

**Welcome to Assessment for NLP Scientist role: Sentiment Analysis**

We , at Lilly, deal with lot of structured and unstructured text data. Understanding and analysing textual data, and inferring context is an important aspect of natural language processing. Your role will require you to have a strong working knowledge for

* pre-processing textual data,
* different NLP techniques, and the appropriate situation for each technique
* applied statistics

A total of 2 questions on each of the above topics are listed below, and you have a total of 1 week to work on these questions. At the end of the week, please email your responses and outputs back to us.

All the best!

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Exercise 1:**

Attached along with this document is a data file (Data.csv). It contains approximately 15k tweets which have been classified as positive, neutral and negative. As part of this exercise we are going to evaluate approach to train a model to identify sentiment.

Using attached data as training set , develop both a **machine learning model and a deep learning model** to predict sentiment of text.

Outputs expected:

1. Create and share your Jupyter notebook files for model training. You can choose to use Google Colab.
   1. Explain your choice of algorithms and approach you used for your models.
   2. Add comments wherever appropriate.

**Exercise 2:**

Given a clinical case sheet which describes all the activities and milestones for a patient. An example sentence in this document could be “23/3/20 – Pt came in with severe headache, advised oral acetaminophen”. Write a generic Python code to identify and parse out all dates that could be available in such documents.

Good Luck!!!