

Data - Quora question pairs dataset



Experimental Preprocessing

- 1) Tokenizing
- 2) Handling stopwords
- 3) Stemming
- 4) Lemmatizing



Feature Engineering

- * Cosine Similarity
- * Common words etc.
- * POS



Modeling / parameter tuning

Possible model choices.

- 1) Logistic Regression
- 2) Random Forest
- 3) XGBoost



Model performance evaluation.



Ensembling models.



Final model