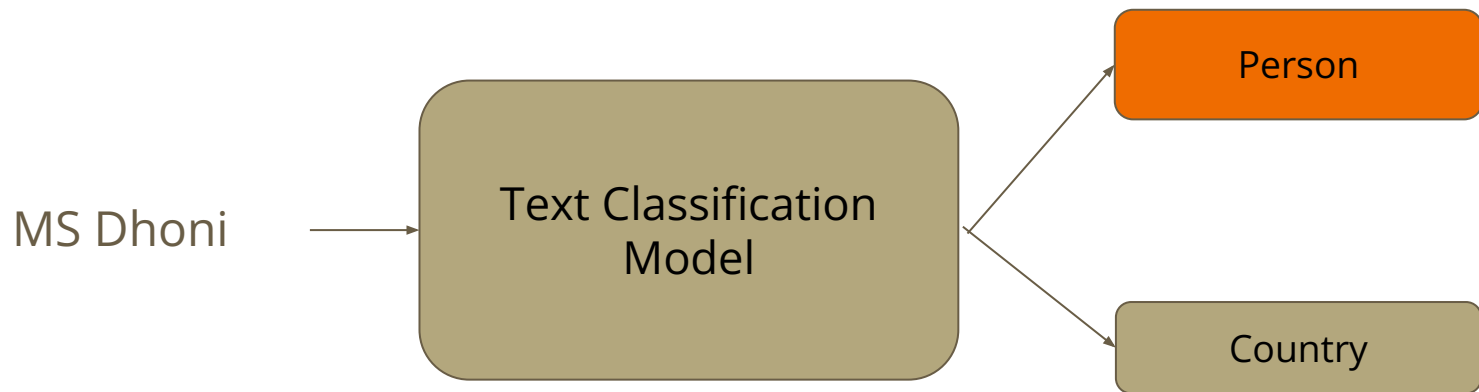
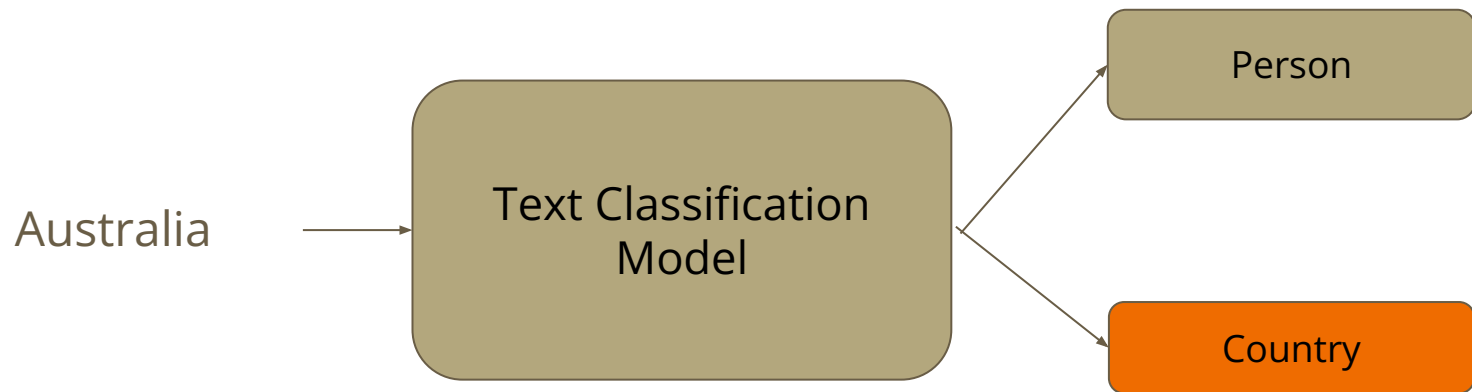

Bidirectional Encoder Representations from Transformers (BERT)

— Design Lab (CS69202) —

Text Classification Problem

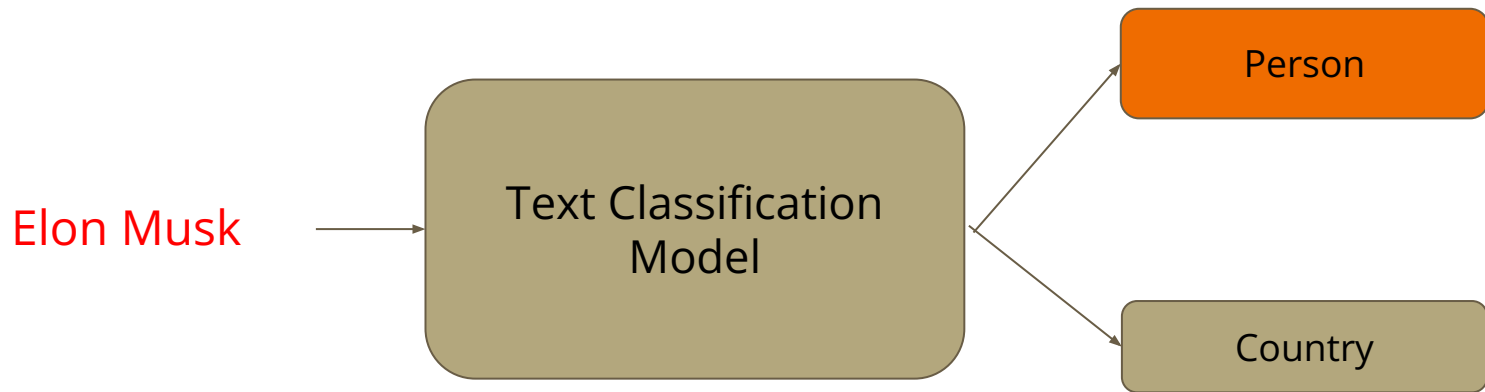


Text Classification Problem



Text Classification Problem

How are they similar??



How can we capture similarity between two words??

Let's look at a housing problem:-

H1

Bedrooms - 3
Area - 1850 sq. ft.
Bathrooms - 2

H2

Bedrooms - 3
Area - 1700 sq. ft.
Bathrooms - 2

H3

Bedrooms - 10
Area - 7500 sq. ft.
Bathrooms - 10

Looks similar??

How can we capture similarity between two words??

Let's look at the initial problem:-

MS Dhoni

Can move - 1
Healthy - 0.9
Location - 0
Has two eyes - 1
Has government - 0

Elon Musk

Can move - 1
Healthy - 0.6
Location - 0
Has two eyes - 1
Has government - 0

Australia

Can move - 0
Healthy - 0.7
Location - 1
Has two eyes - 0
Has government - 1

Looks similar??

How can we capture similarity between two words??

Let's look at their feature vector:-



Looks similar??

- These are word embeddings.
- In real life models can figure out the features on their own

What is BERT?

- BERT is an encoder type transformer that can generate **contextualized** embeddings for an entire sentence.
- Contextualized -
 - S1 - He did not get a **fair** treatment.
 - S2 - Fun **Fair** will be held in NY this summer.
 - S3 - He deserves **unbiased** treatment.

Coding

- You can download the model from hugging face.
(<https://huggingface.co/google-bert>)
- Only difference in dataset class, model definition and training loop.

Resources

1. BERT → <https://jalammar.github.io/illustrated-bert/>
2. Transformer → <https://jalammar.github.io/illustrated-transformer/>
3. BERT code → <https://www.sabrepc.com/blog/Deep-Learning-and-AI/text-classification-with-bert>
4. AutoModel for Sequence Classification → <https://huggingface.co/transformers/v3.3.1/training.html>

BERT (extra)

- BERT trained by Google on 2500 million words on wikipedia.
- Trained using two training process-
 - Masked Language Model
 - Next Sentence Prediction

Transformer (extra)

- Transformers utilize a self-attention mechanism that allows each element in an input sequence to focus on different parts of the sequence during processing. This enables the model to capture contextual relationships efficiently, making them highly effective for tasks involving sequential data like natural language processing.
- For more information refer - <https://arxiv.org/abs/1706.03762>