

# NLP For Economists

## NLP and Machine Learning - Topic Overview

Sowmya Vajjala

Munich Graduate School of Economics - LMU Munich

Guest Course, October 2020

# Goals

- ▶ give an overview of some NLP and Machine learning methods potentially useful to you
- ▶ note on installing libraries

Note: just a short video giving an overview of the topic.

# Topics

1. Regular expressions are used to do pattern based information extraction from data.
2. Corpus collection approaches if we have to collect our own (annotated) data
3. Exploratory analysis of corpus to understand its distribution
4. Text classification: to categorize data and assign (known set of) labels
5. Information extraction: key phrase extraction, named entity recognition
6. Topic modeling: uncovering latent topics in a large corpus
7. Text summarization: making a summary automatically

# Libraries

1. NLTK, and additional stuff with nltk downloader - <https://www.nltk.org/install.html>
2. sklearn - <https://scikit-learn.org/0.16/install.html>
3. gensim - <https://radimrehurek.com/gensim/>
4. spacy, with any one English model (small. medium. large) - <https://spacy.io/usage>
5. lime - <https://pypi.org/project/lime/> to interpret text classifier predictions

# Assignment3

1. 2 questions - 1 requires writing small bits of code.
2. Goals:
  - ▶ To get you started on thinking about the relevance of these methods for your own research.
  - ▶ To familiarize you with using existing NLP tools

# Conclusion

1. Treatment of individual topics is introductory - not very advanced.
2. I will provide links to additional readings wherever possible.
3. Code examples and publicly available github repo links will be provided too.
4. While doing them, start thinking about how these topics are relevant to economics research.