



# A BITE OF NLP

*Yu-Han Chen*

# About Me

NLP Data Engineer @Itemize

Data Scientist/ Teaching Assistant @NYCDSA

Management and Systems @NYU

Programming newbie, NLP enthusiast, baking lover

# What's NLP

Natural Language Processing

A field of computer science and AI concerned with the interactions between computers and human languages

# NLP Recipe

5 minutes • Serves 40

**Objective:** Document classifier

## Inputs

- Text data

## Preprocessing

- Lowercasing
- Stopwords removal
- Lemmatization
- Tokenization

## Feature Extraction

- Count Vector
- TF-IDF

# Google Cloud Vision

OCR (Optical Character Recognition)

Extract texts, labels, locations, logos, faces etc.

[Demo code](#)

# Document Classification

Definition of each category

Ground truth data

Algorithm selection

Model evaluation metrics

Confidence score

# Extra Resources

[NLTK Tutorial](#)

[Spacy, Gensim Tutorial](#)

[Coursera: deeplearning.ai](#)

[Oxford: Deep Learning](#)

[Stanford: Natural Language Processing with Deep Learning](#)

# Contact

Yu-Han Chen

yhc441@nyu.edu

LinkedIn: [yuhanchen36](#)

GitHub: [yuyuhan0306](#)

