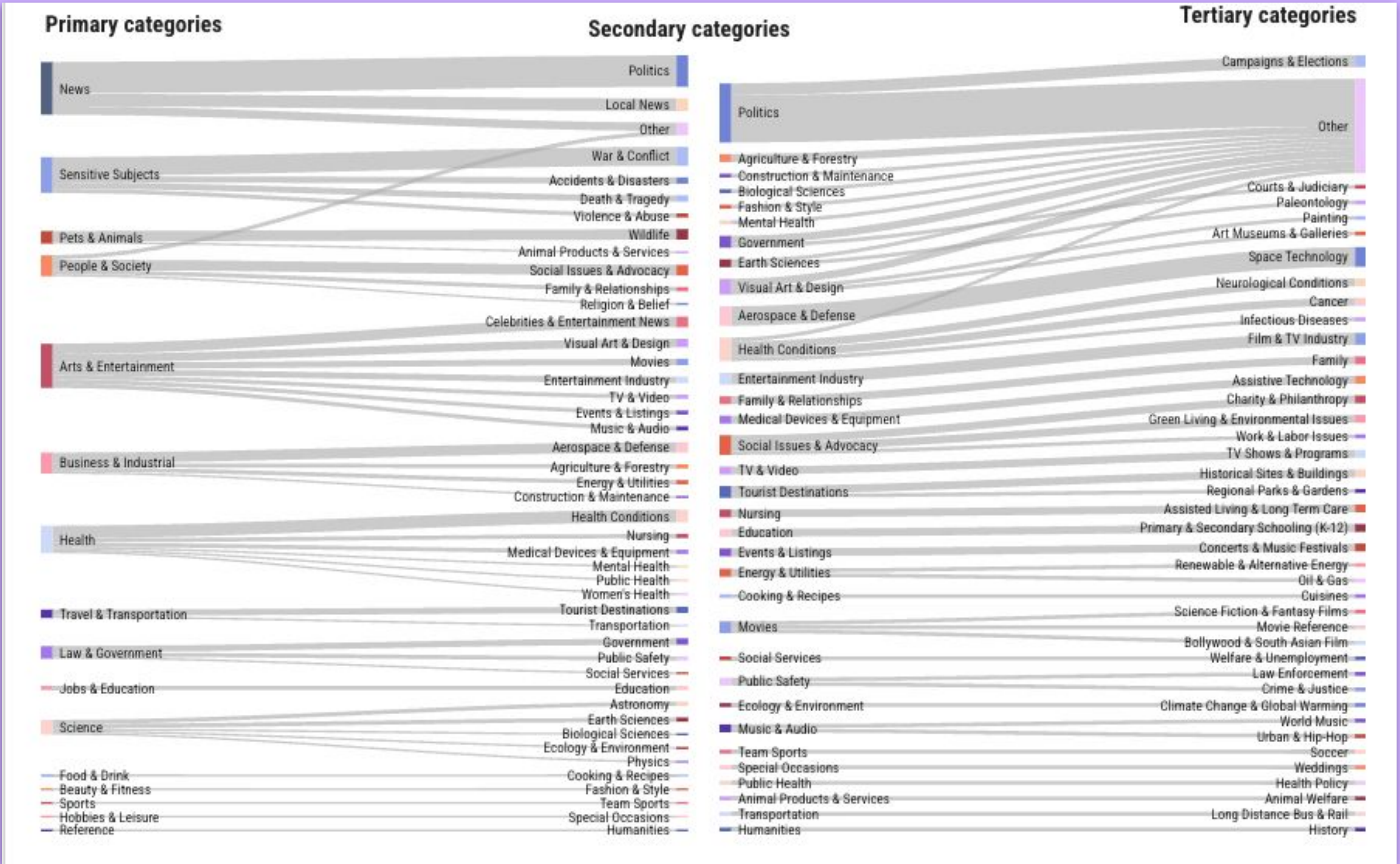




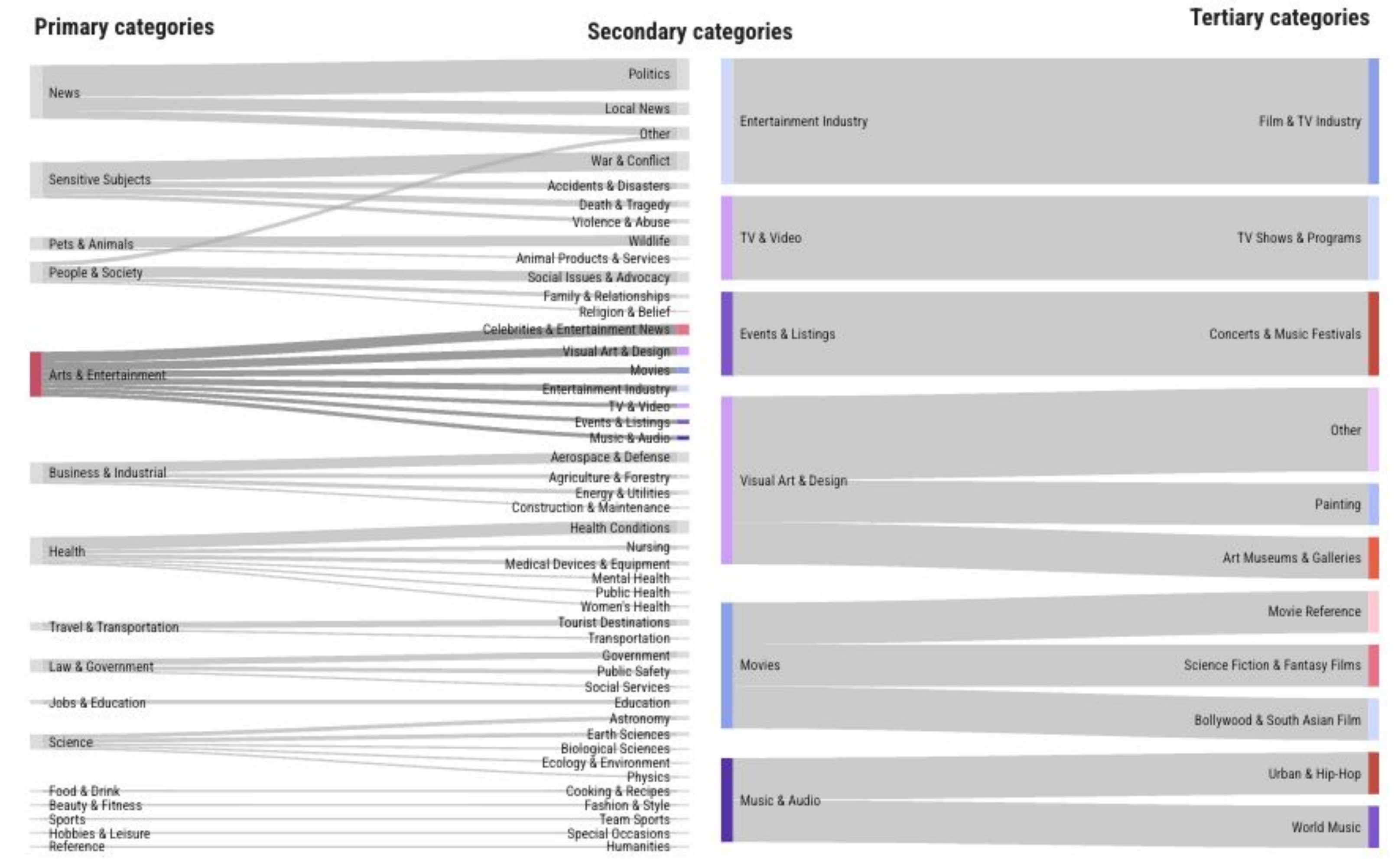
Plug-and-play template in Looker Studio

MLFORSEO
Machine Learning for SEO

www.mlforseo.com



Primary categories
Secondary categories
Tertiary categories
News
Sensitive Subjects
Pets & Animals
People & Society
Arts & Entertainment
Business & Industrial
Health
Travel & Transportation
Law & Government
Jobs & Education
Science
Food & Drink
Beauty & Fitness
Sports
Hobbies & Leisure
Reference
Politics
Local News
Other
War & Conflict
Accidents & Disasters
Death & Tragedy
Violence & Abuse
Wildlife
Animal Products & Services
Social Issues & Advocacy
Family & Relationships
Religion & Belief
Celebrities & Entertainment News
Visual Art & Design
Movies
Entertainment Industry
TV & Video
Events & Listings
Music & Audio
Aerospace & Defense
Agriculture & Forestry
Energy & Utilities
Construction & Maintenance
Health Conditions
Nursing
Medical Devices & Equipment
Mental Health
Public Health
Women's Health
Tourist Destinations
Transportation
Government
Public Safety
Social Services
Education
Astronomy
Earth Sciences
Biological Sciences
Ecology & Environment
Physics
Cooking & Recipes
Fashion & Style
Team Sports
Special Occasions
Humanities
Politics
Agriculture & Forestry
Construction & Maintenance
Biological Sciences
Fashion & Style
Mental Health
Government
Earth Sciences
Visual Art & Design
Aerospace & Defense
Health Conditions
Entertainment Industry
Family & Relationships
Medical Devices & Equipment
Social Issues & Advocacy
TV & Video
Tourist Destinations
Nursing
Education
Events & Listings
Energy & Utilities
Cooking & Recipes
Movies
Social Services
Public Safety
Ecology & Environment
Music & Audio
Team Sports
Special Occasions
Public Health
Animal Products & Services
Transportation
Humanities
Campaigns & Elections
Other
Courts & Judiciary
Paleontology
Painting
Art Museums & Galleries
Space Technology
Neurological Conditions
Cancer
Infectious Diseases
Film & TV Industry
Family
Assistive Technology
Charity & Philanthropy
Green Living & Environmental Issues
Work & Labor Issues
TV Shows & Programs
Historical Sites & Buildings
Regional Parks & Gardens
Assisted Living & Long Term Care
Primary & Secondary Schooling (K-12)
Concerts & Music Festivals
Renewable & Alternative Energy
Oil & Gas
Cuisines
Science Fiction & Fantasy Films
Movie Reference
Bollywood & South Asian Film
Welfare & Unemployment
Law Enforcement
Crime & Justice
Climate Change & Global Warming
World Music
Urban & Hip-Hop
Soccer
Weddings
Health Policy
Animal Welfare
Long Distance Bus & Rail
History



Primary categories
Secondary categories
Tertiary categories
News
Sensitive Subjects
Pets & Animals
People & Society
Arts & Entertainment
Business & Industrial
Health
Travel & Transportation
Law & Government
Jobs & Education
Science
Food & Drink
Beauty & Fitness
Sports
Hobbies & Leisure
Reference
Politics
Local News
Other
War & Conflict
Accidents & Disasters
Death & Tragedy
Violence & Abuse
Wildlife
Animal Products & Services
Social Issues & Advocacy
Family & Relationships
Religion & Belief
Celebrities & Entertainment News
Visual Art & Design
Movies
Entertainment Industry
TV & Video
Events & Listings
Music & Audio
Aerospace & Defense
Agriculture & Forestry
Energy & Utilities
Construction & Maintenance
Health Conditions
Nursing
Medical Devices & Equipment
Mental Health
Public Health
Women's Health
Tourist Destinations
Transportation
Government
Public Safety
Social Services
Education
Astronomy
Earth Sciences
Biological Sciences
Ecology & Environment
Physics
Cooking & Recipes
Fashion & Style
Team Sports
Special Occasions
Humanities
Entertainment Industry
TV & Video
Events & Listings
Visual Art & Design
Movies
Music & Audio
Film & TV Industry
TV Shows & Programs
Concerts & Music Festivals
Other
Painting
Art Museums & Galleries
Movie Reference
Science Fiction & Fantasy Films
Bollywood & South Asian Film
Urban & Hip-Hop
World Music

✨ Use the filters to identify content groups per URL, or pages that contain a certain keyword. You can also filter the page per classification label, using REGEX to view multiple content groups, or filter out low-confidence categories.

URL
Equals
Enter a value

Content
Equals
Enter a value

Classification Label
Equals
Enter a value

Confidence
40% 100%

URL	Classification Label	Confidence	Primary category	Secondary Category	Tertiary category	Quaternary category
https://www.bbc.com/news/world-us-canada-68395414	/News/Politics/Campaigns & Electio...	97%	News	Politics	Campaigns & Elections	null
https://www.bbc.com/news/world-us-canada-68388154	/News/Politics/Other	88%	News	Politics	Other	null
https://www.bbc.com/news/world-us-canada-68387546	/News/Politics/Campaigns & Electio...	96%	News	Politics	Campaigns & Elections	null
https://www.bbc.com/news/world-middle-east-68395173	/Sensitive Subjects/War & Conflict	100%	Sensitive Subjects	War & Conflict	null	null
https://www.bbc.com/news/world-europe-guernsey-68380482	/Arts & Entertainment/Visual Art & D...	62%	Arts & Entertainment	Visual Art & Design	Painting	null
https://www.bbc.com/news/world-europe-68395030	/News/Politics/Other	100%	News	Politics	Other	null
https://www.bbc.com/news/world-europe-68393412	/News/Politics/Other	85%	News	Politics	Other	null
https://www.bbc.com/news/world-europe-68384341	/Sensitive Subjects/War & Conflict	97%	Sensitive Subjects	War & Conflict	null	null
https://www.bbc.com/news/world-europe-68359252	/Sensitive Subjects/War & Conflict	95%	Sensitive Subjects	War & Conflict	null	null
https://www.bbc.com/news/world-europe-68322527	/Sensitive Subjects/War & Conflict	98%	Sensitive Subjects	War & Conflict	null	null
https://www.bbc.com/news/world-europe-68248740	/News/Politics/Other	100%	News	Politics	Other	null

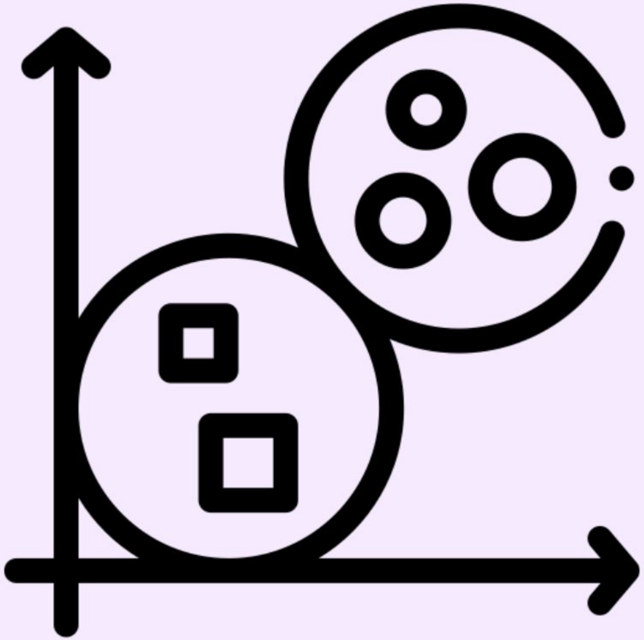
MLFORSEO

With Open AI's GPT4 or ChatGPT, **results are a hit or miss**.

But why? 🤔

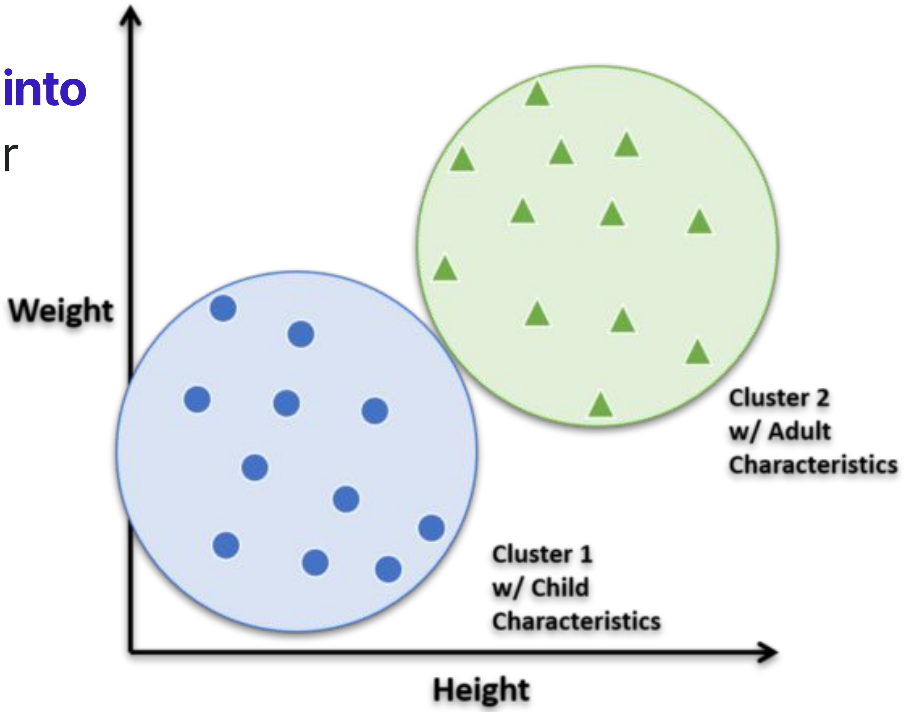
General purpose model
Not efficient in large datasets, Unreliable (unpredictable) results
Non-replicable - output different every time, even if data doesn't change.
Generative AI, trained on a wide variety of general-purpose text but isn't fine-tuned on specific datasets for text classification unless explicitly prompted
Great for creative outputs

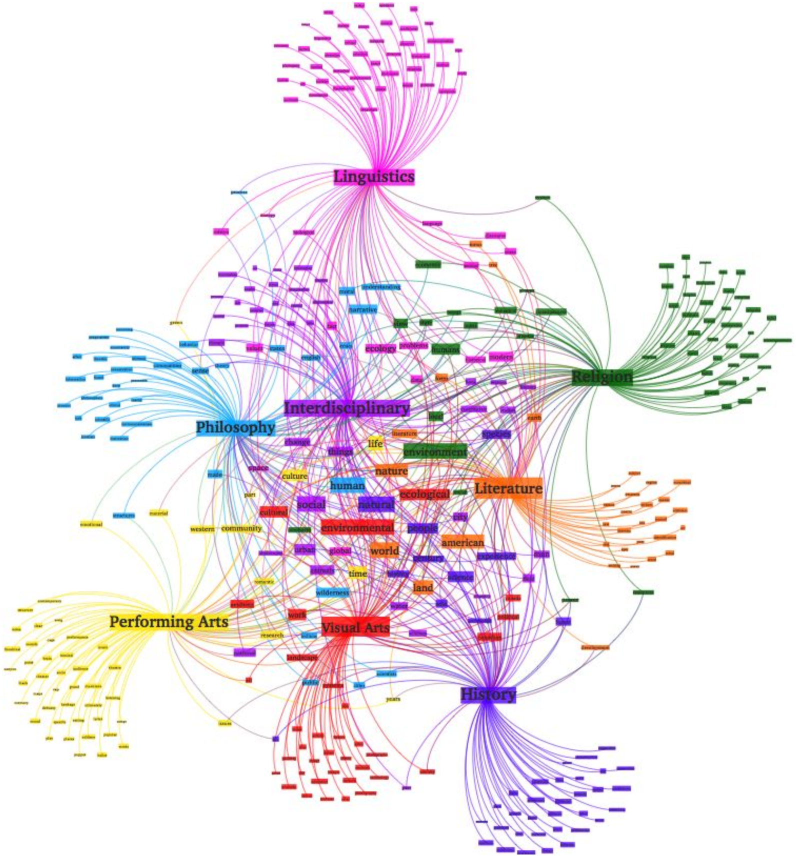
Purpose-built API, specifically designed for tasks like text classification, sentiment analysis, and entity recognition
Efficient and reliable
Output same every time, unless data changes.
Pre-trained Supervised ML model, trained on vast amounts of labeled data related specifically to text classification tasks
Great for tasks that require precision



www.mlforseo.com

Clustering is partitioning an **unlabeled dataset into groups** of similar objects.





Linguistics
Religion
Interdisciplinary
Philosophy
Literature
Performing Arts
Visual Arts
History
ecology
life
environment
nature
culture
human
ecological
social
natural
environmental
people
american
world
time
land
community
western

Watch the details later.

I've recorded a step-by-step tutorial on doing **topic modelling** using a no-code, publicly-available, web-based app using LDA.

Topic models

Topic models

A	B	C	D	E	F	G	H	I	J	K
	health mental staff workplace home support group working ehl students	corporate volunteering giving purpose social responsibility grawehr stéphanie	nonprofit volunteers time content media support volunteer nonprofits share form	alaya platform data user services users policy information conditions general	csr business social companies strategy responsibility corporate initiatives	volunteering program employees corporate giving programs matching benefits	people company time it's back feel start make mission that's	employees impact engage purpose activities make platform community	engagement employees work team engaged good teams find make virtual	nonprofits season donors carmen amell nonprofit fundraising make strategy story
[illegible]	0.00%	17.65%	-44.77%		-16.00%	-104.73%	-4.67%	9.29%	-5.99%	
[illegible]	17.65%	0.00%	-21.13%	-88.64%	-8.61%	-62.86%	-36.37%	-23.42%	27.18%	6.80%
[illegible]	-44.77%	-21.13%	0.00%	9.88%	-26.01%	-58.78%	12.52%	16.87%	-56.55%	70.50%
[illegible]		-88.64%	9.88%	0.00%	-30.66%		-157.96%	-53.92%		
[illegible]	-16.00%	-8.61%	-26.01%	-30.66%	0.00%	60.18%	54.64%	26.11%	-0.59%	-218.54%
[illegible]	-104.73%	-62.86%	-58.78%		60.18%	0.00%	35.22%	30.17%	-0.59%	-77.01%
[illegible]	-4.67%	-36.37%	12.52%	-157.96%	54.64%	35.22%	0.00%	28.20%	27.18%	-46.26%
[illegible]	9.29%	-23.42%	16.87%	-53.92%	26.11%	30.17%	28.20%	0.00%	-1.53%	-40.31%
[illegible]	-5.99%	27.18%	-56.55%		-0.59%	-0.59%	27.18%	-1.53%	0.00%	-29.35%
[illegible]		6.80%	70.50%		-218.54%	-77.01%	-46.26%	-40.31%	-29.35%	0.00%

Topic to Topic Similarity

Topic Modelling per Page

Topic to Topic Similarity

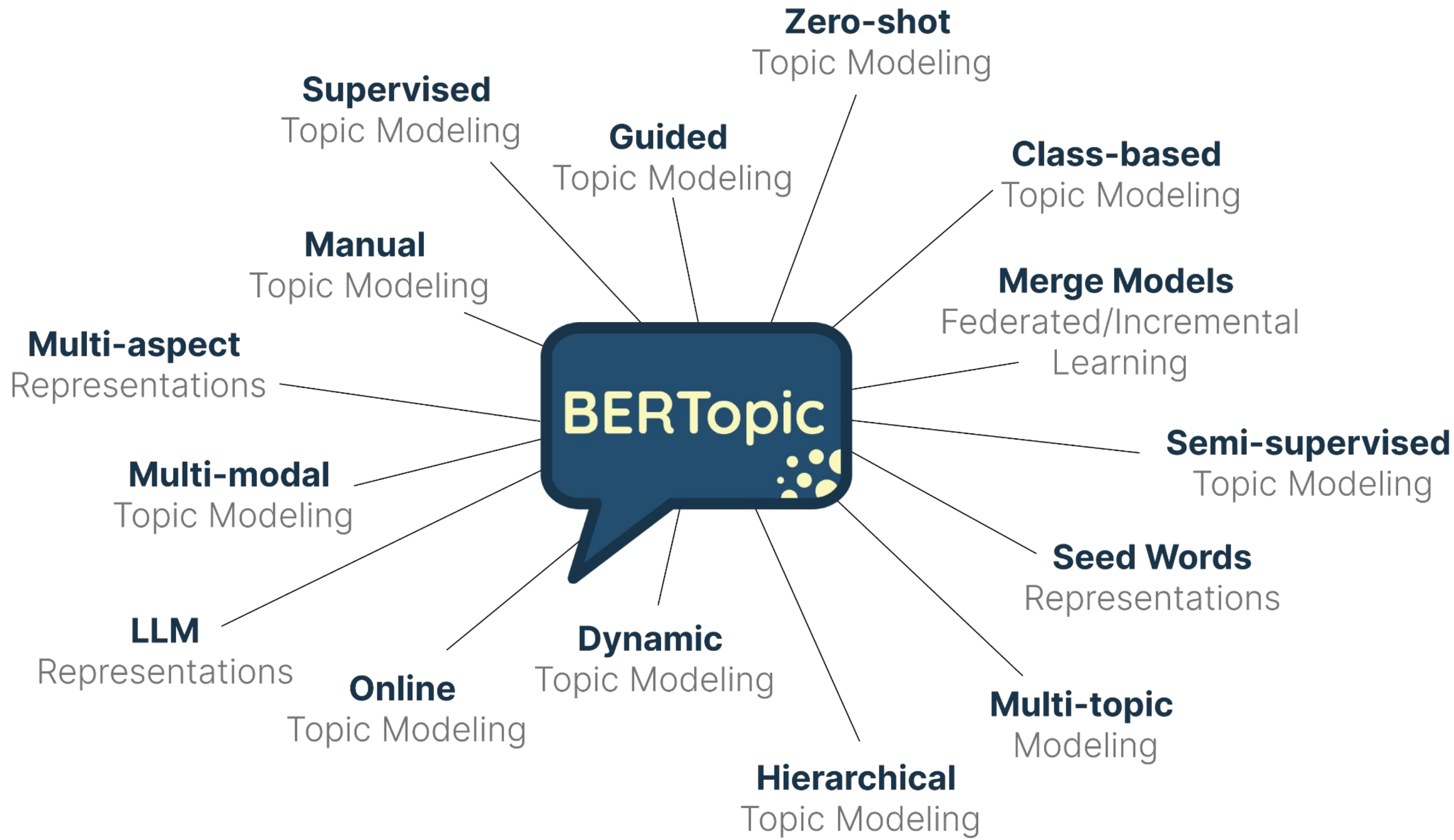
Page Info

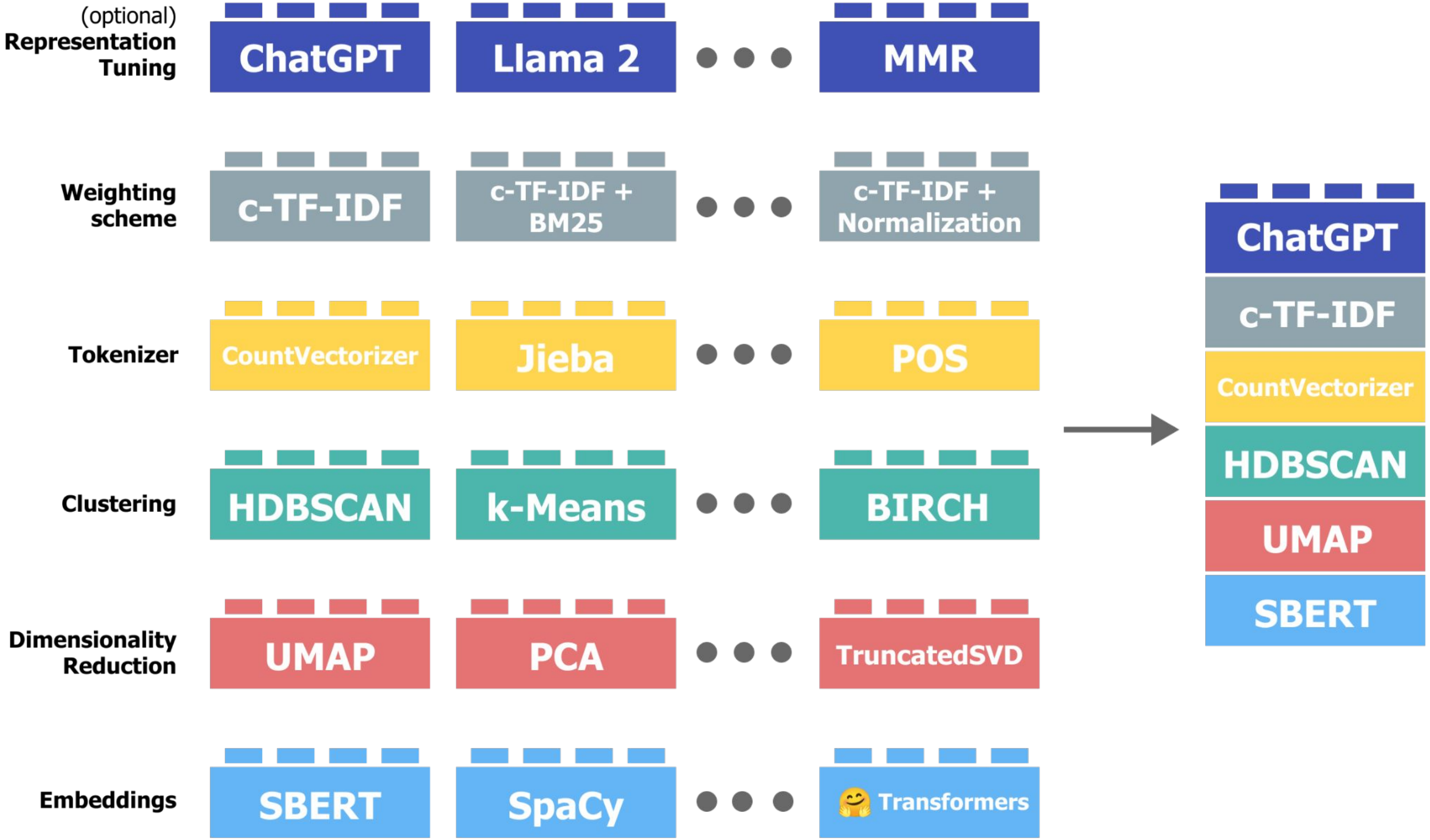
Topic models

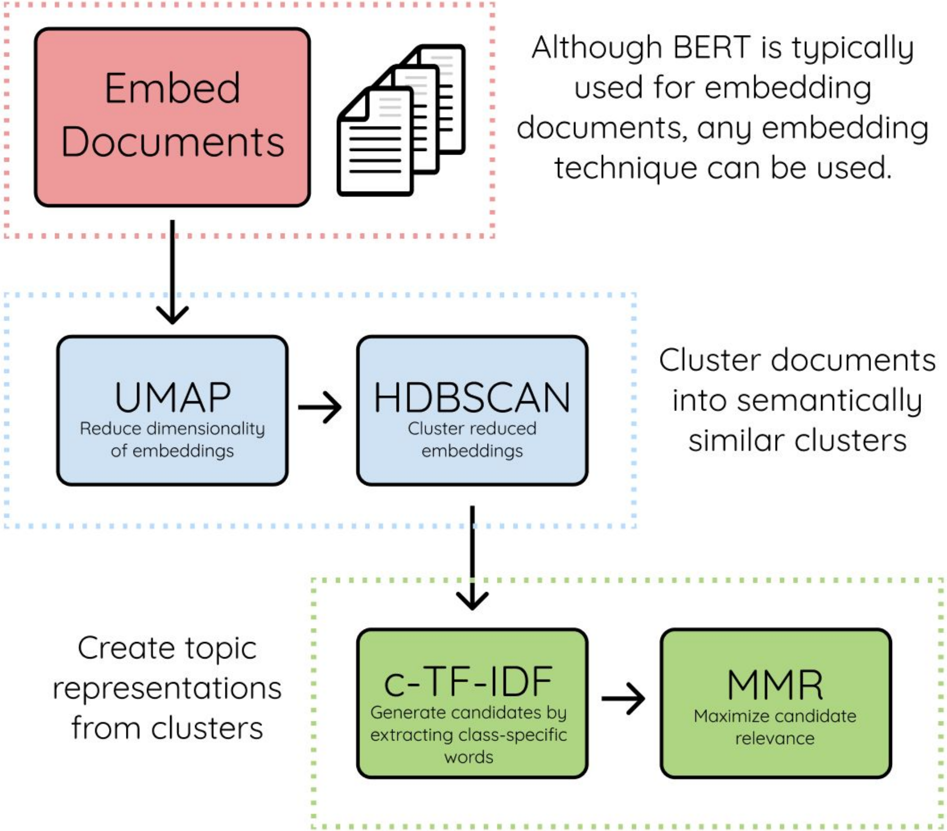
Content Export	Address	normalised title										
			0.00%	0.00%	50.00%	0.00%	0.00%	50.00%	0.00%	0.00%	0.00%	0.00%
			0.00%	0.00%	1.49%	0.37%	0.00%	20.07%	1.86%	33.09%	0.00%	0.00%
			1.33%	0.00%	2.00%	0.36%	5.74%	18.55%	6.20%	10.51%	7.64%	0.00%
			0.00%	0.00%	0.00%	0.00%	0.60%	14.83%	1.40%	38.48%	1.20%	0.00%
			0.00%	48.53%	0.74%	0.00%	0.00%	0.00%	0.00%	0.00%	8.09%	4.04%
			0.00%	0.00%	9.38%	0.00%	0.00%	0.00%	0.00%	3.13%	0.00%	53.13%
			0.00%	3.51%	15.59%	2.10%	0.86%	2.73%	6.63%	5.22%	0.00%	8.96%
			2.99%	0.48%	4.31%	1.08%	4.55%	1.20%	7.19%	6.23%	15.81%	0.96%
			6.50%	3.58%	3.17%	0.41%	6.81%	7.16%	7.98%	4.35%	15.05%	0.10%
			9.84%	7.81%	9.12%	1.74%	3.91%	0.00%	5.64%	9.41%	2.03%	0.00%
			0.09%	1.29%	0.76%	0.09%	25.68%	5.38%	8.90%	5.47%	4.09%	0.00%
			4.12%	1.17%	9.48%	1.58%	1.37%	2.68%	5.36%	4.53%	0.76%	15.52%
			1.40%	0.97%	7.97%	2.37%	3.47%	4.50%	5.36%	4.81%	1.58%	7.61%
			0.00%	2.53%	27.09%	2.61%	2.61%	0.00%	3.15%	6.45%	2.23%	8.21%
			0.00%	49.14%	0.00%	0.00%	0.00%	0.00%	0.69%	0.00%	9.62%	5.84%
			0.00%	5.18%	8.83%	6.91%	7.49%	1.92%	4.41%	15.36%	0.58%	0.58%
			0.00%	0.00%	43.24%	0.00%	0.00%	0.00%	2.70%	0.00%	0.00%	18.92%
			0.00%	19.63%	8.89%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	32.59%
			0.00%	62.77%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
			7.07%	51.52%	0.00%	0.00%	0.00%	0.00%	0.34%	0.00%	5.39%	1.68%
			4.34%	1.42%	5.54%	0.78%	2.35%	0.28%	10.73%	3.98%	0.57%	13.15%
			0.00%	54.29%	2.86%	2.86%	2.86%	0.00%	0.00%	0.00%	0.00%	0.00%
			1.93%	4.16%	12.01%	1.97%	6.22%	13.46%	4.72%	4.25%	2.19%	0.00%
			2.59%	4.25%	3.22%	0.00%	5.97%	3.14%	7.63%	12.74%	12.50%	0.39%
			2.03%	1.92%	1.05%	0.17%	23.55%	8.31%	14.42%	2.56%	3.43%	0.00%

+ ≣ Search Intent Matching ▾ Topic to Topic Similarity ▾ Topic Modelling per Page ▾ Other Linking Opportunities (3N ▾ Cluster 1 ▾ Cluster 2 ▾ Cluster 3

Topic Modelling per Page ▾


BERTopic
Zero-shot Topic Modeling
Supervised Topic Modeling
Guided Topic Modeling
Class-based Topic Modeling
Manual Topic Modeling
Merge Models Federated/Incremental Learning
Multi-aspect Representations
Semi-supervised Topic Modeling
Multi-modal Topic Modeling
Seed Words Representations
LLM Representations
Dynamic Topic Modeling
Online Topic Modeling
Multi-topic Modeling
Hierarchical Topic Modeling


(optional)
Representation Tuning
ChatGPT
Llama 2
MMR
Weighting scheme
c-TF-IDF
c-TF-IDF + BM25
c-TF-IDF + Normalization
Tokenizer
CountVectorizer
Jieba
POS
Clustering
HDBSCAN
k-Means
BIRCH
Dimensionality Reduction
UMAP
PCA
TruncatedSVD
Embeddings
SBERT
SpaCy
Transformers
ChatGPT
c-TF-IDF
CountVectorizer
HDBSCAN
UMAP
SBERT

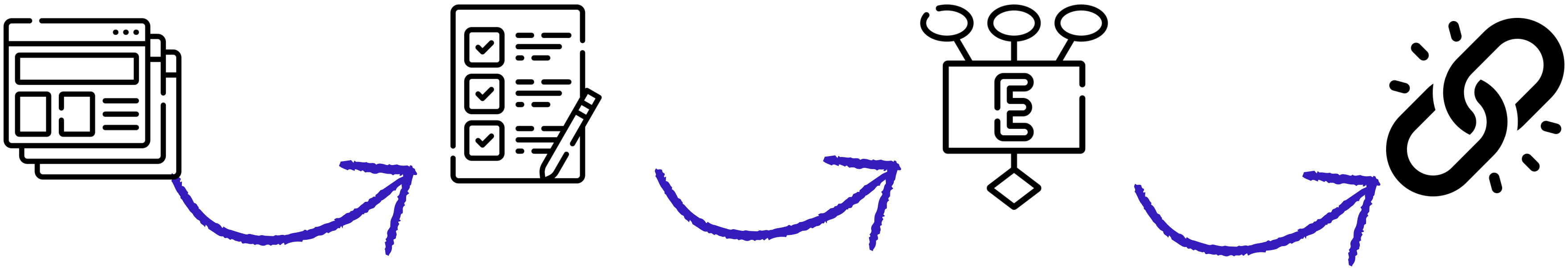


Although BERT is typically used for embedding documents, any embedding technique can be used.

Cluster documents into semantically similar clusters

Create topic representations from clusters

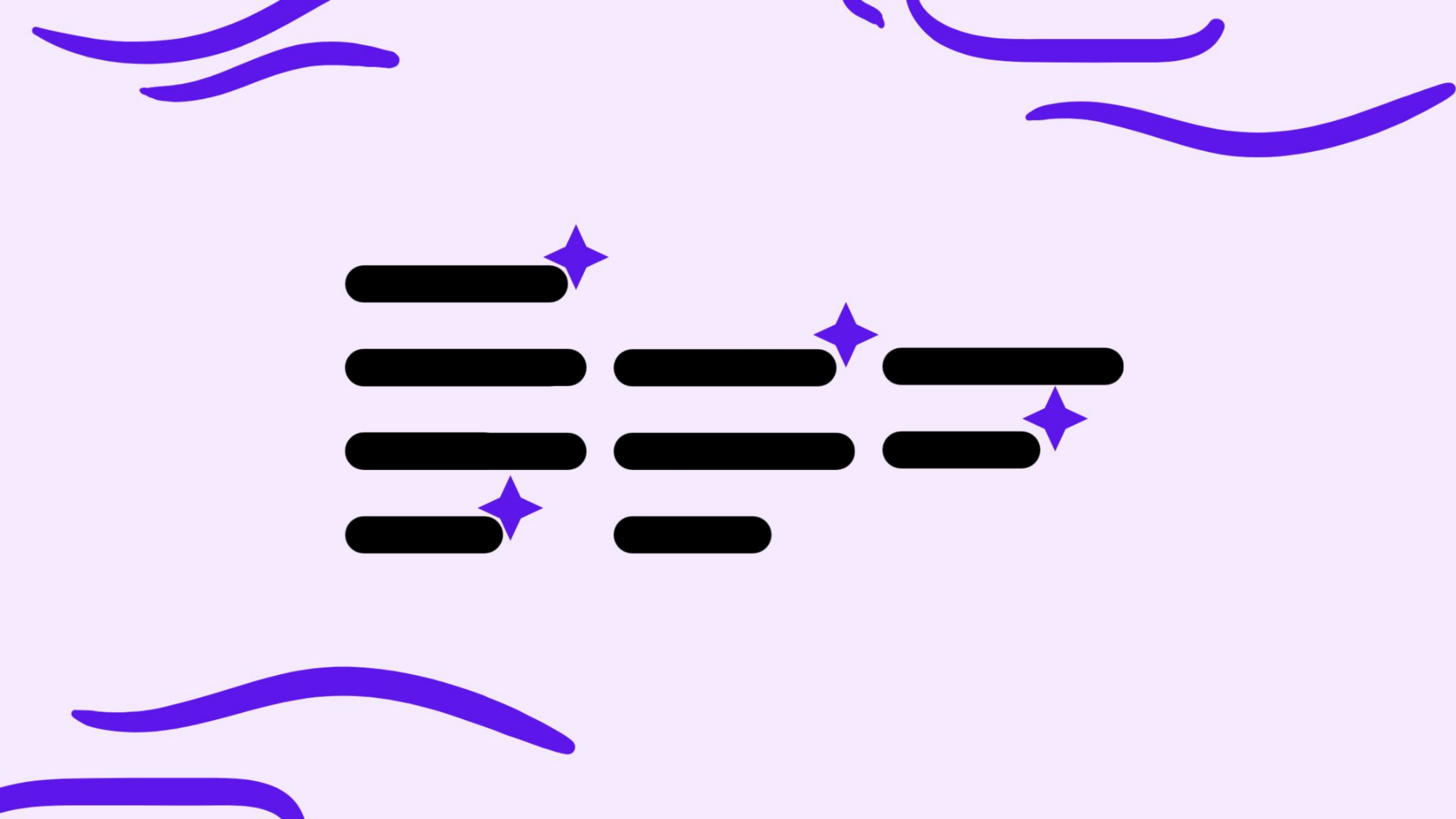
BERTopic

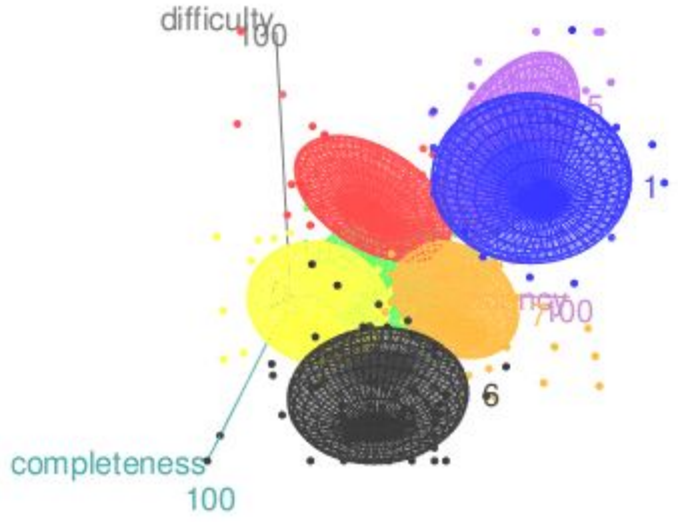


Identify content categories (classification) + topics + subtopics

Extract entities

- Interlink pages that **mention the same subtopic** more than 30%
- Interlink pages mentioning the same or **semantically related entities**
- Improve **content categories and tag** systems

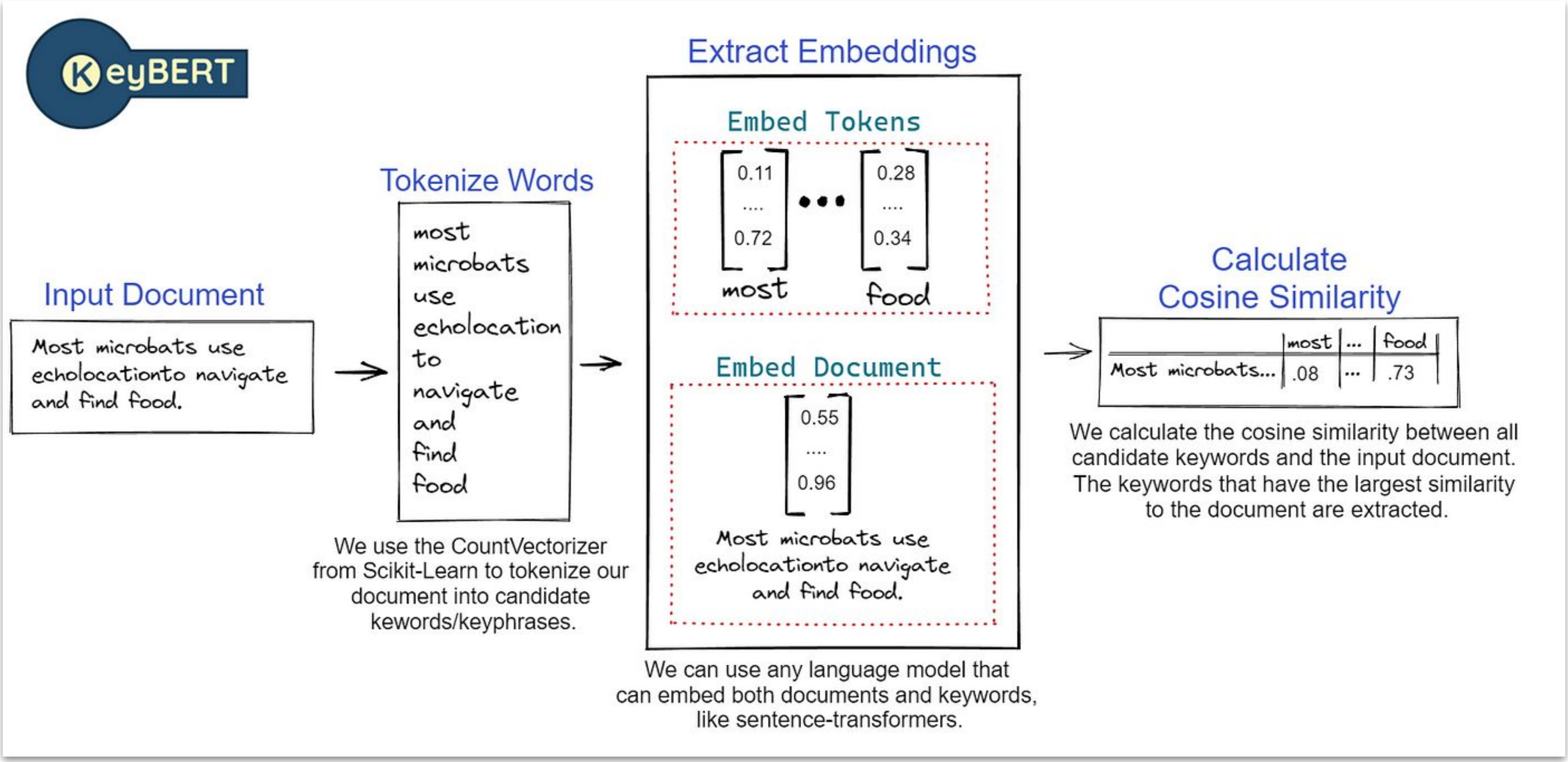


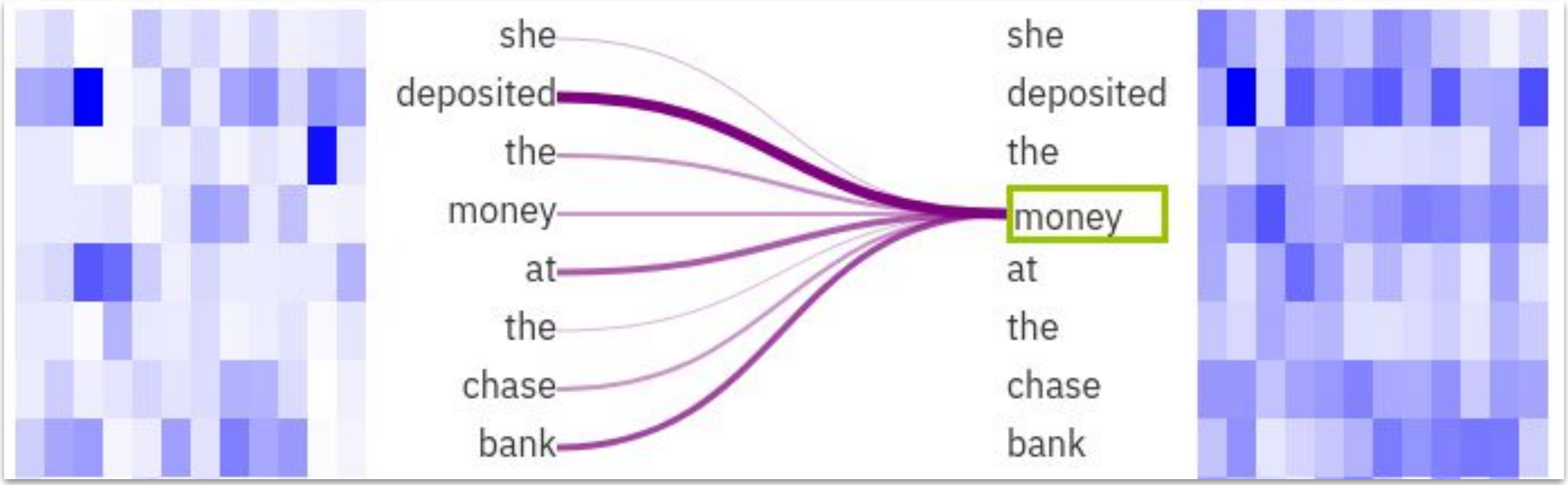


difficulty
100
5
1
100
6
completeness
100

KeyBERT

Although there are already many methods available for keyword generation (e.g., Rake, YAKE!, TF-IDF, etc.) I wanted to create a very basic, but powerful method for extracting keywords and keyphrases. This is where **KeyBERT** comes in! Which uses BERT-embeddings and simple cosine similarity to find the sub-phrases in a document that are the most similar to the document itself.


KeyBERT
Input Document
Most microbats use echolocationto navigate and find food.
Tokenize Words
most
microbats
use
echolocation
to
navigate
and
find
food
We use the CountVectorizer from Scikit-Learn to tokenize our document into candidate kewords/keyphrases.
Extract Embeddings
Embed Tokens
0.11
....
0.72
most
0.28
....
0.34
food
Embed Document
0.55
....
0.96
Most microbats use echolocationto navigate and find food.
We can use any language model that can embed both documents and keywords, like sentence-transformers.
Calculate Cosine Similarity
most
...
food
Most microbats...
.08
...
.73
We calculate the cosine similarity between all candidate keywords and the input document. The keywords that have the largest similarity to the document are extracted.


she
deposited
the
money
at
the
chase
bank
she
deposited
the
money
at
the
chase
bank

www.youtube.com › watch

How to Easily Find Keywords in a Document with KeyBERT in ...

... **analysis**. Extracting **Keywords** with **KeyBERT**: Dive into the code to extract **keywords** from each text using the **KeyBERT** model. Conclusion ...

YouTube · Python Tutorials for Digital Humanities · Aug 21, 2023

Great intro video for beginners

```python
#Basic usage - keyword extraction
from keybert import KeyBERT

doc = """
         Supervised learning is the machine learning task of learning a function that
         maps an input to an output based on example input-output pairs. It infers a
         function from labeled training data consisting of a set of training examples.
         In supervised learning, each example is a pair consisting of an input object
         (typically a vector) and a desired output value (also called the supervisory signal).
         A supervised learning algorithm analyzes the training data and produces an inferred function,
         which can be used for mapping new examples. An optimal scenario will allow for the
         algorithm to correctly determine the class labels for unseen instances. This requires
         the learning algorithm to generalize from the training data to unseen situations in a
         'reasonable' way (see inductive bias).
      """
kw_model = KeyBERT()
keywords = kw_model.extract_keywords(doc)
```

```python
#n-gram specified keyword extraction
kw_model.extract_keywords(doc, keyphrase_ngram_range=(1, 1), stop_words=None)
[('learning', 0.4604),
 ('algorithm', 0.4556),
 ('training', 0.4487),
 ('class', 0.4086),
 ('mapping', 0.3700)]
```

```
[('learning', 0.4604),
 ('algorithm', 0.4556),
 ('training', 0.4487),
 ('class', 0.4086),
 ('mapping', 0.37)]
```

```python
#Basic usage - keyword extraction
from keybert import KeyBERT

doc = """
         Supervised learning is the machine learning task of learning a function that
         maps an input to an output based on example input-output pairs. It infers a
         function from labeled training data consisting of a set of training examples.
         In supervised learning, each example is a pair consisting of an input object
         (typically a vector) and a desired output value (also called the supervisory signal).
         A supervised learning algorithm analyzes the training data and produces an inferred function,
         which can be used for mapping new examples. An optimal scenario will allow for the
         algorithm to correctly determine the class labels for unseen instances. This requires
         the learning algorithm to generalize from the training data to unseen situations in a
         'reasonable' way (see inductive bias).
      """
kw_model = KeyBERT()
keywords = kw_model.extract_keywords(doc)
```

```python
#n-gram specified keyword extraction
kw_model.extract_keywords(doc, keyphrase_ngram_range=(2, 2), top_n=10, stop_words=None)
```

```
[('supervised learning', 0.6779),
 ('signal supervised', 0.6152),
 ('in supervised', 0.6124),
 ('labeled training', 0.6013),
 ('learning function', 0.5755),
 ('learning algorithm', 0.5632),
 ('machine learning', 0.5555),
 ('training data', 0.5271),
 ('learning task', 0.5121),
 ('training examples', 0.4668)]
```

```python
#Basic usage - keyword extraction
from keybert import KeyBERT

doc = """
         Supervised learning is the machine learning task of learning a function that
         maps an input to an output based on example input-output pairs. It infers a
         function from labeled training data consisting of a set of training examples.
         In supervised learning, each example is a pair consisting of an input object
         (typically a vector) and a desired output value (also called the supervisory signal).
         A supervised learning algorithm analyzes the training data and produces an inferred function,
         which can be used for mapping new examples. An optimal scenario will allow for the
         algorithm to correctly determine the class labels for unseen instances. This requires
         the learning algorithm to generalize from the training data to unseen situations in a
         'reasonable' way (see inductive bias).
      """
kw_model = KeyBERT()
keywords = kw_model.extract_keywords(doc)
```

```python
#highlight keywords in the document
keywords = kw_model.extract_keywords(doc, highlight=True)
```

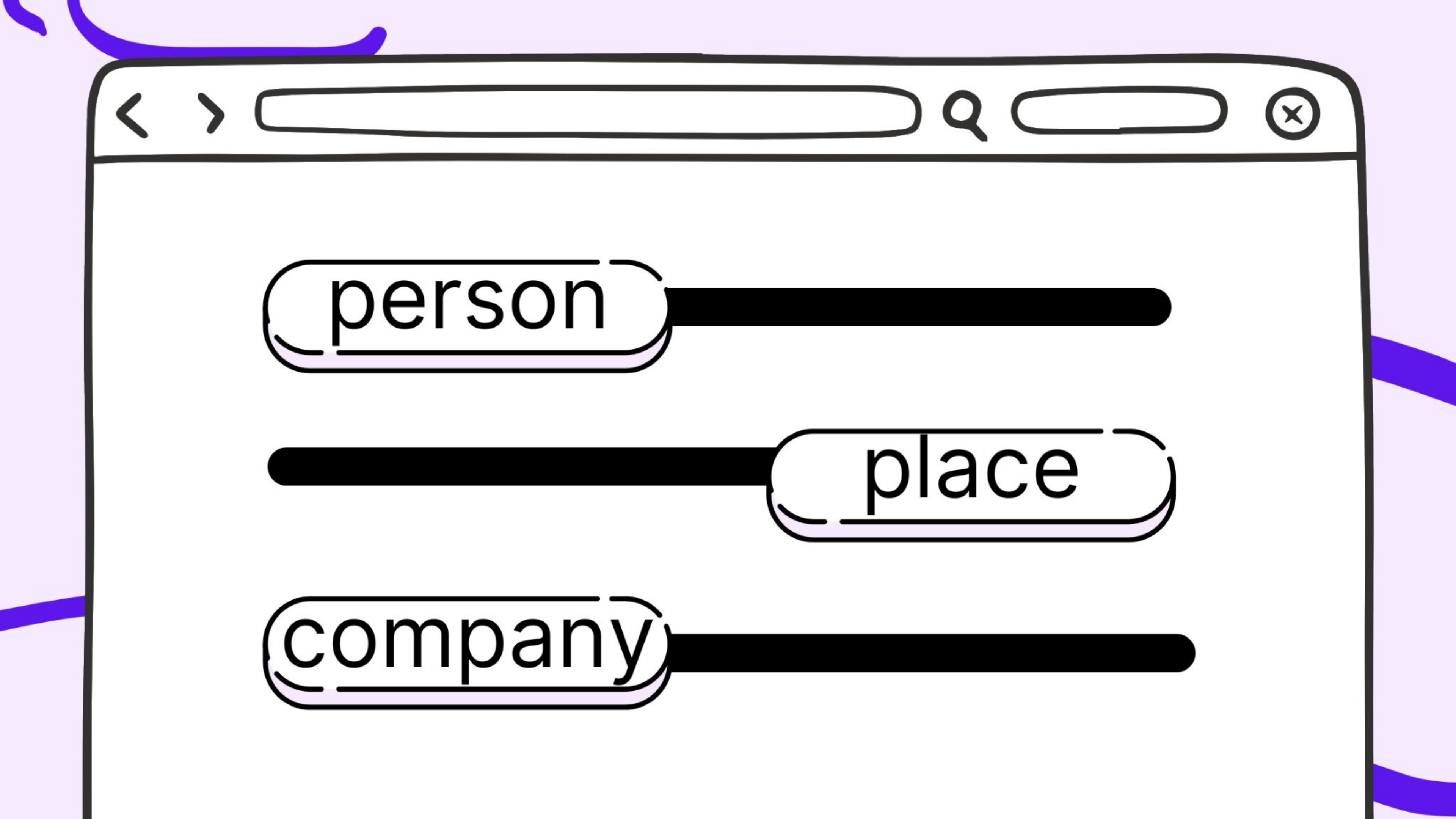
Supervised learning is the machine learning task of learning function that maps an input to an output based on example input output pairs It infers function from labeled training data consisting of set of training examples In supervised learning each example is pair consisting of an input object typically vector and desired output value also called the supervisory signal supervised learning algorithm analyzes the training data and produces an inferred function which can be used for mapping new examples An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances This requires the learning algorithm to generalize from the training data to unseen situations in reasonable way see inductive bias

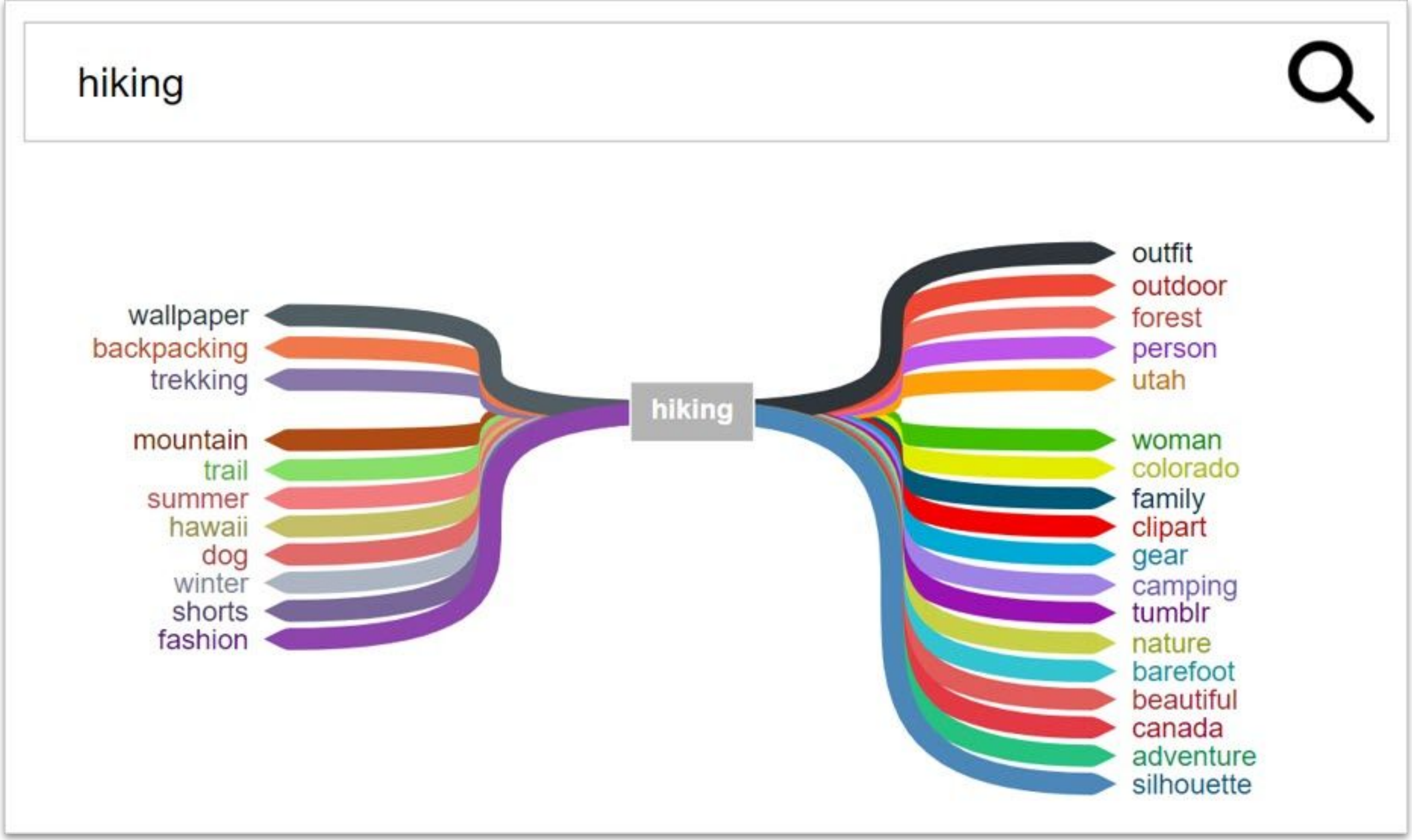
...but what you want to do is cluster keywords in sheets

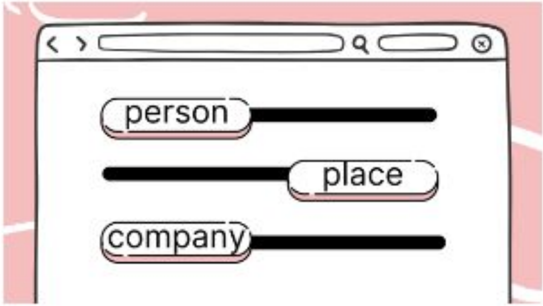
```python
# Function to apply KeyBERT on the 'Keywords' column
def apply_keybert(df):
    if 'Keywords' not in df.columns:
        print("Error: The dataframe must contain a column named 'Keywords'.")
        return None

    # Create new columns for unigrams and bigrams
    def extract_ngram(text, ngram_range):
        # Extract keywords with specified ngram range, handle the case where no keywords are found
        keywords = kw_model.extract_keywords(text, keyphrase_ngram_range=ngram_range, stop_words='english')
        return keywords[0][0] if keywords else ""  # Return the keyword or an empty string if none found

    # Apply to the 'Keywords' column
    df['Core (1-gram)'] = df['Keywords'].apply(lambda x: extract_ngram(x, (1, 1)) if len(x) > 0 else "")
    df['Core (2-gram)'] = df['Keywords'].apply(lambda x: extract_ngram(x, (2, 2)) if len(x) > 0 else "")

    return df

# Main function to upload the file and apply the transformations
def main():
    df = load_dataframe()

    if df is not None:
        # Apply KeyBERT to extract keywords
        df_with_keybert = apply_keybert(df)

        if df_with_keybert is not None:
            # Show the modified dataframe
            print(df_with_keybert.head())
            # Save the modified dataframe to a new CSV
            df_with_keybert.to_csv('keywords_with_keybert.csv', index=False)
            print("File saved as 'keywords_with_keybert.csv'.")
            files.download('keywords_with_keybert.csv')
```

	A	B	C
1	**Keywords**		
1670	sustainable energy articles		
1671	sustainable energy business		
1672	sustainable energy business ideas		
1673	sustainable energy companies to invest in		
1674	sustainable energy finance		
1675	sustainable energy futures		
1676	sustainable energy ideas		
1677	sustainable energy industry		
1678	sustainable energy investment funds		
1679	sustainable energy investment funds		
1680	sustainable energy production ideas		
1681	sustainable energy products		
1682	sustainable energy sector		
1683	sustainable engineering projects		
1684	sustainable esg investing		
1685	sustainable fashion business		
1686	sustainable fashion business ideas		
1687	sustainable fashion business plan		
1688	sustainable finance and investment		
1689	sustainable finance podcast		
1690	sustainable financial investments		
1691	sustainable food business		

A	B	C
Keywords	**Core (1-gram)**	**Core (2-gram)**
sustainable energy articles	sustainable	sustainable energy
sustainable energy business	sustainable	energy business
sustainable energy business ideas	sustainable	energy business
sustainable energy companies to invest in	sustainable	energy companies
sustainable energy finance	sustainable	energy finance
sustainable energy futures	sustainable	energy futures
sustainable energy ideas	sustainable	sustainable energy
sustainable energy industry	sustainable	sustainable energy
sustainable energy investment funds	investment	energy investment
sustainable energy investment funds	investment	energy investment
sustainable energy production ideas	sustainable	sustainable energy
sustainable energy products	sustainable	sustainable energy
sustainable energy sector	sustainable	sustainable energy
sustainable engineering projects	sustainable	sustainable engineering
sustainable esg investing	esg	esg investing
sustainable fashion business	fashion	sustainable fashion
sustainable fashion business ideas	fashion	sustainable fashion
sustainable fashion business plan	fashion	sustainable fashion
sustainable finance and investment	investment	sustainable finance
sustainable finance podcast	podcast	finance podcast
sustainable financial investments	investments	sustainable financial
sustainable food business	sustainable	sustainable food


person
place
company


hiking
wallpaper
backpacking
trekking
mountain
trail
summer
hawaii
dog
winter
shorts
fashion
hiking
outfit
outdoor
forest
person
utah
woman
colorado
family
clipart
gear
camping
tumblr
nature
barefoot
beautiful
canada
adventure
silhouette



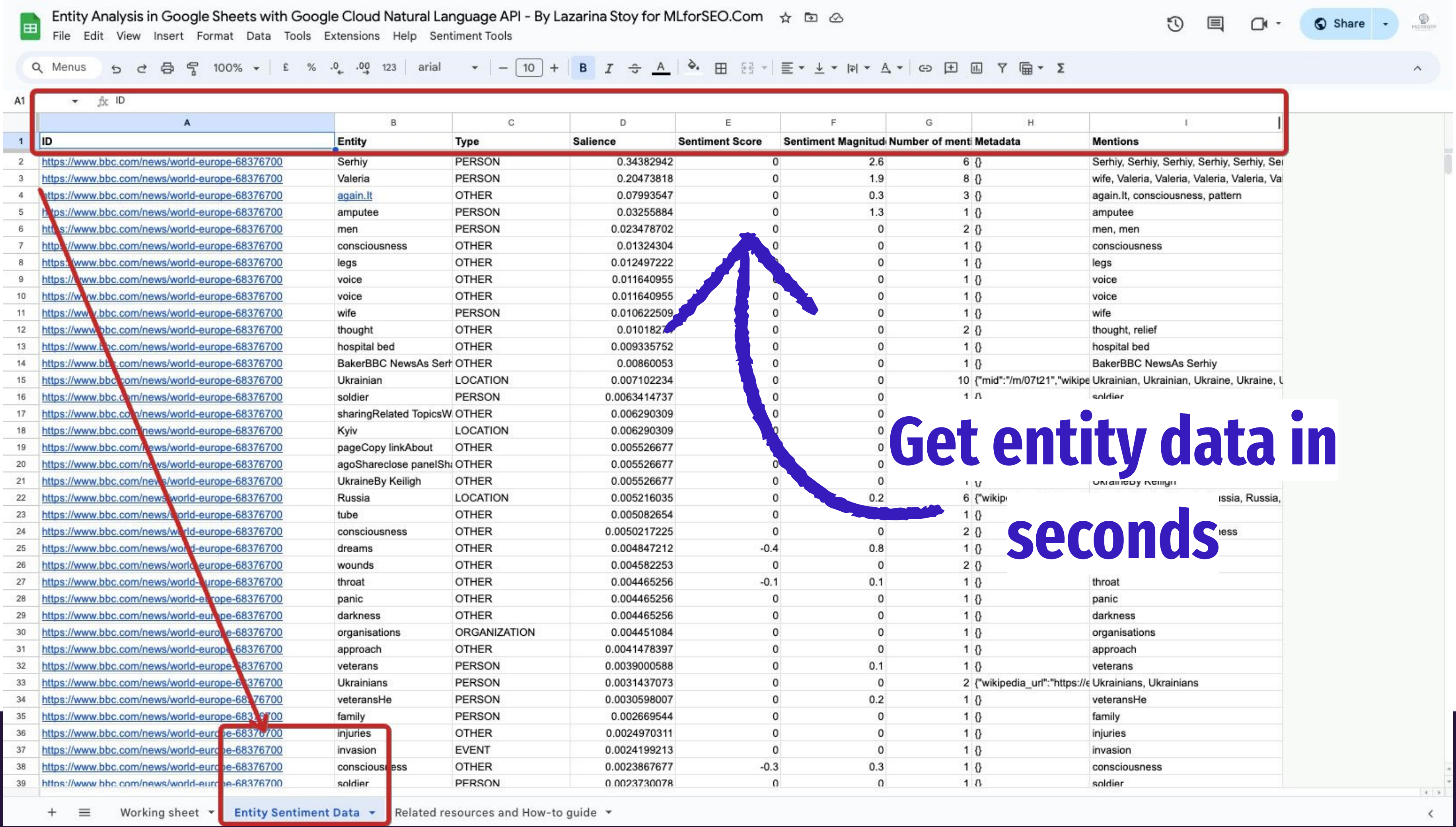
Entity analysis

How to do Entity Extraction with Google's Natural Language API in Google Sheets (Apps Script)

Lazarina Stoy. · Mar 27, 2024

Process will take no more than 20 minutes

Run the script via the menu to extract entities from content



	A	B	C	D	E	F	G	H	I
1	ID	Entity	Type	Salience	Sentiment Score	Sentiment Magnitud	Number of ment	Metadata	Mentions
2	https://www.bbc.com/news/world-europe-68376700	Serhiy	PERSON	0.34382942	0	2.6	6	{}	Serhiy, Serhiy, Serhiy, Serhiy, Serhiy, Ser
3	https://www.bbc.com/news/world-europe-68376700	Valeria	PERSON	0.20473818	0	1.9	8	{}	wife, Valeria, Valeria, Valeria, Valeria, Va
4	https://www.bbc.com/news/world-europe-68376700	again.It	OTHER	0.07993547	0	0.3	3	{}	again.It, consciousness, pattern
5	https://www.bbc.com/news/world-europe-68376700	amputee	PERSON	0.03255884	0	1.3	1	{}	amputee
6	https://www.bbc.com/news/world-europe-68376700	men	PERSON	0.023478702	0	0	2	{}	men, men
7	https://www.bbc.com/news/world-europe-68376700	consciousness	OTHER	0.01324304	0	0	1	{}	consciousness
8	https://www.bbc.com/news/world-europe-68376700	legs	OTHER	0.012497222		0	1	{}	legs
9	https://www.bbc.com/news/world-europe-68376700	voice	OTHER	0.011640955		0	1	{}	voice
10	https://www.bbc.com/news/world-europe-68376700	voice	OTHER	0.011640955	0	0	1	{}	voice
11	https://www.bbc.com/news/world-europe-68376700	wife	PERSON	0.010622509	0	0	1	{}	wife
12	https://www.bbc.com/news/world-europe-68376700	thought	OTHER	0.0101827	0	0	2	{}	thought, relief
13	https://www.bbc.com/news/world-europe-68376700	hospital bed	OTHER	0.009335752	0	0	1	{}	hospital bed
14	https://www.bbc.com/news/world-europe-68376700	BakerBBC NewsAs Serh	OTHER	0.00860053	0	0	1	{}	BakerBBC NewsAs Serhiy
15	https://www.bbc.com/news/world-europe-68376700	Ukrainian	LOCATION	0.007102234	0	0	10	{"mid":"/m/07t21","wikipe	Ukrainian, Ukrainian, Ukraine, Ukraine, U
16	https://www.bbc.com/news/world-europe-68376700	soldier	PERSON	0.0063414737	0	0	1	{}	soldier
17	https://www.bbc.com/news/world-europe-68376700	sharingRelated TopicsW	OTHER	0.006290309	0	0			
18	https://www.bbc.com/news/world-europe-68376700	Kyiv	LOCATION	0.006290309		0			
19	https://www.bbc.com/news/world-europe-68376700	pageCopy linkAbout	OTHER	0.005526677		0			
20	https://www.bbc.com/news/world-europe-68376700	agoShareclose panelSh	OTHER	0.005526677	0	0			
21	https://www.bbc.com/news/world-europe-68376700	UkraineBy Keiligh	OTHER	0.005526677	0	0	1	{}	UkraineBy Keiligh
22	https://www.bbc.com/news/world-europe-68376700	Russia	LOCATION	0.005216035	0	0.2	6	{"wikip	ssia, Russia,
23	https://www.bbc.com/news/world-europe-68376700	tube	OTHER	0.005082654	0	0	1	{}	
24	https://www.bbc.com/news/world-europe-68376700	consciousness	OTHER	0.0050217225	0	0	2	{}	ness
25	https://www.bbc.com/news/world-europe-68376700	dreams	OTHER	0.004847212	-0.4	0.8	1	{}	
26	https://www.bbc.com/news/world-europe-68376700	wounds	OTHER	0.004582253	0	0	2	{}	
27	https://www.bbc.com/news/world-europe-68376700	throat	OTHER	0.004465256	-0.1	0.1	1	{}	throat
28	https://www.bbc.com/news/world-europe-68376700	panic	OTHER	0.004465256	0	0	1	{}	panic
29	https://www.bbc.com/news/world-europe-68376700	darkness	OTHER	0.004465256	0	0	1	{}	darkness
30	https://www.bbc.com/news/world-europe-68376700	organisations	ORGANIZATION	0.004451084	0	0	1	{}	organisations
31	https://www.bbc.com/news/world-europe-68376700	approach	OTHER	0.0041478397	0	0	1	{}	approach
32	https://www.bbc.com/news/world-europe-68376700	veterans	PERSON	0.0039000588	0	0.1	1	{}	veterans
33	https://www.bbc.com/news/world-europe-68376700	Ukrainians	PERSON	0.0031437073	0	0	2	{"wikipedia_url":"https://e	Ukrainians, Ukrainians
34	https://www.bbc.com/news/world-europe-68376700	veteransHe	PERSON	0.0030598007	0	0.2	1	{}	veteransHe
35	https://www.bbc.com/news/world-europe-68376700	family	PERSON	0.002669544	0	0	1	{}	family
36	https://www.bbc.com/news/world-europe-68376700	injuries	OTHER	0.0024970311	0	0	1	{}	injuries
37	https://www.bbc.com/news/world-europe-68376700	invasion	EVENT	0.0024199213	0	0	1	{}	invasion
38	https://www.bbc.com/news/world-europe-68376700	consciousness	OTHER	0.0023867677	-0.3	0.3	1	{}	consciousness
39	https://www.bbc.com/news/world-europe-68376700	soldier	PERSON	0.0023730078	0	0	1	{}	soldier

Get entity data in seconds

- Your content
- competitors’ content
- Your competitors’ YouTube content
- your SERP data versus competitors’
- first-party data
- UGC
- social mentions
- Your internal link text anchors

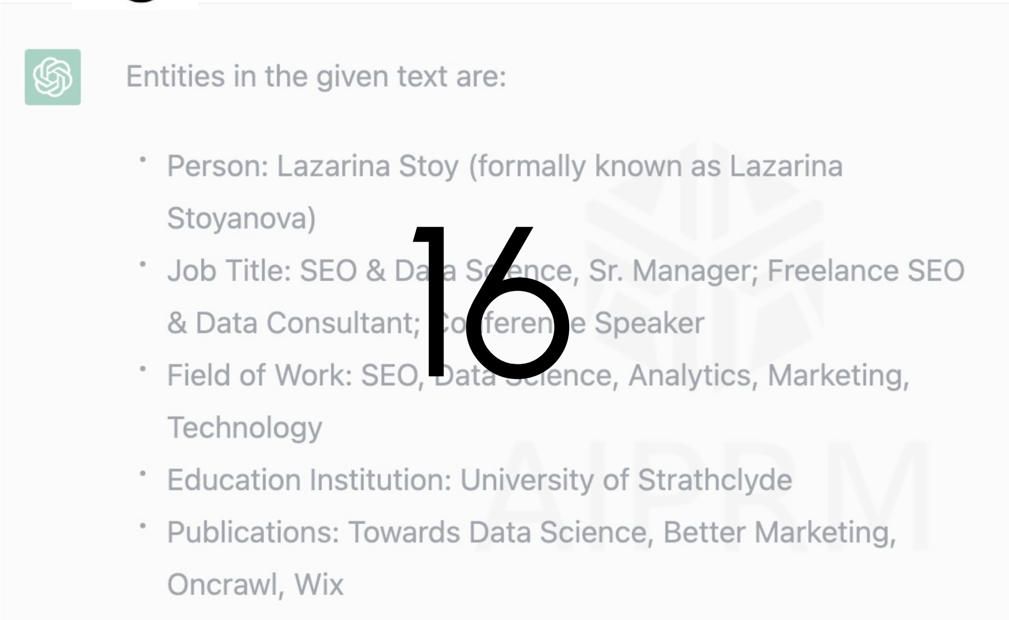
…

Possible data points

Does it matter *how* you do it?

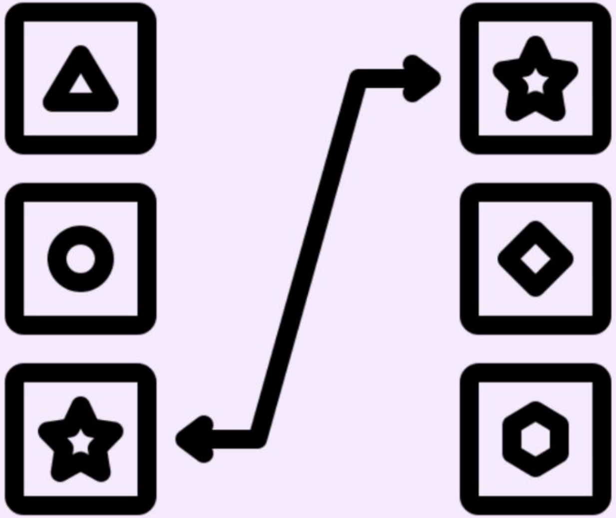
Yes, use a tailored model.

53



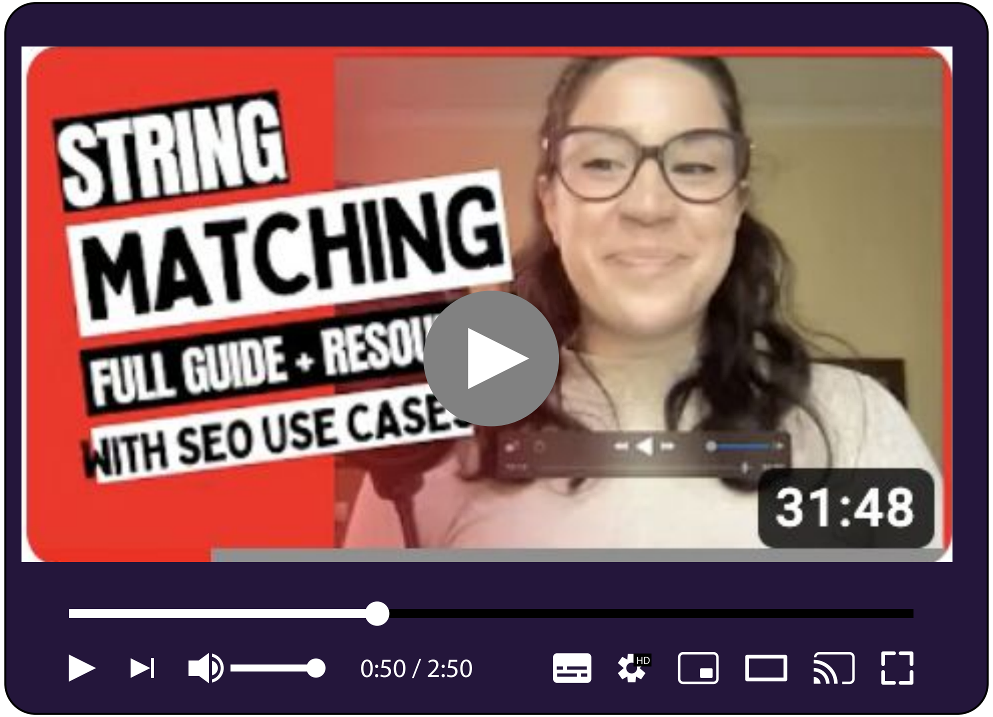
16

Model	Benefits	Limitations
Google Cloud Natural Language API	• Recognizes entities and provides a score for prominence, importance, number of mentions • Syntax analysis, including dependency trees, part-of-speech tagging • Sentiment analysis in entities (in context), and of entire documents or texts • Scalable • Easy to use with multiple integrations possible (including Google Sheets)	• Overstuffing of entities recognised – e.g., singular and plural forms
GPT-4	• Great for one-offs and quick analyses • Can identify entities with some limitations • Can do syntax analysis with some limitations	• Prone to hallucinations, e.g., will pull out words or entities that are not in the text • Limited entity recognition and syntax analysis • Sometimes false category attribution • Limited scalability • Much slower comparatively



Fuzzy matching is a quick and dirty way for calculating

similarity between two strings



WATCH THE DETAILS LATER

I've recorded a step-by-step tutorial on using **fuzzy matching** for things like:

- Identifying link opportunities
- String Similarity Analysis
- redirect mapping of URLs

String similarity & redirect mapping

H1 match

URL	URL2	Similarity
https://www.example.com/destinations/egypt/	https://www.example.com/egypt/	0.9333
https://www.example.com/destinations/italy/rome/	https://www.example.com/italy/rome	0.8667
https://www.example.com/destinations/portugal/algarve/	https://www.example.com/portugal/algarve/	0.9333
https://www.example.com/destinations/spain/	https://www.example.com/spain/	0.9333
https://www.example.com/destinations/spain/balearics/ibiza/	https://www.example.com/spain/balearics/ibiza/	0.9333
https://www.example.com/destinations/spain/costa-brava/benidorm/	https://www.example.com/spain/costa-brava/benidorm/	0.9333

Page title match

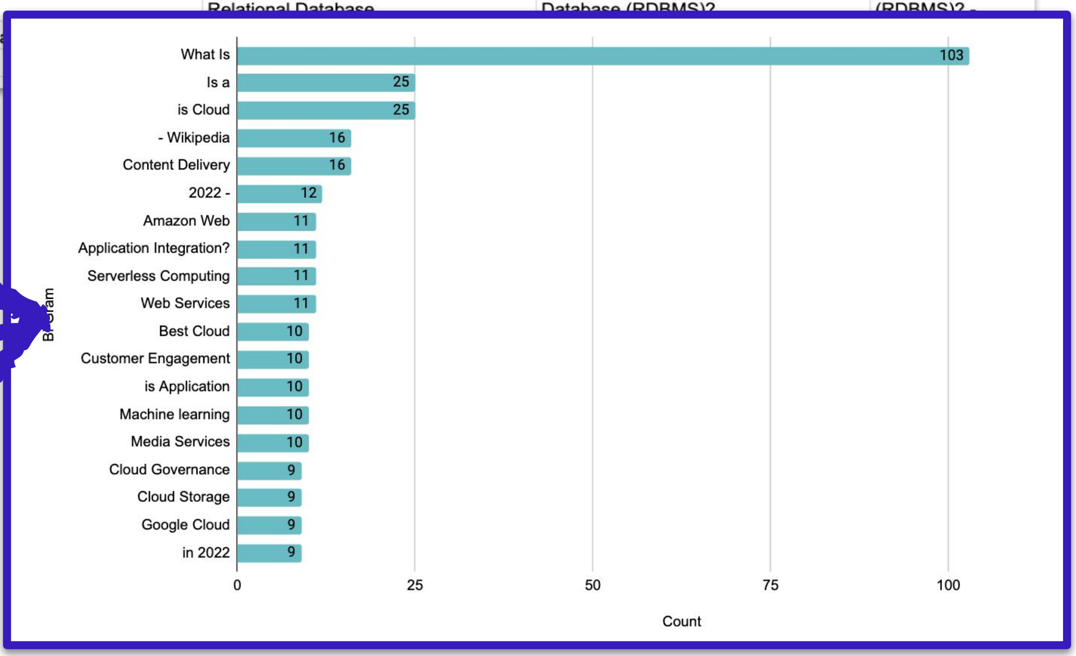
URL	URL2	Similarity
https://www.example.com/destinations/egypt/	https://www.example.com/egypt/	0.9091
https://www.example.com/destinations/italy/rome/	https://www.example.com/italy/rome	0.9231
https://www.example.com/destinations/portugal/algarve/	https://www.example.com/portugal/algarve/	0.9231
https://www.example.com/destinations/spain/	https://www.example.com/spain/	0.9000
https://www.example.com/destinations/spain/balearics/ibiza/	https://www.example.com/spain/balearics/ibiza/	0.9286
https://www.example.com/destinations/spain/costa-brava/benidorm/	https://www.example.com/spain/costa-brava/benidorm/	0.9375

URL match

URL	URL2	Similarity
https://www.example.com/destinations/egypt/	https://www.example.com/egypt/	0.8745
https://www.example.com/destinations/italy/rome/	https://www.example.com/italy/rome	0.9111
https://www.example.com/destinations/portugal/algarve/	https://www.example.com/portugal/algarve/	0.8745
https://www.example.com/destinations/spain/	https://www.example.com/spain/	0.8820
https://www.example.com/destinations/spain/balearics/ibiza/	https://www.example.com/spain/balearics/ibiza/	0.8910
https://www.example.com/destinations/spain/costa-brava/benidorm/	https://www.example.com/spain/costa-brava/benidorm/	0.9158

```
=transpose(getngrams(F2, 2))
```

I	J	K	L	M	
bi-grams in title					
Enterprise application	application integration	integration -	- Wikipedia		
What Is	Is a	a Database	Database -	- Oracle	
What is	is a	a Relational	Relational Database	Database (RDBMS)?	(RDBMS)? -
What Is	Is A	A Non-Relationa			
What is	is a	a CDN?			



N-gram analysis:
Understanding language used within the high-performing articles for your terms can be beneficial for building content briefs.

A more advanced use case of n-grams and language analysis to identify opportunities for Structured Data

From blog post

Table Of Contents

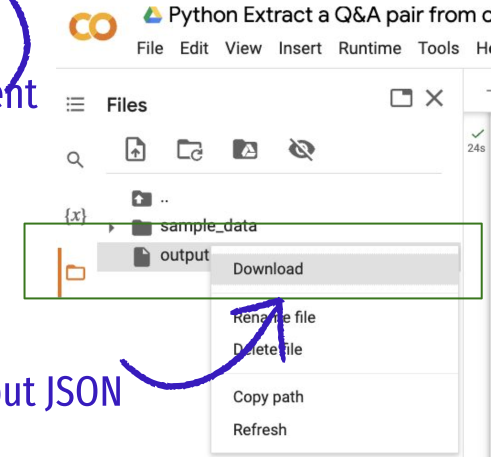
- A beginner-friendly SEO guide, based on blogs that convert
- Link your blog from the main navigation and other important menus
    - Why should you link your blog from the main menu?
    - Why should you link your blog from the footer menu?
- Choose a blog location, which promotes site authority
    - What should you choose – hosting the blog on a subdomain or subfolder?
    - Should you include tags and category names in your blog URLs?
- Use a mix of broad and specific topic tags, keep them relevant to user intent
    - What are the benefits of using a mix of broad and specific topic tags in a blog?
    - How many categories and tags should a blog have?
- Use titles and headings to increase CTR and blog engagement
    - How to optimally use the H1 heading tag?
- Improve User Experience with indicators for reading time, content difficulty, and content type
    - Should you publish news and other content (e.g. press releases) as part of your blog?
- Takeaway

+ Code + Text

```python
# Tokenize text into sentences
sentences = nltk.sent_tokenize(text)

# Initialize FAQ Schema dictionary
faq_schema = {'@context': 'https://schema.org', '@type': 'FAQPage', 'mainEntity': []}

# Loop through sentences to find questions and their answers
for i in range(len(sentences)):
    if is_question(sentences[i]):
        # Extract question
        question = re.sub(r'[^\w\s]', '', sentences[i]).strip()
        # Extract answer
        j = i+1
        answer = ''
        while j < len(sentences) and not is_question(sentences[j]):
            answer += ' ' + sentences[j]
            j += 1
        answer = answer.strip()
        # Add question-answer pair to FAQ Schema
        faq_schema['mainEntity'].append({'@type': 'Question', 'name': question, 'acceptedAnswer': {'@type': 'Answer', 'text': answer}})

# Save FAQ Schema to JSON file
with open('output.json', 'w') as f:
    json.dump(faq_schema, f, indent=4)
print(faq_schema)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
Paste the webpage content: ch engines and visitors.
```

Run the cell, paste the content



Download the output JSON

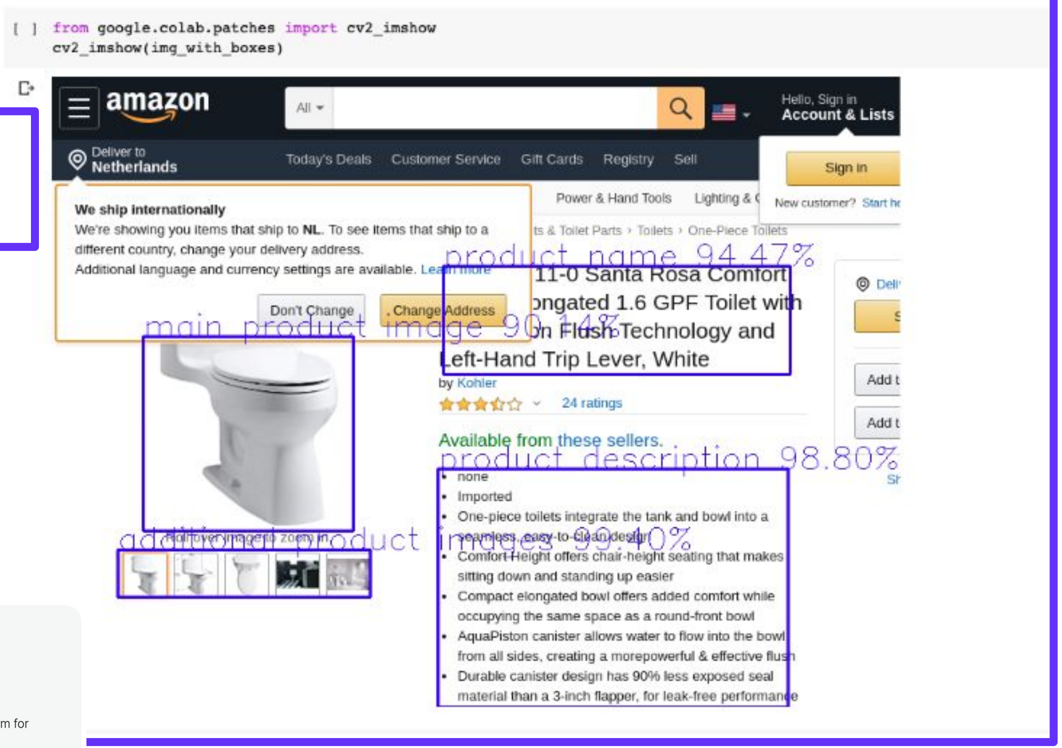
```python
import nltk
nltk.download('punkt')
import re
import json

# Function to check if a sentence is a question that ends with a question mark
def is_question(sentence):
    question_words = ['what', 'when', 'where', 'which', 'who', 'whom', 'whose', 'why', 'how', 'is', 'will', 'should']
    if sentence.split()[0].lower() in question_words and sentence.strip().endswith('?'):
        return True
    else:
        return False


# Read text file
text =input("Paste the webpage content: ")

# Tokenize text into sentences
sentences = nltk.sent_tokenize(text)

# Initialize FAQ Schema dictionary
faq_schema = {'@context': 'https://schema.org', '@type': 'FAQPage', 'mainEntity': []}

# Loop through sentences to find questions and their answers
for i in range(len(sentences)):
    if is_question(sentences[i]):
        # Extract question
        question = re.sub(r'[^\w\s]', '', sentences[i]).strip()
        # Extract answer
        j = i+1
        answer = ''
        while j < len(sentences) and not is_question(sentences[j]):
            answer += ' ' + sentences[j]
            j += 1
        answer = answer.strip()
        # Add question-answer pair to FAQ Schema
        faq_schema['mainEntity'].append({'@type': 'Question', 'name': question, 'acceptedAnswer': {'@type': 'Answer', 'text': answer}})

# Save FAQ Schema to JSON file
with open('output.json', 'w') as f:
    json.dump(faq_schema, f, indent=4)
print(faq_schema)
```


```json
{
    "@context": "https://schema.org",
    "@type": "FAQPage",
    "mainEntity": [
        {
            "@type": "Question",
            "name": "Why should you link your blog from the main menu",
            "acceptedAnswer": {
                "@type": "Answer",
                "text": "The main navigation is typically the primary point of entry for website users and should contain all significant sections, includi
            }
        },
        {
            "@type": "Question",
            "name": "Why should you link your blog from the footer menu",
            "acceptedAnswer": {
                "@type": "Answer",
                "text": "Placing a link to your blog in the footer ensures accessibility from every page of your website, providing a seamless user experie
            }
        },
        {
            "@type": "Question",
            "name": "What should you choose  hosting the blog on a subdomain or subfolder",
            "acceptedAnswer": {
                "@type": "Answer",
                "text": "One option is to integrate your blog into your website\u2019s domain by creating a subdirectory or otherwise -subfolder, such as y
            }
        },
        {
            "@type": "Question",
            "name": "Should you include tags and category names in your blog URLs",
            "acceptedAnswer": {
                "@type": "Answer",
                "text": "Including tags and category names in your blog URLs can have some benefits, but it may not be necessary in all cases. Including ta
            }
        },
        {
            "@type": "Question",
            "name": "What are the benefits of using a mix of broad and specific topic tags in a blog",
            "acceptedAnswer": {
```


→ Script tokenizes the text, discovers the questions, and pulls the answers

→ Script organises these into a schema dictionary, which is saved as a JSON file

AutoML



VIP CONTRIBUTOR

Hamlet Batista

CEO at RankSense

Hamlet Batista is CEO and founder of RankSense, an agile SEO platform for online retailers and manufacturers. He holds US ...

Read Full Bio

Content Moderation

How to do content moderation with Google's Natural Language API in Google Sheets (Apps Script)

Lazarina Stoy. · May 8, 2024

Syntax analysis

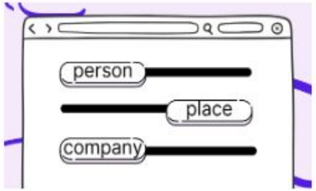
How to do Syntax Analysis with Google's Natural Language API in Google Sheets (Apps Script)

Lazarina Stoy. · Apr 23, 2024

Sentiment Analysis

How to do Sentiment Analysis with Google's Natural Language API in Google Sheets (Apps Script)

Lazarina Stoy. · Apr 22, 2024



Entity analysis

How to do Entity Extraction with Google's Natural Language API in Google Sheets (Apps Script)

Lazarina Stoy. · Mar 27, 2024

Text Classification

How to do Text Classification with Google's Natural Language API in Google Sheets (Apps Script)

Lazarina Stoy. · Mar 27, 2024

.

The Content Moderation API module automatically analyzes text for **inappropriate** or **undesirable** content, helping you to maintain a **clean** and **professional data set** **without manually reviewing each entry**.

But wait...

It can detect if a topic is YMYL 👀

What is YMYL?

YMYL is another acronym from Google's Search Quality Guidelines, which stands for Your Money, Your Life. Examples of YMYL topics or pages are ones that can impact a person's future happiness, health, financial stability, or safety.

Toxic	Insult	Profanity	Derogatory	Sexual	Death, Harm & Tragedy	Violent	Firearms & Weapons
Content that is rude, disrespectful, or unreasonable.	Insulting, inflammatory, or negative comment towards a person or a group of people.	Obscene or vulgar language such as cursing.	Negative or harmful comments targeting identity and/or protected attributes.	Contains references to sexual acts or other lewd content.	Human deaths, tragedies, accidents, disasters, and self-harm.	Describes scenarios depicting violence against an individual or group, or general descriptions of	Content that mentions knives, guns, personal weapons, and accessories such as ammunition, holsters, etc.

Public Safety	Health	Religion & Belief	Illicit Drugs	War & Conflict	Politics	Finance	Legal
Services and organizations that provide relief and ensure public safety.	Human health, including: Health conditions, diseases, and disorders Medical therapies, medication,	Belief systems that deal with the possibility of supernatural laws and beings; religion, faith, belief, spiritual	Recreational and illicit drugs; drug paraphernalia and cultivation, headshops, etc. Includes medicinal use of	War, military conflicts, and major physical conflicts involving large numbers of people. Includes discussion of	Political news and media; discussions of social, governmental, and public policy.	Consumer and business financial services, such as banking, loans, credit, investing, and insurance.	Law-related content, including law firms, legal information, primary legal materials, paralegal

Happiness

Health

Financial stability

Safety

Text Moderation in Google Sheets with Google Cloud Natural Language API - By Lazarina Stoy for MLforSEO.Com

File Edit View Insert Format Data Tools Extensions Help

E19 | 5.4254804%

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	**ID**	**Content**	**Length**	**Toxic**	**Insult**	**Profanity**	**Derogatory**	**Sexual**	**Death, Harm & Tragedy**	**Violent**	**Firearms & Weapons**	**Public Safety**	**Health**	**Religion & Belief**	**Illicit Drugs**	**War & Conflict**	**Politics**	**Finance**	**Legal**
2	*Place a unique identifier here - author name, URL, ID, anything you can use to trace back the comment to a*	*Paste the content y*		Content that is rude, disrespectful, or unreasonable.	Insulting, inflammatory, or negative comment towards a person or a group of people.	Obscene or vulgar language such as cursing.	Negative or harmful comments targeting identity and/or protected attributes.	Contains references to sexual acts or other lewd content.	Human deaths, tragedies, accidents, disasters, and self-harm.	Describes scenarios depicting violence against an individual or group, or general descriptions of	Content that mentions knives, guns, personal weapons, and accessories such as ammunition, holsters, etc.	Services and organizations that provide relief and ensure public safety.	Human health, including: Health conditions, diseases, and disorders Medical therapies, medication,	Belief systems that deal with the possibility of supernatural laws and beings; religion, faith, belief, spiritual	Recreational and illicit drugs; drug paraphernalia and cultivation, headshops, etc. Includes medicinal use of	War, military conflicts, and major physical conflicts involving large numbers of people. Includes discussion of	Political news and media; discussions of social, governmental, and public policy.	Consumer and business financial services, such as banking, loans, credit, investing, and insurance.	Law-related content, including law firms, legal information, primary legal materials, paralegal
3	WebLinkr	I see a lot from SG	364	5.23%	1.89%	1.15%	1.10%	1.17%	8.54%	5.28%	1.82%	6.37%	9.63%	13.70%	5.74%	6.60%	5.99%	10.23%	10.09%
4	Lanky-Football857	So.. I know a bit ab	539	6.36%	3.66%	1.59%	0.89%	0.90%	5.97%	1.71%	0.00%	1.07%	5.02%	5.39%	3.66%	0.63%	1.93%	1.49%	0.29%
5	Altruistic_Angle590	According to GS St	3282	13.60%	7.74%	5.05%	5.21%	2.82%	18.98%	5.28%	7.69%	6.37%	10.31%	52.90%	5.74%	6.60%	11.22%	23.08%	17.69%
6	ptadisbanded	Google updates se	2251	13.60%	7.74%	5.05%	3.64%	1.50%	8.26%	2.81%	7.69%	5.18%	3.50%	15.11%	5.74%	5.26%	3.46%	88.55%	9.38%
7	Ok-Rule7537	I am new to SEO. T	724	18.47%	9.59%	8.30%	5.21%	10.77%	18.98%	23.78%	10.53%	9.79%	17.86%	52.90%	29.41%	12.20%	16.67%	17.70%	10.09%
8	Krollwut	Last September, ou	900	17.69%	9.59%	9.66%	4.99%	5.76%	25.00%	23.78%	37.85%	8.70%	21.38%	35.23%	29.41%	14.19%	23.77%	21.43%	10.09%
9	Maslakovic	Key information:	861	7.97%	5.43%	5.05%	3.30%	2.91%	7.35%	5.28%	7.69%	10.64%	7.11%	11.50%	5.74%	12.20%	11.22%	19.23%	42.17%
10	TheRealDrNeko	how do we do this?	152	7.69%	9.59%	9.66%	4.99%	13.77%	0.60%	1.63%	1.82%	1.07%	3.22%	13.70%	0.96%	1.04%	0.50%	8.20%	1.20%
11	squad1984	Hey guys!	737	9.08%	4.32%	5.05%	1.85%	2.91%	8.54%	2.81%	7.69%	5.18%	7.11%	26.77%	5.74%	12.20%	11.22%	17.70%	10.09%
12	Saiyyidi	What do you think	216	4.36%	1.45%	0.79%	0.48%	0.24%	0.05%	0.00%	0.00%	0.00%	0.57%	0.25%	0.00%	0.05%	0.00%	0.68%	0.00%
13	aashirvad_seo	Suppose we have	158	1.46%	1.04%	0.29%	0.36%	0.24%	0.23%	0.25%	0.00%	0.14%	0.14%	1.92%	0.23%	1.04%	0.28%	0.32%	1.20%
14	hookages	Hi everyone! I just	211	1.88%	1.20%	0.52%	0.39%	0.24%	0.16%	0.00%	0.00%	0.14%	2.40%	0.29%	0.23%	0.05%	0.06%	3.03%	1.20%
15	Marian_97c	Hello,	1615	7.69%	2.26%	2.06%	1.60%	1.50%	1.61%	1.63%	1.82%	1.07%	2.40%	5.39%	3.66%	1.04%	5.08%	4.22%	1.20%
16	DAMJim	Google Search Cor	257	1.46%	1.04%	0.29%	0.36%	0.24%	0.21%	0.00%	0.00%	0.50%	0.64%	0.25%	0.23%	0.05%	0.06%	0.68%	0.00%
17	ExplanationSuper2	I regularly use Mail	335	1.67%	1.04%	0.37%	0.39%	0.25%	0.52%	0.25%	0.00%	0.53%	2.40%	1.92%	0.96%	1.04%	0.84%	1.49%	5.56%
18	Can19977	Hi all,	1691	15.05%	7.01%	5.05%	4.37%	5.41%	8.54%	18.67%	10.53%	5.18%	10.31%	35.23%	20.00%	14.19%	11.22%	21.43%	10.09%
19	Ancient_Bathroom!	Hello everyone,	866	11.08%	5.43%	5.05%	3.30%	3.80%	11.54%	12.80%	9.56%	6.37%	10.31%	41.18%	5.74%	14.19%	25.48%	23.08%	17.69%
20	Obvious_Substanc	As the title says, w	200	5.03%	2.26%	1.56%	0.89%	1.01%	8.26%	1.45%	1.82%	1.07%	3.50%	11.50%	3.66%	5.26%	3.46%	1.49%	3.17%
21	jacoob_john	anybody know abo	25	3.48%	2.26%	2.06%	1.60%	1.50%	4.27%	1.45%	0.00%	2.06%	2.63%	5.39%	3.66%	1.04%	0.84%	3.03%	1.20%
22	_bobbyfischer	Im having trouble g	420	12.54%	13.05%	12.19%	5.21%	13.77%	25.00%	19.51%	42.86%	10.64%	34.75%	41.18%	18.75%	20.90%	33.13%	23.08%	17.69%
23	GiniMiniManeMo	Hi all,	395	18.47%	10.56%	10.72%	7.40%	15.73%	25.00%	23.78%	50.00%	14.08%	15.93%	52.90%	18.75%	21.84%	34.69%	17.70%	17.69%
24	Kitchen_Knee_740	Hi, I've created a W	1590	11.08%	6.14%	5.05%	1.85%	1.50%	8.26%	2.81%	7.69%	5.18%	7.11%	14.29%	5.74%	5.26%	5.99%	8.20%	9.38%
25	vijaydigi07	My keyword is best	144	1.88%	1.45%	0.60%	0.48%	0.25%	0.60%	0.25%	0.00%	0.50%	1.35%	1.25%	0.23%	0.05%	0.28%	7.97%	0.29%
26	N4riN4ri	I help develop and	1023	20.98%	16.00%	12.19%	9.17%	13.77%	11.54%	23.78%	10.53%	9.79%	15.93%	52.90%	20.00%	14.19%	33.13%	10.23%	17.69%
27	SE_Ranking	Last week, our tear	4154	10.50%	5.43%	5.05%	3.30%	2.82%	8.54%	5.28%	7.69%	5.18%	4.05%	19.75%	5.74%	6.60%	11.22%	21.43%	10.09%
28	Diligent_Response	Hey everyone! I ha	399	7.69%	3.66%	2.06%	1.85%	1.50%	7.35%	2.81%	1.82%	2.06%	12.46%	15.11%	5.74%	5.26%	9.37%	7.97%	10.09%
29	Vivissiah	I am saying this to	725	36.05%	19.68%	18.54%	6.32%	53.61%	35.79%	92.51%	42.86%	25.68%	33.33%	43.10%	49.64%	14.19%	23.77%	10.23%	42.17%
30	AutoModerator	There's no strict lim	1577	6.04%	3.66%	1.59%	1.60%	1.50%	82.50%	48.32%	1.82%	2.06%	21.38%	35.23%	5.74%	6.60%	11.22%	4.13%	9.38%
31	its_leslievanilla	There's a serious p	74	33.34%	9.59%	10.72%	1.85%	53.61%	35.79%	83.70%	31.25%	9.79%	34.75%	46.15%	88.14%	12.20%	58.22%	23.08%	26.67%
32	Celestial_Ram	I've never commen	81	6.04%	4.33%	2.40%	0.52%	1.50%	0.60%	1.45%	0.00%	1.07%	1.35%	15.11%	0.96%	1.04%	0.28%	0.68%	0.29%
33	SlavePrincessVibe	Tbh, I didn't even r	356	42.12%	35.36%	49.12%	21.18%	19.48%	18.98%	33.33%	75.00%	16.67%	23.19%	85.62%	36.11%	20.90%	39.19%	34.15%	43.52%
34	jason-jason2000	Like, is there an ide	1448	36.05%	19.68%	28.71%	15.13%	55.57%	18.98%	33.33%	7.69%	5.18%	42.27%	52.90%	5.74%	5.26%	3.46%	4.22%	3.17%
35	Negotiation_Previo	I thought it was iror	68	40.08%	28.50%	48.25%	1.60%	19.48%	11.54%	27.83%	7.69%	6.37%	15.93%	12.21%	36.11%	6.60%	7.14%	8.20%	10.09%
36	dexamphetamines	Or did I miss a perr	61	2.60%	1.45%	0.87%	0.48%	0.90%	3.01%	2.81%	0.00%	5.18%	9.63%	1.25%	5.74%	1.04%	3.46%	8.20%	85.81%
37	PhoonTFDB	Just the usual com	113	6.04%	3.60%	0.60%	0.89%	0.30%	1.61%	0.26%	0.00%	0.14%	0.64%	5.39%	0.96%	0.63%	0.06%	3.03%	0.29%
38	Jealous_Author_43	i've been so angry	2531	47.19%	46.10%	54.06%	37.23%	38.35%	33.56%	55.56%	7.69%	5.18%	12.46%	15.11%	29.41%	1.04%	5.08%	3.03%	3.17%
39	Superb_Ad1765	OP doesn‚Äôt ever	138	2.60%	1.62%	9.66%	1.10%	0.30%	0.52%	2.81%	0.00%	1.07%	0.54%	0.65%	0.23%	1.04%	0.06%	0.32%	0.00%
40																			

A	B
ID	hookages
Content	Hi everyone! I just want to ask for any tips on how to write meta descriptions for sellers or manufacturers on an e-commerce website. Should I add our keywords to it or not? Any tips would be greatly appreciated
Length	211
Toxic	1.88%
Insult	1.20%
Profanity	0.52%
Derogatory	0.39%
Sexual	0.24%
Death, Harm & Tragedy	0.16%
Violent	0.00%
Firearms & Weapons	0.00%
Public Safety	0.14%
Health	2.40%
Religion & Belief	0.29%
Illicit Drugs	0.23%
War & Conflict	0.05%
Politics	0.06%
Finance	3.03%
Legal	1.20%

A	B
ID	Vivissiah
Content	I am saying this to remind all, there is a zero tolerance for any violence wishing, wanting or the likes on anyone no matter who or what they are. Are the incels wishing violence? Still zero tolerance. Are they wishing rape? Still zero tolerance to wish similar on them. It is all zero tolerance. Even implied such will not be tolerated and is on zero tolerance and this includes jail jokes involving soaps or the likes.
Length	725
Toxic	36.05%
Insult	19.68%
Profanity	18.54%
Derogatory	6.32%
Sexual	53.61%
Death, Harm & Tragedy	35.79%
Violent	92.51%
Firearms & Weapons	42.86%
Public Safety	25.68%
Health	33.33%
Religion & Belief	43.10%
Illicit Drugs	49.64%
War & Conflict	14.19%
Politics	23.77%
Finance	10.23%
Legal	42.17%

Possible data points

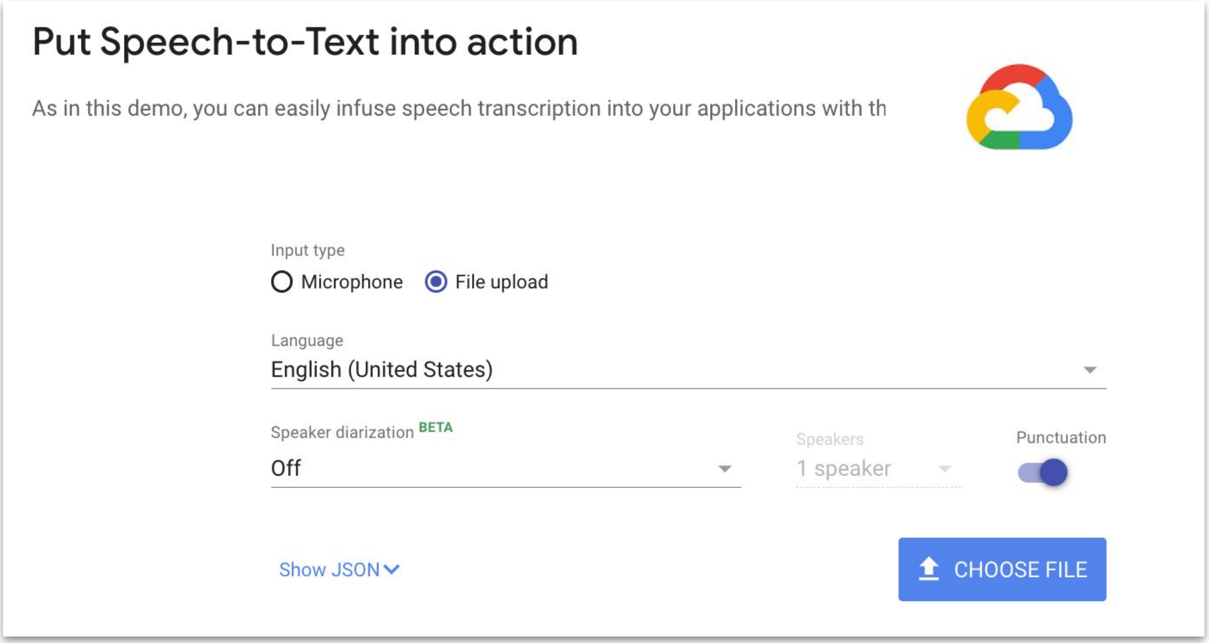
- Your content
- Competitors' content
- YouTube video transcripts (monetization)
- social media mentions
- comments on your website
- community posts

…


abcd

Good videos on YouTube but not on the blog?

→ **Transcribe.**



Approach	Suitable for	Limitation	Tools
No-code	• Beginners • Non-technical	• Limited scalability	happyscribe descript **Restream**
Programmatic	• Intermediate • A little bit more technical • API-savvy	• Time • Adoption costs - learning, dev resources	• Google Cloud Speech-to-Text • Amazon Transcribe • OpenAI Whisper - *but bear in mind it sucks for anything over 2-3 minutes, and has no small language support.*

Caveat

What I'm not saying ❌

- Spam your blog with auto-transcribed content
- Fire your content team
- Scrape competitors' Youtube videos, transcribe, and traffic to the moon, *bro* 🥲

What I am saying ✅

- Bridge gaps between different teams, if enterprise
- Make content work harder, especially if you're already producing webinars, live streams, etc.
- Use transcription for competitor analysis, not copying

Mix & match with other approaches 🌟

Context:

- SERP Analysis shows Google ranks more videos, but there is still lots of traffic for web pages.
- There are some YouTubers that don't do a good job at blogging, but their channels get a ton of traction

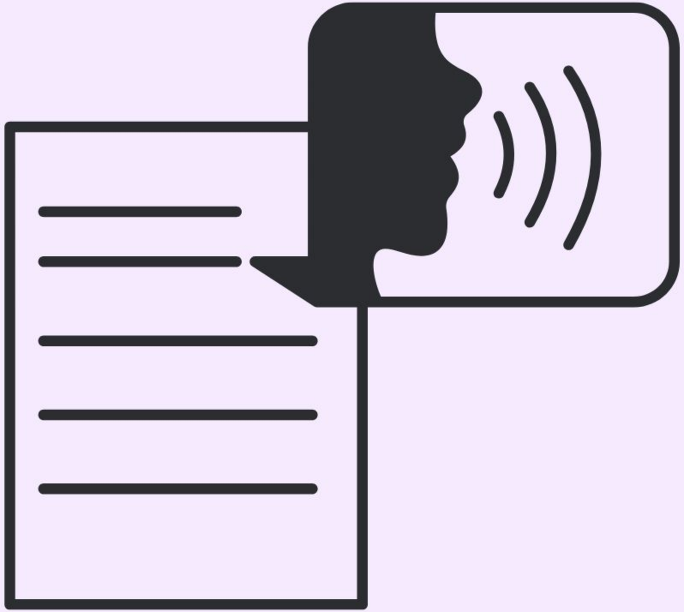
Scrape videos/audio

Transcribe

Identify topics + subtopics

Extract entities

- Compare with website
- Add to topical map
- Find and address information gaps

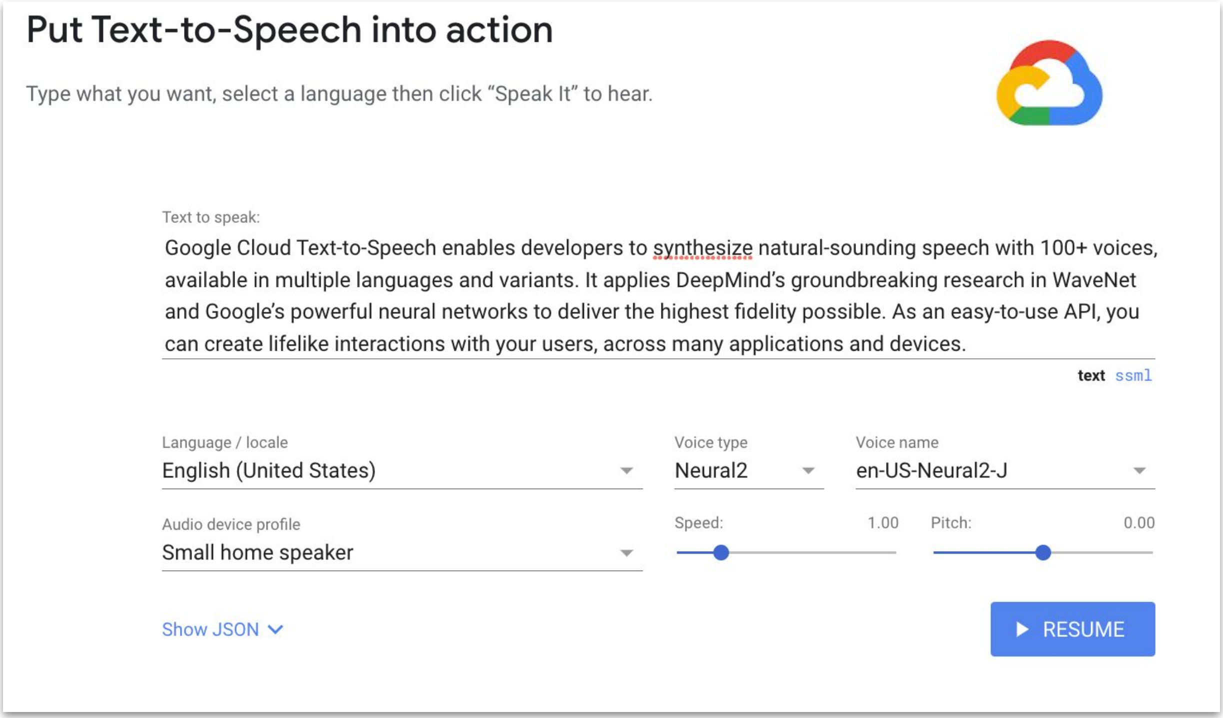


MLFORSEO
Machine Learning for SEO

www.mlforseo.com

You have a library of high-performing tutorials but no presence on YouTube/TikTok?

→Scale production with text to speech.



Approach	Suitable for	Limitation	Tools
No-code	• Beginners • Non-technical	• Limited scalability	Speechify, MURF.AI
Programmatic	• Intermediate • A little bit more technical • API-savvy	• Time and other adoption costs	• Google Cloud Text-to-Speech API • Amazon Polly - Text To Speech AI Tool • OpenAI GPT4o (soon)

What I'm not saying ❌

- Spam YouTube with AI generated trash
- You can replace video production

What I am saying ✅

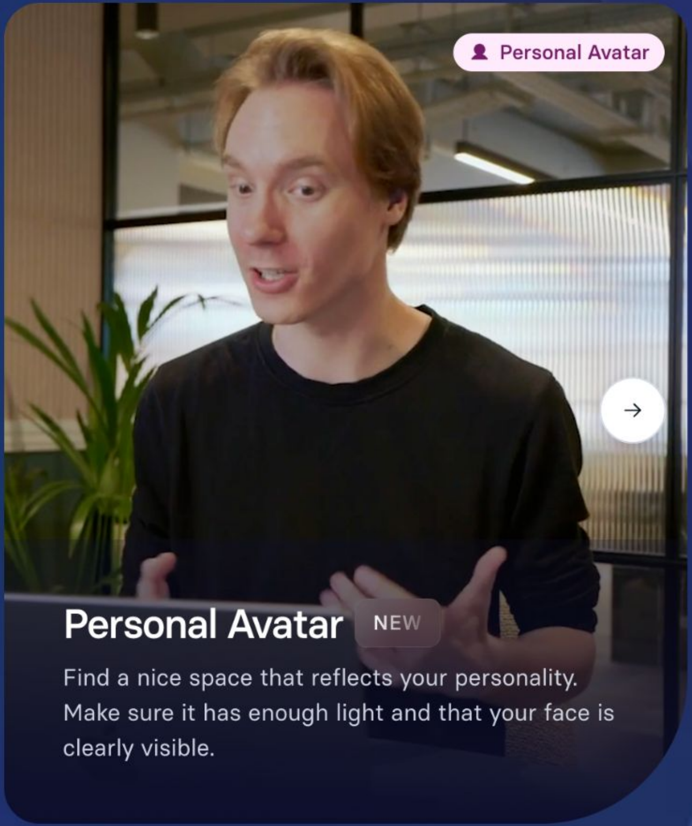
- Certain content formats don't require video and can be made more accessible via audio & stills, like **interviews** or **tutorials**
- To add automation, but still have at least some personalised look/feel, tutorial videos might have avatars, instead of person-to-camera setups to lower costs

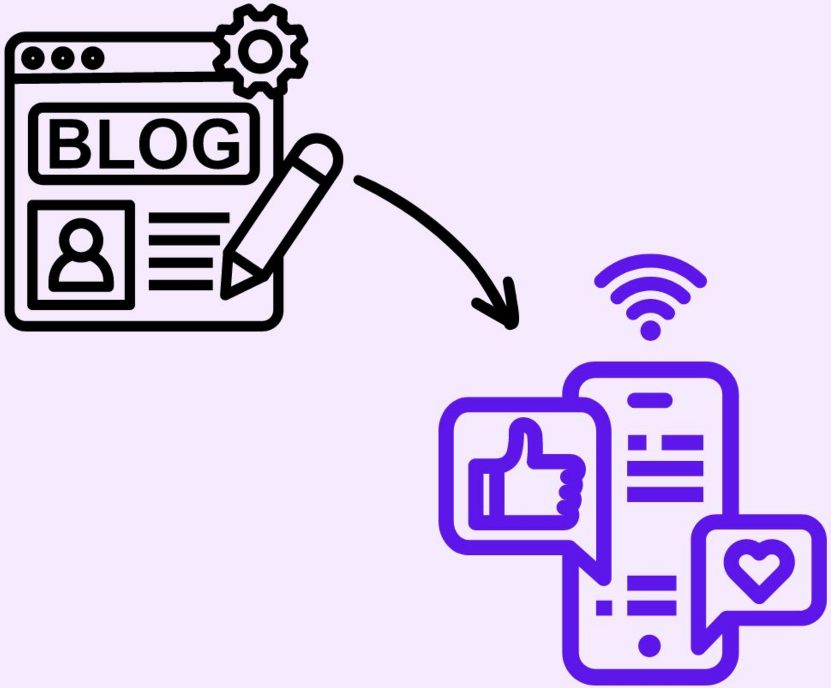
CUSTOM AI AVATARS

Custom Avatar Maker

Using Synthesia's custom avatar maker, choose between a browser-based webcam AI avatar or a professional-quality studio avatar.

Select the type of custom avatar





MLFORSEO

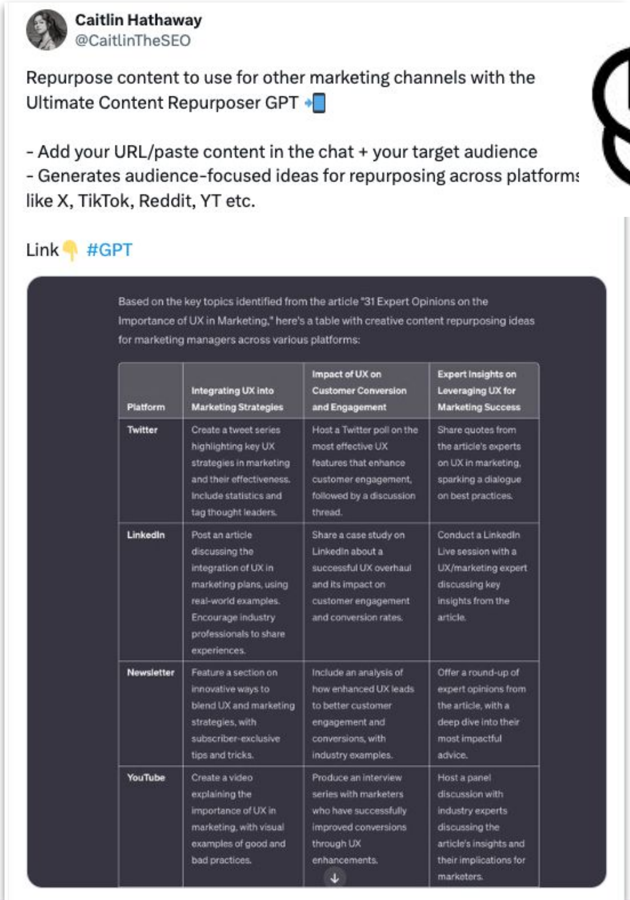
Machine Learning for SEO

www.mlforseo.com

Approach	Suitable for	Limitation	Tools
No-code	• Beginners • Non-technical	• Limited scalability	• ChatGPT • Custom GPTs • ~~Web tools~~ (they're all wrappers of GPT, so not worth it)
Programmatic	• Intermediate • A little bit more technical • API-savvy	• Time and other adoption costs	• GPT4/ GPT4o • Any LLMs • BERT

You have a library of high-performing blog posts but no content distribution?

→ Transform blog posts to insightful posts for social media.



Caitlin Hathaway
@CaitlinTheSEO

Repurpose content to use for other marketing channels with the Ultimate Content Repurposer GPT ➡️📱

- Add your URL/paste content in the chat + your target audience
- Generates audience-focused ideas for repurposing across platforms like X, TikTok, Reddit, YT etc.

Link 👇 #GPT

Based on the key topics identified from the article "31 Expert Opinions on the Importance of UX in Marketing," here's a table with creative content repurposing ideas for marketing managers across various platforms:

Platform	Integrating UX into Marketing Strategies	Impact of UX on Customer Conversion and Engagement	Expert Insights on Leveraging UX for Marketing Success
Twitter	Create a tweet series highlighting key UX strategies in marketing and their effectiveness. Include statistics and tag thought leaders.	Host a Twitter poll on the most effective UX features that enhance customer engagement, followed by a discussion thread.	Share quotes from the article's experts on UX in marketing, sparking a dialogue on best practices.
LinkedIn	Post an article discussing the integration of UX in marketing plans, using real-world examples. Encourage industry professionals to share experiences.	Share a case study on LinkedIn about a successful UX overhaul and its impact on customer engagement and conversion rates.	Conduct a LinkedIn Live session with a UX/marketing expert discussing key insights from the article.
Newsletter	Feature a section on innovative ways to blend UX and marketing strategies, with subscriber-exclusive tips and tricks.	Include an analysis of how enhanced UX leads to better customer engagement and conversions, with industry examples.	Offer a round-up of expert opinions from the article, with a deep dive into their most impactful advice.
YouTube	Create a video explaining the importance of UX in marketing, with visual examples of good and bad practices.	Produce an interview series with marketers who have successfully improved conversions through UX enhancements.	Host a panel discussion with industry experts discussing the article's insights and their implications for marketers.

You have a library of high-performing blog posts but no newsletter?

→Use an LLM to rewrite these into newsletter edition drafts.

You have comprehensive guides or reports in PDF format?

→You can extract key insights, summaries, or actionable tips from these documents and repurpose them into blogs or social posts/ threads.

Overall, **any LLM** would do a great job here.

While we're on the topic of LLMs doing a good job...

Structured data + LLMs = ♡

How to use generative AI with structured data for programmatic SEO

Google Colab (Python) | Intermediate | OpenAI API

Michael Jordan Career Stats Table

	SEASON	TEAM	AGE	GP	GS	MIN	FGM	FGA	FG %	FG3M	FG3A	FG3 %	FTM	FTA	FT %	OREB	DREB	REB	AST	STL	BLK	TOV	PF	PTS
1	1984-85	CHI	22	82	82	3,144	837	1,625	51.5%	9	52	17.3%	630	746	84.5%	167	367	534	481	196	69	291	285	2,313
2	1985-86	CHI	23	18	7	451	150	328	45.7%	3	18	16.7%	105	125	84.0%	23	41	64	53	37	21	45	46	408
3	1986-87	CHI	24	82	82	3,281	1,098	2,279	48.2%	12	66	18.2%	833	972	85.7%	166	264	430	377	236	125	272	237	3,041
4	1987-88	CHI	25	82	82	3,311	1,069	1,998	53.5%	7	53	13.2%	723	860	84.1%	139	310	449	485	259	131	252	270	2,868
5	1988-89	CHI	26	81	81	3,255	966	1,795	53.8%	27	98	27.6%	674	793	85.0%	149	503	652	650	234	65	290	247	2,633
6	1989-90	CHI	27	82	82	3,197	1,034	1,964	52.6%	92	245	37.6%	593	699	84.8%	143	422	565	519	227	54	247	241	2,753
7	1990-91	CHI	28	82	82	3,034	990	1,837	53.9%	29	93	31.2%	571	671	85.1%	118	374	492	453	223	83	202	229	2,580
8	1991-92	CHI	29	80	80	3,102	943	1,818	51.9%	27	100	27.0%	491	590	83.2%	91	420	511	489	182	75	200	201	2,404
9	1992-93	CHI	30	78	78	3,067	992	2,003	49.5%	81	230	35.2%	476	569	83.7%	135	387	522	428	221	61	207	188	2,541
10	1994-95	CHI	32	17	17	668	166	404	41.1%	16	32	50.0%	109	136	80.1%	25	92	117	90	30	13	35	47	457
11	1995-96	CHI	33	82	82	3,090	916	1,850	49.5%	111	260	42.7%	548	657	83.4%	148	395	543	352	180	42	197	195	2,491
12	1996-97	CHI	34	82	82	3,109	920	1,892	48.6%	111	297	37.4%	480	576	83.3%	113	369	482	352	140	44	166	156	2,431
13	1997-98	CHI	35	82	82	3,181	881	1,893	46.5%	30	126	23.8%	565	721	78.4%	130	345	475	283	141	45	185	151	2,357
14	2001-02	WAS	39	60	53	2,089	551	1,324	41.6%	10	53	18.9%	263	333	79.0%	50	289	339	310	85	26	162	119	1,375
15	2002-03	WAS	40	82	67	3,029	679	1,527	44.5%	16	55	29.1%	266	324	82.1%	71	426	497	311	123	39	173	171	1,640

Elias Dabbas

```
role system
content
You are a smart, detail-oriencted, keen NBA Basketball player
analyst.
Please write an introductory text for a profile page of this
Basketball player.
Length: 500 - 800 words.
please stick to the stats provided.
Tone: should be interesting factual intriguing and inviting the
user to dive
into the charts on the page to better get to know the player.
```

Elias Dabbas

MLFORSEO
Machine Learning for SEO

www.mlforseo.com

Nikola Jokic: The Game-Changing Force from the Mile High City

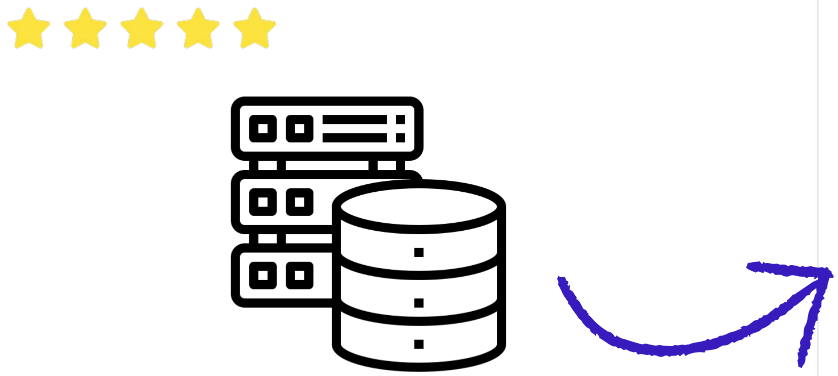
When it comes to the realm of basketball, few players possess the unique blend of skills and basketball IQ that Nikola Jokic brings to the court. As a key player for the Denver Nuggets since the 2015-2016 season, Jokic has not only solidified his place as one of the league's top players but has also become a fan favorite in the city of Denver, Colorado.

Standing at 7 feet tall, Jokic cuts an imposing figure on the court, but it's his finesse and versatility that truly set him apart from his peers. With a stellar field goal percentage of 55.7%, Jokic has proven time and time again that he has the scoring touch to make an impact in any game situation. His ability to stretch the floor as a center, with a reliable mid-range shot and
kes him a nightmare matchup for opposing

Elias Dabbas

MLFORSEO
Machine Learning for SEO

www.mlforseo.com

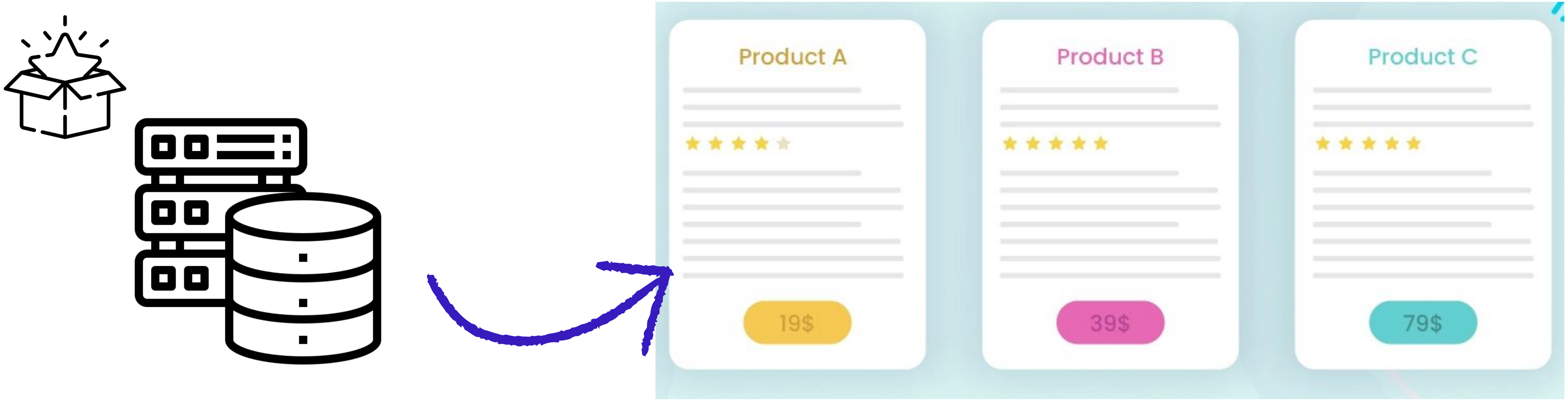


AI Summary

Based on 10 customer ratings

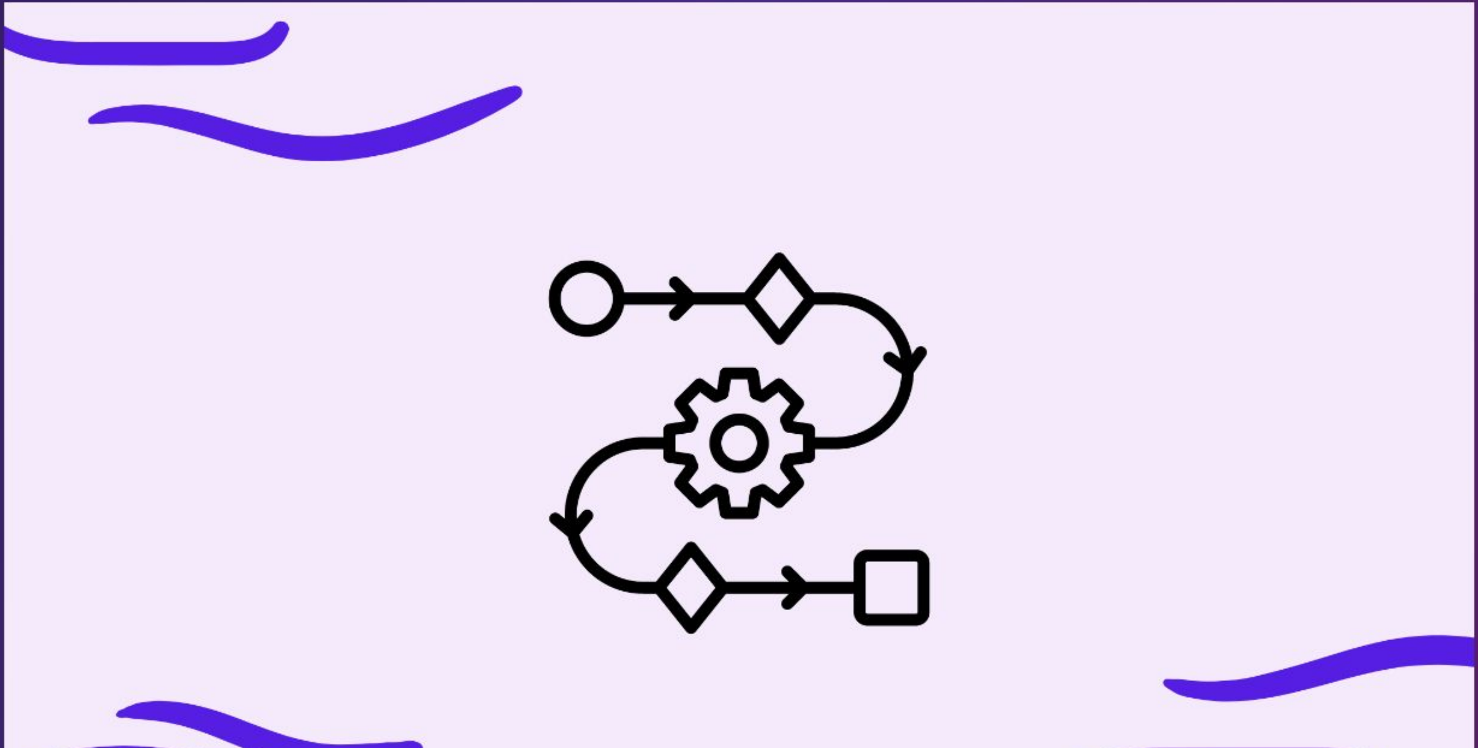
3 out of 5

Customers share a variety of opinions about the Mower3000. The most common pros are its quiet operation, user-friendly app and setup, and autonomous lawn care capabilities. However, some customers have reported issues with uneven cutting, getting stuck, and lost connections to the boundary wire. The battery life also seems insufficient for some users. Despite these challenges, many customers still consider it a good buy, especially for convenience and reduced manual labor.


Product A
19$
Product B
39$
Product C
79$

How to Automatically Optimize your SEO Metadata with FuzzyWuzzy and OpenAI in Google Colab

Beginner FuzzyWuzzy Google Colab (Python) OpenAI API



```
# Execute the request to the Search Console API
response = service.searchanalytics().query(siteUrl=site, body=request).execute()
print("Getting Google Search Console...")

# Parse the JSON response
scDict = defaultdict(list)

for row in response['rows']:
    scDict['page'].append(row['keys'][0] or 0)
    scDict['query'].append(row['keys'][1] or 0)
    scDict['clicks'].append(row['clicks'] or 0)
    scDict['ctr'].append(row['ctr'] or 0)
    scDict['impressions'].append(row['impressions'] or 0)
    scDict['position'].append(row['position'] or 0)

# Create a DataFrame from the parsed data
df = pd.DataFrame(data=scDict)
```

Natzir Turrado Ruiz

Step 3: Scrape metadata with BeautifulSoup

Using BeautifulSoup, we can pull existing SEO metadata from your site, like title, h1 and descriptions.

```
# Function to extract metadata from a URL
def get_meta(url):
    try:
        response = requests.get(url)
        encoding = chardet.detect(response.content)[‘encoding’]
        if encoding:
            page_content = response.content.decode(encoding)
        else:
            page_content = response.content
        soup = BeautifulSoup(page_content, ‘html.parser’)
        title = soup.find(‘title’).get_text() if soup.find(‘title’) else ‘No title’ # Get the title
        meta = soup.select(‘meta[name=”description”]’)[0].attrs[“content”] if
soup.select(‘meta[name=”description”]’) else ‘No meta description’ # Get the meta description
        h1 = soup.find(‘h1’).get_text() if soup.find(‘h1’) else ‘No h1’ # Get the first h1
        return title, meta, h1
    except Exception as e:
        return ‘Error’, ‘Error’, ‘Error’

# Apply the function and add the results to the DataFrame
df[‘title’], df[‘meta’], df[‘h1’] = zip(*df[‘page’].apply(get_meta))
df
```

Natzir Turrado Ruiz

Step 5: Compute similarity using Fuzzy matching

In this step, we use the fuzzywuzzy library to measure how closely the cleaned SEO metadata (titles, meta descriptions, and headers) matches the top-performing search queries. This helps us identify areas where the content might not be optimized for relevant search terms.

We use token_set_ratio from the fuzzywuzzy library, which compares strings based on their content, ignoring the order and repeated words. This method is ideal for analyzing how well the cleaned text matches the search queries, as it provides a robust similarity score.

```
columns = [‘title_clean’, ‘meta_clean’, ‘h1_clean’]

for col in columns:
    similarity = []
    for index, row in df.iterrows():
        sim = fuzz.token_set_ratio(row[‘query_clean’], row[col])
        similarity.append(sim)
    df[f‘{col}_similarity’] = similarity

# Rename columns for clarity
df.rename(columns=lambda x: x.replace(‘_clean_similarity’, ‘_similarity’) if x.endswith(‘_clean_similarity’) else x,
inplace=True)
columns_to_drop = [col for col in df.columns if ‘_clean’ in col]
df.drop(columns=columns_to_drop, inplace=True)
```

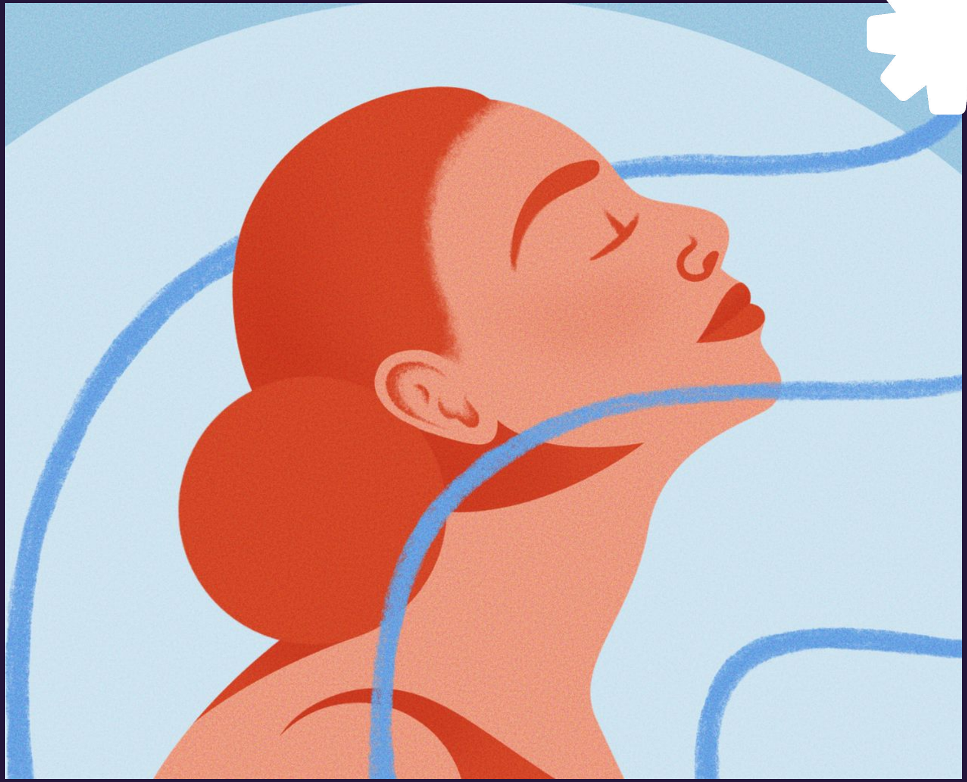
Natzir Turrado Ruiz

Step 6: Generate new titles using OpenAI

	page	query	clicks	ctr	impressions	position	title	meta	h1	title_similarity	meta_similarity	h1_similarity	new_title
0	https://www.analistaseo.es/conversion/modelo-p...	modelo de probabilidad de elaboración	76	7.826982	971	2.74	Modelo de Probabilidad de Elaboración y Persua...	El Modelo de la Probabilidad de Elaboración, t...	Cómo usar el Modelo de la Probabilidad de Elab...	100	100	100	nan
1	https://www.analistaseo.es/posicionamiento-bus...	buscador semantico	50	7.485030	668	2.46	Qué es un buscador semántico	Un buscador semántico es aquel que no da enlac...	Qué es un buscador semántico	100	100	100	nan
5	https://www.analistaseo.es/posicionamiento-bus...	navboost	25	4.901961	510	37.94	Los algoritmos de Google al descubierto. Cómo ...	\nEn este artículo nos adentramos en el funcio...	Los algoritmos de Google al descubierto. Cómo ...	12	11	12	Navboost: Algoritmos de Google y Documentos Fi...
6	https://www.analistaseo.es/google-api-indexing...	google colab indexing api	21	0.264517	7939	7.90	API Indexing Test with Google Colab	No meta description	API Indexing Test	100	20	83	nan
7	https://www.analistaseo.es/posicionamiento-bus...	cynefin	16	0.410572	3897	9.30	Aplicando el marco de Cynefin en la toma de de...	\nTomar decisiones acertadas en el volátil mun...	Aplicando el marco de Cynefin en la toma de de...	100	9	100	nan
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1399	https://www.analistaseo.es/conversion/neuromar...	neuromarketing barcelona	0	0.000000	3	77.00	Neuromarketing en eShow Barcelona 2014	El 95% de nuestros pensamientos, emociones y a...	Neuromarketing en eShow Barcelona 2014	100	18	100	nan
1497	https://www.analistaseo.es/posicionamiento-bus...	análisis seo	0	0.000000	1	9.00	Artículos de Posicionamiento en Buscadores (SEO)	No meta description	Artículos de SEO	40	7	56	Análisis SEO: Artículos de Posicionamiento en ...
1728	https://www.analistaseo.es/posicionamiento-bus...	logeados	0	0.000000	1	78.00	Búsqueda Segura en Google (SSL Search)	Desde ayer todos los usuarios que se identifiq...	Búsqueda Segura en Google (SSL Search)	24	10	24	Búsqueda Segura de Google para Usuarios Logeados
1773	https://www.analistaseo.es/posicionamiento-bus...	link rel alternate hreflang	0	0.000000	60	64.42	Caso de éxito SEO Internacional con HrefLang	El caso de éxito SEO que os enseño a continuac...	Caso de éxito SEO Internacional con HrefLang	50	46	50	Éxito Internacional con link rel alternate hre...
1560	https://www.analistaseo.es/posicionamiento-bus...	ctr organico	0	0.000000	5	90.00	Google AdWords y SEO - ¿Por qué sube el tráfic...	¿Por qué sube el tráfico orgánico cuando invie...	Google AdWords y SEO ¿Por qué sube el tráfico ...	80	80	80	nan

89 rows × 13 columns

Natzir Turrado Ruiz



03. What you need to grow 🌱

Mindset + Community + Resources

MLFORSEO

Machine Learning for SEO

Subscribe

Email

You can unsubscribe anytime. For more details, review our Privacy Policy.

I'm not a robot

reCAPTCHA

Privacy - Terms

Subscribe

ML/AI news

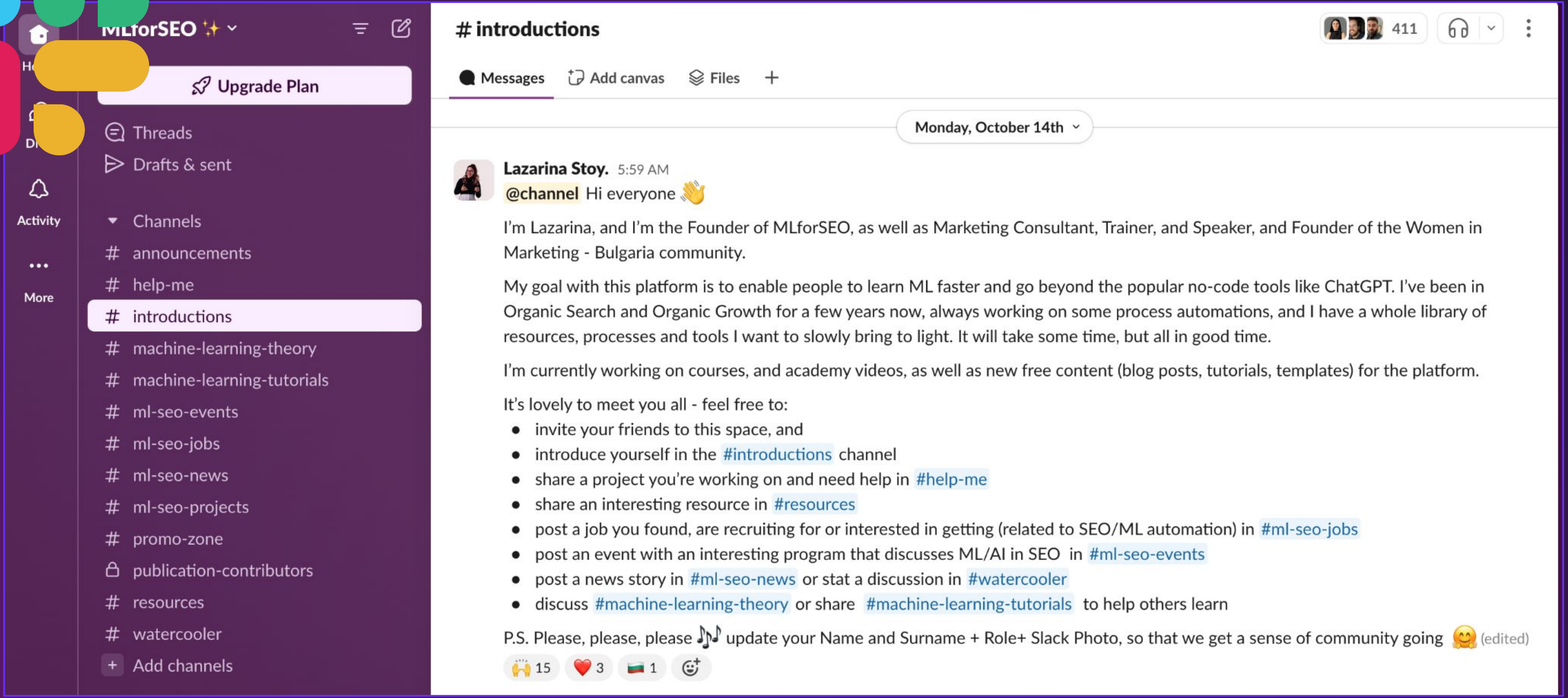
All the latest news and developments in the ML/AI industry that are relevant to Organic Search marketers.

Expert Commentary

Expert commentary, tips, and tricks on doing the most with ML without sacrificing executional quality or the human touch

Content updates

All the latest content updates from our blog, resources, online courses, academy, and experts.



www.mlforseo.com

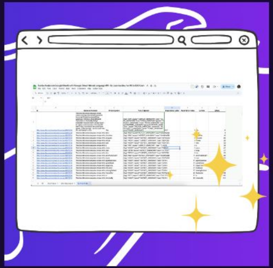
FREE CODE SAMPLES, SHEETS TEMPLATES, AND DATA VISUALISATION DESIGNS

View all our resources

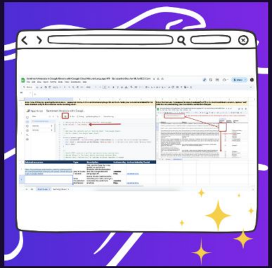
MLforSEO Templates

Kickstart machine learning implementation with our collection of templates, featuring Google Sheets templates with AppScript, Looker Studio dashboard templates, and a range of coding scripts and notebooks. Tailored for efficiency and effectiveness, these resources are designed with beginners in mind, ensuring you start your machine learning journey on the right foot.

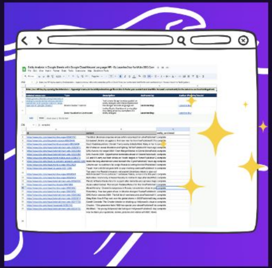
Content Moderation with Google Cloud Natural Language API (Google Sheets Template and Apps Script)



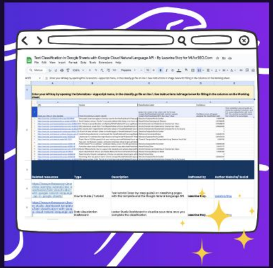
Syntax Analysis with Google Cloud Natural Language API (Google Sheets Template and Apps Script)



Sentiment Analysis with Google Cloud Natural Language API (Google Sheets Template and Apps Script)



Entity Analysis with with Google Cloud Natural Language API (Google Sheets Template and Apps Script)



Text Classification with with Google Cloud Natural Language API (Google Sheets Template with Apps Script)

Text Classification with Google Cloud Natural Language API (Looker Studio Template)

LEARN THE HOW IN MACHINE LEARNING

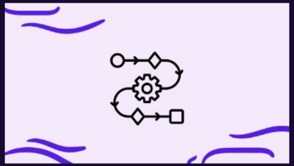
Machine Learning Tutorials

View all machine learning tutorials

Straightforward machine learning tutorials and how-to guides, ideal for beginners. Learn how to implement an API, or train your own machine learning model from scratch, using popular tools and technologies. Each tutorial includes all the resources you need, plus step-by-step guidance.

Content Creation

How to use generative AI with structured data for programmatic SEO



Onpage SEO

How to Automatically Optimize your SEO Metadata with FuzzyWuzzy and OpenAI in Google Colab

Audio Transcription

How to transcribe audio with OpenAI's Whisper API in Google Colab (Python)

✧ academy.mlforseo.com ✧

MLforSEO

Machine Learning for SEO

what I do my website

seo@lazarinastoy.com

social handle

best way to get in touch

TECH SEO CONNECT

Thank you for listening. ♡

Find my deck at

lazarinastoy.com/conference-decks-and-presentations/