<https://www.kaggle.com/code/hafiznouman786/duplicate-question-pair>

1. Adding more features:
   1. Length of Questions
   2. Number of Words in Both Questions
   3. Number of Common words
   4. Number of Total Words
   5. Word Share Calculation :: Word Share = (word common)/(word total)
2. Data Viz
3. Data Prep
4. NLTK packages and Lemmatization
5. Word2Vec
6. LSTM ( RNN)
7. Model Training

<https://www.kaggle.com/code/currie32/predicting-similarity-tfidfvectorizer-doc2vec>

1. Data Pre-processing
2. Method1 -> TFIDF – Use Cosine similarity to determine if 2 questions are similar
3. Method2 -> Doc2Vec - Precision, Recall, F1 Score

<https://www.kaggle.com/code/arathee2/predictive-modelling-a-simple-approach>

Judgment based on common words present in both sentences.

Models Used:

1. Predictive Modelling – Baseline Model
2. Logistic Regression
3. Decision Trees
4. Random Forest

Extra Idea: Removing stopwords for improved accuracy.

Research Paper:

Bilateral Multi-Perspective Matching (BiMPM) Model.