Peem Lerdputtipongporn

36757: Advance Data Analysis I

Last updated: 31st January 2022

Due Date: 4th February 2022

**Statistics advisor(s):** Nynke Niezink and Larry Wasserman (Department of Statistics and Data Science, CMU).

**External scientists names and affiliations:** Rema Padman (Heinz College of Information Systems and Public Policy, CMU); Anjana Susarla (Broad College of Business, Michigan State University); Xiao Liu (WP Carey Business School, Arizona State University).

**Abstract**

Youtube offers alternative means to engage users and promote their health literacy through health videos. Ideal health videos should be understandable, relevant, actionable, and medically accurate. Traditionally, assessing the extent to which each video satisfies these each of the criteria requires physicians manually label according to guidelines such as PEMAT (Patient Education Material Assessment Tools), USPHS, among others. Two major drawbacks of manual labelling are 1) it does not scale well with large dataset; 2) it does not assess video-specific features associated with quality, preventing content creators from improving their videos. Liu et. al. (2021) have proposed methods to classify video understandability according to PEMAT as characterize medical information on a binary scale using augmented intelligence approach. Our work complements that of Liu et. al. by operationalizing PEMAT-based guideline on actionability and domain expert assessment on accuracy via machine learning algorithms. Our analysis shows that actionable and medically accurate videos tend to come from reputable institutions and contain many medical terms presented using simple language. Implications for research and practice are discussed.