**Introduction:**

As we all know Quora is a platform to ask and answer questions on different type of topics like business, technology, innovation, personal life, and sports…etc. Quora is a driven by community to share the knowledge among the groups. Any one can answer any question and any one can answer in their own style.

The main issue these kinds of platforms face is same question is being asked multiple times though the answer may already exist to it. To improve user experience, it is suggested to reduce the count of duplicate questions. With millions of questions being asked daily on Quora it’s difficult to identify duplicate questions manually.

The prime focus of this project is to identify duplicate questions on Quora using machine learning and NLP capabilities so that we can improve user experience by saving users time and resources which in turn improves overall platform efficiency.

**Area of Study:**

This project comes under the domain of NLP and ML As the input data is of type textual information, we leverage the NLP techniques to solve this problem. Using the NLP techniques, we extract features from textual question data and then we apply predictive ML models to train our model on the features extracted.

**Motivation:**

Being a Quora user, many times felt difficulty to find relevant information related to my question. Identifying duplicate questions not just solves difficulties faced by users, but also enables the ways to give aggregated information to a question so that user experience will be much better with efficiency. It is kind of a practical application where NLP and ML capabilities can be utilized a solve a real word problem which also gives an opportunity to understand and utilize these techniques.

**Significance:**

This project important because it can change the ways to handle duplicate questions handled in the question answering platforms. It targets the long-term outcomes of the platforms in terms of underlying resources being used for this platform and also user experience as already mentioned. Moreover, we can extend this methodology to other platforms like stack overflow, Reddit…etc.

**Contribution:**

**Problem Statement:**

Redundant questions clutter the Quora platform so need to build an automated system to identify the duplicate question pairs. We hypothesize that it is possible to create a predictive model to identify two questions are similar or not based on various features extracted using NLP. Then creating a predictive model with high accuracy in identifying duplicate questions.