**Idea #1**

**Background:**

Many social media, news, e-commerce websites allow users to share their comments with others. This type of information sharing helps people stayed informed and brings others that are geographically remote closer together.

**Problem:**

However, some people use these website features to post inappropriate messages, links or scam for their own ill-intentioned motives. It devalues the communication channel and ruins users’ opportunities to share information with others.

**Objective:**

Develop an algorithm to identify whether posts are spams or actual comments by an individuals. Client can use this algorithm to flag, filter and remove spams from their website posts.

**Sample Data:**

YouTube Spam Collection Data Set --- Contains total of 1,956 comments from 5 music video pages. The data is labeled and the spam vs non-spam ratio is about 50/50.

< <http://archive.ics.uci.edu/ml/datasets/YouTube+Spam+Collection> >

**Other related application:**

Identification of spam emails.

**Idea #2**

Lending Club Data Set --- Predict whether individuals would default on the loan or how much he/she would pay back based on previous loan data.

< <https://www.lendingclub.com/info/download-data.action> >

**Idea #3**

Census Income Data Set --- Predict whether individuals have income exceeds $50K/yr based on Census data.

< <https://archive.ics.uci.edu/ml/datasets/census+income> >