Duc Le

CS 614

Although I am not currently participating in any research of my own, I did find this article helpful for possible future research opportunities and personal projects. A lot of school and personal projects of mine related to data science have required external data sets. And in my experience, it was always a challenge finding an existing data set that could adequately answer the proposed research question. Last semester in CS 530 (Data Mining), for our final project, we wanted to determine the outcome of COVID-19 patients, predict whether they die or recover, and the probabilities of each outcome given certain features. The algorithms used were Random Forest and Logistic Regression and we had a difficult time finding a fitting data set that properly contained all the things we wanted. The Iterative Process steps in Table 1 provides a helpful step-by-step approach when performing secondary analysis. Particularly, refining the research question and evaluate the appropriateness of the original sample, design and measures.