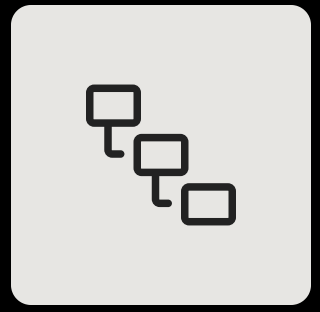
PROJECT PROPOSAL PRESENTATION

Second Year Project

Baseline recap

Model - Recurrent Neural Network

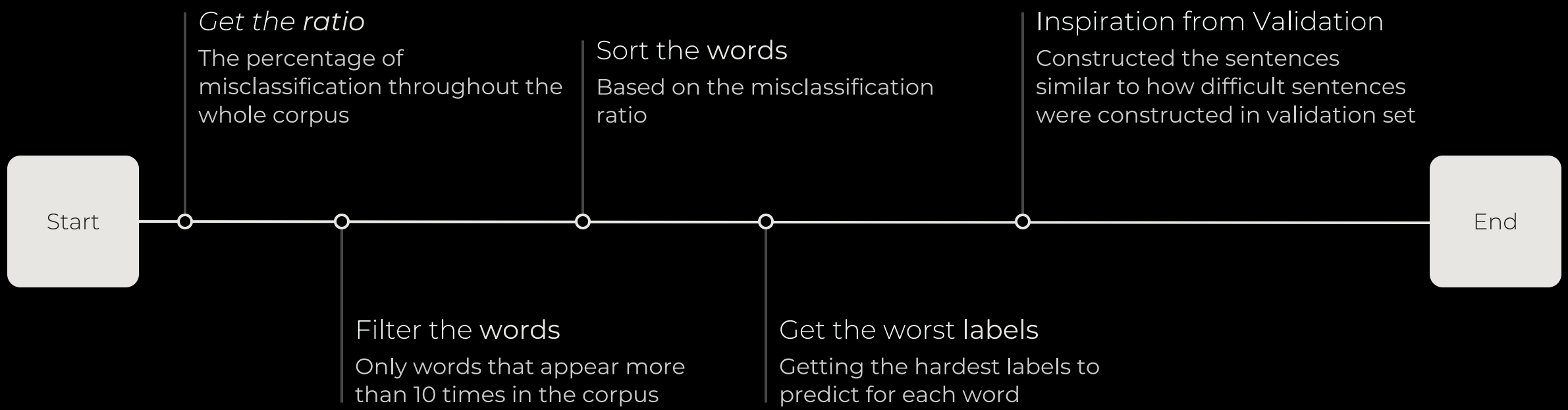
- Optimizer - Adam -
- Loss function - Categorical Cross Entropy -
- Initialized with n - amount of neurons, where n = vocabulary size
- hidden layer - 25 neurons - using 'tanH'
- output layer - 6 neurons (labels)



Feature Extraction

- Encoding used to pre-process the data for input
- Padding used to match the longest sentence

Making difficult sentences



Research Question

How can feature extraction be used to improve and adapt our baseline model?

Approaches to answer our RQ

Without adding new data, how can we make full use of the existing data ?

PoS - tagging

- categorizing the words
- context
- definition of the word

Model improvement

- tuning hyperparameters
- tuning model parameters
- exploring other models (maybe?)

Other properties

- Frequency
- Embedding

Word length

First letter