

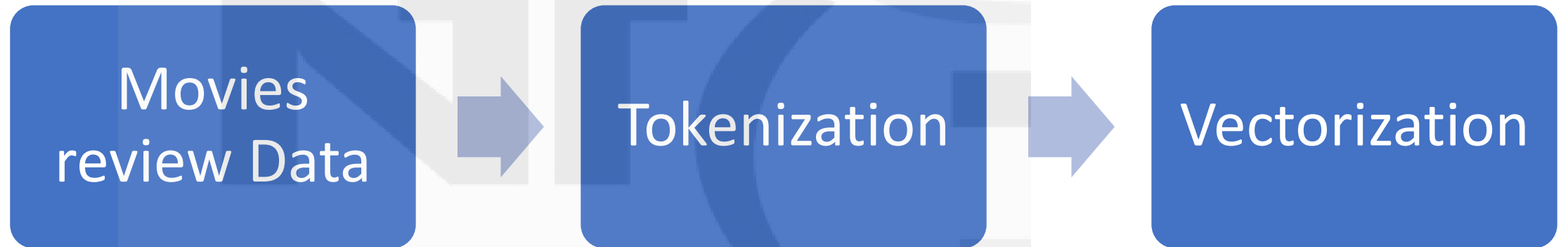


Sentiment Analysis

MUKESH KUMAR

- Clean the data but do not do lemmatization/Stemming because sequence of data is very important

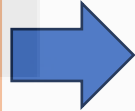
Data Preprocessing in Movies Review Data



Data Preprocessing

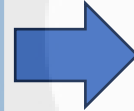
Movies Review Data

Awesome movie..
Great movie..
What a movie..



Tokenization

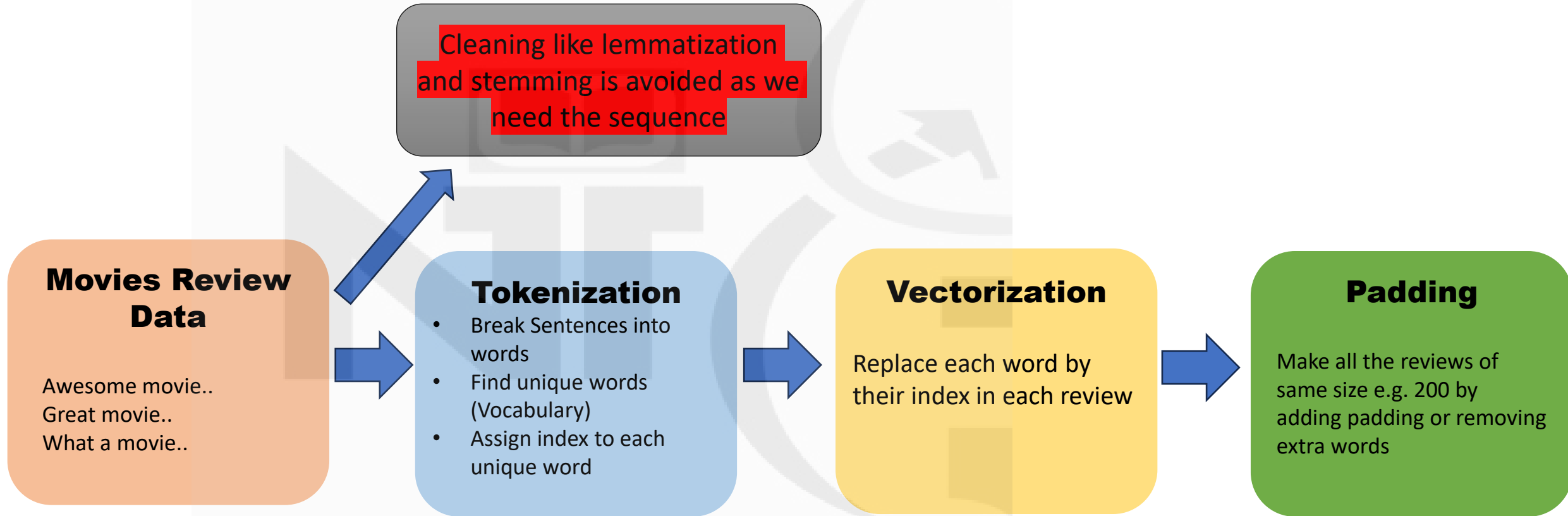
- Break Sentences into words
- Find unique words (Vocabulary)
- Assign index to each unique word



Vectorization

Replace each word by their index in each review

Data Preprocessing

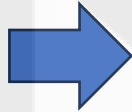


Mode Building

- First layer in the model will be embedding layer as we need to first transform the data to word embeddings/Vectors.

Padding

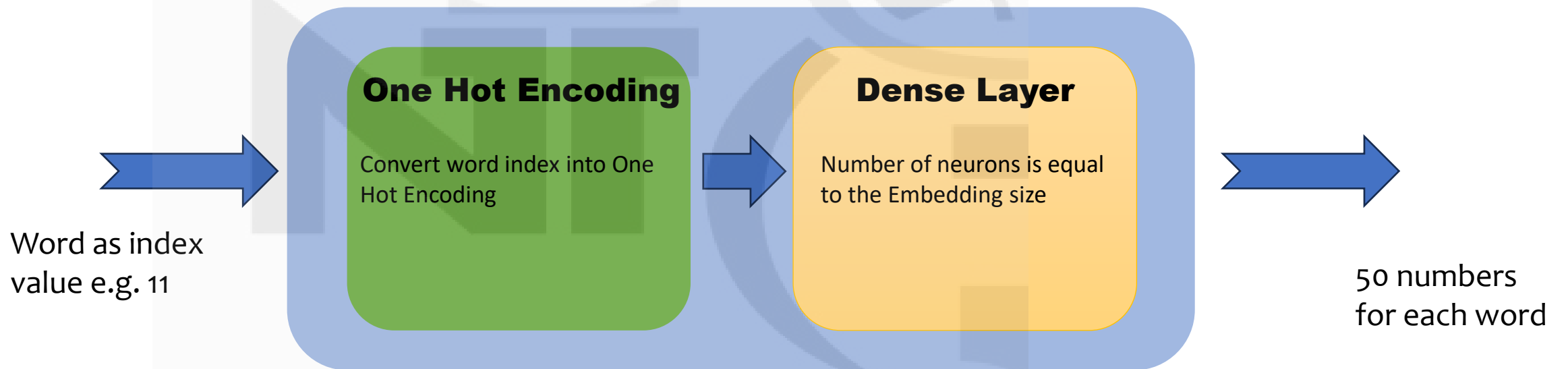
Make all the reviews of same size e.g. 200 by adding padding or removing extra words



Embedding Layer

Replace each word with multiple numbers(word embedding)

Embedding Layer



MODEL BUILDING

Embedding Layer

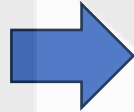
- Output of embedding layer is a 50 number vector
- Every word in the vocab gets converted into a 50 number vector
- Each of review contains 300 words ,when we feed a review to this embedding layer we will get [300,50], each of the 300 words represented by a 50 number vector

Mode Building

- Next layer will be LSTM

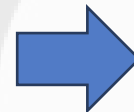
Padding

Make all the reviews of same size e.g. 200 by adding padding or removing extra words



Embedding Layer

Replace each word with multiple numbers(word embedding)



LSTM

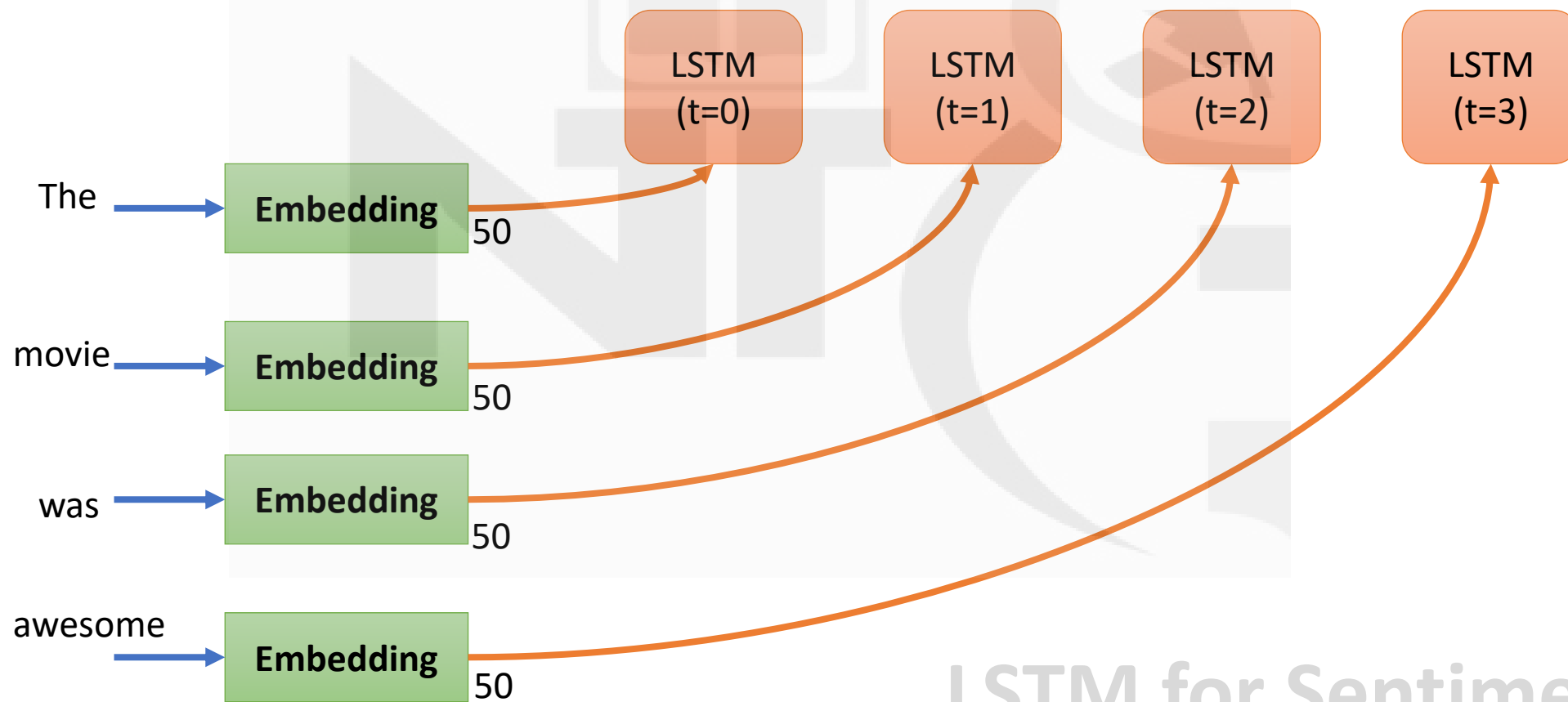
Build understanding of review / sequence

How many time LSTM will run???

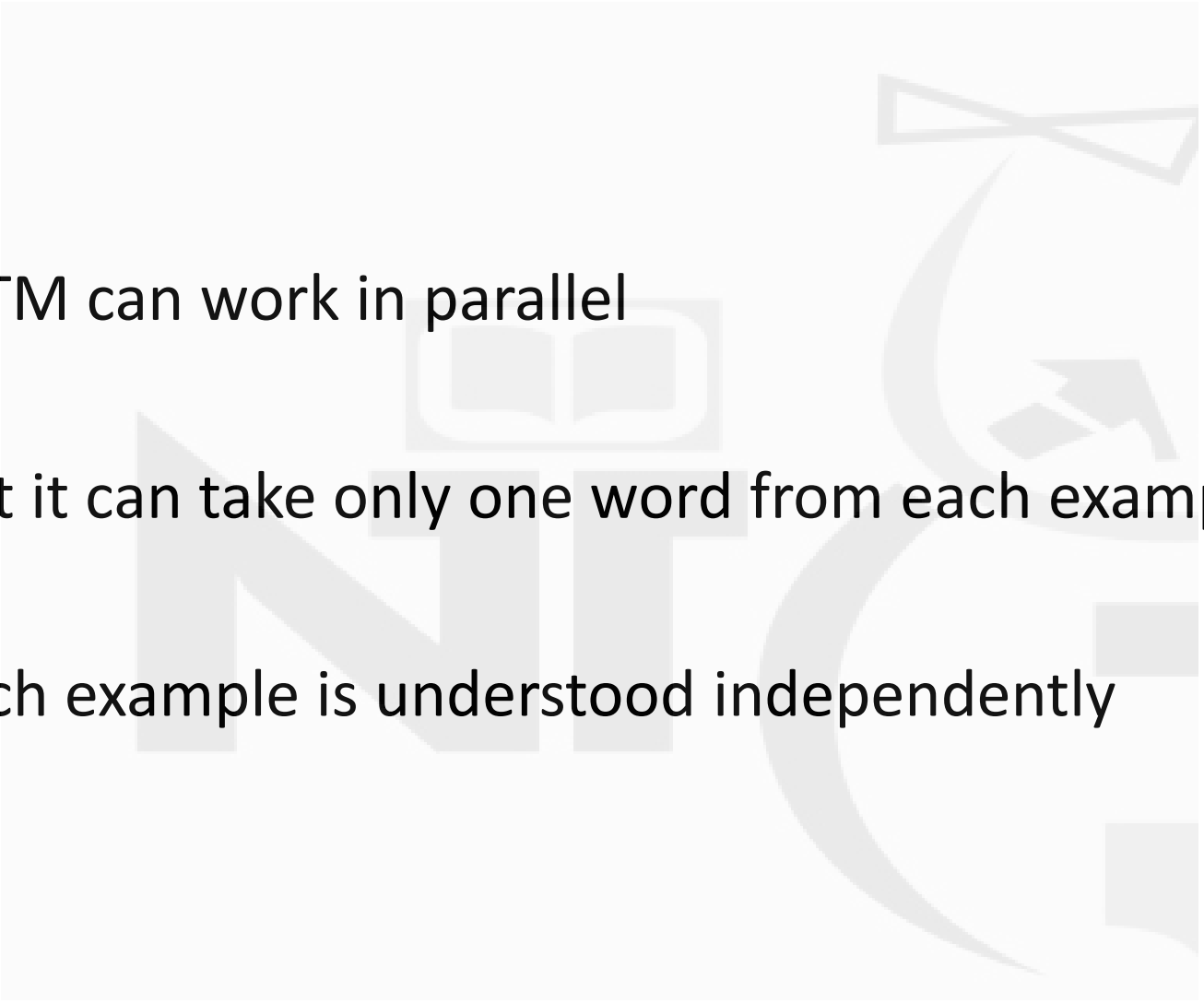
How many times will LSTM run per review?

- LSTM will run 300 times per review as each review has 300 words

Let's say if review has 4 words

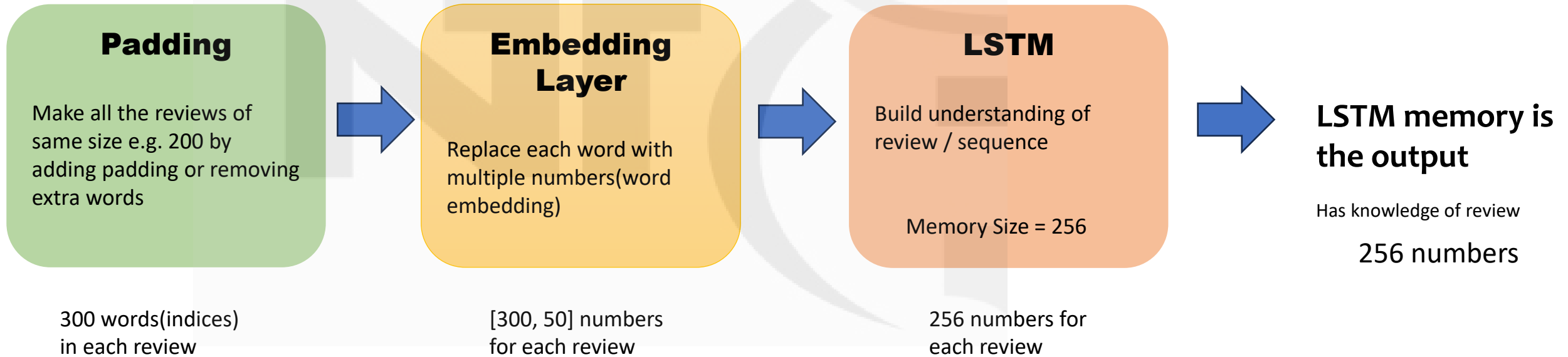


LSTM for Sentiment Analysis

- 
- LSTM can work in parallel
 - But it can take only one word from each example at a time
 - Each example is understood independently

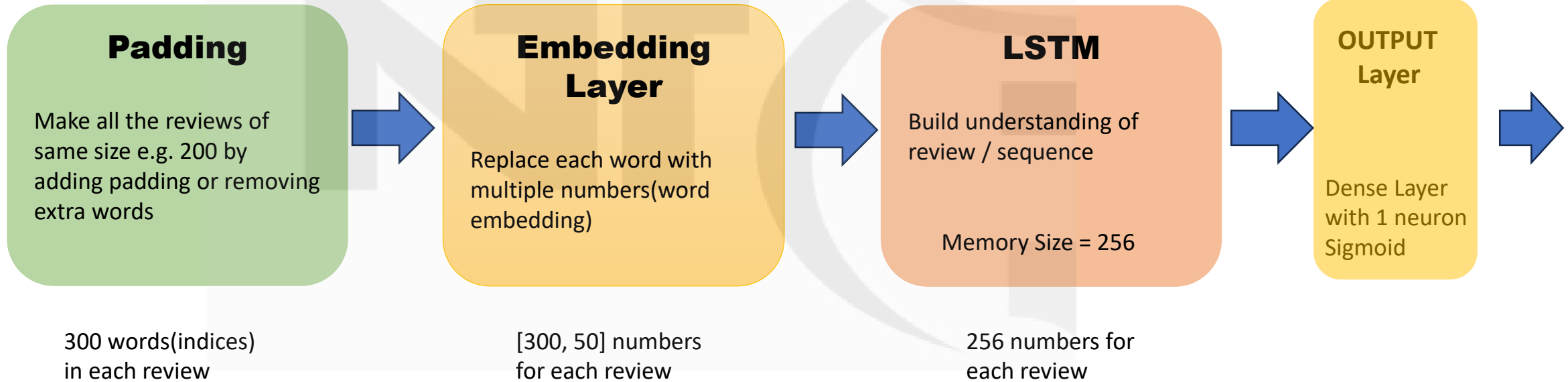
Mode Building

- Output is the memory size of LSTM



Next layer?

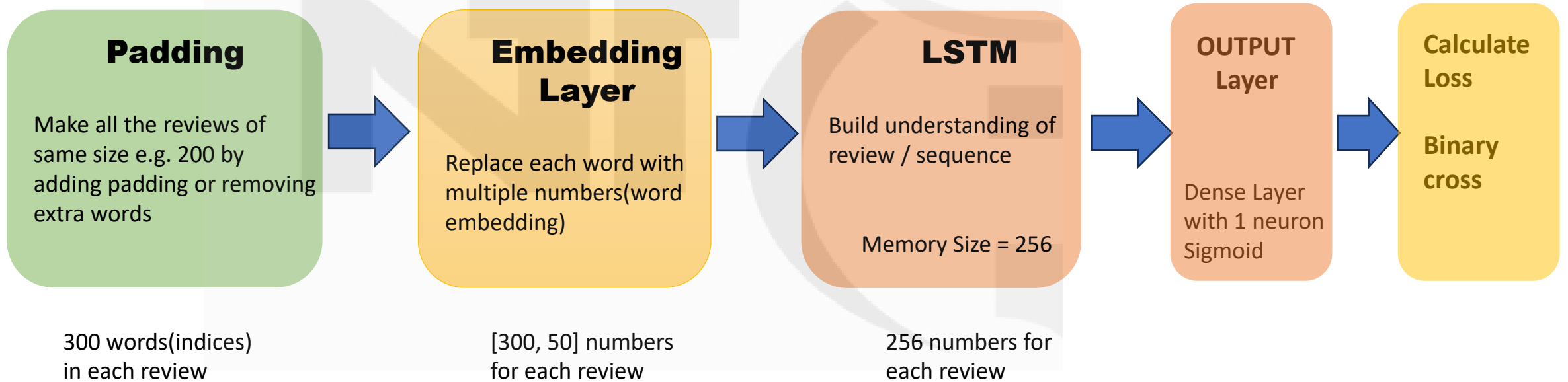
- Next layer will be output layer



MODEL BUILDING

Next?

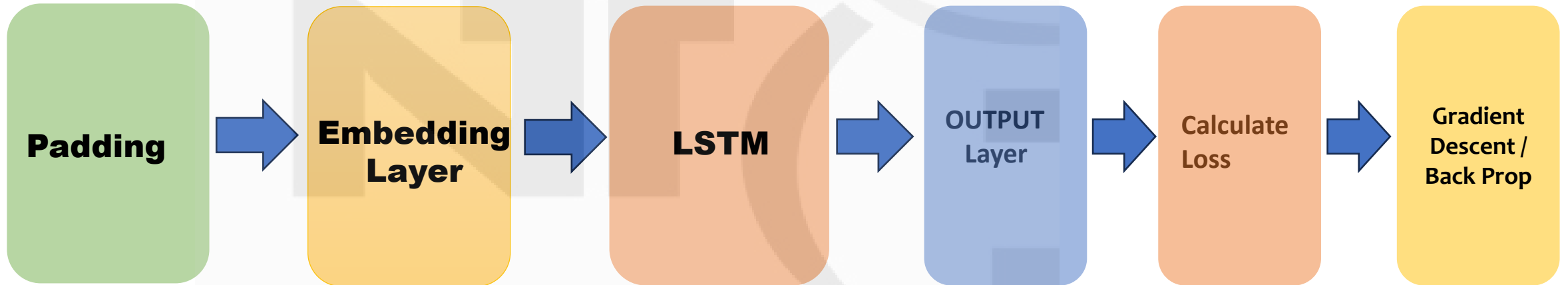
- We calculate the loss



MODEL BUILDING

Next?

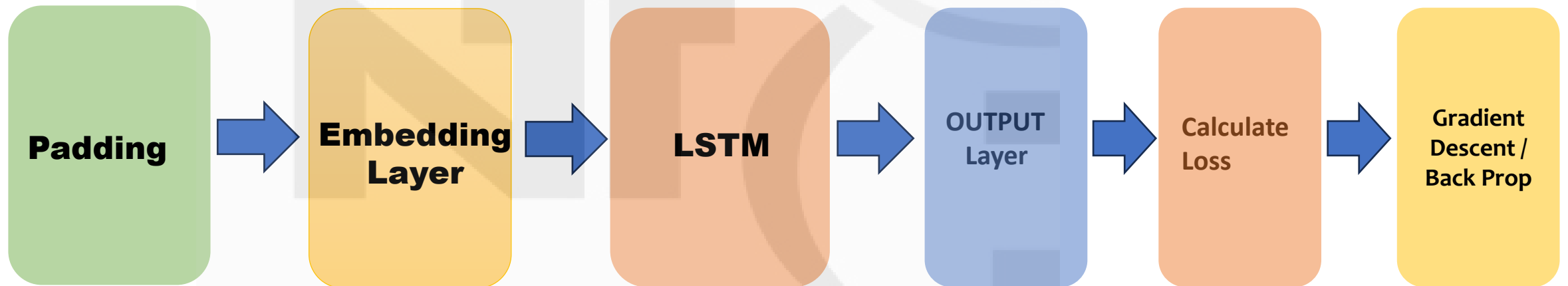
- Gradient Descent



MODEL BUILDING

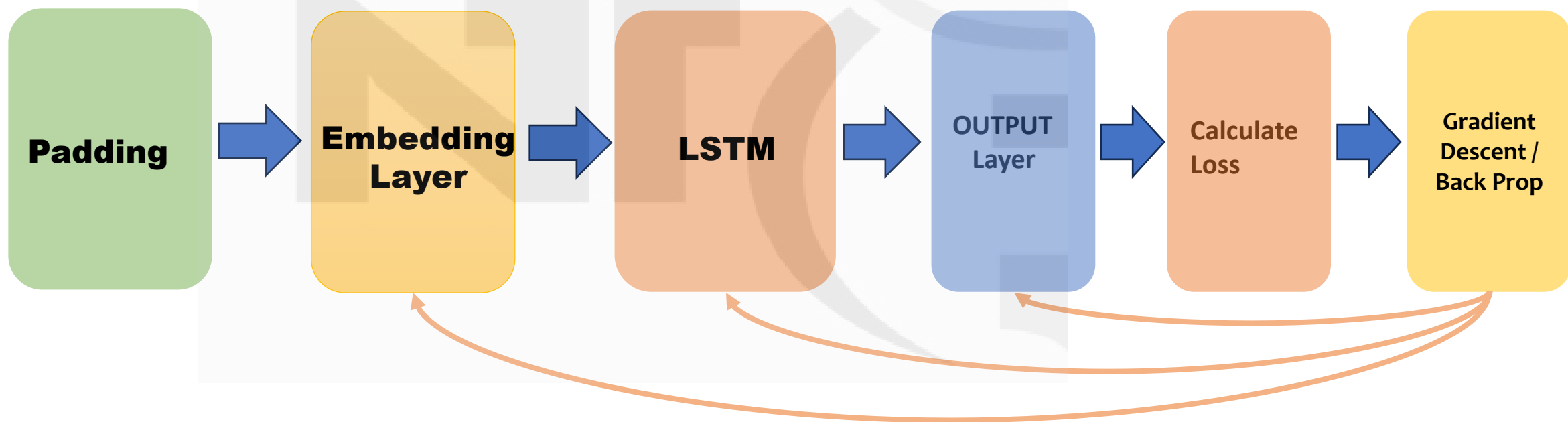
Which layers will have weights

- Gradient Descent



MODEL BUILDING

Update weights during backprop



MODEL BUILDING