Machine Learning Techniques – Problem Set 1

WHAT?

I am going to use Trending YouTube Video Statistics data from Kaggle website. This dataset consists of viewership habits for the most famous videos on Youtube. Each region's data is in a separate file. Data includes the video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count.

I will work with data for the Great Britain on 2017 and 2018. Data has 16 columns and 38 916 rows.

WHY?

Categorising YouTube videos based on their comments and statistics. I will create a classification model, that will detect which category of the video depends on viewership habits and statistical details.

HOW?

The Decision Tree Classifier will be used for building a model. A **decision tree** or a **classification tree** is a **tree** in which each internal (non-leaf) node is labeled with an input feature.

OUTCOME DISCUSSION

Based on our outcome we can increase classification accuracy of the video content across the platform and provide better customer experience.