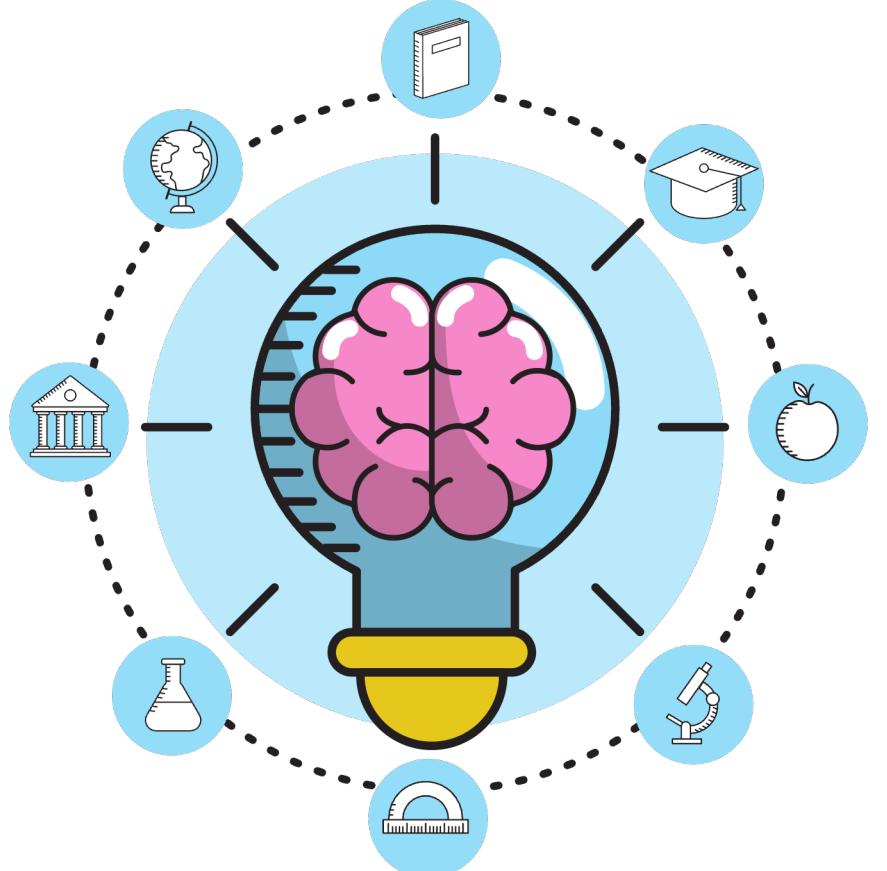


Dec 29, 2021
8:00 a.m.

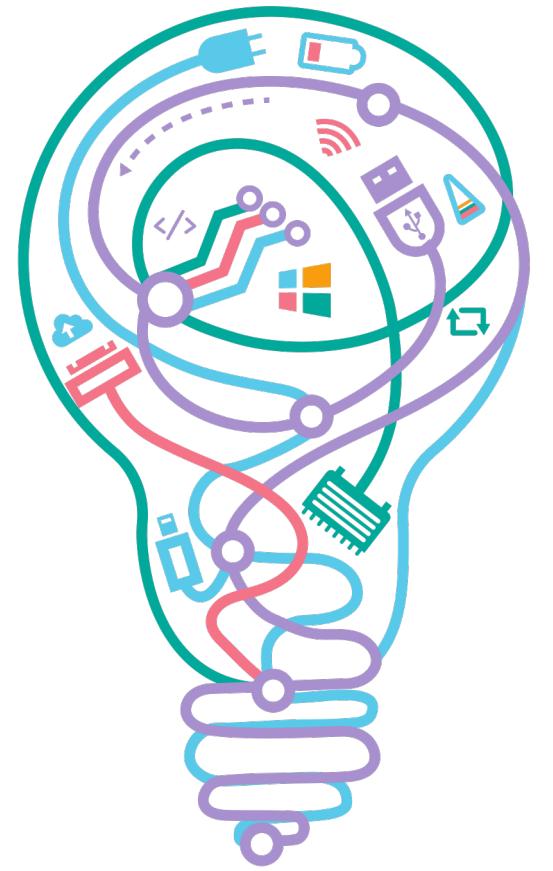
DATA SCIENCE FIELDS

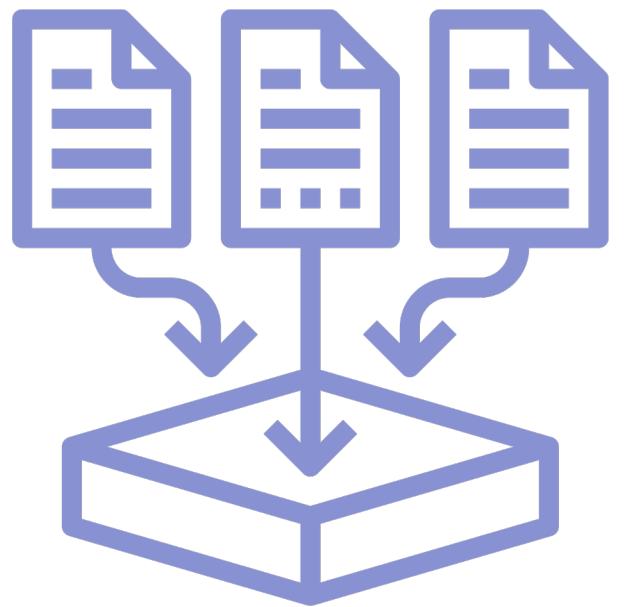
Bushra AlOlit,
Mukhtar bin al hamad
Mohammed alzabyedi

Dr.Mejdal ALQahtani



Unsupervised learning
Natural-Language-processing

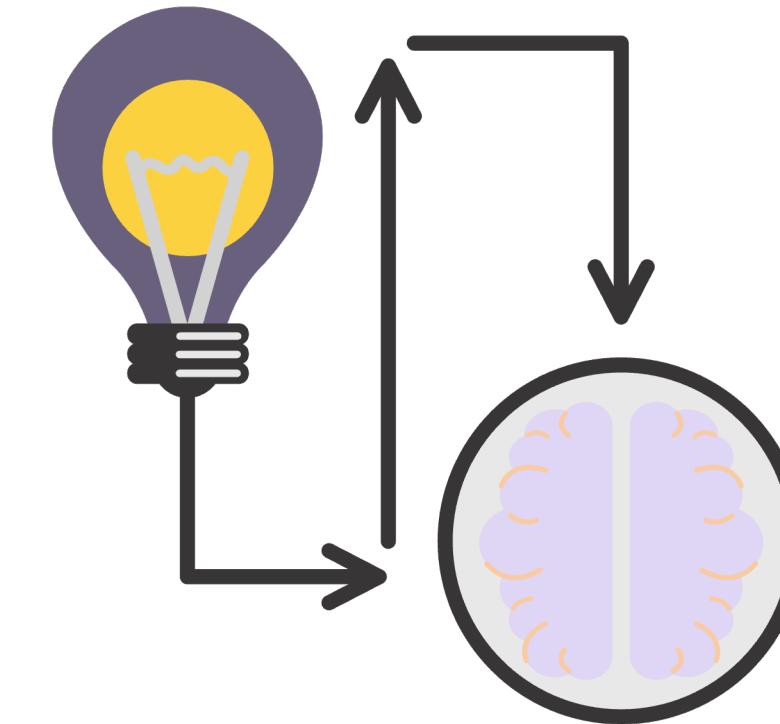




Introduction

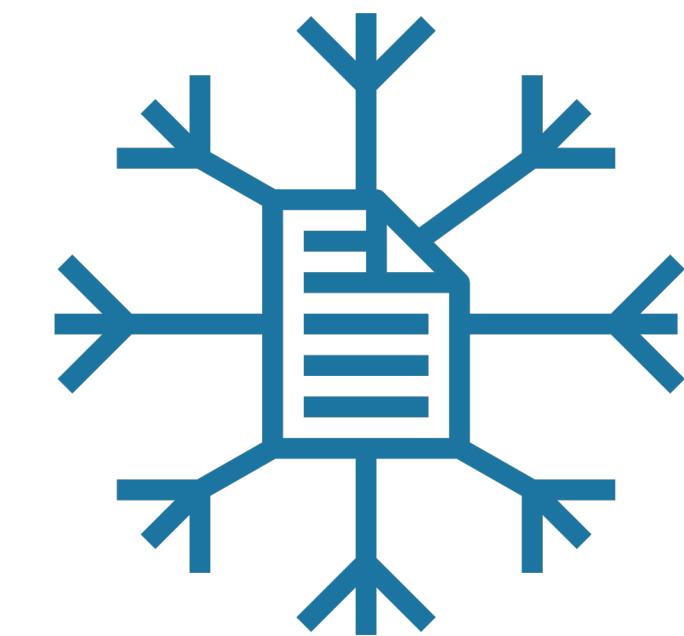
The importance

Of this project comes from the importance of the Data Science

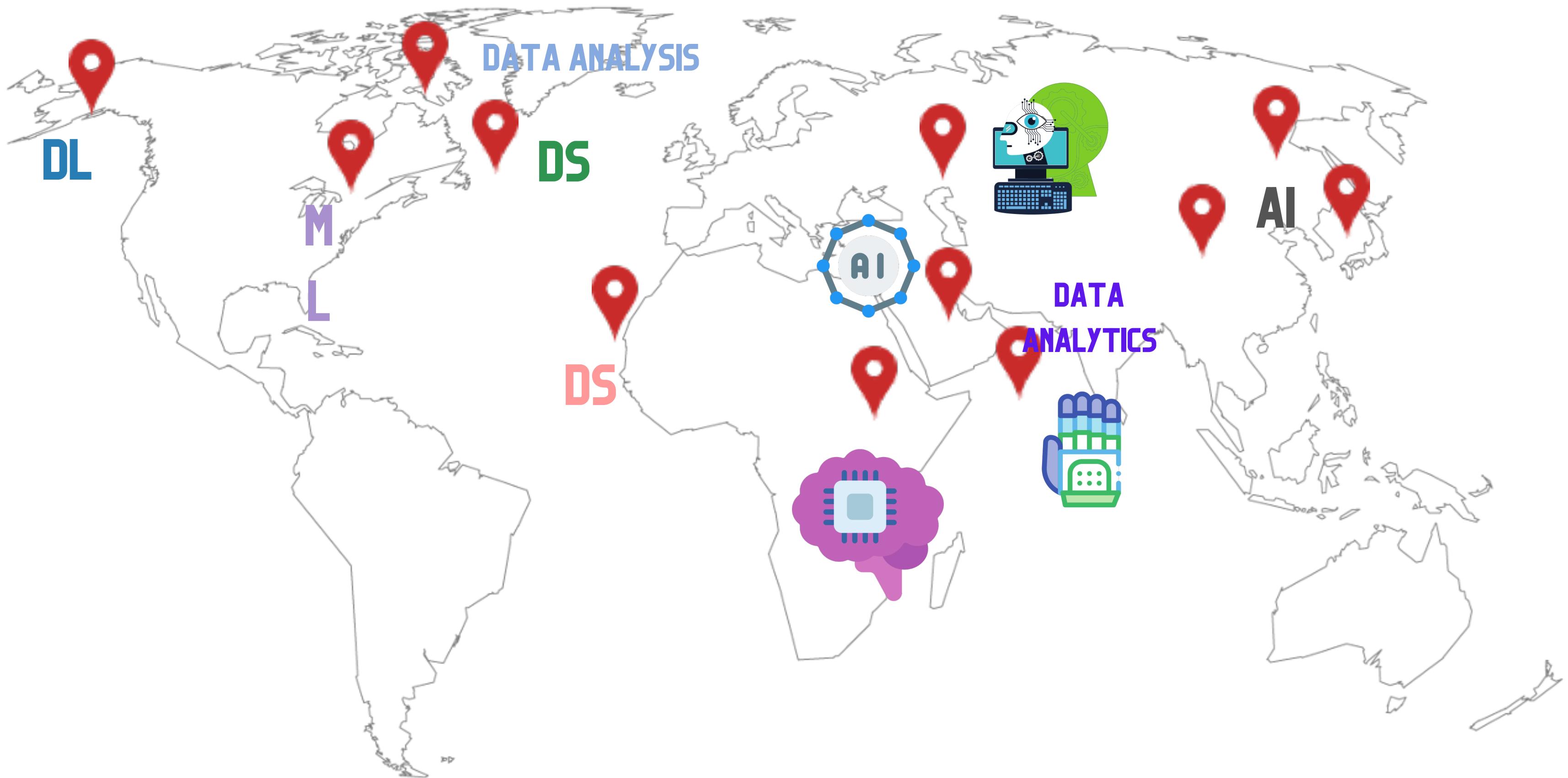


The objective is

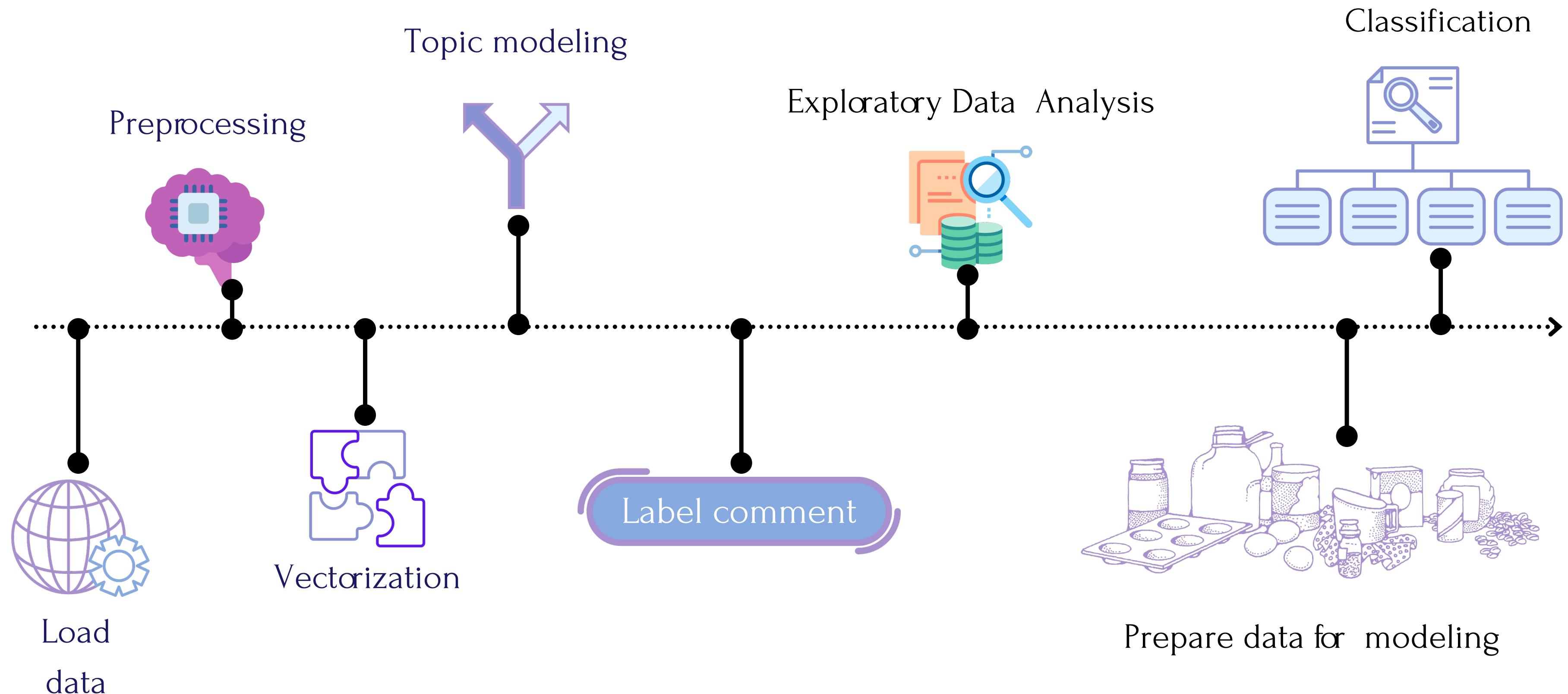
to help institutions recognize and identify their customers in which order to improve their services.

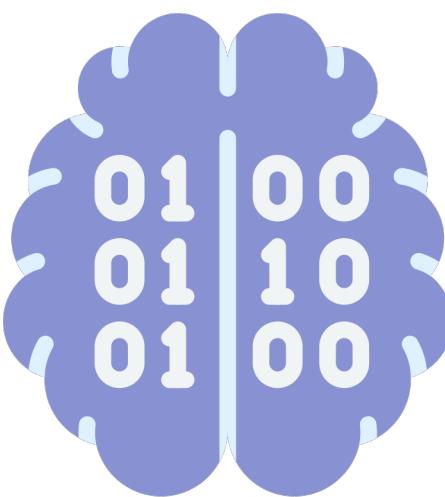


Field Map



Methodology





Data Set

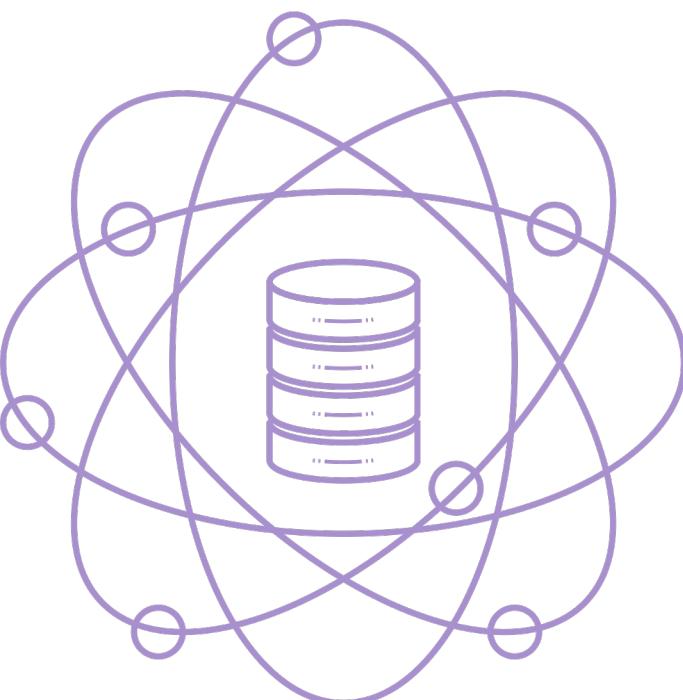
	stars	comment	book_url
0	5.0	Let me say at the outset of this review, that ...	https://www.amazon.com/dp/1492041130
1	1.0	Its Capitalism at its best. I read the book fr...	https://www.amazon.com/dp/1492041130
2	5.0	If you're looking to get started with data sci...	https://www.amazon.com/dp/1492041130
3	4.0	In general, this is a good book. It covers a w...	https://www.amazon.com/dp/1492041130
4	5.0	I was an academic data scientist decades befor...	https://www.amazon.com/dp/1492041130
...
20642	3.0	This book offers nice information and thoughts...	https://www.amazon.com/dp/0980200482
20643	5.0	Am disabled, retired RN, but always wished I w...	https://www.amazon.com/dp/0980200482
20644	3.0	At one point I considered library work as oppo...	https://www.amazon.com/dp/0980200482
20645	5.0	Overall, I think this is an excellent resource...	https://www.amazon.com/dp/0980200482
20646	5.0	GREAT	https://www.amazon.com/dp/0980200482

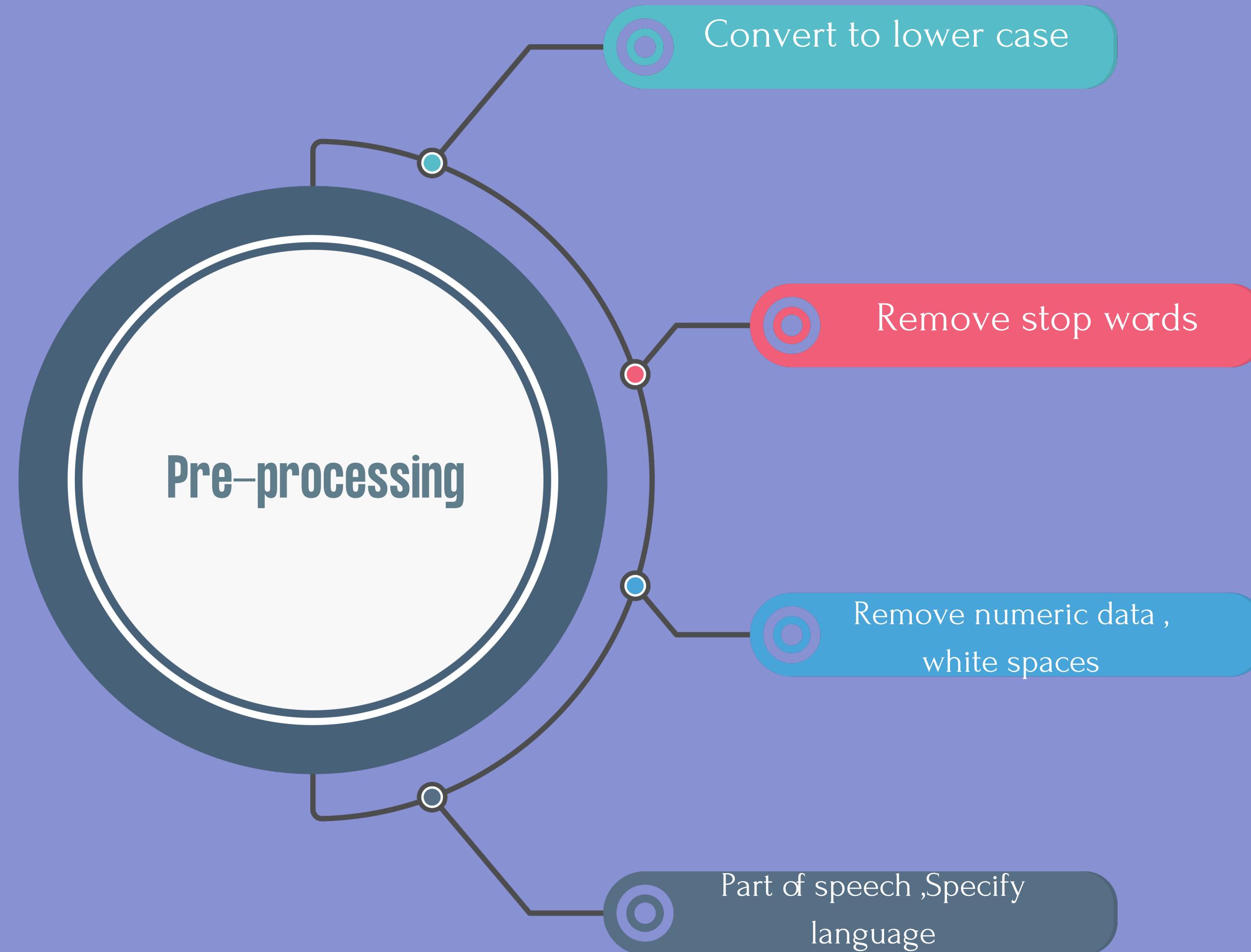
20647 rows × 3 columns

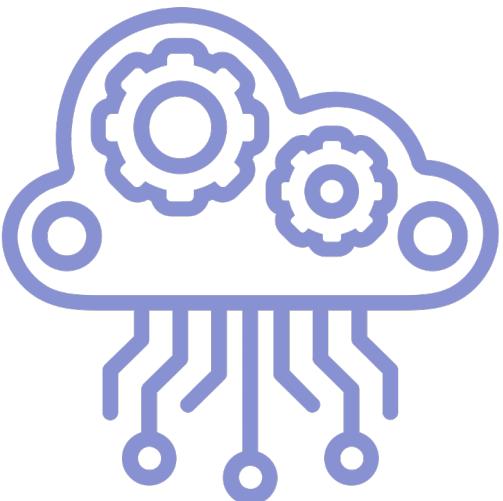


	comment
0	let me say at the outset of this review that ...
1	its capitalism at its best i read the book fr...
2	if you re looking to get started with data sci...
3	in general this is a good book it covers a w...
4	i was an academic data scientist decades befor...
...	...
20641	as someone who doesn t know much about library...
20642	this book offers nice information and thoughts...
20643	am disabled retired rn but always wished i w...
20644	at one point i considered library work as oppo...
20645	overall i think this is an excellent resource...

19470 rows × 1 columns

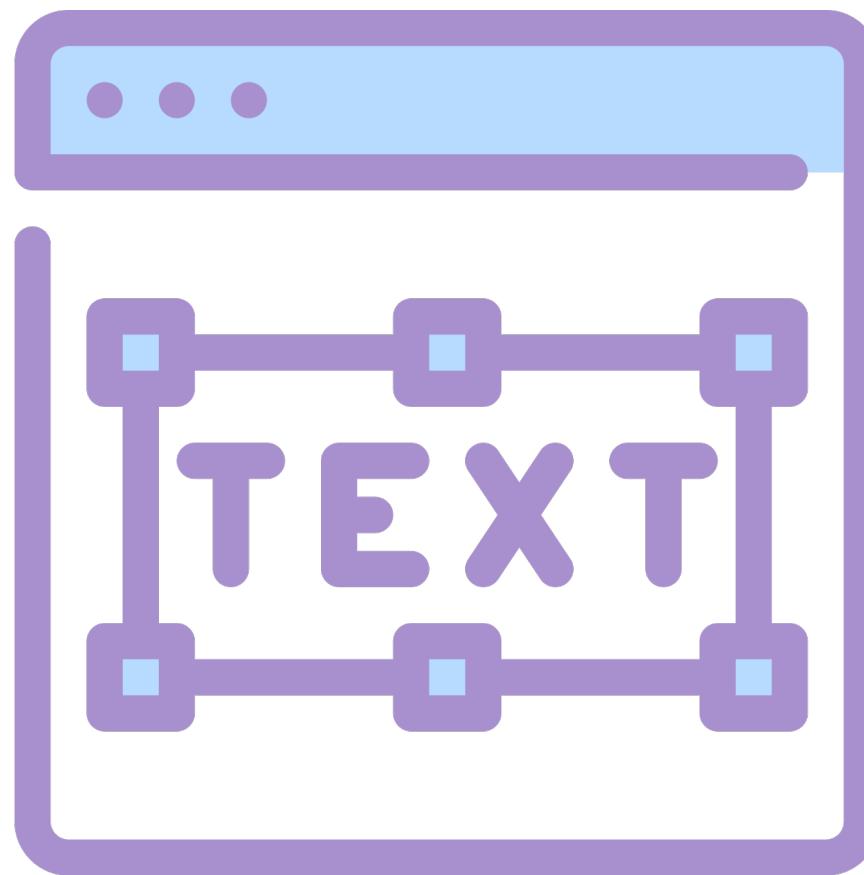
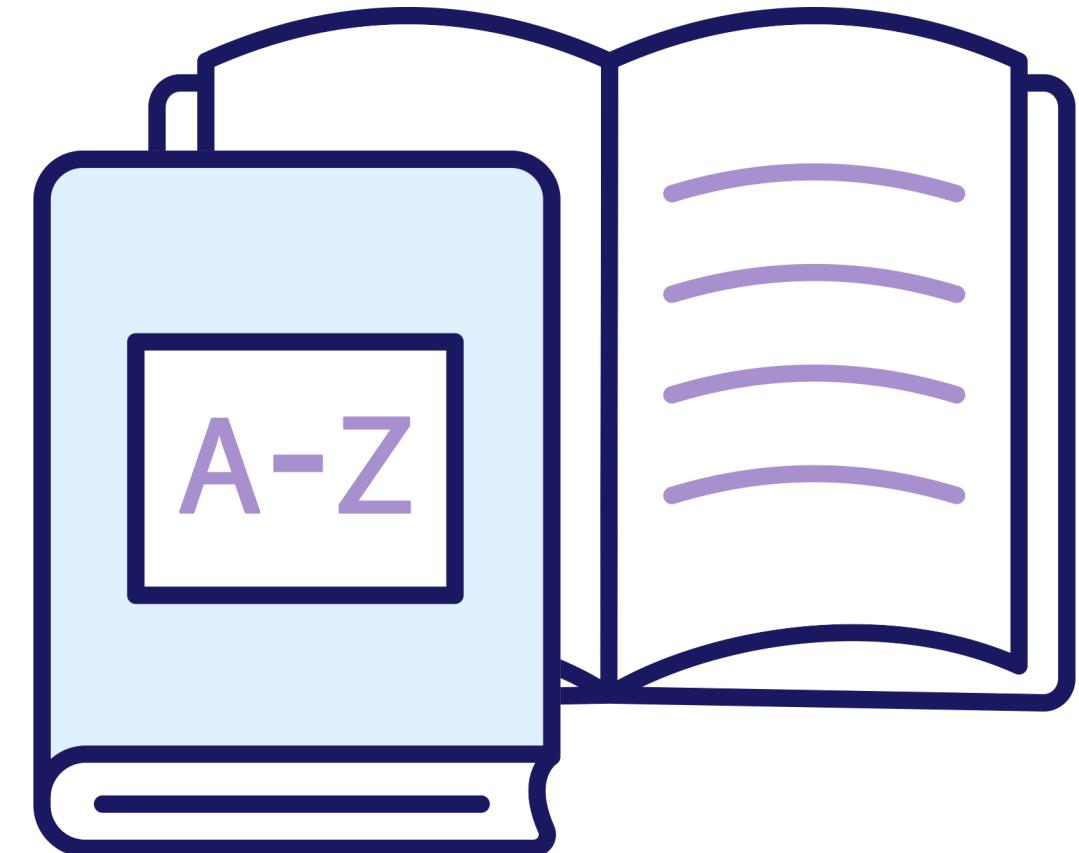


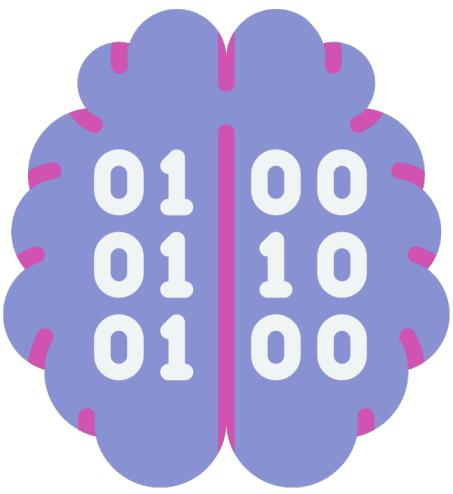




Remove stop word

- ◎ We used a nltk stop words
- ◎ Common words
- ◎ add new stop word after topic model

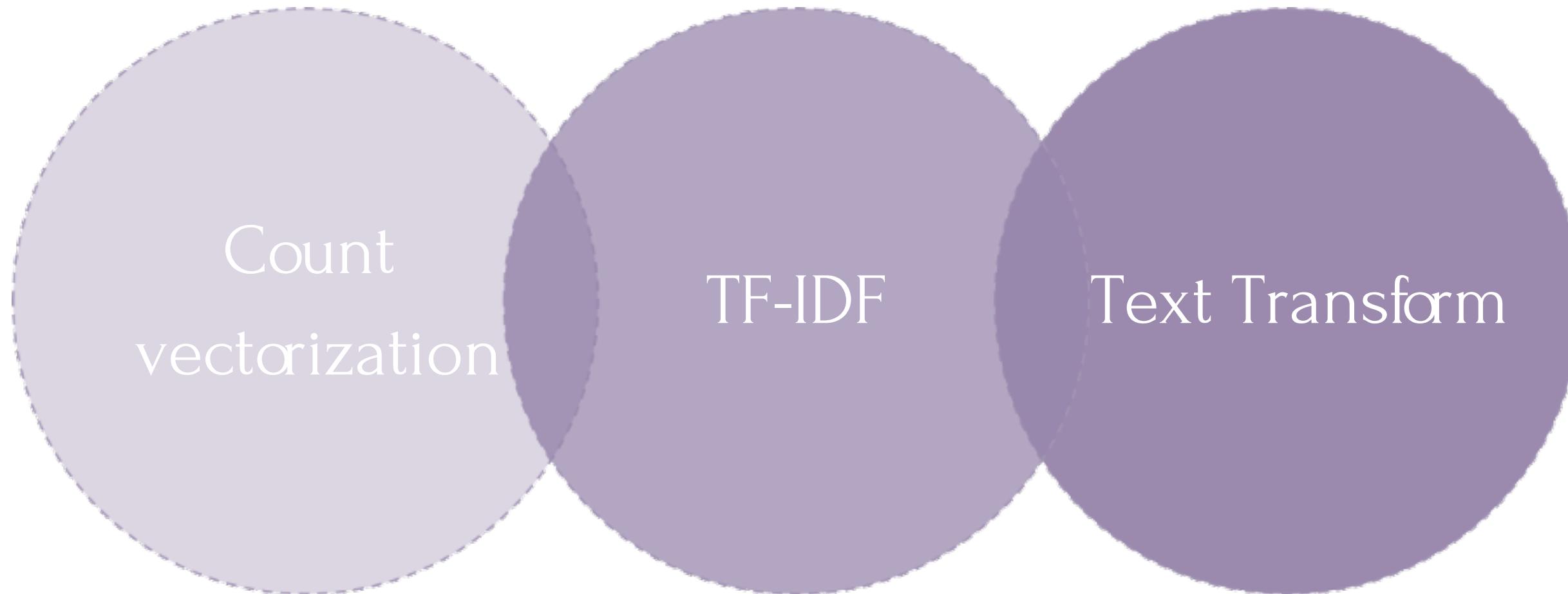


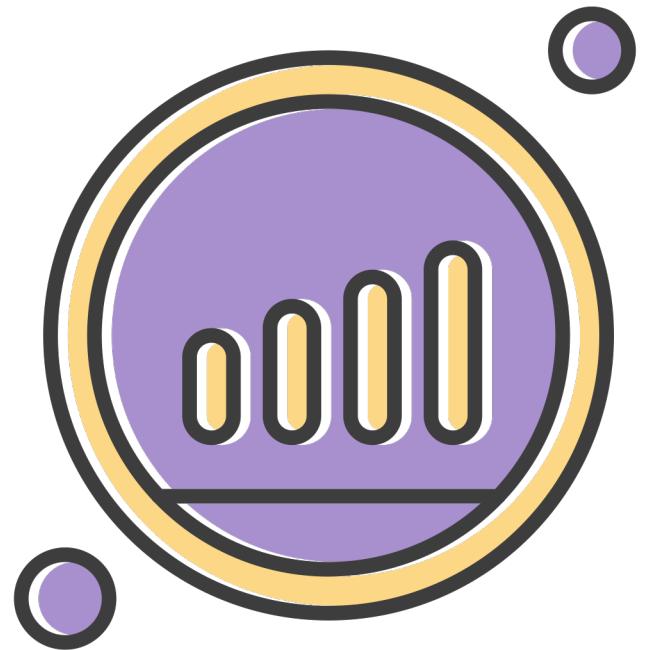


Word Cloud after clean



Vectorization

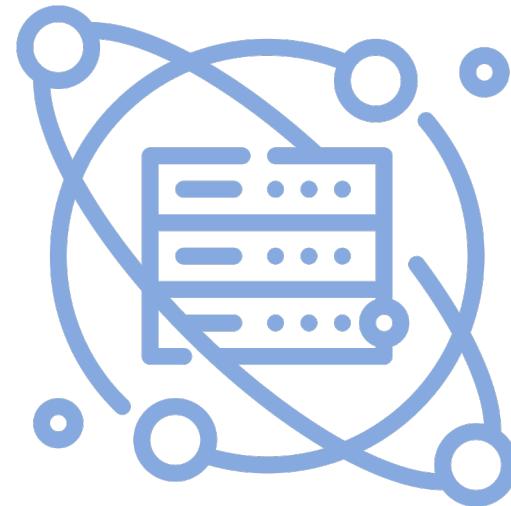




Topic Modeling

- LSA (components=5)
- NMF (components=5)
- LDA(components=5)





NMF and TF-IDF Vectorizer is best models

AI

information
really
real life problem
simulate

ML

learn
explain
great
help

DL

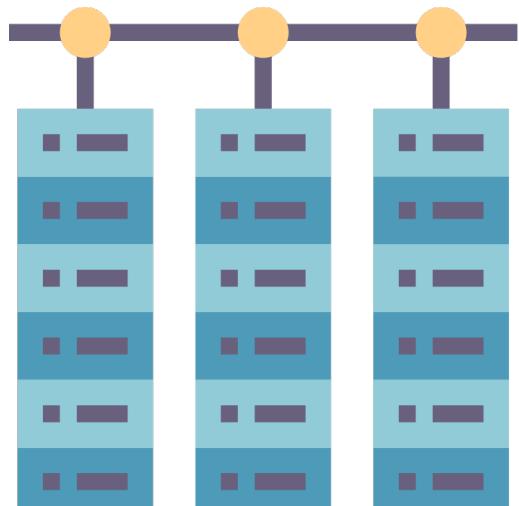
big data
data mining
technique
algorithm
statistic

DS

simple
easy
business
concept

Data analysis

analysis
material
useful
clear
topic



AI Field

reference
still
using
life
want
page
theory
without
learn
different
content
text
really
learning
give
something
information
interesting
data
author
come
year
science
useful
little
every
problem
concept
part
question
world
topic
know
better
best
great
provide
feel
material
algorithm
algorithim
statistic
explanation
thing
understand
help
study
written
review
python
written
cover
information

ML Field

explain
written content life
best provide python learning
love know basic
problem great' thing dashboard
great must year
interesting text
knowledge every
help story
useful still material
topic learn bought
science cover
information give
algorithm
using looking detail
great

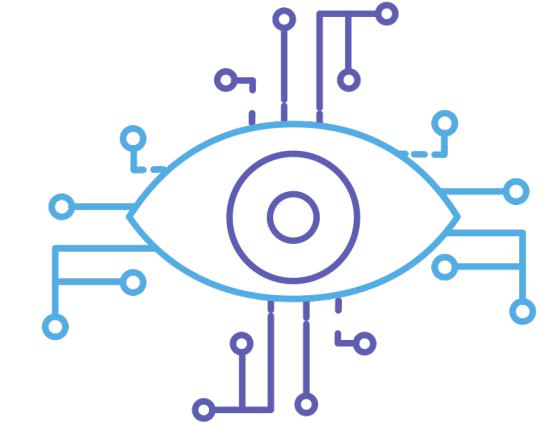
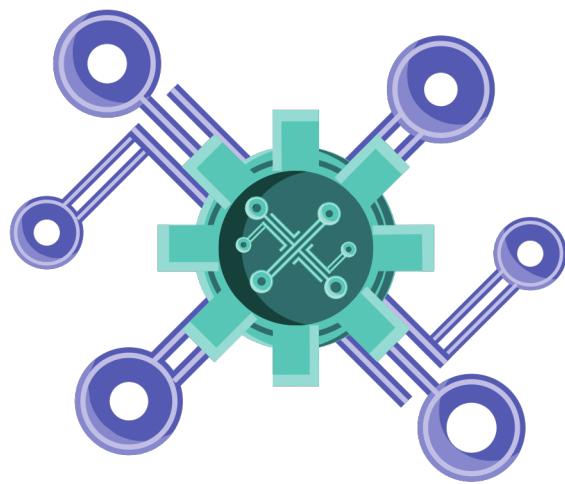
DL Field

DS Field

easy follow
basic interesting material
learning help
language better helpful
want algorithm
recommend python text
information made
concept practical thing
topic know life clear provide
really learn
easy understand
best statistic give
informative little understanding
little statistic give
informative business written easy
understanding useful
written science

Data analysis Field

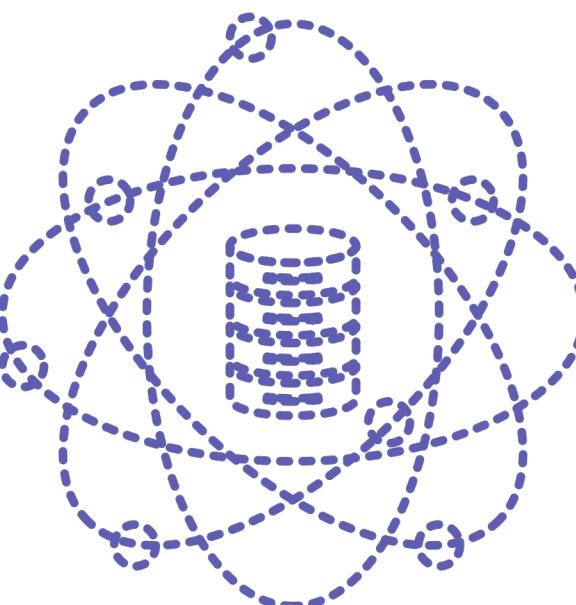
The word cloud is centered around the word "excellent" in a large, bold, yellow-green font. Other prominent words include "data" (dark teal), "information" (green), "science" (dark blue), "reference" (light green), "great" (purple), "understand" (dark purple), "explain" (blue), "theory" (yellow-green), "really" (light green), "topic" (yellow-green), "provide" (dark teal), "problem" (light green), "edition" (teal), "better" (green), "must" (light green), "statistic" (yellow-green), "understanding" (teal), "useful" (purple), "detail" (teal), "arrived" (yellow-green), "algorithm" (yellow-green), "recommend" (teal), "material" (light green), "level" (light green), "want" (purple), "easy" (light green), "best" (light green), "text" (dark blue), "using" (light green), "business" (dark blue), "method" (light green), "learn" (dark blue), "give" (light green), "condition" (yellow-green), "content" (teal), "condition'" (teal), "clear" (yellow-green), and "condition" (teal).



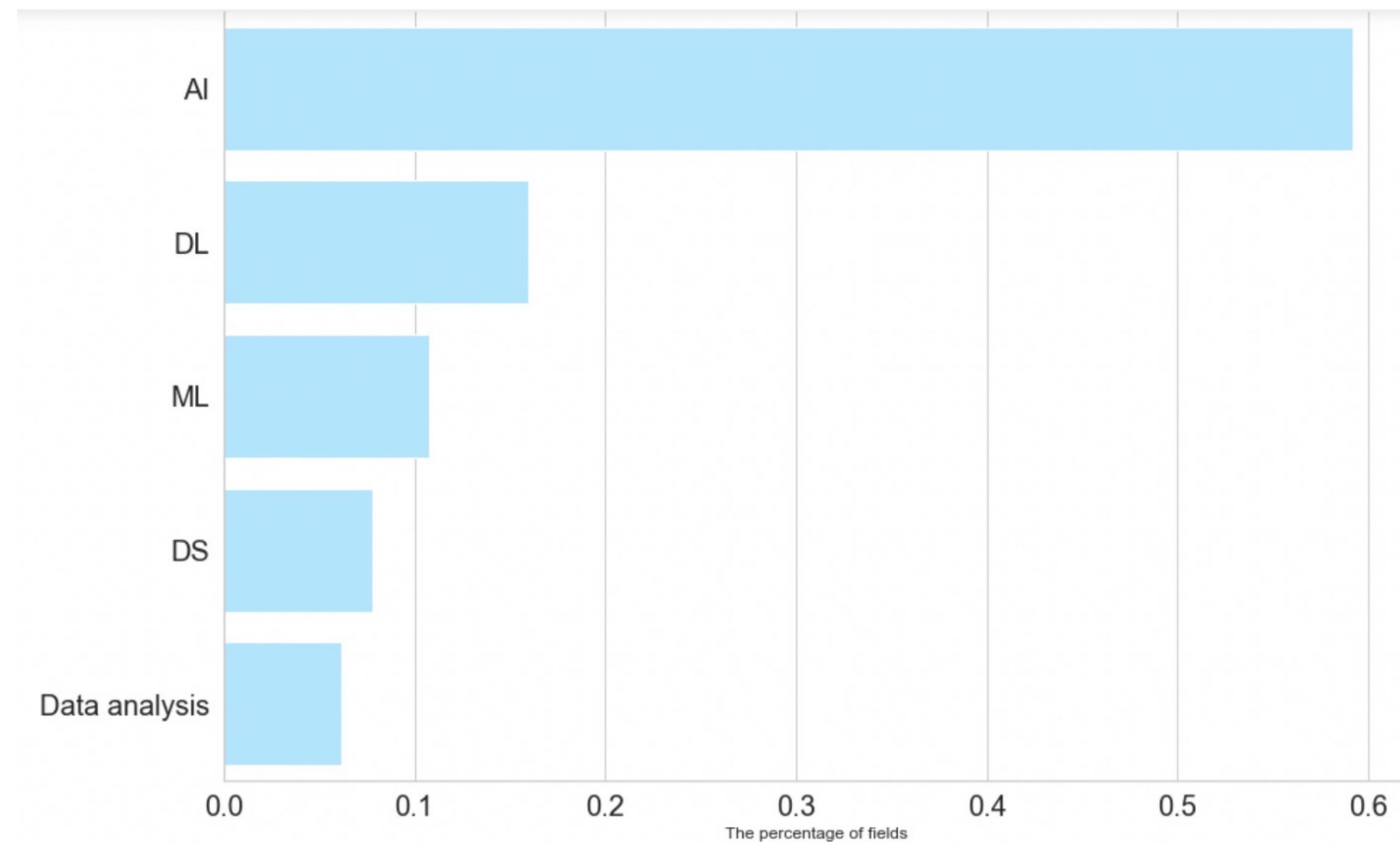
Label Comment

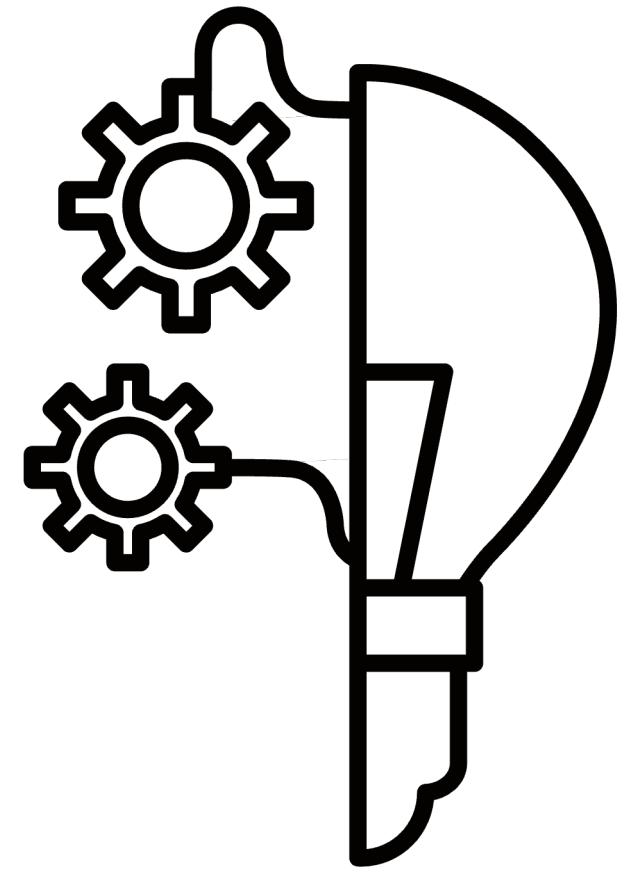
	index	comment	topics	Field	cluster
0	0	review sheer level topic engineer year enginee...	2	DL	1
1	1	capitalism best nearby bookstore reviewer gray...	0	AI	1
2	2	looking data science confused material video s...	2	DL	1
3	3	cover wide variety topic go enough detail trou...	0	AI	1
4	4	academic data scientist decade coined love mot...	2	DL	1
...
17492	20641	know library information science except intere...	0	AI	1
17493	20642	offer nice information thought becoming librari...	0	AI	1
17494	20643	disabled retired wished librarian perhaps enco...	0	AI	1
17495	20644	considered library opposed literacy profit lib...	0	AI	1
17496	20645	excellent pursuing librarian library school ac...	4	Data analysis	1

17497 rows × 5 columns

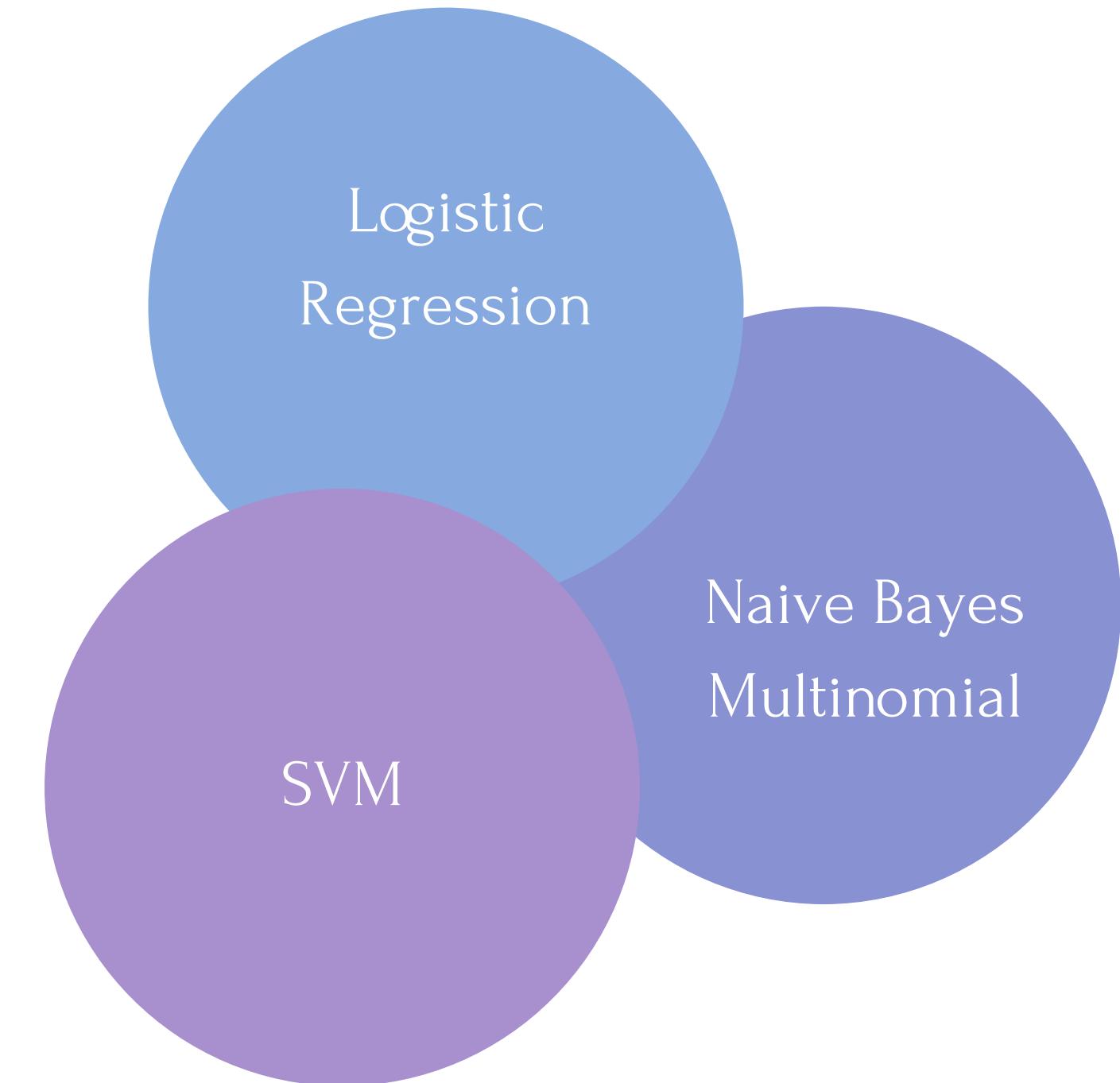


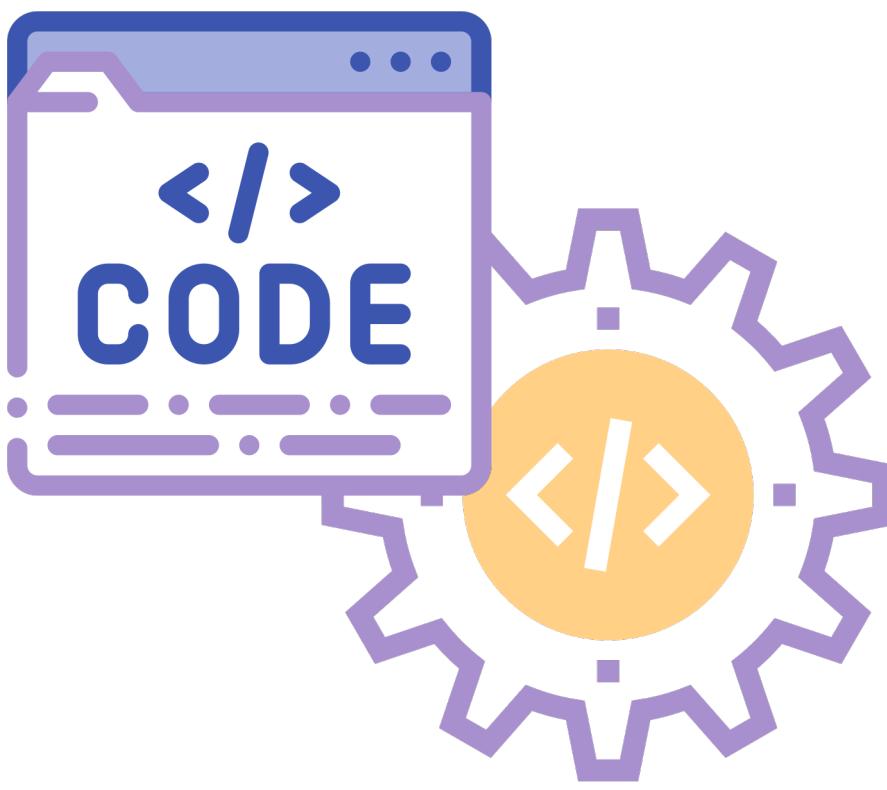
Label Comment





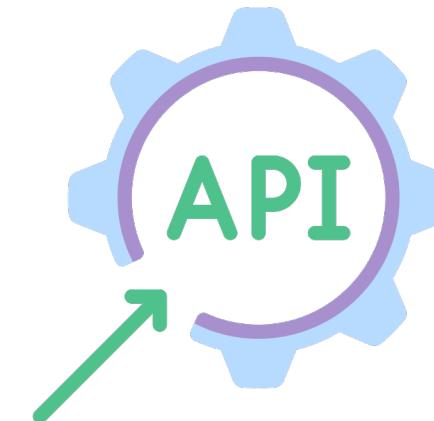
Classification model





Model name	Accuracy	F1 score
Logistic Regression	0.9055	0.9055
Naive Bayes	0.5999	0.5999
SVM	0.9361	0.9361

Flask

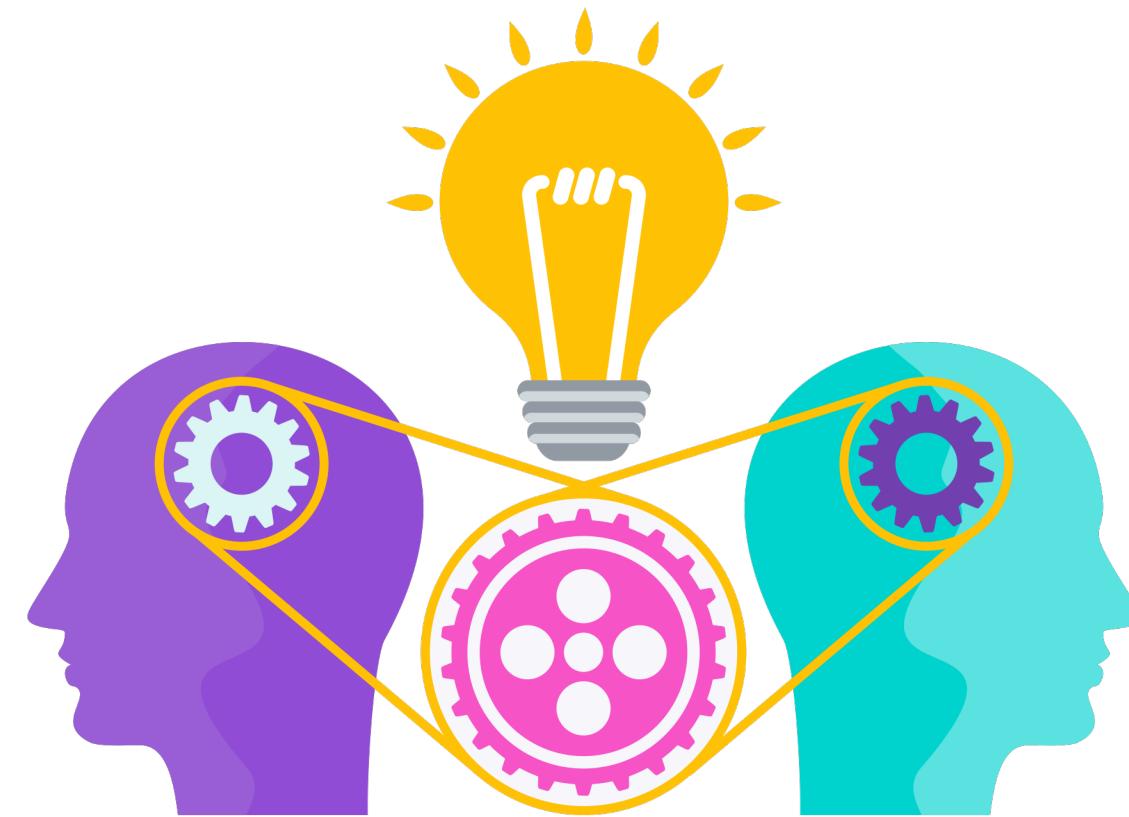


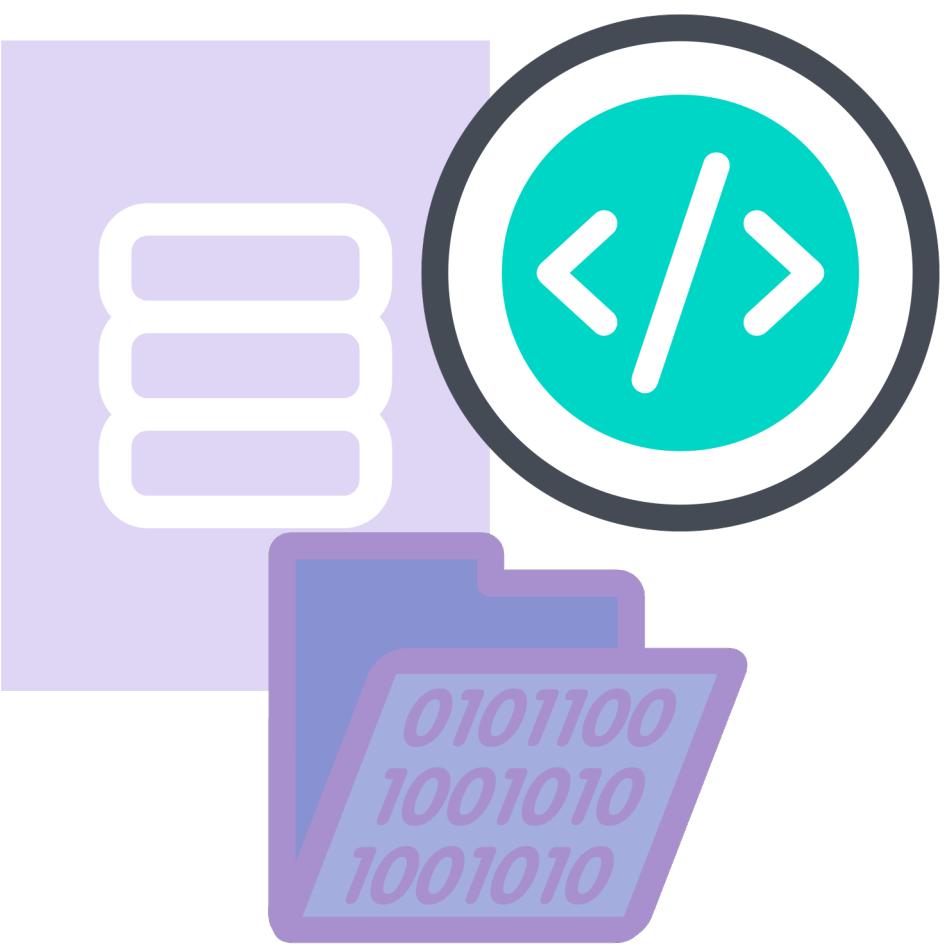
- ◎ Save the Model
- ◎ Create a function that takes the comment review and extracts the field
- ◎ Build website using Flask API <http://127.0.0.1:5000/>



Conclusion

- ◎ We analyzed 20,647 comment
- ◎ We were able to create a model capable of analyzing comment and detecting their field with an accuracy of 90%.





-We aspire to improve the model results to the best result

We apply a deep learning chatbot in DS field