\*readme get finalized after class on Thursday\*

Additional pre-processing

* Tokenizing
* Vectorizing (word embedding)

Types of applicable machine learning models (one model per team member)

\*each model will have it’s own pre-processing\*

1. RNN - LSTM (could use keras library)- Jay
   1. Features – text, tokenized words, learned representations of text
2. CNN – Myles (could also use Keras)
   1. Features – word embeddings – could use bert embeddings
3. Transformers – BERT (bidirectional encoder representations from transformers)
   1. Features – contextual word embeddings, attention mechanisms
   2. Nick
4. Bag of Words – represents text as a vector of word counts (binary counts)
5. Reinforcement learning
6. Support vector machines ?

Possible evals to add – create tables/plots – visualizing data

* R-Squared - Michael
* Accuracy
* Cross-validation
* Sparse categorical cross entropy - Archan
* Cross-entropy loss