Exploring Text Data: An Intro to NLP

Lucas Kelly

www.lucaskellydataportfolio.weebly.com

www.github.com/lucaskelly49

https://www.linkedin.com/in/lucaskelly49/

OBJECTIVES:

Be the end of this lesson you should be able to...

tokenize text data

remove stop words

create a frequency distribution and showcase it using a visual

be aware of the next steps of an NLP project

What is NLP?

Definition:

Analytics tasks that deal with natural human language
in the form of text or speech.

Use Cases:

Can tweets the relate to world-wide disasters be
detected in order to improve relief efforts?

STUDY
HARD!

√123

NLP in the Real World

	id	keyword	location	text	target
74	107	accident	North Carolina	I-77 Mile Marker 31 South Mooresville Iredell...	1
78	113	accident	North Carolina	I-77 Mile Marker 31 to 40 South Mooresville I...	1
172	247	ambulance	Jackson	http://t.co/AY6zzcUpnz Twelve feared killed in...	1
174	249	ambulance	NaN	Twelve feared killed in Pakistani air ambulanc...	1
177	253	ambulance	NaN	Twelve feared killed in Pakistani air ambulanc...	1

Definition
Analytics tasks that deal with natural human language in the form of text or speech.

THE GENERAL FLOW OF AN NLP PROJECT

1

DATA CAPTURE

Twitter, e-mails, customer reviews

2

FEATURE EXTRACTION

Tokenized, remove stop words, frequency distributions, bigrams, mutual information score

3

MODEL

Traditional modelling, Machine Learning, Deep-Learning

Thank You!

www.lucaskellydataportfolio.weebly.com

www.github.com/lucaskelly49

https://www.linkedin.com/in/lucaskelly49/

Links and References:

Link to video:

https://www.youtube.com/watch?v=u7cQt1AETSQ&feature=youtu.be

Link to repo:

https://github.com/lucaskelly49/NLP-RealorNot-DisasterTweets

Kaggle Competition:

https://www.kaggle.com/c/nlp-getting-started/overview

Data Crunch Podcast:

https://datacrunchcorp.com/podcast/social-media-and-machine-learning/