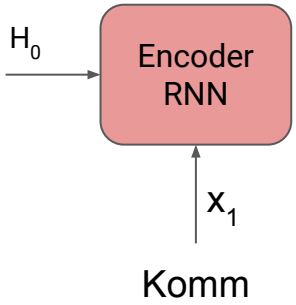
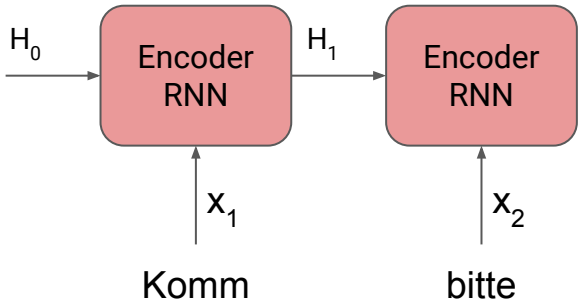
Understanding Transformer Architecture

Recap



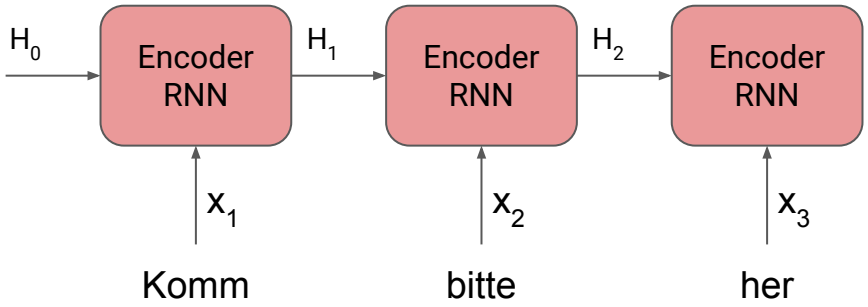
ENCODER

Recap



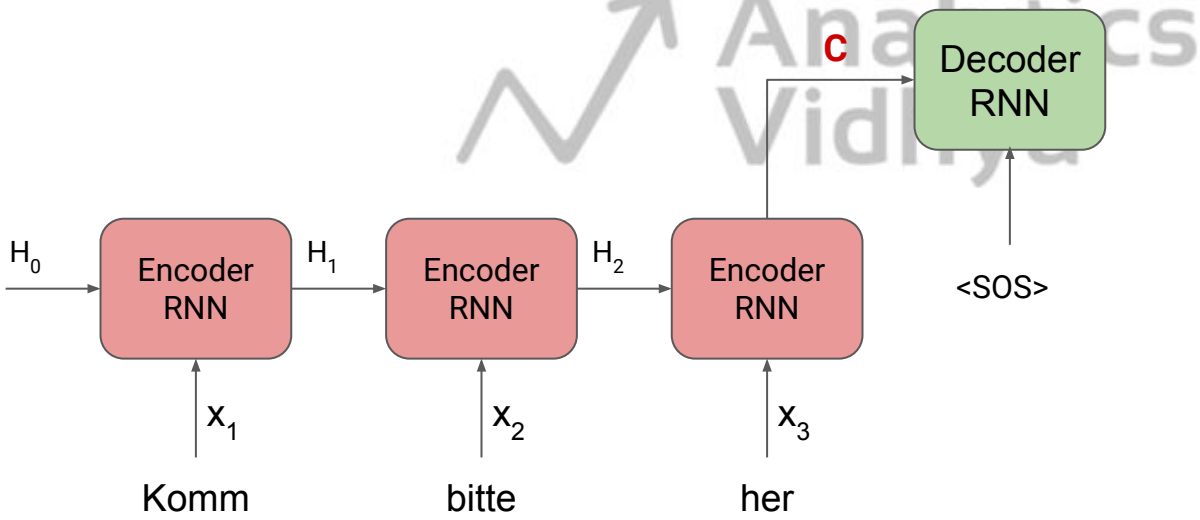
ENCODER

Recap



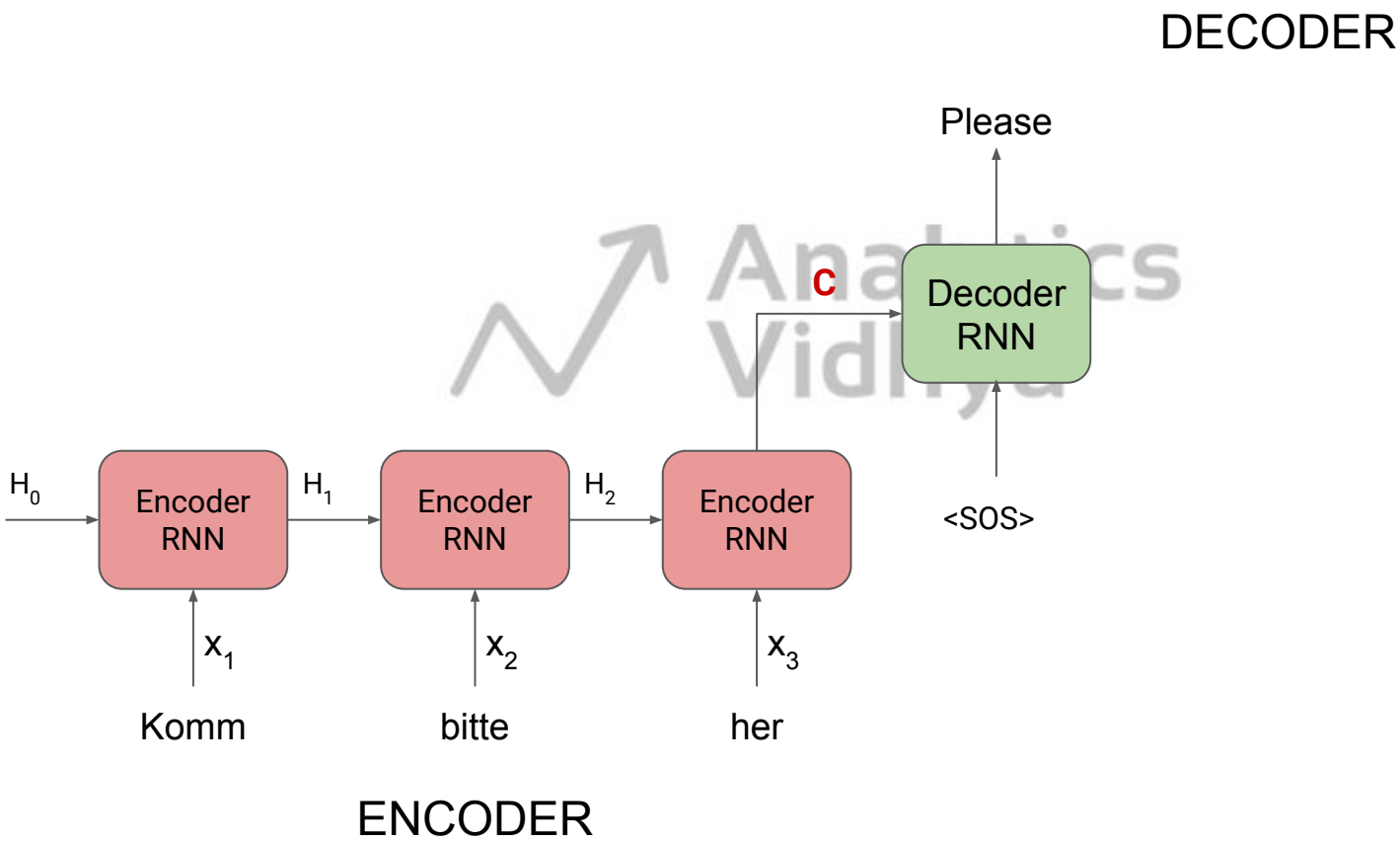
ENCODER

Recap

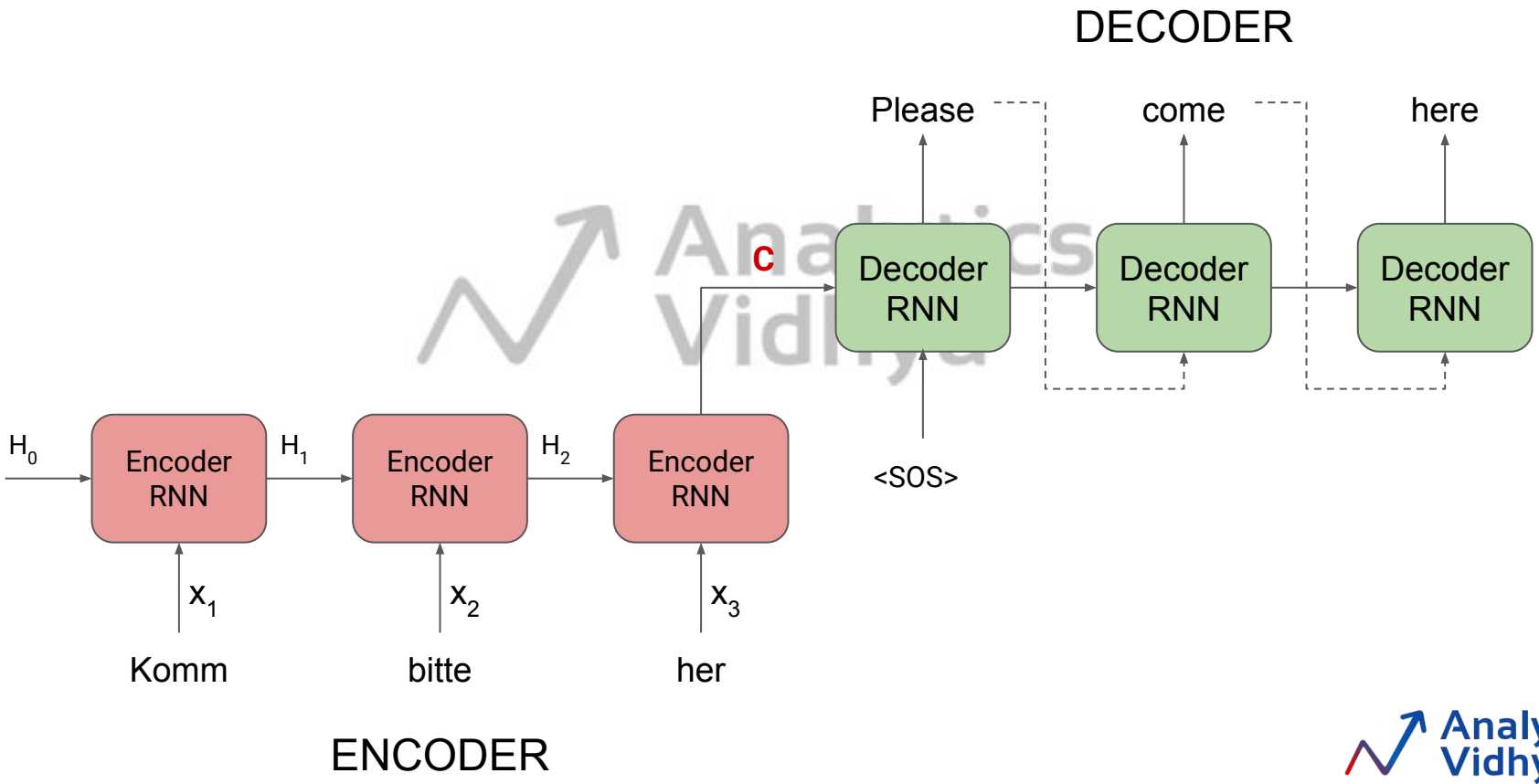


ENCODER

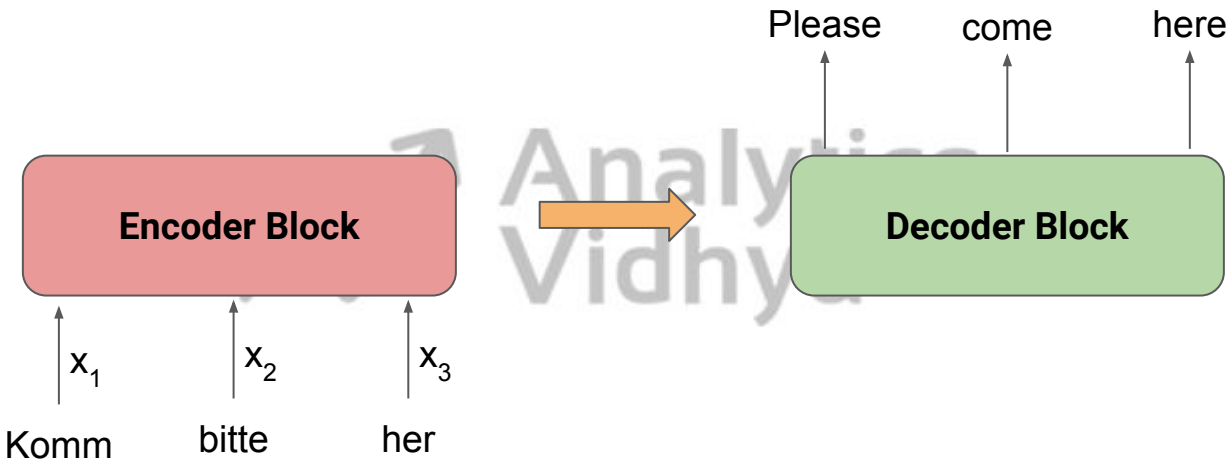
Recap



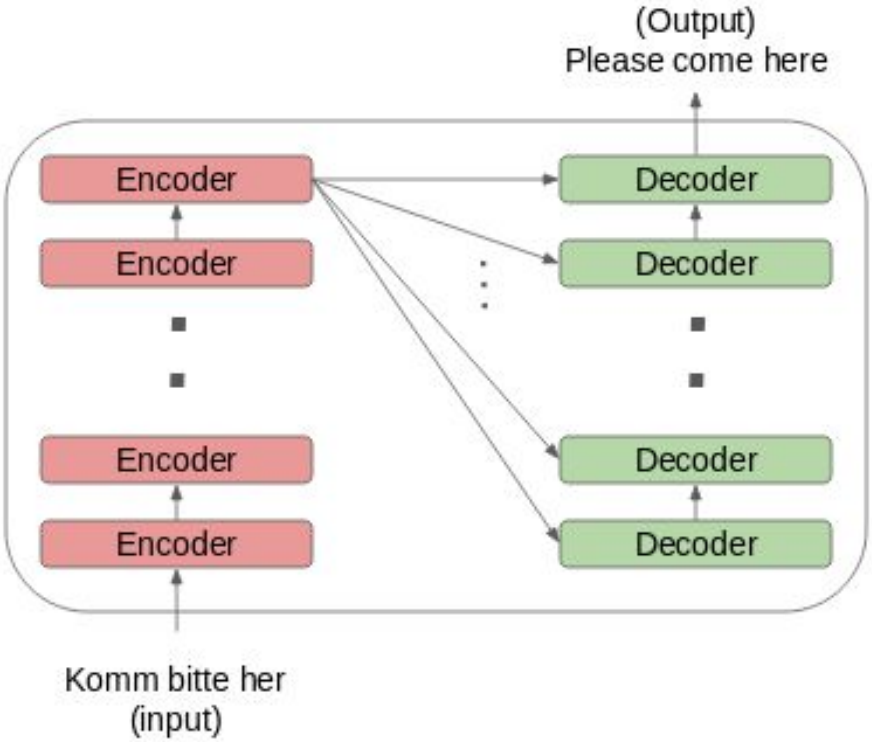
Recap



Transformer Architecture

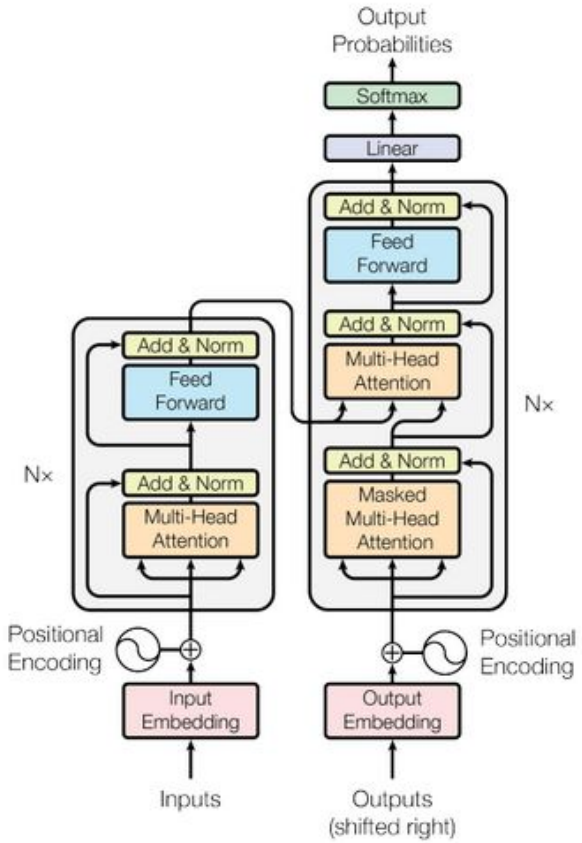


Transformer Architecture



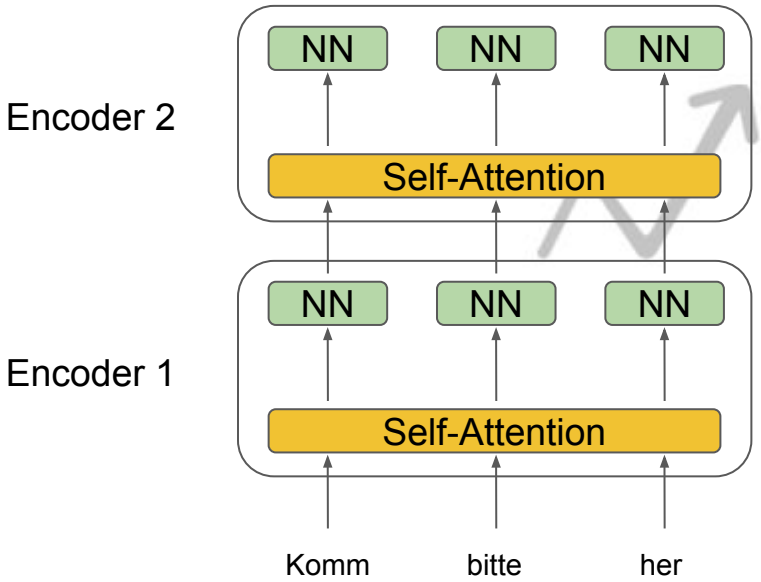
Analytics Vidhya

Transformer Architecture

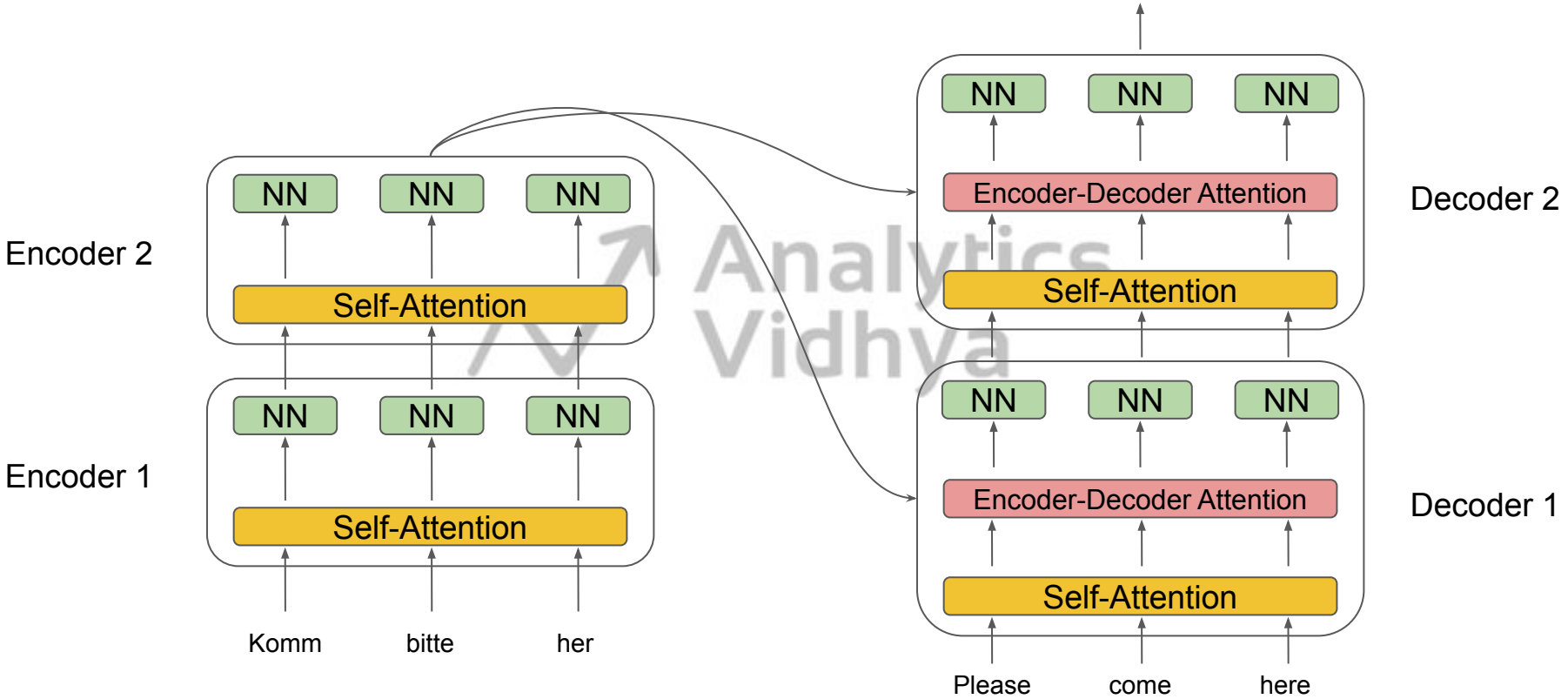


Analytics Vidhya

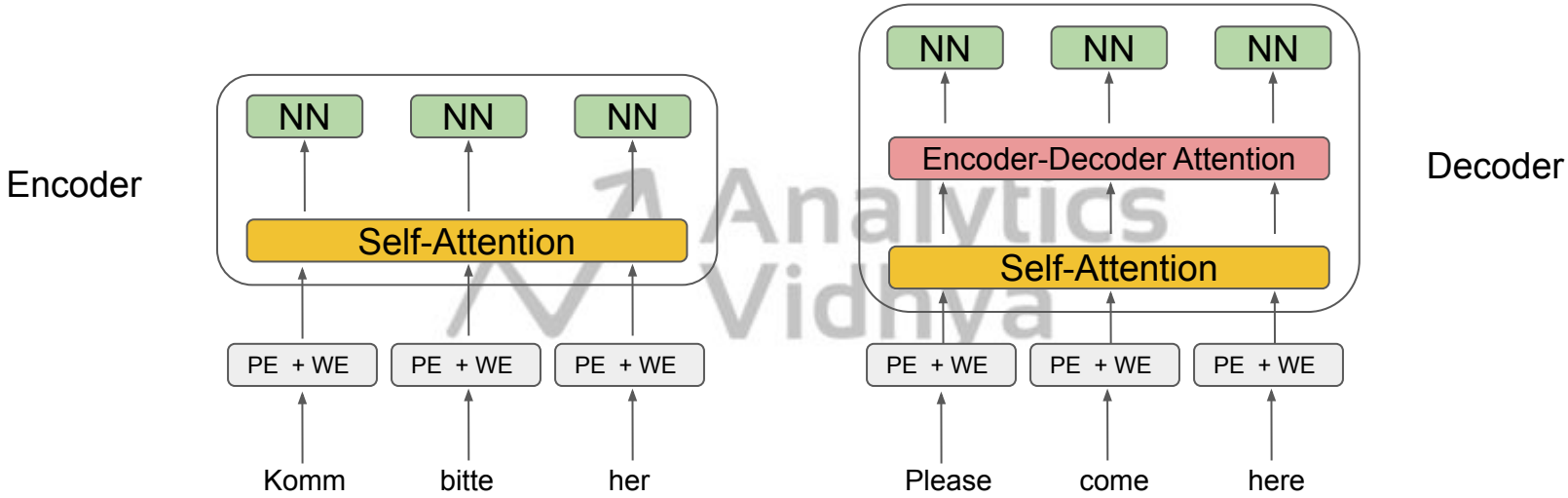
Encoder-Decoder in Transformer



Encoder-Decoder in Transformer



Positional Embeddings in Transformer

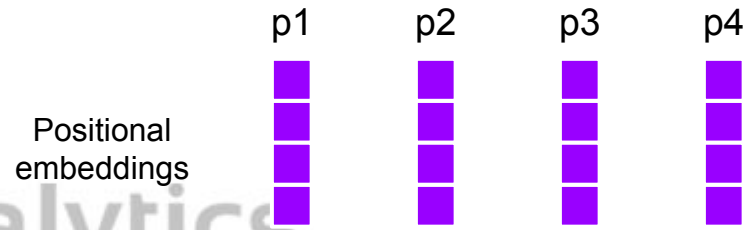


PE - Positional Embedding

WE - Word Embedding

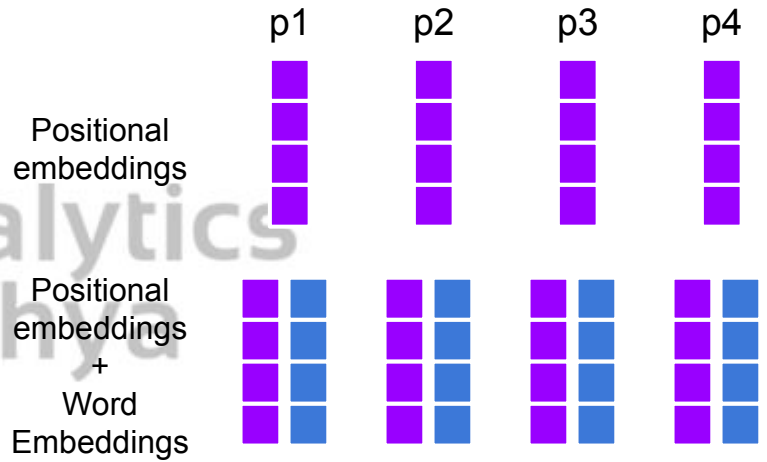
Positional Embeddings in Transformer

- Positional embeddings (PE) encode the order of tokens into fixed size vectors



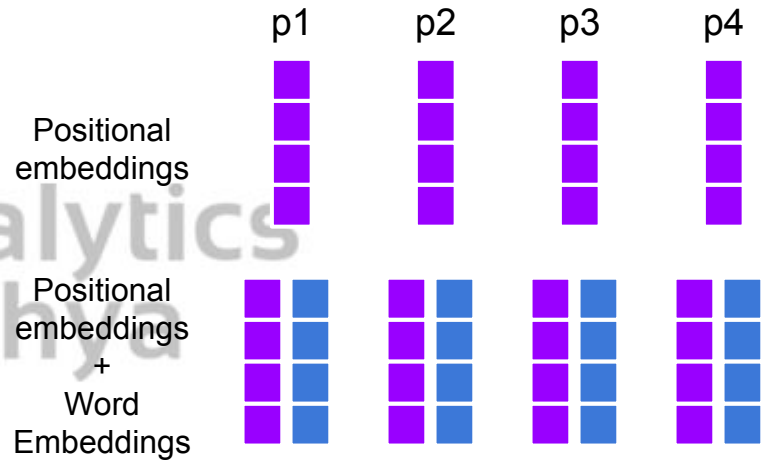
Positional Embeddings in Transformer

- Positional embeddings (PE) encode the order of tokens into fixed size vectors
- Combined with word embeddings and passed to encoder and decoder layers



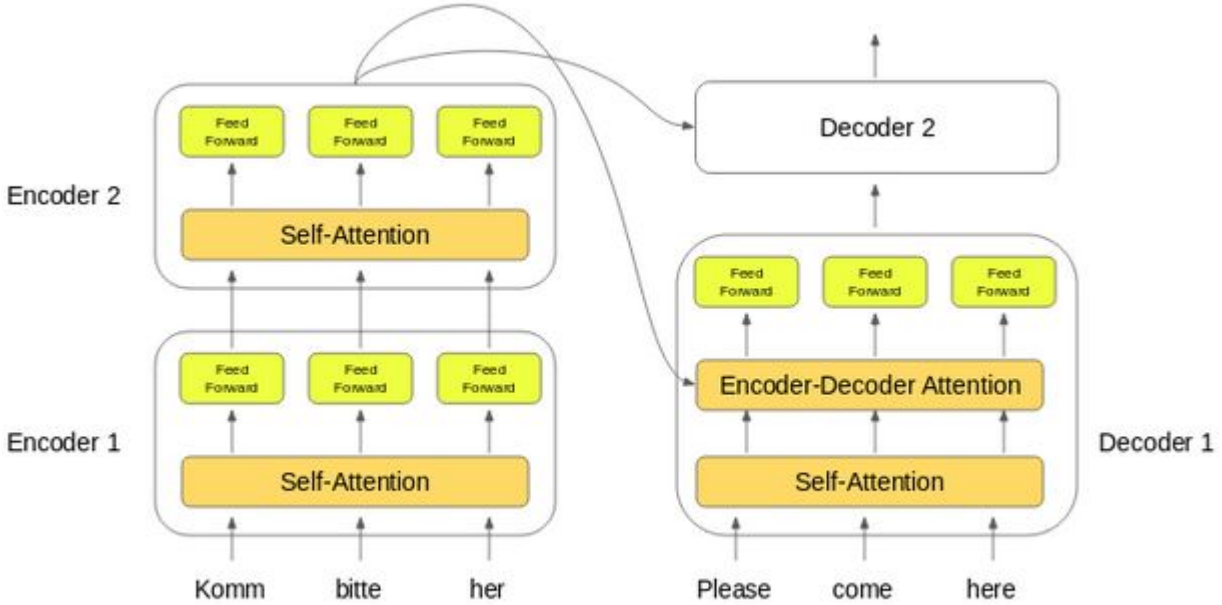
Positional Embeddings in Transformer

- Positional embeddings (PE) encode the order of tokens into fixed size vectors
- Combined with word embeddings and passed to encoder and decoder layers
- PE are generated by two functions



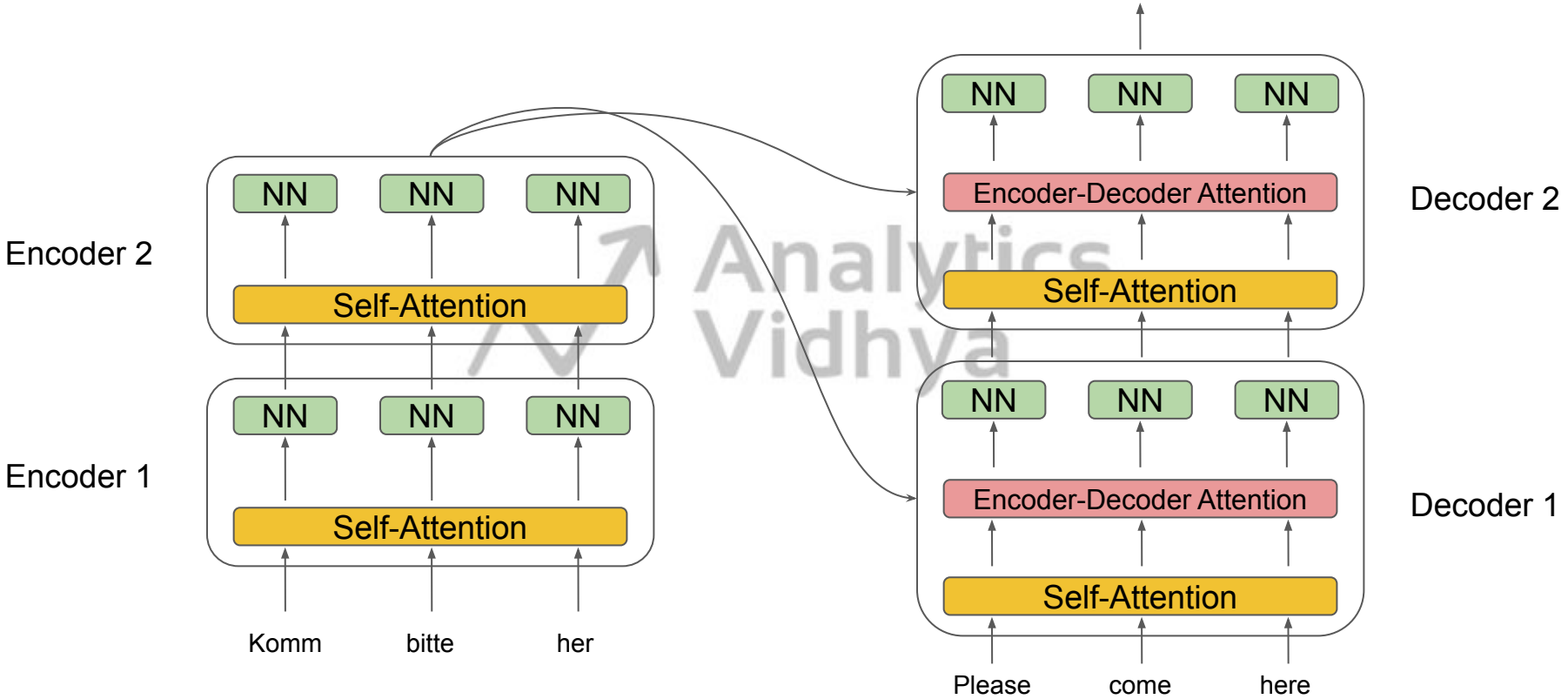
- $PE(p, 2i + 1) = \cos(p/10000^{2i / emb\_dim})$
- $PE(p, 2i) = \sin(p/10000^{2i / emb\_dim})$

Encoder-Decoder in Transformer



Analytics Vidhya

Encoder-Decoder in Transformer

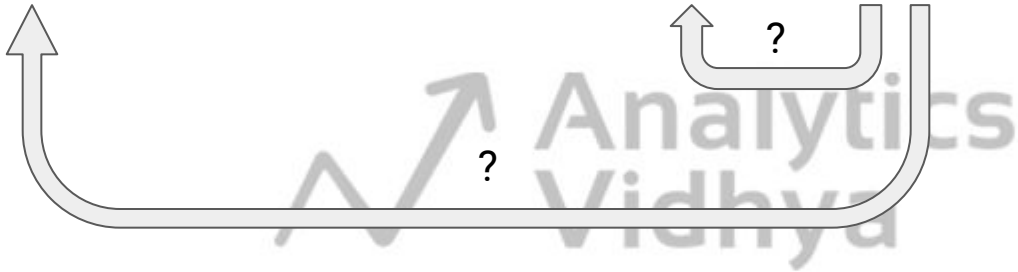


Intuition behind Self Attention

“The kids were scared of the lions, so they left right away.”

Intuition behind Self Attention

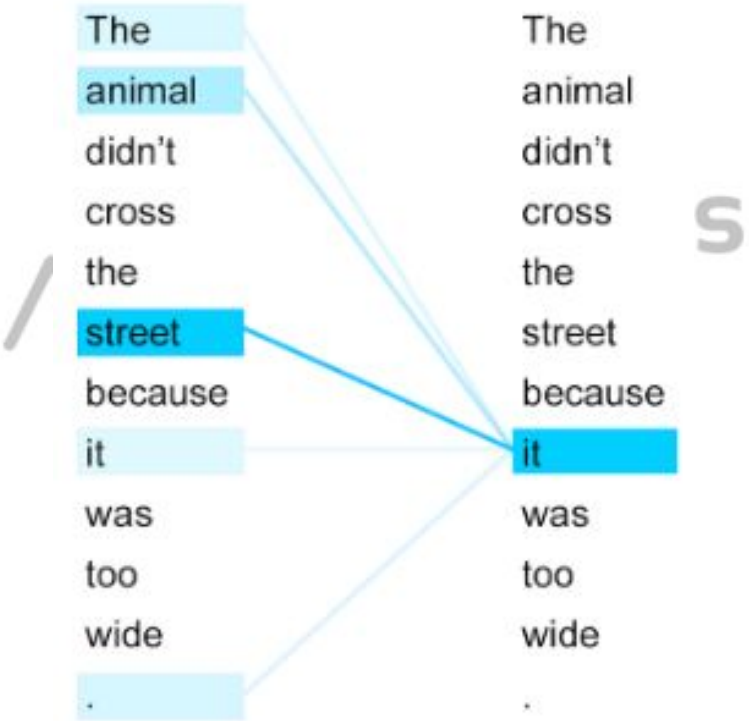
“The kids were scared of the lions, so they left right away.”



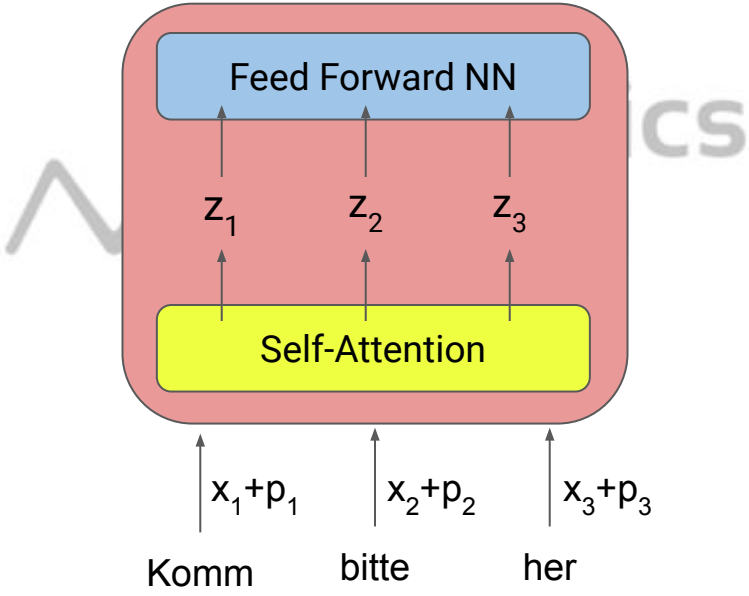
Intuition behind Self Attention

“The kids were scared of the lions, so they left right away.”

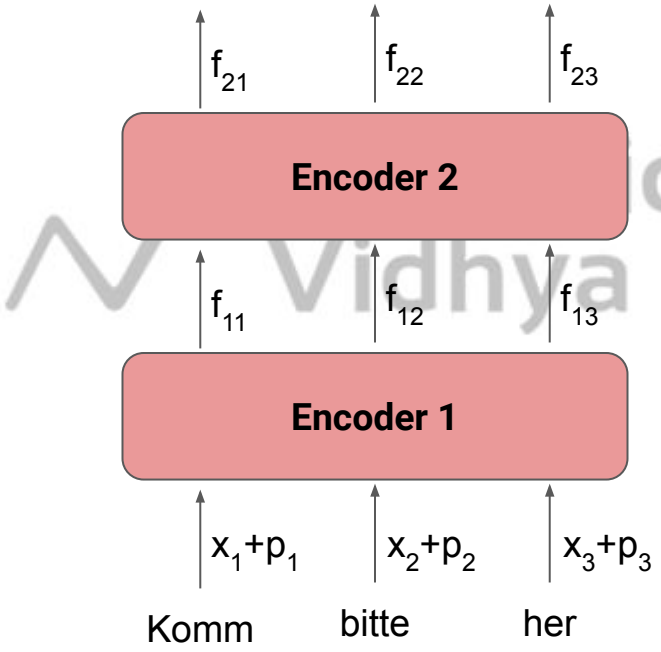
Intuition behind Self Attention



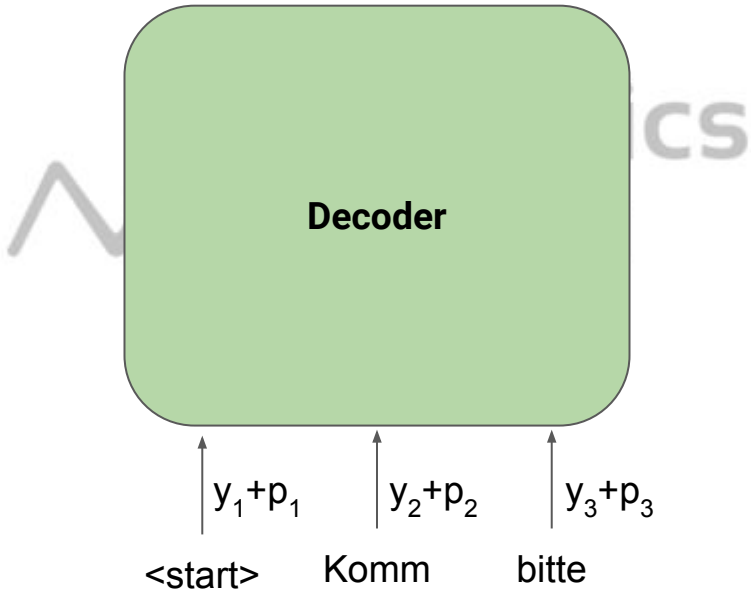
Encoder Layer in Transformer



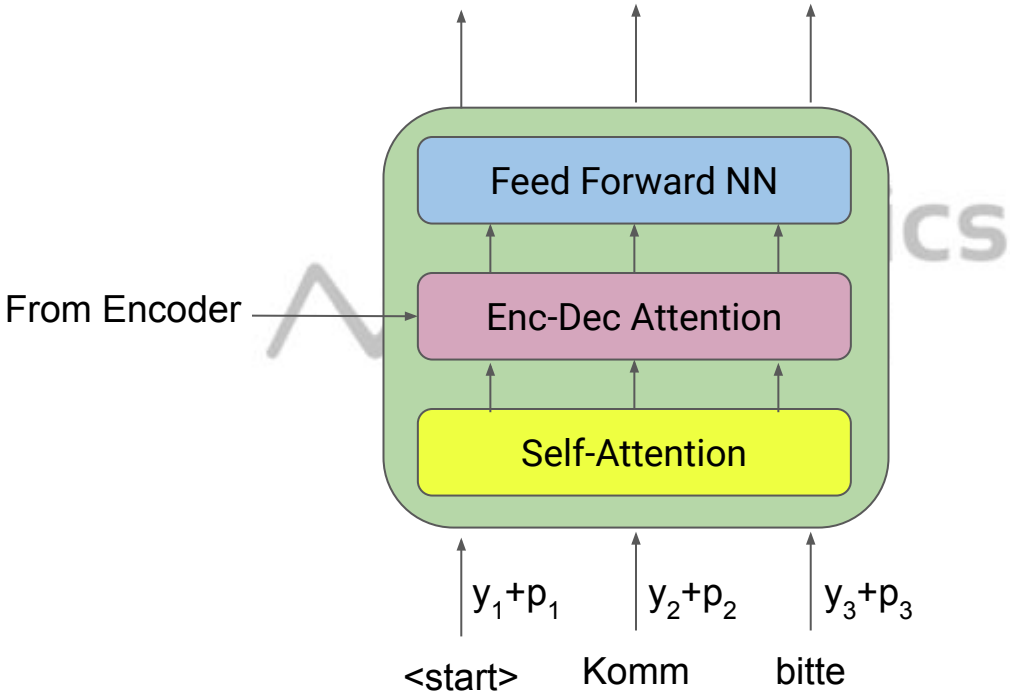
Encoder Layer in Transformer



Decoder Layer in Transformer



Decoder Layer in Transformer



Thank You

Analytics
Vidhya