Done by:

Alaa AL-Ghamdi

Arwa Alolyani

Nadia Hajrasi

Quora Question Pairs

Can you identify question pairs that have the same intent?

Introduction:

Quora is a place to gain and share knowledge—about anything. It's a platform to ask questions and connect with people who contribute unique insights and quality answers. This empowers people to learn from each other and to better understand the world.

Dataset:

The Quora dataset consists of over 400,000 lines of potential question duplicate pairs. Each line contains IDs for each question in the pair, the full text for each

question, and a binary value that indicates whether the line truly contains a duplicate pair.

Data fields :

- id - the id of a training set question pair
- qid1, qid2 - unique ids of each question (only available in train.csv)
- question1, question2 - the full text of each question
- is_duplicate - the target variable, set to 1 if question1 and question2 have essentially the same meaning, and 0 otherwise

Goal:

The goal of this competition is to predict which of the provided pairs of questions contain two questions with the same meaning. The ground truth is the set of labels that have been supplied by human experts.

Question:

1- What is the number of words in each question?

2- What is the percentage of similarity between the questions?

3- What is the main topic of the question?

4- How many frequently asked questions?

5- What are the most frequently asked questions?

Tools:

- **1. Natural Language Toolkit (NLTK)**
- **2. TextBlob**
- **3. CoreNLP**
- **4. Gensim**
- **5. spaCy**
- **6. polyglot**
- **7. scikit–learn**
- **8. Pattern**