# Tutorial 11 - NLP

Victoria Ajila, MASc Computer Engineering

Carleton University

Monday 6th December, 2021

# Disclaimer: Tutorials are Recorded and Posted

**Privacy Preservation:**

- Ask questions in the chat[1]
- Keep video off

**Note:** If the above *hinders your ability to learn* $\wedge$ *violates your privacy*, please let me/Dr. Green know ASAP and video will be post-processed accordingly.

[1] I encourage unmuted/voice-based questions at any time, but know that this isn't explicitly privacy-preserving

# leFinal Tutorial

## All Good Things (Come to an End) - Nelly

# ML Weekly

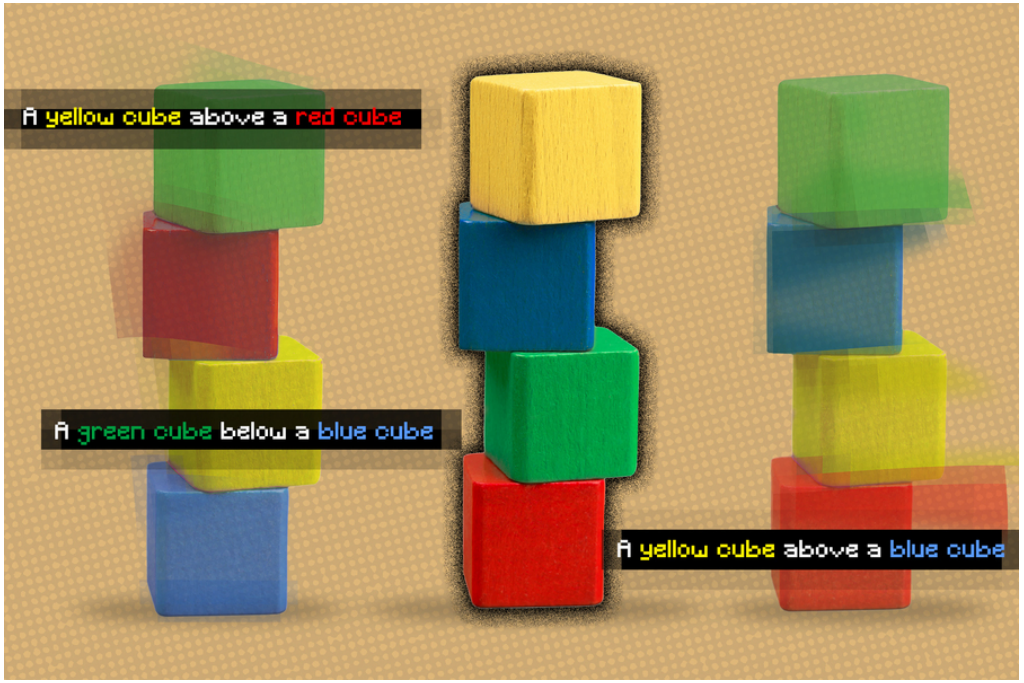
## Recent news events from the ML community

# ML Weekly

1. **(RL)** Everyday robots are (slowly) leaving the lab

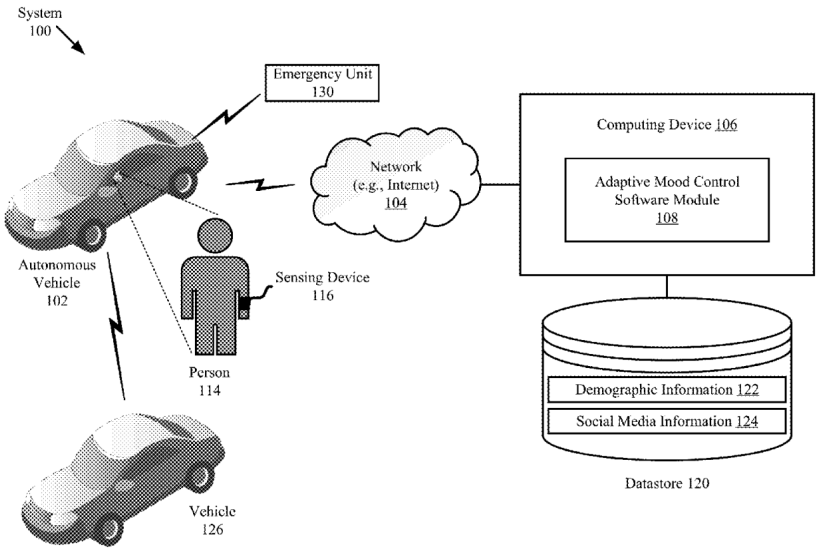
# ML Weekly

1. **(RL)** Everyday robots are (slowly) leaving the lab
2. **(CV)** Artificial intelligence that understands object relationships



# ML Weekly

1. **(RL)** Everyday robots are (slowly) leaving the lab
2. **(CV)** Artificial intelligence that understands object relationships
3. **(AI)** Invention uses machine-learned human emotions to 'drive' autonomous vehicles



# NLP Intuition

## Sentence Tokenization

### Original Text:

Introduction to supervised and unsupervised machine learning (ML), including deeper knowledge of several algorithms of each type. Evaluation and quantification of predictive performance of ML systems. Use of one or more ML development environments.

### Tokenized Sentences:

0: Introduction to supervised and unsupervised machine learning (ML), including deeper knowledge of several algorithms of each type'

1: Evaluation and quantification of predictive performance of ML systems.

3: Use of one or more ML development environments.

# NLP Intuition

## Word Tokenization

Original Sentence:

Evaluation and quantification of predictive performance of ML systems

Tokenized Words:

[“Evaluation”, “and”, “quantification ”, “of ”, “predictive”, “performance ”, “of ”, “ML”, “systems”]

# NLP Intuition

## Lemmatization

"The task of removing inflectional endings only and to return the base dictionary form of a word which is also known as a lemma. Lemmatization is another technique for reducing words to their normalized form. But in this case, the transformation actually uses a dictionary to map words to their actual form."

Input:

"seen"

Output:

"see"

# NLP Intuition

## Stemmer

"The process of reducing inflected (or sometimes derived) words to a base form (*e.g.*, "close" will be the root for "closed", "closing", "close", "closer" etc.). Stemming yields similar results as lemmatization, but does so on grounds of rules, not a dictionary."

Input:

"closed" or ""closing", "close", "closer"

Output:

"close"

# NLP Intuition

## Removal of Stop Words

Original Input:

["Evaluation", "and", "quantification ", "of ", "predictive", "performance ", "of ", "ML", "systems"]

After stop word removal:

["Evaluation", "quantification ", "predictive", "performance", "ML", "systems"]

# Into the Notebooks we Go...

We will cover one notebook today!

1. Natural Language Processing

# Tutorial 11 - NLP

Victoria Ajila, MASc Computer Engineering

Carleton University

Monday 6th December, 2021