Capstone 2 Goal: Learn Text Classification and Natural Language Processing

Option 1: Classifying Genetic Mutations

<https://www.kaggle.com/c/msk-redefining-cancer-treatment>

* Description – Cancer center looking into personalized medicine
* Background – Researchers & oncologists manually annotate whether mutations are drivers (contribute to tumor growth) or passengers (neutral mutations) based on clinical text or observations
* Goal – Classify mutations from a variable range of 1-9
* Cases – 3,321

Option 2: Toxic Comment Classification Challenge

<https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge>

* Description – Conversation AI team founded by Jigsaw and Google aim to help online discussions become more productive and respectful
* Background – Provided wikipedia comments
* Goal – Classify toxicity from 6 variables: toxic, severely toxic, oscene, threat, insult, identity hate
* Cases – 95,851

Option 3: How ISIS Uses Twitter

<https://www.kaggle.com/fifthtribe/how-isis-uses-twitter>

* Description – Develop effective counter-messaging measures against violent extremists in the US and abroad
* Background – Provided tweets from ISIS supporters around the world
* Goal – Develop social network cluster analysis, ID peak moments, keyword analysis
* Cases – 17,000