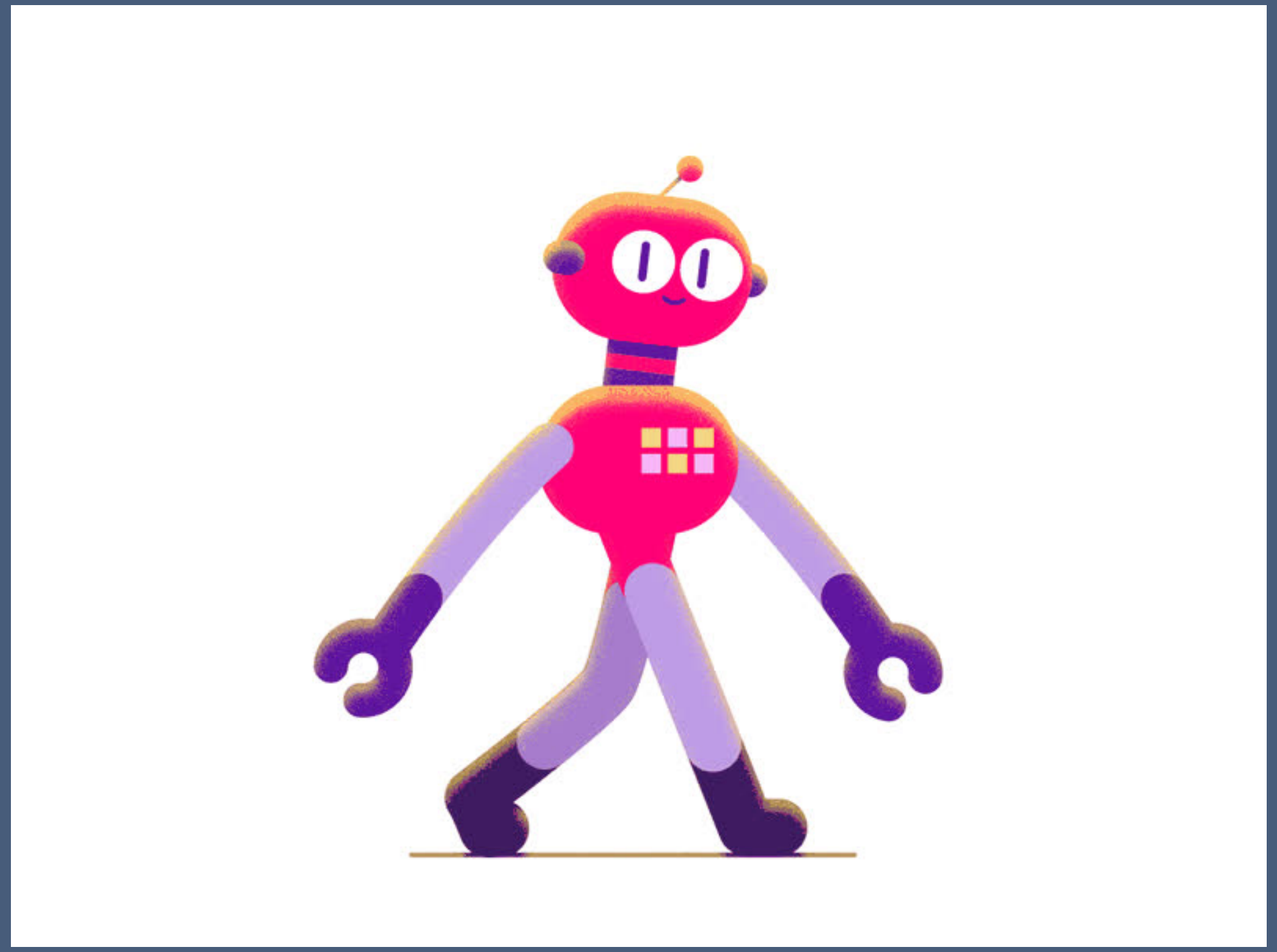
Bilbo the bot

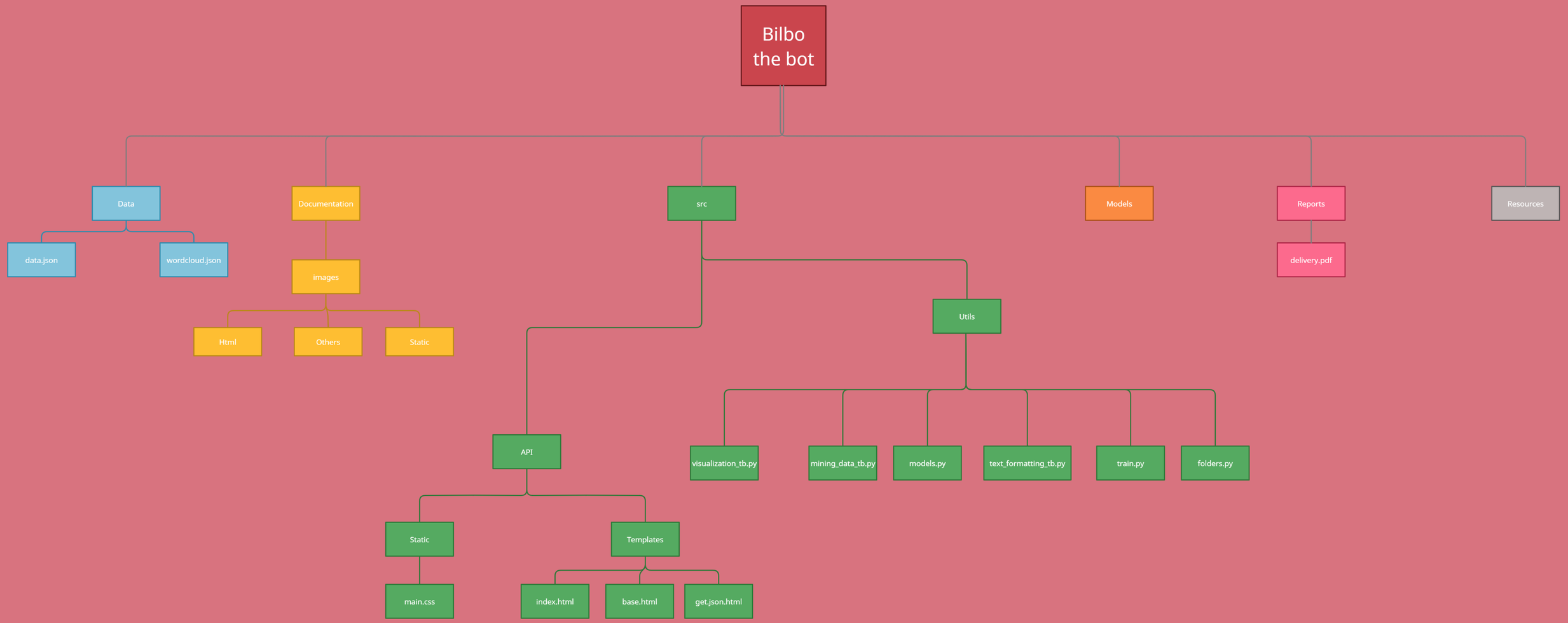


a fancy NLP project by @leosanchezsoler

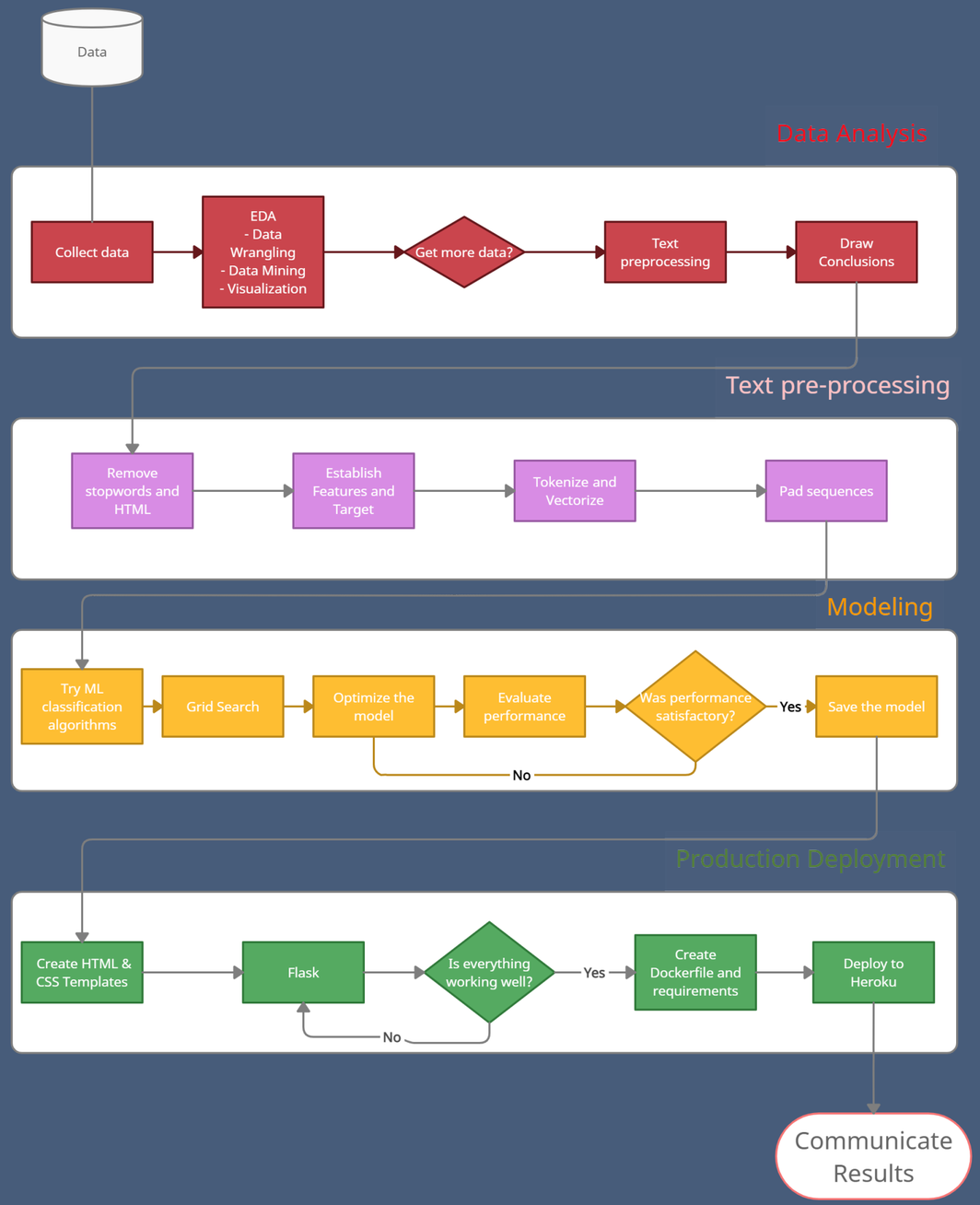
Only thing to do! On we go!

- Bilbo Baggins -

Folder structure


Bilbo the bot
Data
data.json
wordcloud.json
Documentation
images
Html
Others
Static
src
API
Static
main.css
Templates
index.html
base.html
get.json.html
Utils
visualization_tb.py
mining_data_tb.py
models.py
text_formatting_tb.py
train.py
folders.py
Models
Reports
delivery.pdf
Resources

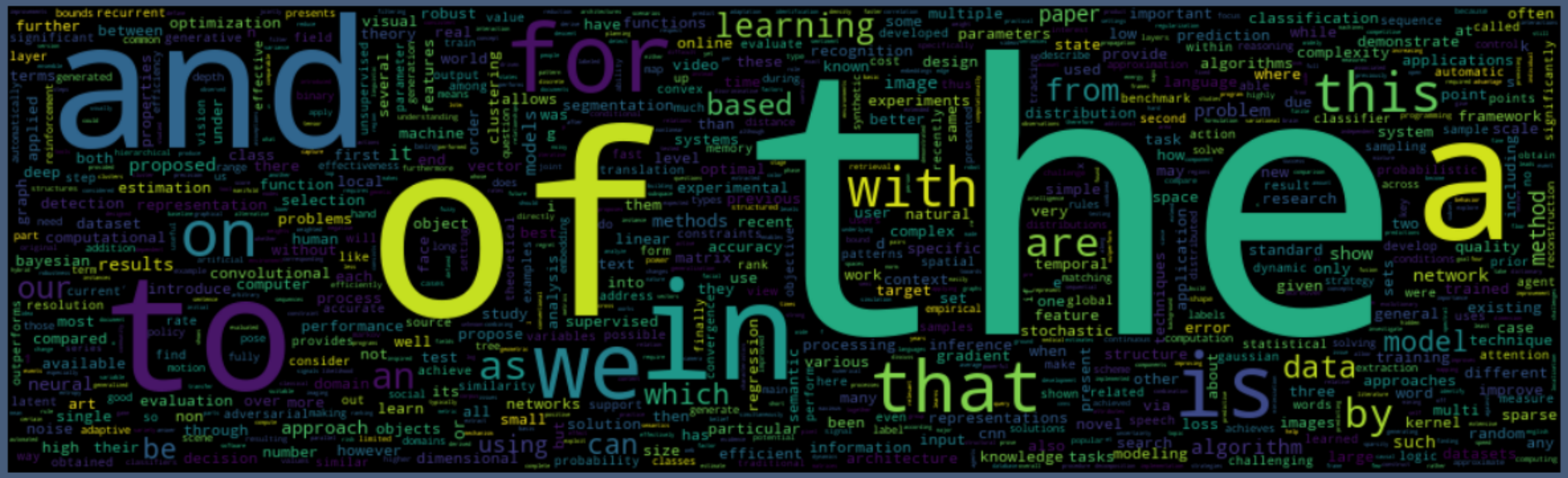
Project steps

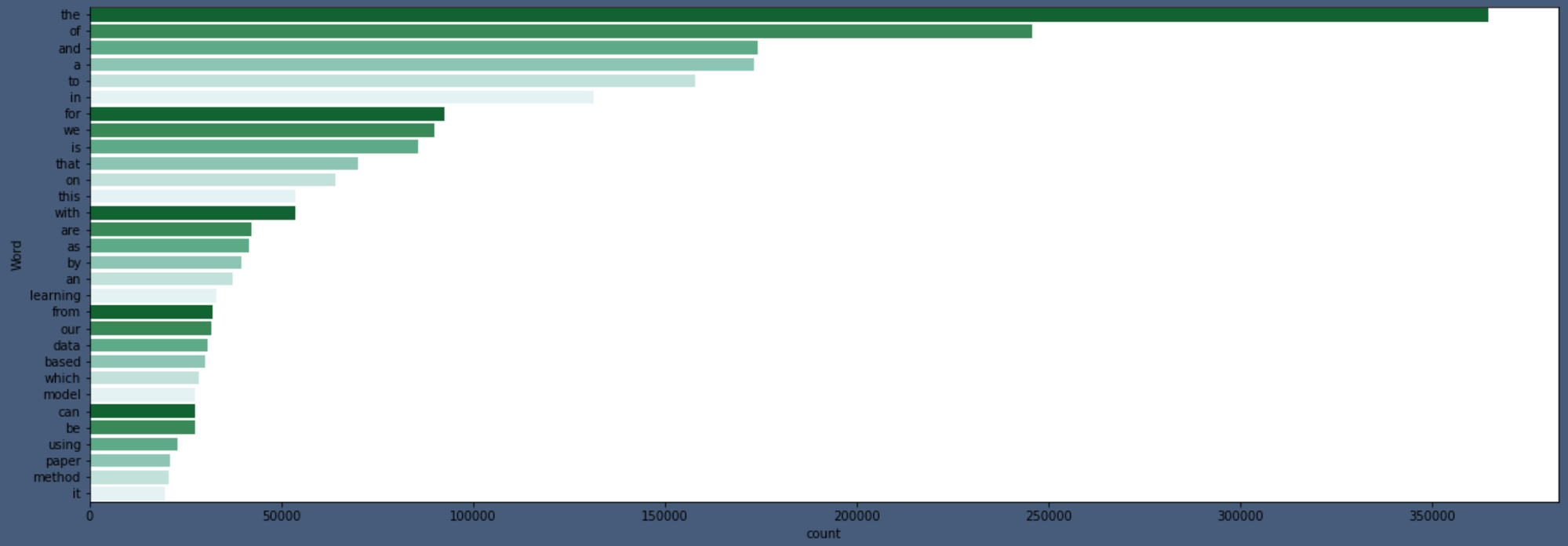


Text preprocessing

Before

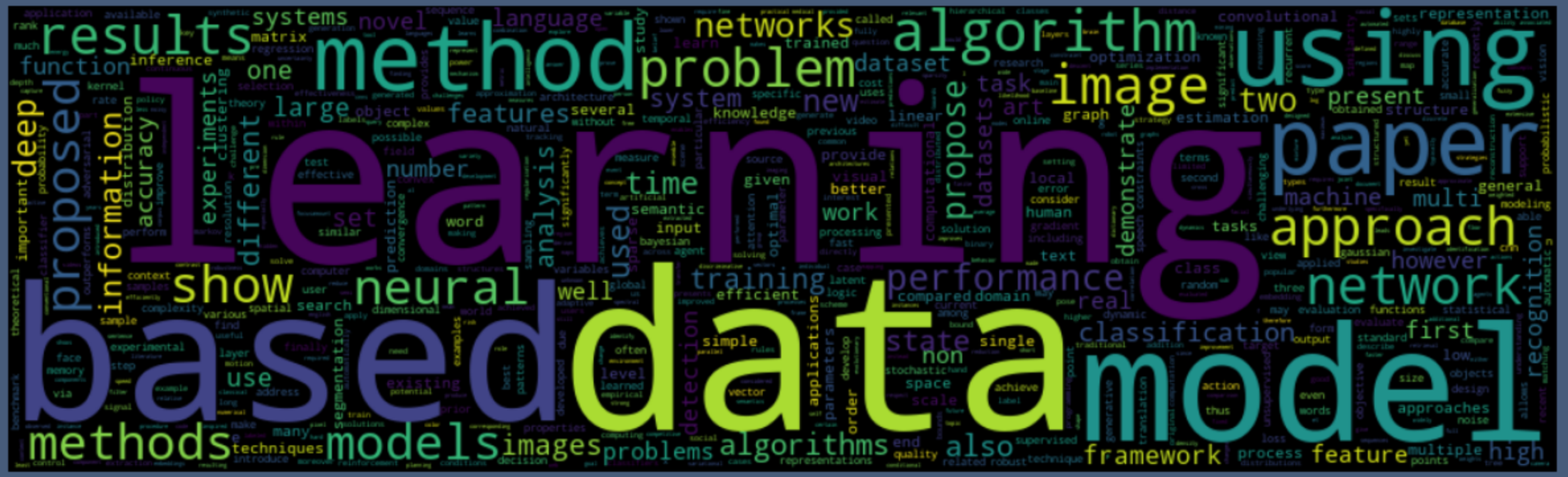
-Stopwords included-


and
of
the
a
to
in
for
we
is
that
with
on
this
learning
are
as
by
data
model
which
using
can
be
our
an
from
based
paper
method
network
algorithm


the
of
and
a
to
in
for
we
is
that
on
this
with
are
as
by
an
learning
from
our
data
based
which
model
can
be
using
paper
method
it
Word
0
50000
100000
150000
200000
250000
300000
350000
count

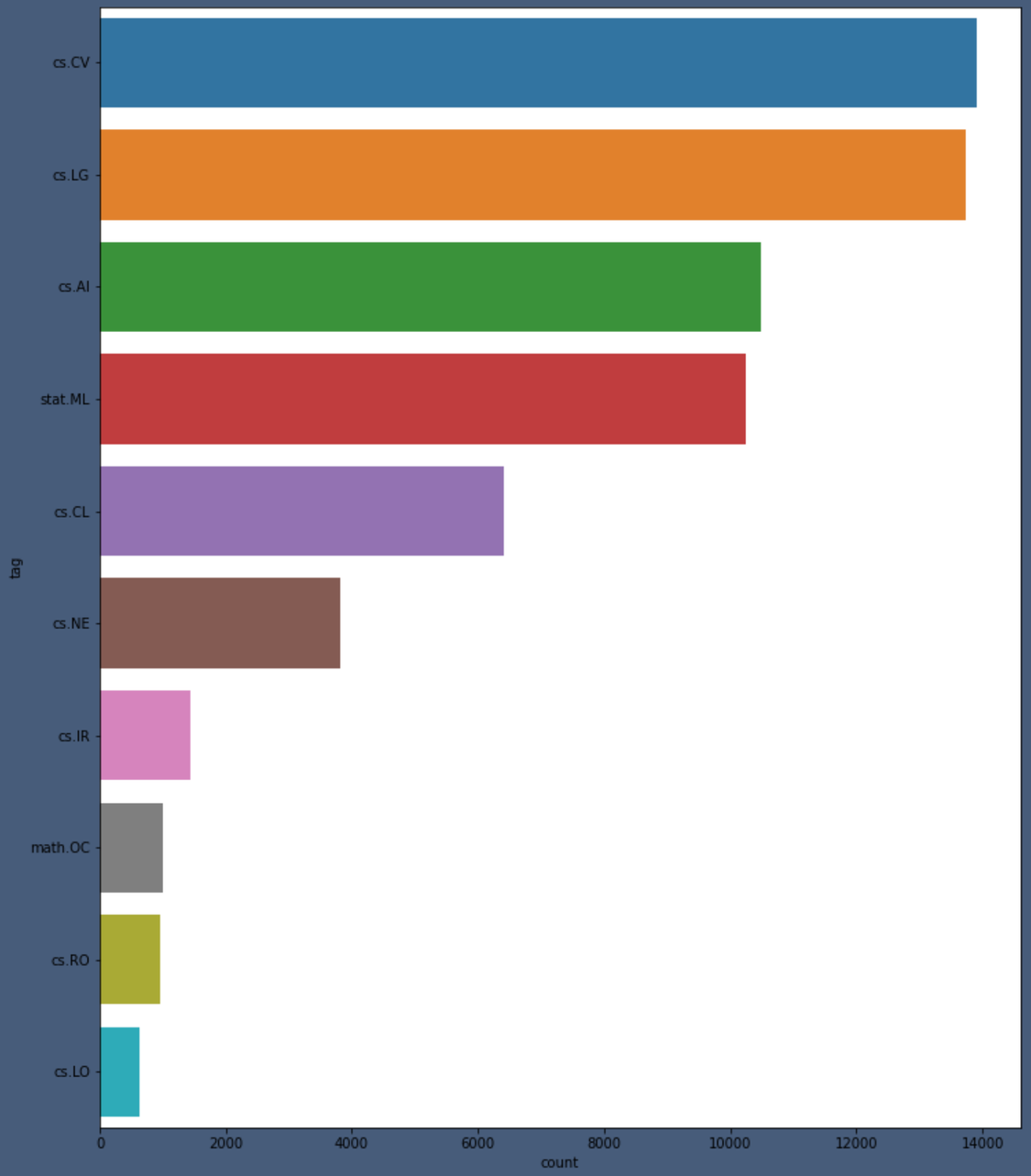
After

-Without stopwords-


learning
based
data
model
using
paper
method
algorithm
image
approach
problem
results
network
show
models
methods
neural
proposed
performance
networks
propose
two
deep
new
time
information
used
algorithms
also
state
Word
0
5000
10000
15000
20000
25000
30000
count

Frequent tags

-Data Science topics-



Top 3 topics

Computer Vision

Machine Learning

Artificial Intelligence

Algorith used

-Cosine Similarity-

Definition

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space. It is defined to equal the cosine of the angle between them, which is also the same as the inner product of the same vectors normalized to both have length 1.

Definition

The cosine of 0° is 1, and it is less than 1 for any angle in the interval(0, pi] radians. It is thus a judgement of orientation and not magnitude: two vectors with the same orientation hace a cosine similarity of 1, two vectors oriented 90° relative to each orher have a similarity of 0, and two vectors siametrically opposed have a similarity of -1...

Definition

...independent of their magnitude, The cosine similarity is particularly used in positive space, where the outcome is neatly bounded in [0, 1,]. The name derives from the term 'direction cosine': in this case, unit vectors are maximally 'similar' if they're parallel and maximally 'dissimilar' if they're orthogonal (perpendicular).

Definition

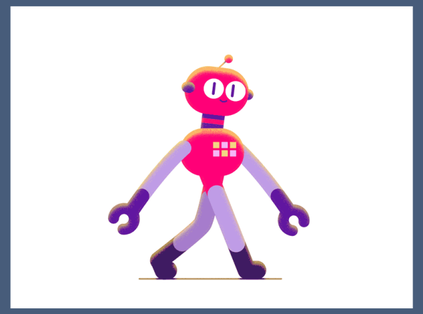
This is analogous to the cosine, which is unity (maximum value) when the segments subtend a zero angle and zero (uncorrelated) when the segments are perpendicular.

Once again... thanks, Wikipedia!

Deploy to production

This is what the web looks like

Bilbo the bot



Bilbo is pleased to help you in your Journey as a Data Scientist

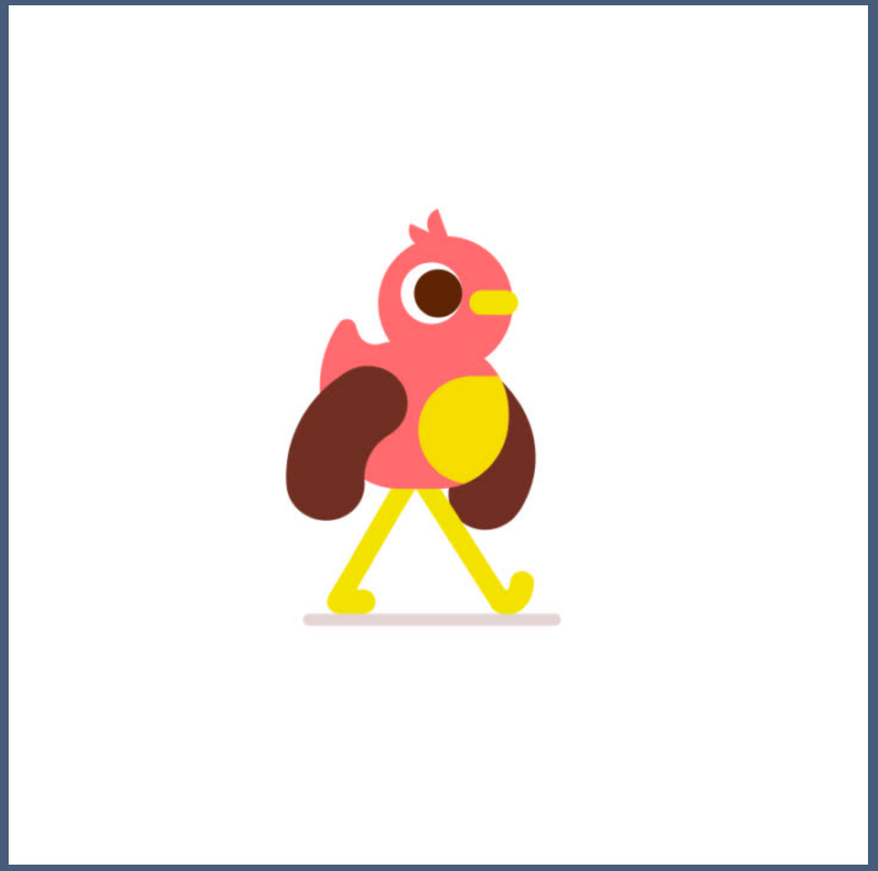
What do you want to do?

submit

Get Json

See the source code

Next steps



Deploy Bilbo to other apps

Make it conversation-like

Feed Bilbo with more data

Last but not least!

say hello to bilbo:

bilbo-the-bot.herokuapp.com

Any questions?

Good night, and good luck!

-Edward R. Murrow-

github.com/leosanchezsoler/bilbo-the-bot

linkedin.com/in/leonardosanchezsoler/