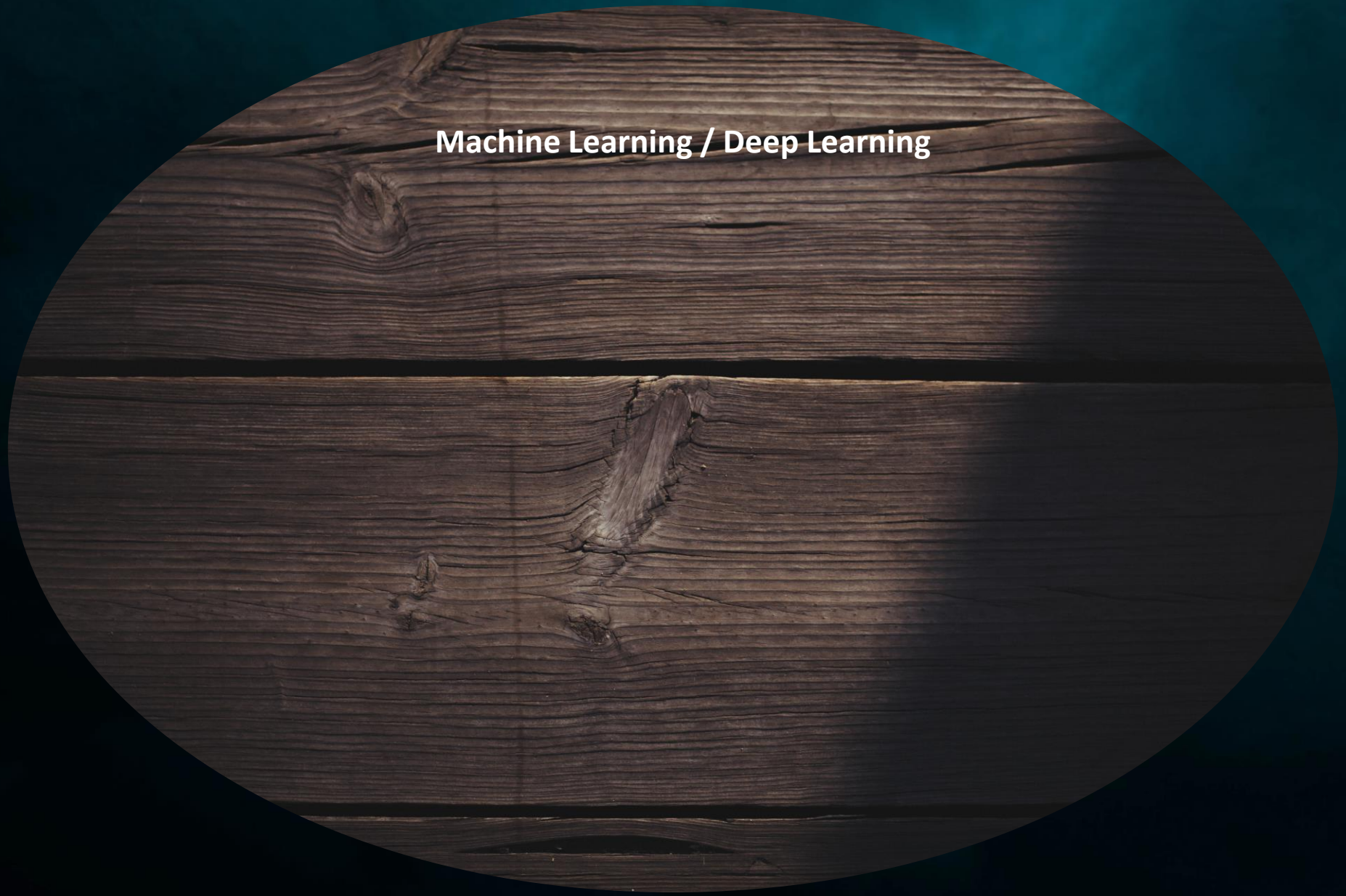
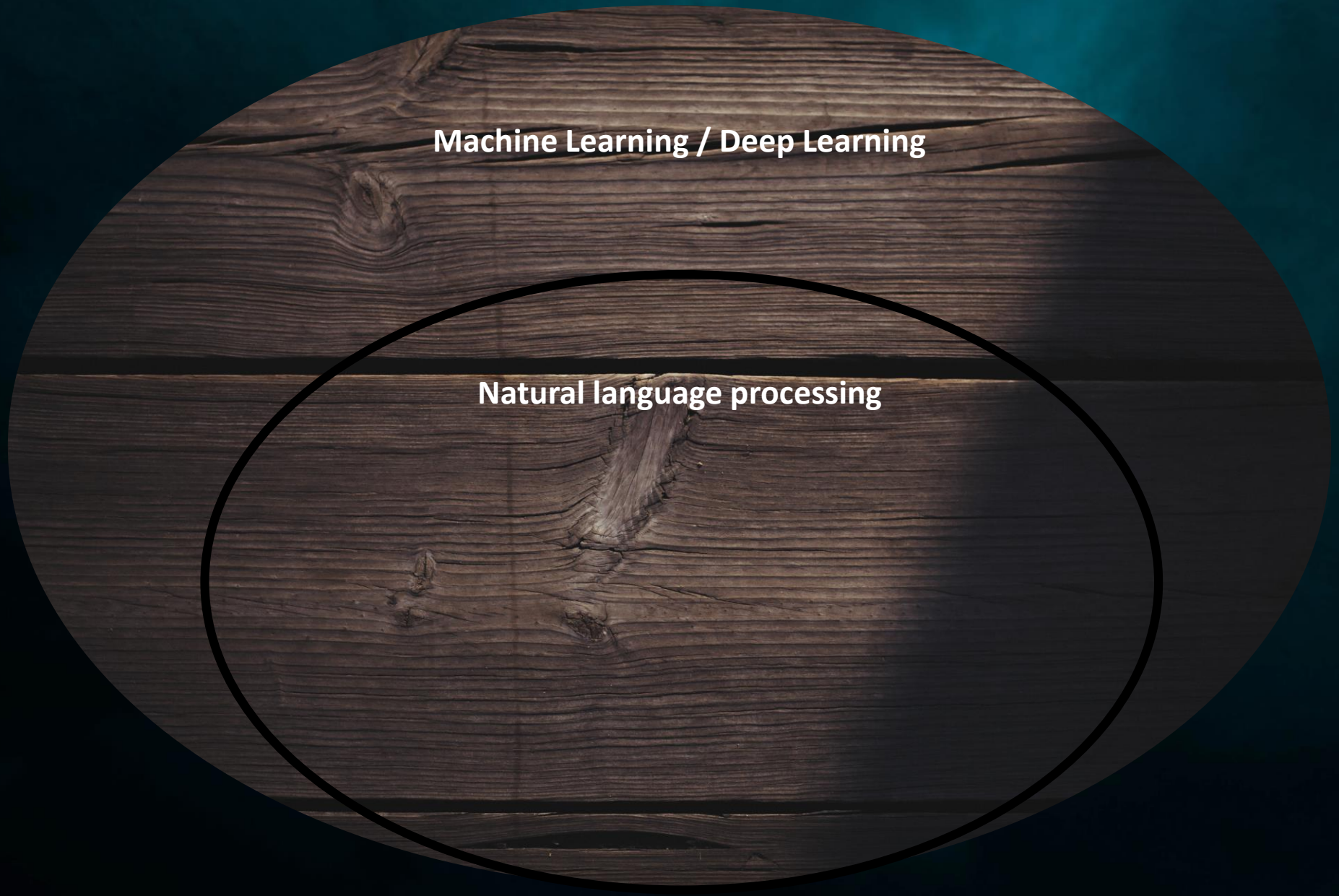
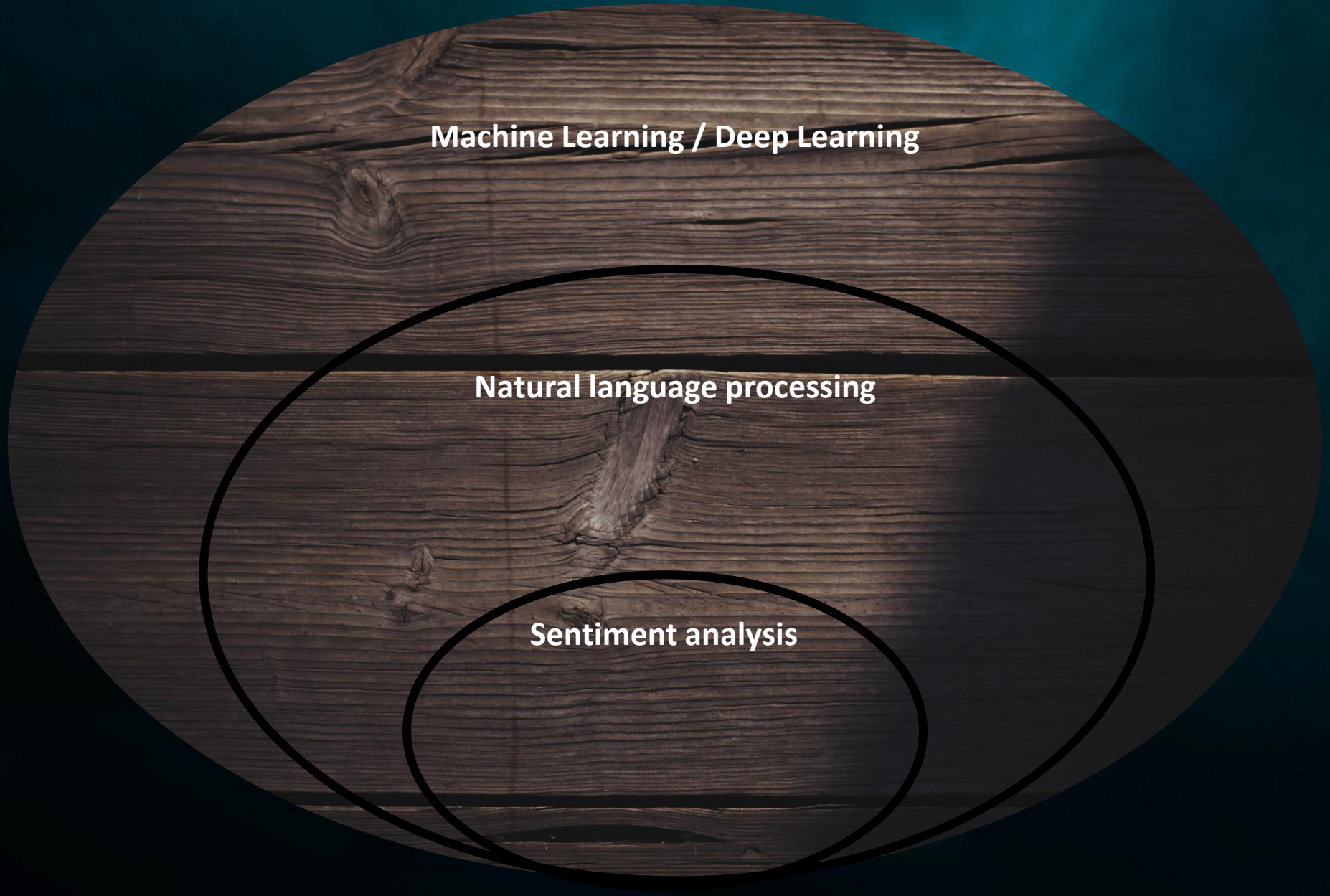
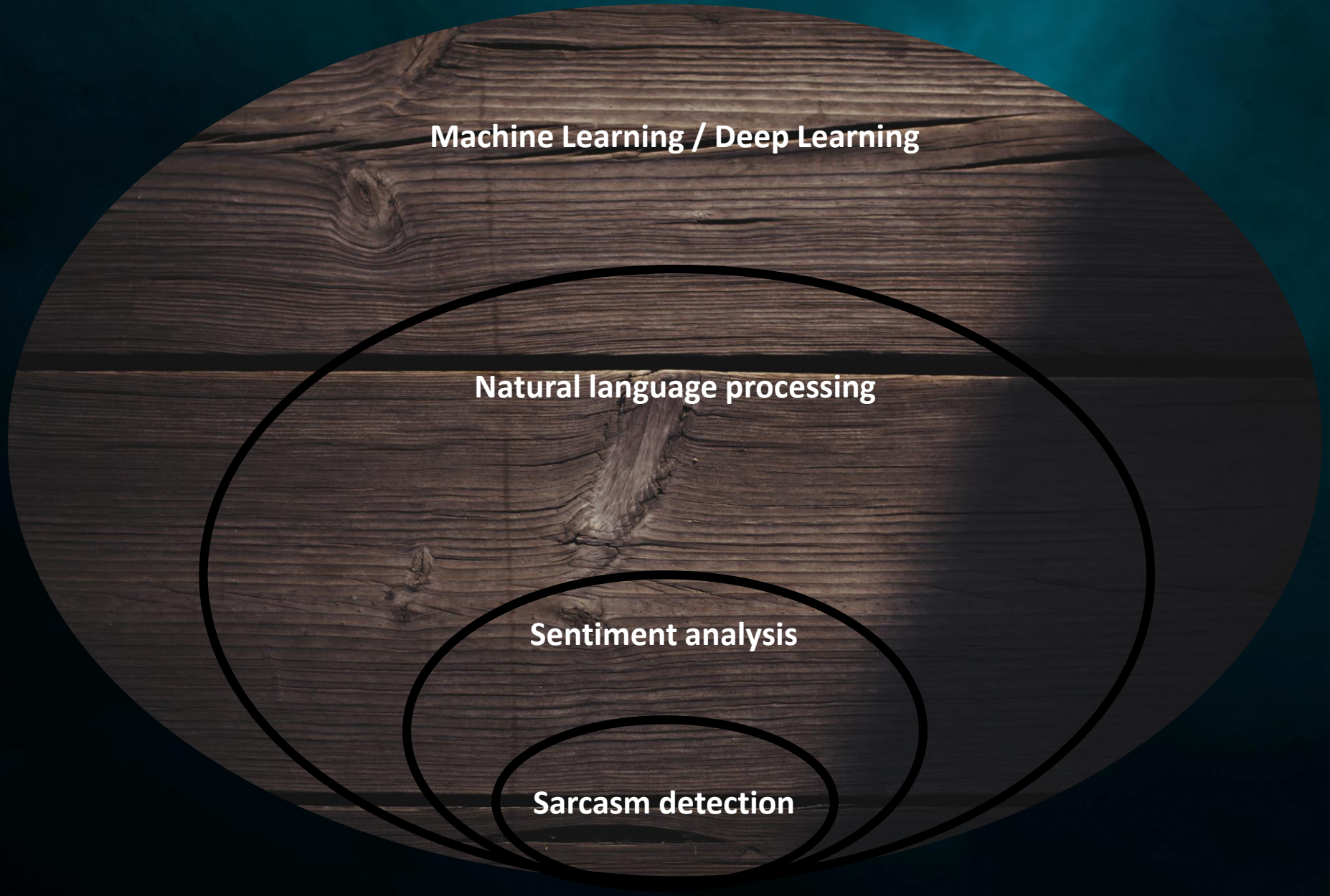
Sarcasm detection

Ane Berasategi


Machine Learning / Deep Learning


Machine Learning / Deep Learning
Natural language processing


Machine Learning / Deep Learning
Natural language processing
Sentiment analysis


Machine Learning / Deep Learning
Natural language processing
Sentiment analysis
Sarcasm detection

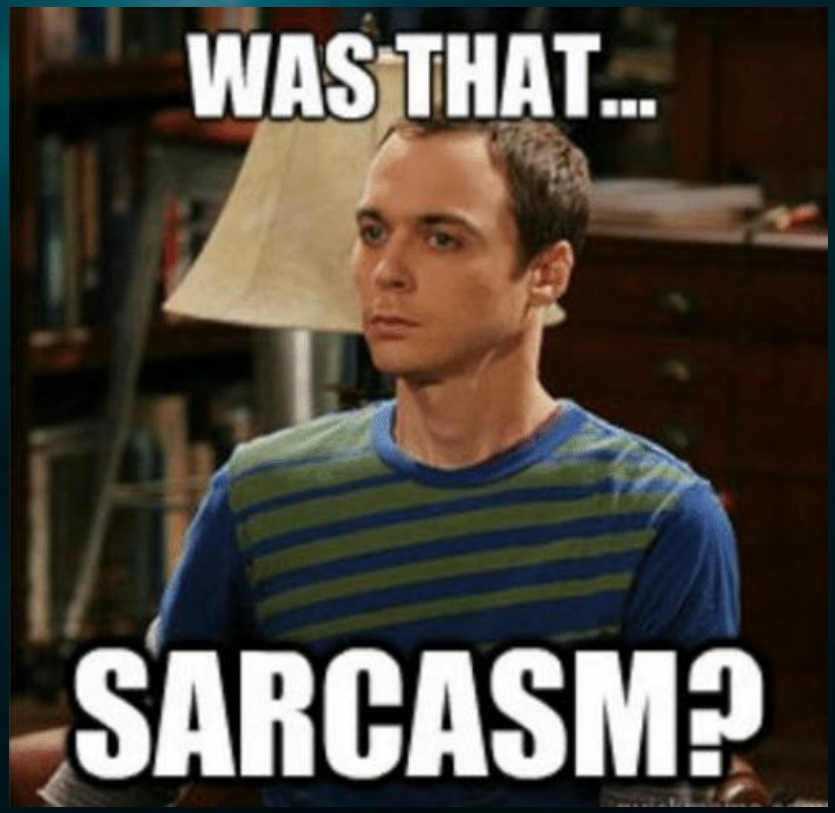
- Task: detect if a text is sarcastic or not
- But what is **sarcasm**?

Possible definitions:

- The **opposite of what you mean** in order to make fun of someone
- It has an **implied negative** sentiment, but a **positive surface** sentiment

Possible definitions:

- The **opposite of what you mean** in order to make fun of someone
- It has an **implied negative** sentiment, but a **positive surface** sentiment
- Intended sarcasm != detected sarcasm



Plan

Part 1: Overview

- Datasets overview
- Feature overview
- Model overview

Part 2: Analysis

- Analysis
- Current stand
- Future research

2.1. Datasets overview

Key aspects to consider in datasets:

2.1. Datasets overview

Key aspects to consider in datasets:

- Size
  - DL models need big datasets but they are difficult to obtain
  - Manually annotated: ~10k examples, automatically extracted: ~50k

2.1. Datasets overview

Key aspects to consider in datasets:

- Size
  - DL models need big datasets but they are difficult to obtain
  - Manually annotated: ~10k examples, automatically extracted: ~50k
- Balance: 50-50%? 20-80%?

2.1. Datasets overview

Key aspects to consider in datasets:

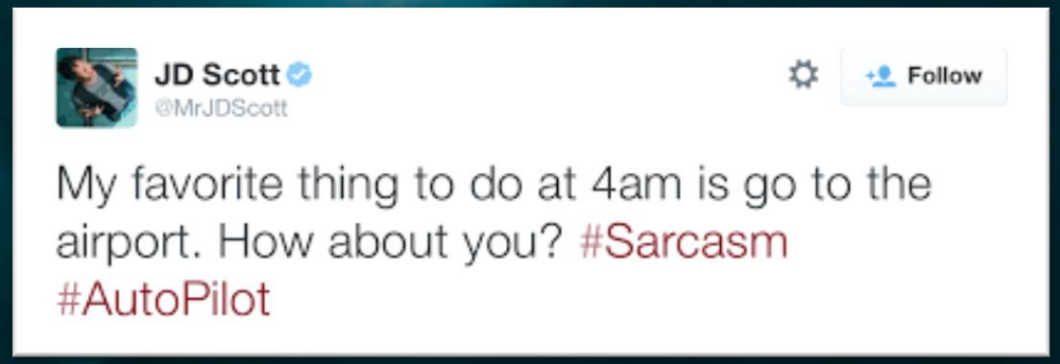
- Size
  - DL models need big datasets but they are difficult to obtain
  - Manually annotated: ~10k examples, automatically extracted: ~50k
- Balance: 50-50%? 20-80%?
  - Sarcasm is rare, not 50% of our interactions
  - Is it better to represent reality?
  - or add more sarcastic examples to help the model generalize?

2.1. Datasets overview

Key aspects to consider in datasets:

- Size
  - DL models need big datasets but they are difficult to obtain
  - Manually annotated: ~10k examples, automatically extracted: ~50k
- Balance: 50-50%? 20-80%?
  - Sarcasm is rare, not 50% of our interactions
  - Is it better to represent reality?
  - or add more sarcastic examples to help the model generalize?
- Nature of the examples
  - Intended sarcasm vs. perceived sarcasm

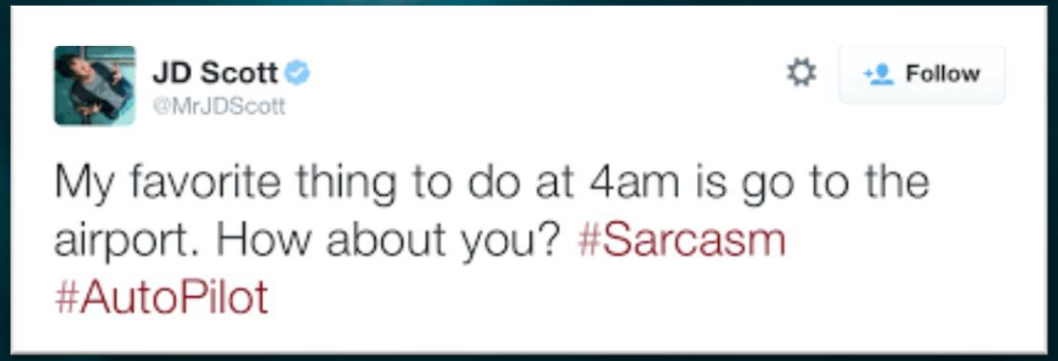
2.1. Two Twitter datasets



2.1. Two Twitter datasets

Riloff et al.[1]:

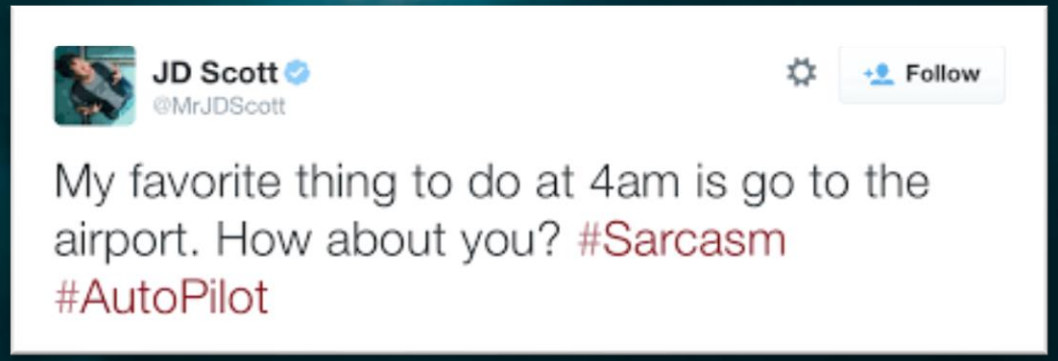
- 35k tweets with #sarcasm
- 140k random tweets



2.1. Two Twitter datasets

Riloff et al.[1]:

- 35k tweets with #sarcasm
- 140k random tweets



Ptáček et al.[2]:

- 7k tweets manually annotated
- 7k random tweets

2.2. Feature overview

We have the tweets, what do we do with them?

Can we provide more information/features to help the models?

2.2. Feature overview

We have the tweets, what do we do with them?
Can we provide more information/features to help the models?

- Cleaning: remove URLs, emojis, images

2.2. Feature overview

We have the tweets, what do we do with them?
Can we provide more information/features to help the models?

- Cleaning: remove URLs, emojis, images
- Lexical features: remove stopwords, word frequency, POS tagging

2.2. Feature overview

We have the tweets, what do we do with them?
Can we provide more information/features to help the models?

- Cleaning: remove URLs, emojis, images
- Lexical features: remove stopwords, word frequency, POS tagging
- User embeddings[6]: stylometric and personality features from the user

2.3. Model overview

- Until ~2017 mostly ML models
  - They need manual engineering
- From 2017 LSTMs, Transformers
  - They capture long-range dependencies better

Plan

Part 1: Overview

- Datasets overview
- Feature overview
- Model overview

Part 2: Analysis

- Analysis
- Current stand
- Future research

3.2. Analysis[7]

Paper analysing previous work, 2 questions:

3.2. Analysis[7]

Paper analysing previous work, 2 questions:

- Is the user embedding predictive of the sarcastic nature of the tweet?

3.2. Analysis[7]

Riloff: large, imbalanced, hashtag-based
Ptacek: smaller, balanced, manual annotation

Paper analysing previous work, 2 questions:

- Is the user embedding predictive of the sarcastic nature of the tweet?
- What are the performance differences in Riloff vs. Ptacek datasets?

3.2. Analysis[7]

Riloff: large, imbalanced, hashtag-based
Ptacek: smaller, balanced, manual annotation

Paper analysing previous work, 2 questions:

- Is the user embedding predictive of the sarcastic nature of the tweet?
- What are the performance differences in Riloff vs. Ptacek datasets?

Their findings:

- Users have a disposition to (non) sarcasm

3.2. Analysis[7]

Riloff: large, imbalanced, hashtag-based
Ptacek: smaller, balanced, manual annotation

Paper analysing previous work, 2 questions:

- Is the user embedding predictive of the sarcastic nature of the tweet?
- What are the performance differences in Riloff vs. Ptacek datasets?

Their findings:

- Users have a disposition to (non) sarcasm
- Models perform better on Riloff, not so well on Ptacek

3.2. Analysis[7]

Riloff: large, imbalanced, hashtag-based
Ptacek: smaller, balanced, manual annotation

Paper analysing previous work, 2 questions:

- Is the user embedding predictive of the sarcastic nature of the tweet?
- What are the performance differences in Riloff vs. Ptacek datasets?

Their findings:

- Users have a disposition to (non) sarcasm
- Models perform better on Riloff, not so well on Ptacek
- Riloff: intended, Ptacek: perceived sarcasm

3.4. Current stand

- Twitter, Reddit and forum post datasets
- User embeddings as features
- Performance way below human performance

3.4. Current stand

- Twitter, Reddit and forum post datasets
- User embeddings as features
- Performance way below human performance
- False negatives (sarcastic texts not detected)

3.4. Current stand

- Twitter, Reddit and forum post datasets
- User embeddings as features
- Performance way below human performance
- False negatives (sarcastic texts not detected)
  - world knowledge is lacking

3.5. Future research?

- Bigger datasets: but very time intensive
- Incorporate world knowledge: but how?
- Something else?

References

1. Sarcasm as Contrast between a Positive Sentiment and Negative Situation, Riloff et al., 2013
2. Sarcasm Detection on Czech and English Twitter, Ptáček et al., 2014
3. Identifying sarcasm in twitter: A closer look, González-Ibáñez et al., 2015
4. Harnessing Context Incongruity for Sarcasm Detection, Joshi et al., 2015
5. Detecting Sarcasm is Extremely Easy ;-), Parde and Nielsen, 2018
6. CASCADE: Contextual Sarcasm Detection in Online Discussion Forums, Hazarika et al., 2018
7. Exploring Author Context for Detecting Intended vs Perceived Sarcasm, Oprea and Magdy, 2019
8. Towards Multimodal Sarcasm Detection (An Obviously Perfect Paper), Castro et al., 2019

Sarcasm detection

Ane Berasategi