

# WORD SENSE DISAMBIGUATION WITH NATURAL LANGUAGE PROCESSING

Hải Nguyễn Âu

Galvanize - San Francisco  
Cohort 16 – 2016

**Email:** hainguyenau@gmail.com

**Github:** [https://github.com/hainguyenau/Word\\_Sense](https://github.com/hainguyenau/Word_Sense)

**LinkedIn:** <https://www.linkedin.com/in/hai-nguyen-au-672481a8> .

Question:

Why are movie stars so cool?



Answer.

Because they have lots of *fans*.



# Homographs



←Cool→



←Fan→



# The Data

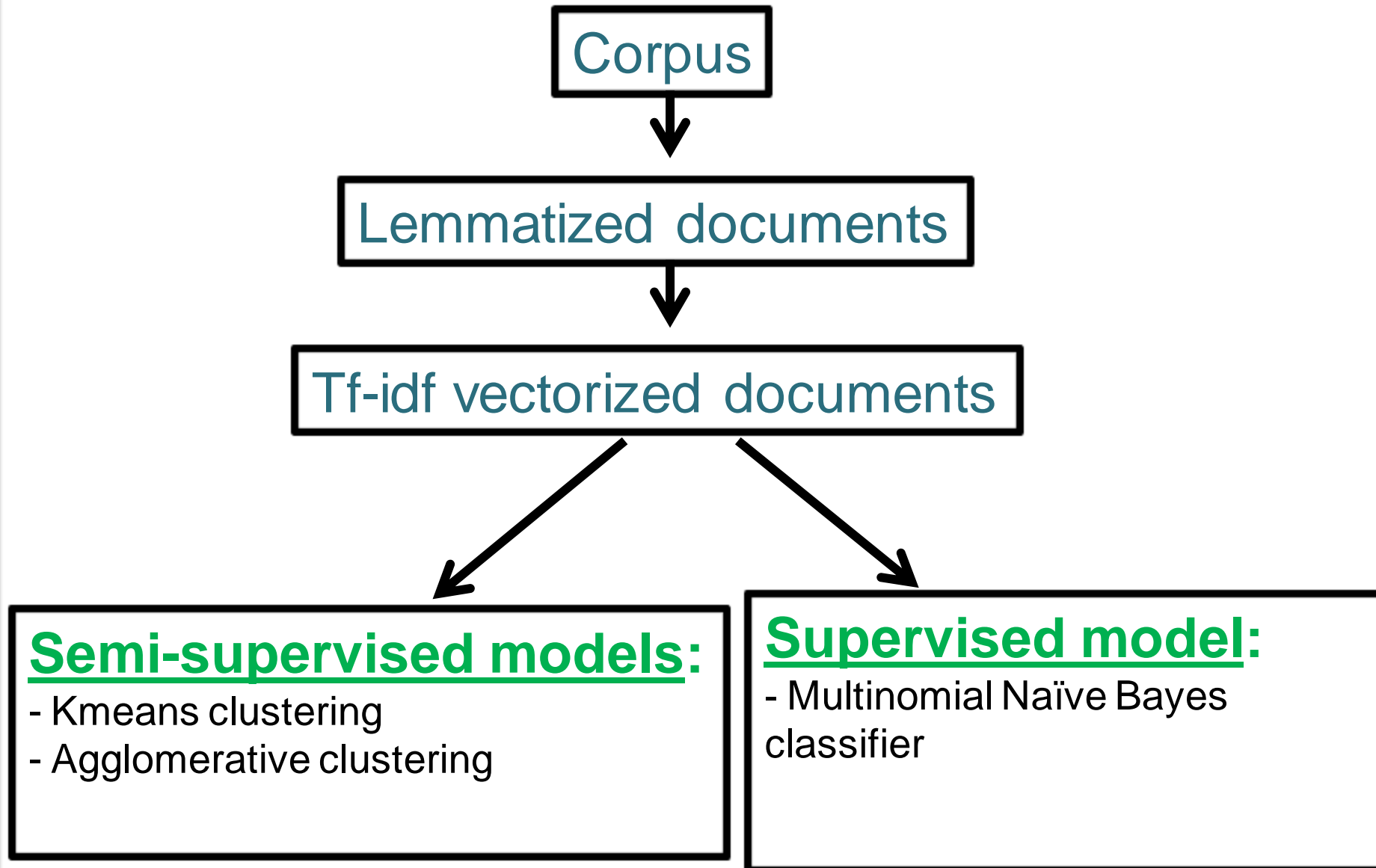
## ❖ Datasets:

- ❑ Senseval corpus.
- ❑ Each file contains sentences from different meanings of one word.
- ❑ The meaning of the word is labeled.

## ❖ “Hard”:

- ❑ HARD<sub>1</sub>: Difficult
- ❑ HARD<sub>2</sub>: Strong and vigorous
- ❑ HARD<sub>3</sub>: Resisting weight or pressure

# My Approach





# The “Best” Model

## Multinomial Naïve Bayes Classifier

- Simple and effective.
- Cross-Validation Results :

Word	Accuracy	Precision	Recall
“HARD”	72%	74%	80%
“INTEREST”	75%	73%	75%
“LINE”	72%	74%	72%

**Naïve, yet effective →**



# Applications

- Extend to other languages, work with non-latin languages.
- Build on top of known algorithms (such as nltk's Lesk)
- Help computers have more critical thinking and even discussion.





# THANK YOU !!!

## Questions and Answers

Hải Nguyễn Âu

Galvanize - San Francisco  
Cohort 16 – 2016

**Email:** hainguyenau@gmail.com

**Github:** [https://github.com/hainguyenau/Word\\_Sense](https://github.com/hainguyenau/Word_Sense)

**LinkedIn:** <https://www.linkedin.com/in/hainguyenau>