**R & Python Programming Assignment**

**Note: It’s mandatory to attempt all the parts.**

1. Using Python how can you fetch/extract data from one website (for the whole website, in addition to home pages) and present it on another Dashboard for Internal Use?

2. Using Arima Or Vector Autoregression (VAR), we want you to prepare an Automated Data Forecasting Sheet for attached GCC Construction Chemicals Market. We want it in such a way that only if we have a enter a few parameters the whole ME Sheet is prepared automatically.

3. Use the [Head and Neck Cancer Medication Data](https://umich.instructure.com/files/1614350/download?download_frd=1) to to apply NLP/TM methods and investigate the corpus.

Use the MEDICATION\_SUMMARY to construct a VCorpus object

Clean the VCorpus object.

Build a document term matrix (DTM).

Add a column to indicate early and later stage according to seer\_stage, refer to [Chapter 7](http://www.socr.umich.edu/people/dinov/2017/Spring/DSPA_HS650/notes/07_NaiveBayesianClass.html).

Use the DTM to construct a word cloud for early stage, later stage and the complete archive.

Interpret the results of the three generated word clouds.

Compute the TF-IDF(Term Frequency - Inverse Document Frequency).

Apply LASSO on the unweighted and weighted DTM respectively and evaluate the results according to AUC.

Try cosine similarity transformation, apply LASSO and compare the results.

Use other measures such as “class” for cv.glmnet().

Does it appear that these automated machine learning methods understand well human language?