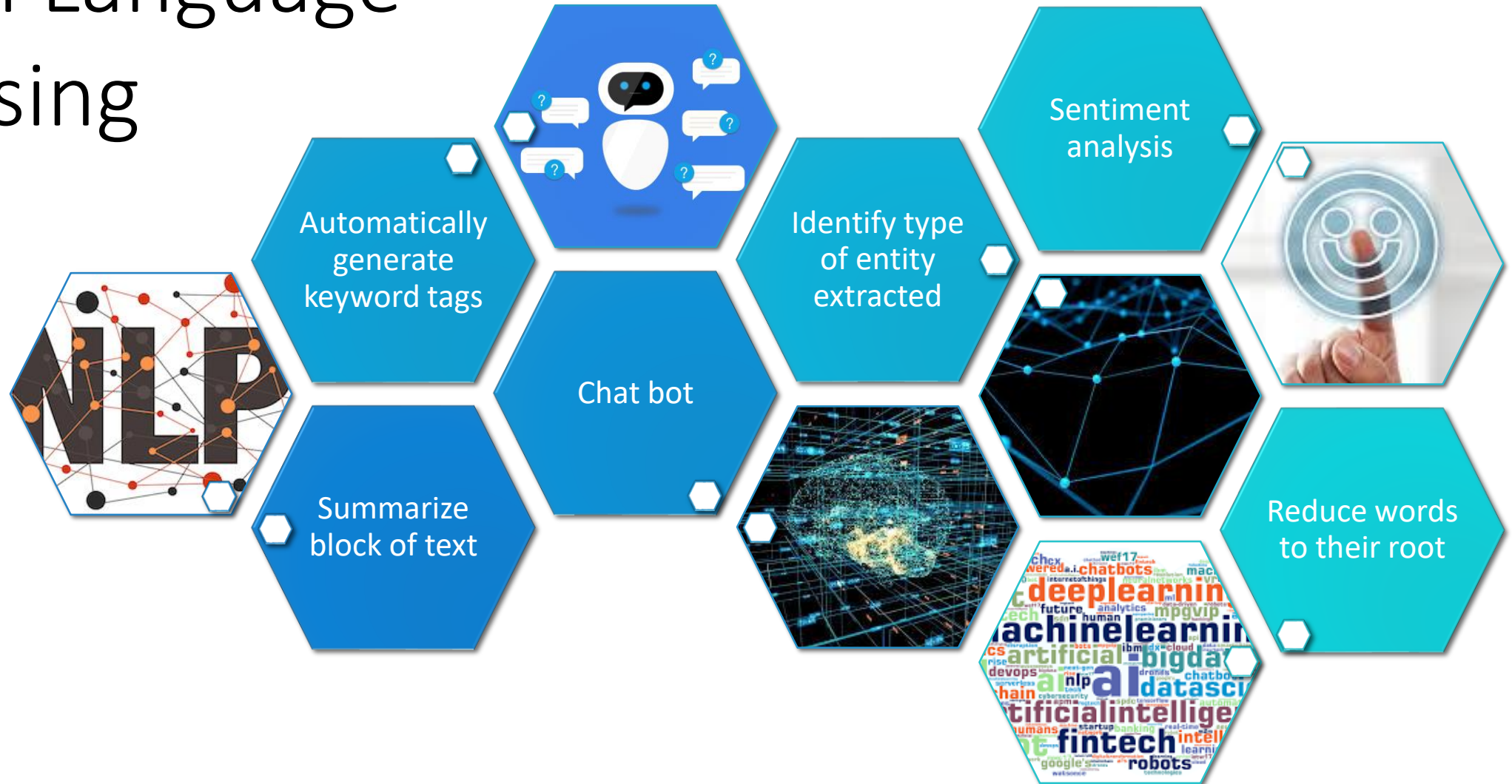


Sentiment analysis

TATIANA

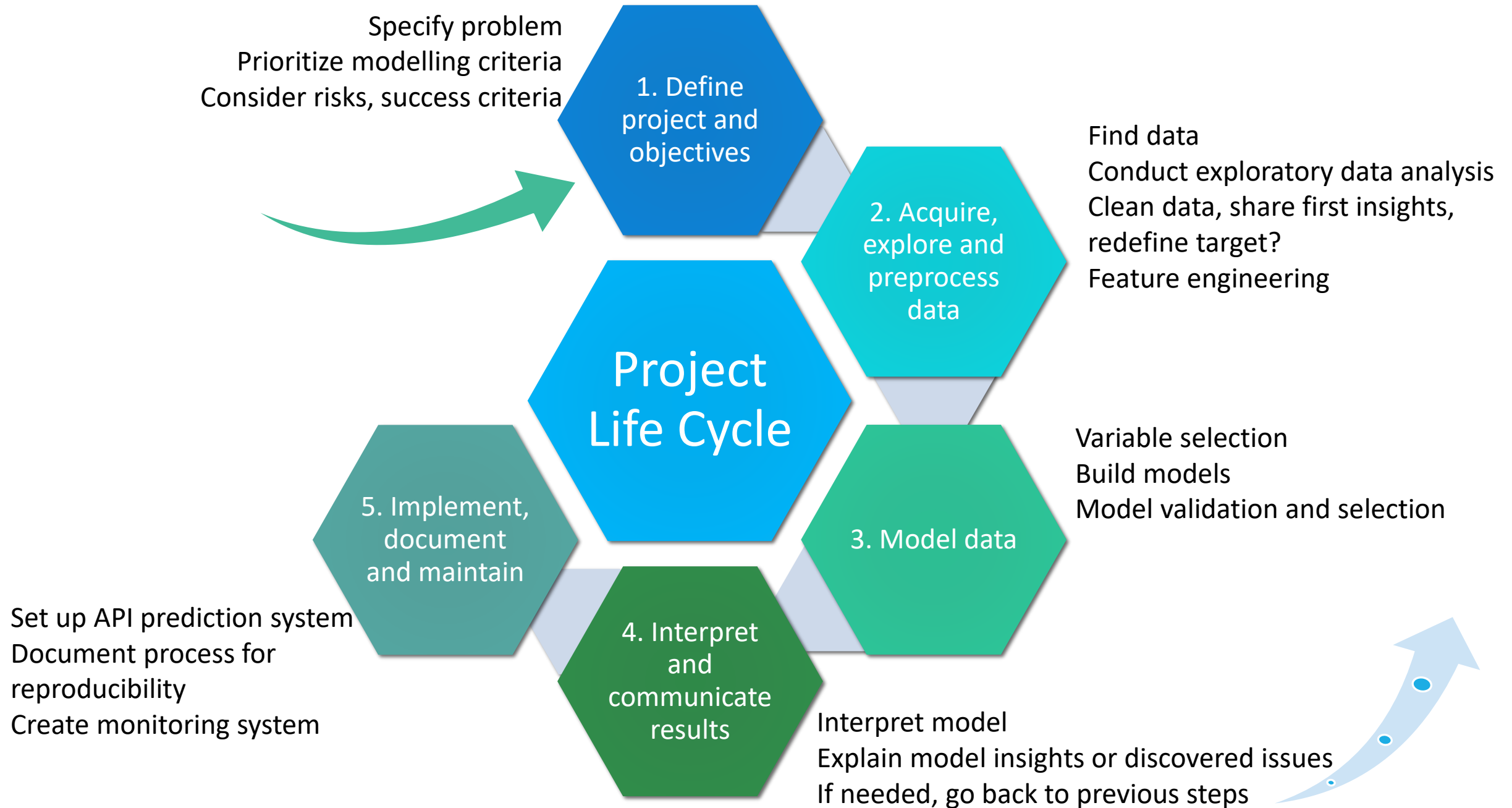
Natural Language Processing





Sentiment analysis

- Unbelievably disappointing
- Full of zany characters and richly applied satire, and some great plot twists
- This is the greatest screwball comedy ever filmed
- It was pathetic. The worst part about it was the boxing scenes.



Project

Movie review prediction



Build a movie review sentiment classifier



15,423 movie reviews with a score from 1 to 10 from IMDB

Data preprocessing



Remove punctuation, duplicated letters, numbers



Normalization = to *normalize* the words in the corpus by trying to convert all of the different forms of a given word into one
Stemming and Lemmatization



Stemming: basic rules to chop off the ends of words.



Lemmatization: identify the part-of-speech of a given word and then apply more complex rules to transform the word into its true root

Feature extraction

Features:

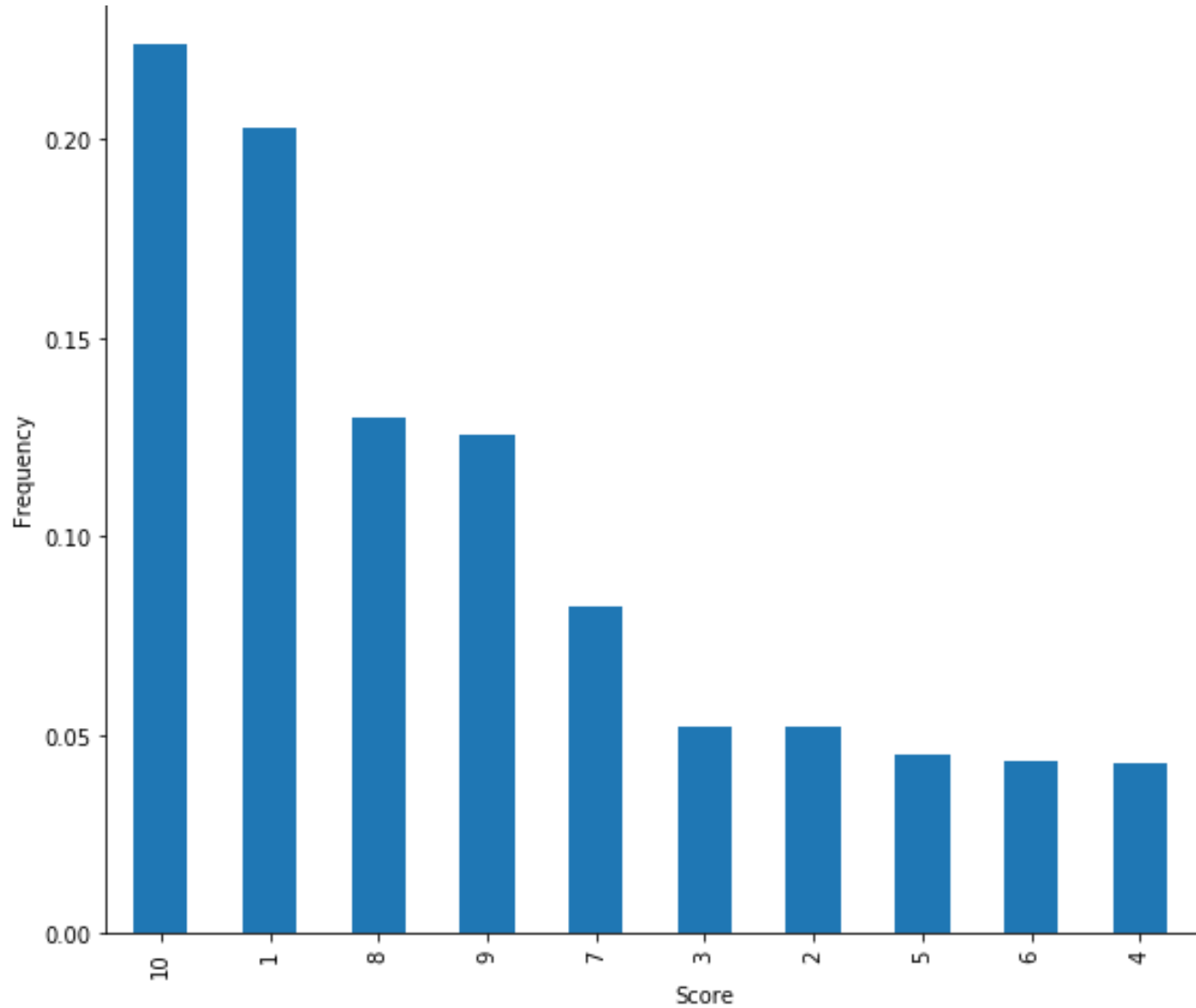
- words (bag-of-words)
- **n-grams**
- parts-of-speech (e.g. Adjectives and adjective-adverb combinations)
- opinion words (lexicon-based: dictionary or corpus)
- valence intensifiers and shifters (for negation); modal verbs; syntactic dependency

Feature selection based on

- frequency
- information gain
- odds ratio (for binary-class models)
- mutual information

Feature weighting

- term presence or term frequency
- frequency-inverse document frequency (TF-IDF)
- term position : e.g. title, first and last sentence(s)



Target value

Unbalanced classes

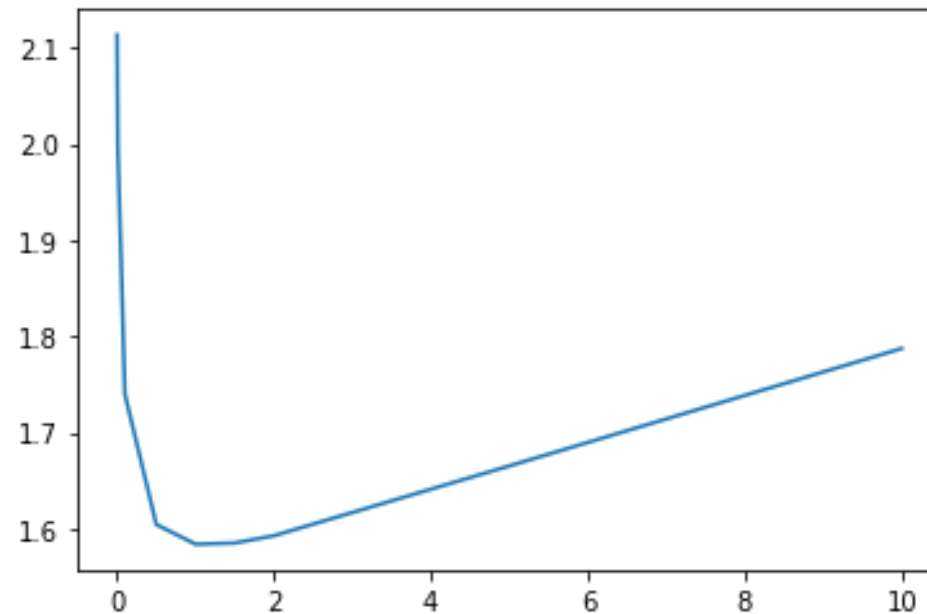
More prone towards positive scores

After preprocessing, **68774** words in the dictionary

Classification / Regression

Baseline algorithms based on cross
validation with 5 folds

Ridge	1.911
-------	-------



Selection of optimal
parameters **Alpha** in
Ridge regression

Optimal parameters

with en- and bi-grams
and set of English stop
words

SVM

- $C = 6$
- `class_weight='balanced'`
- `loss='squared_hinge'`
- `multi_class='ovr'`

1.3381

GradientBoosting

- `Subsample = 0.8`
- Need to adjust all parameters

1.6504

Ridge

- $\alpha = 0.45$

1.5486

SGD

- `Max_iter = 600`
- `Tol = 1e-4`

1.6904

Predictive words in SVC

('worst', 5.40054035348061)

('wast', 4.163555236735492)

('aw', 3.8850219620992594)

('terribl', 3.6038151303303647)

('garbag', 3.162175429687793)

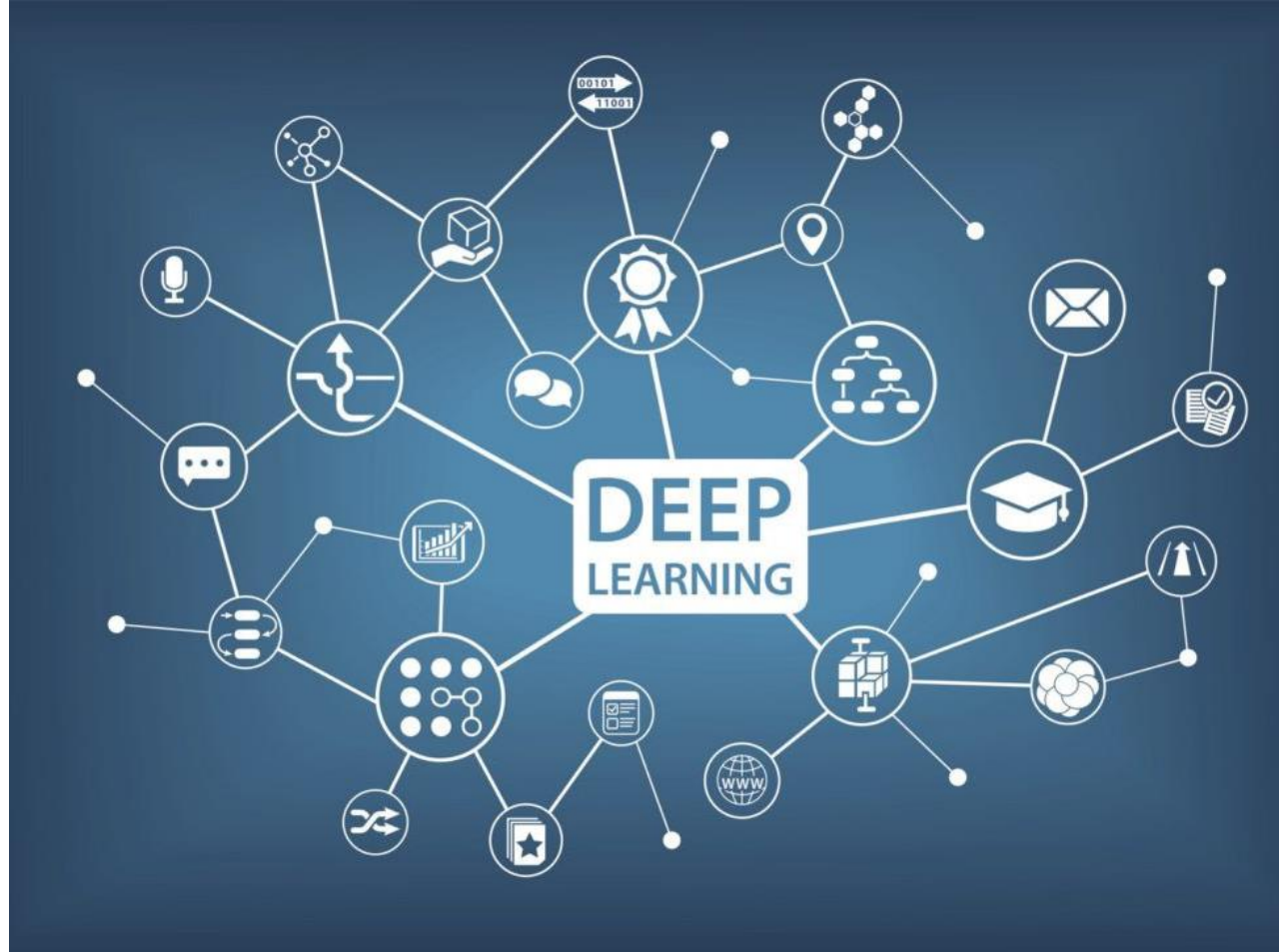
('great', -2.174787957631803)

('pretti', -1.9693168731975181)

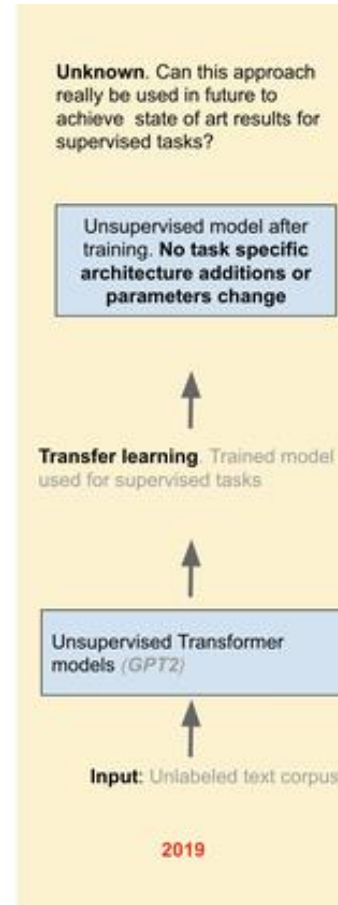
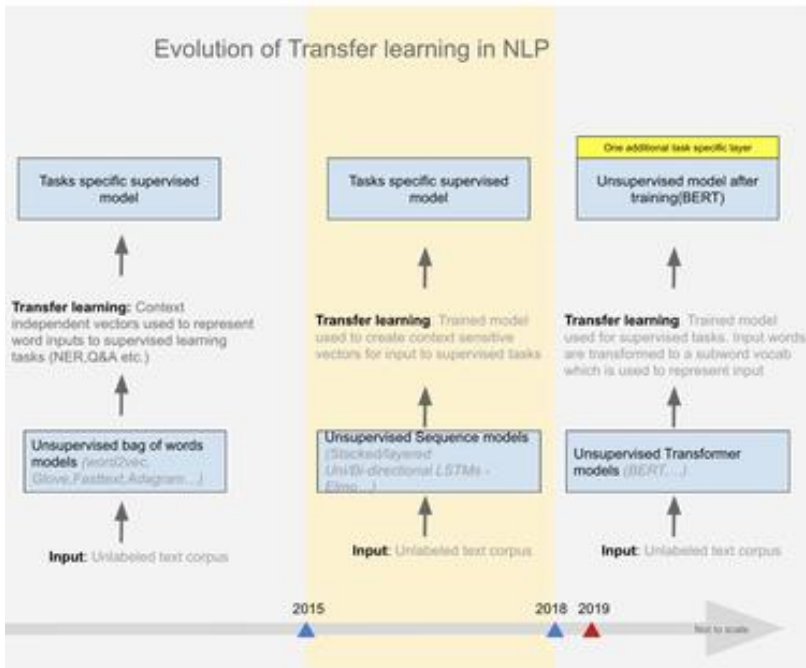
('veri', -1.872858223683766)

('best', -1.8588912444230878)

('enjoy', -1.8385690106182409)



Deep learning in NLP



Top NLP models

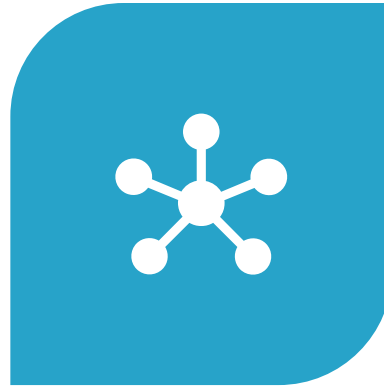
Bert, Bidirectional
Encoder
Representations
from Transformers

GPT-2, based on
generative pre-
trained model
(GPT)

Deep Learning



KTRAIN - BERT




KTRAIN – NBSVM



KERAS -LSTM

LSTM

Layer (type)	Output Shape	Param #	
===== (Embedding)	(None, 48, 300)	4119900	embedding_11
			lstm_21 (LSTM)
(None, 48, 128)	219648		
			lstm_22 (LSTM)
(None, 64)	49408		
			dense_21
(Dense) (None, 100)	6500		
			dropout_11
(Dropout) (None, 100)	0		
			dense_22
(Dense) (None, 5)	505		
===== 4,395,961	Trainable params: 4,395,961	Non-trainable params: 0	Total params:



The best way
to predict the
future is

to create it

Peter Drucker

1

Create app.py with a selected classifier

2

Call app.py, train a model and open <http://127.0.0.1:5000> in your browser

3

Input English text and check a predicted score



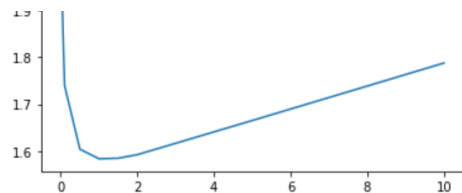
Deployed application

tprojects.pythonanywhere.com



Summary

- SVM gives fast and relatively accurate classifier for sentiment analysis task
- Deployed app at tprojects.pythonanywhere.com
- BERT, LSTM and Gradient Boosting require long training time and bigger memory, but these methods could give better results after adjustment
- Many libraries and labelled datasets for binary classification
- Explainable AI



Generate pipeline and set of parameters for grid search.

```
In [10]: > def make_pipeline(vectorizer, transformer, classifier):
>         return Pipeline([
>             ('vectorizer', vectorizer),
>             ('transformer', transformer),
>             ('classifier', classifier)
>         ])

In [11]: > params_grid_vectorizer = {
>     'vectorizer_max_df' : [0.85, 0.95, 1],
>     'vectorizer_ngram_range' : [(1, 1), (1, 2), (1, 3)]
>     'vectorizer_stop_words' : ['english', None]
> }
```

```
C:\Users\aleff\Documents\PythonScripts\Untitled Folder\myapp>python app3.py
C:\Users\aleff\Anaconda3\lib\site-packages\sklearn\externals\joblib\__init__.py:15: DeprecationWarning:
externals.joblib is deprecated in 0.21 and will be removed in 0.23. Please import this function
from joblib, which can be installed with: pip install joblib. If this warning is raised when
importing joblib, you may need to re-serialize those models with scikit-learn 0.21+.
warnings.warn(msg, category=DeprecationWarning)
* Serving Flask app "app3" (lazy loading)
* Environment: production
WARNING: This is a development server. Do not use it in a production deployment.
Use a production WSGI server instead.
* Debug mode: on
* Restarting with stat
C:\Users\aleff\Anaconda3\lib\site-packages\sklearn\externals\joblib\__init__.py:15: DeprecationWarning:
externals.joblib is deprecated in 0.21 and will be removed in 0.23. Please import this function
from joblib, which can be installed with: pip install joblib. If this warning is raised when
importing joblib, you may need to re-serialize those models with scikit-learn 0.21+.
warnings.warn(msg, category=DeprecationWarning)
* Debugger is active!
* Debugger PIN: 299-287-796
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```



Avengers: Endgame (2019)

User Reviews

+ Review this title

8,410 Reviews

☐ Hide Spoilers Filter by Rating: Show All Sort by: Helpfulness

★ 10/10

An experience you'll gonna remember forever.

raudafitriani 24 April 2019

Warning: Spoilers

4,243 out of 6,455 found this helpful. Was this review helpful? [Sign in](#) to vote.

[Permalink](#)

★ 10/10

I have to admit, I got so emotional all throughout the movie

andiemartinez23 27 November 2019

And some parts brought me to tears. The cast was phenomenal and I think every

<https://www.amazon.com/Google-Pixel-Memory-F>

Showing 1-10 of 791 reviews



Ian Lucas

★★★★★ **The perfect phone for 99% of people**

May 16, 2019

Color: Clearly White | Style: 3a | Pattern: Single

Want the best camera? Want a great looking screen? Want amazing battery life? Want gre 99% of people. Now if all you care about is specs, just walk away, it's not a flagship, but yc regular phone user, takes some photos, watches some videos, strolls through social media phone for you, save your money. Some more pros, Google photos back up is unlimited, 3 t keep pictures completely separate. If you use any type of voice assistant, you'll love havin only features, screen calling in the US is an amazing benifit. Also, if you want to switch fro

697 people found this helpful

Helpful

7 comments

Report abuse



Mike M.

tprojects.pythonanywhere.com

Machine Learning App for

Sentiment Analysis

Enter a review below please (applied for English text)

Having read the first few Harry Potter books before 2001 and hearing about the hype for the first movie, I was excited. I heard there was going to be an all-British cast (which makes sense, right?) and we'd get to see a live version of one of the defining novels of our generation. From what I remember I went with my family and a family friend to see the movie the day after Christmas and was pleasantly amazed. After the movie was over, I watched the credits and discovered some familiar names (the late Alan Rickman, Sister Act's Maggie Smith, James Bond 007's Robbie Coltrane, and Star Wars' Warwick Davis); others not so familiar (the kids, some of whom had their debut). But it was a good movie and was a party of colors and sights for all to see. This is easily my favorite of all the Harry Potter films. The catalyst of the movie series!

[Predict](#)

The aim for this project is to build up an application which will predict the sentiment of an input text. 15,000 IM removing emoticons, numbers, URLs. Several models were evaluated. The minimal MAE for this data was obtai The trained model is used for the application to make a prediction. This app might be used for sentiment analysis performance or game reviews.

tprojects.pythonanywhere.com/predict

Predicted sentiment of the entered text:

[10]

Scaled from 1 (the most negative) up to 10 (the most positive)