YouTube text analytics: Imported necessary libraries and removing stop words, pre-processing the transcript text and tokenizing before calculating tf-idf from a Count vectorizer. Then built a logistic regression model to classify data

Face detection.ipynb: Consists of python code to detect human faces using OpenCV. The code extracts 2000 frames of images of YouTube videos and detects if a human face is available or not.
If at least not 3 human faces are seen in frames of 0 - 2000-time frames with intervals of 100, then the video is deemed to be animated.

YouTube comment sentiment analysis.R: Comment sentiment analysis code consists of comment extraction part from videos by connecting to YouTube API. This code also includes a syuzhet package used to identify the sentiment scores for each video comment, visualizing them into aggregated bar charts, finally helpful in identifying overall sentiment values.

Unsupervised Topic Modelling.ipynb: This file consists of preprocessing codes like Stemming, vectorization, Lemmatization and creating a dictionary for LDA.

This file also consists of gensim model and LDA codes for running an unsupervised topic modelling. Visualization codes for LDA include pyLDAvis and word frequency distribution charts. This file also consists of dominant topic identifying codes.

Topic modelling final.ipynb - Topic modelling consisted of preparing a bag of words for 4 different topics of “Symptoms”, ”Prevention”, ”Treatment” and “Remedies”
that are picked up from the reference papers. Each transcript is then subjected to supervised modelling where percentage of the transcript covering the 4 topics is shown by the model.

*This code also consists of an automating code to scrape video transcripts from all videos once at a time for the key values given as input.