Assignment 3: Business Intelligence Using Text Mining

1. A dataset of Shark Tank episodes is made available. It contains 495 entrepreneurs making their pitch to the VC sharks.
2. You will ONLY use “Description” column for the initial text mining exercise.
3. Step 1:
   1. Extract the text into text corpus and perform following operations:
      1. Create DTM
      2. Use “Deal” as a Dependent Variable
      3. Use CART model and arrive at your CART diagram
      4. Build Logistic Regression Model and find out your accuracy of the model
      5. Build randomForst model and arrive at your varImpPlot
4. Step 2:
   1. Now, add a variable to you analysis called as “ratio”. This variable is “askedfor/valuation”. (This variable is to be added as a column to your dataframe in Step 1.)
   2. Rebuild “New” models- CART, randomForest and Logistic Regression
5. Deliverables: (in a word document)
   1. CART Tree (Before and After) (5 Marks)
   2. RandomForest plot (Before and After) (5 Marks)
   3. Confusion Matrix of Logistic Regression (Before and After) (5 Marks)
   4. (Most important)- Your interpretation in plain simple English not extending more than half a page. (15 Marks)