# ***Assignment Name - Advance Predictive Modelling***

1. **How will you treat text having short cut words (like bcz, u, thr etc…)?**

Answer: - We’ve all seen tweets and many other social media platforms with a plethora of spelling mistakes, used many abbreviated form or basic shortcuts. In this situation, spelling correction is a useful pre-processing step because this also will help us in reducing multiple copies of words. For example, “You” and “u” will be treated as different words even if they are used in the same sense.

To achieve this we will use the textblob library.  Moreover, we cannot always expect it to be accurate so some care should be taken before applying it. There is one function called “Spelling correction” which replace misspelled words in our dataset. For very common usable small shortcuts we can create a list of most frequent shortcuts and can replace by calling that list to replace those words.

1. **Write R and python code to replace “bcz” with “because” in whole text?**

Answer:- R and python code has attached herewith as well as attached separately. .

1. **How do you deal with the English text having Hindi words in between?**

Answer: - In our daily life, we often notice in Tweeter or other social media platforms, people use Hindi words with English texts in between while communicating or reviewing any products or thoughts. If we need to deal with those kind of data we need to process those Hindi words to get any meaningful information. We can use a [phonetic algorithm](https://en.wikipedia.org/wiki/Phonetic_algorithm) like soundex to deal with out-of-vocabulary terms and to try to match them to Hindi words. Then we translate those Hindi words into English.

The process is as follows:

* Transliterate the words in Hindi language script and look up the words in Hindi dictionary for its existence. Check for spelling variations. If found tag them as such.
* Lookup all the words in English dictionary and tag them as such. If there is tie with the other language then use word frequency probability to break the tie.
* Loop up the sentiment scores from sentiwordnet of each Hindi words.
* Finally convert the words to features using ngrams with their corresponding sentiment scores and perform the cross — validation of training and test data to see the accuracy of results.

1. **Write R code to connect with this public API -** [**http://www.omdbapi.com**](http://www.omdbapi.com)

Answer:- R code has attached herewith as well as attached separately. 

1. **What are the different methods to deploy a model into production system?**

Answer: - There are two ways to deploy a solution or script to production system.

* ***Offline Methods: -*** In offline method of deployment we usually schedule a task in our UNIX or Windows Task scheduler which execute in a specific defined time and execute the code and store or send the output by email.
* ***Online Method: -*** In online methods we host our solution or script in cloud. There are some tools called RavoDeploy R, Orange which help to set the script in cloud. On click of a button using UI, it calls for a Function and that connect to the online script to those cloud platforms and execute accordingly and give us outputs.  
   We can use JAVA, .NET, VB or any other language to make a good UI.  
   Sometimes we use TSQL or PL-SQL to connect to Databases to retrieve data. Then we use PMML (Predictive Model Markup Language), Tableau, PowerBI or any other visualization tools to view models or outcomes in a structured form.