Problem Statement

In real-time summarization, we need to predict the event of insights (actions items, decisions, etc.) from streaming data. At a particular moment, we have a portion of text as a chunk from which we need to predict if the portion of text can be an insight or not (binary classifier)   
  
P.S. - You will get to work on the bigger problem of multi-class classification of the same problem and our proprietary algorithm with Instaminutes.  
  
Q1. What algorithm or architecture do you think of which can satisfy the requirement (if any)? (Analogy will suffice), bonus point (optional) if you can think of some innovative method outside of present architectures like stacked LSTM, transformers, etc.

Q2. Any automatic (python script) approach can you think of to label the data for training?  
(Don’t overthink, think of simple approaches that you use as human)

Q3. Please provide the python code which can label a portion of text (maybe sentences) as insights. Please show the output of your code for the text attached as the input.

Expected Solution

You can find a sample output file marked with a portion of text (chuck) that can be part of insights (from our current algorithm)  
  
Your output must contain the list of strings that can be labeled as insights (Boolean true).  
  
**We will need all solution on GitHub and share the link**

Text file (Docx, txt, etc.) explaining Q1, Q2

Python code (Ipynb, py) and txt file as output for Q3

Hint: Do not go for fine-grained chunks initially. Try to label complete sentences.

Timeline

Submit at [connect@instaminutes.com](mailto:connect@instaminutes.com) before 10 pm 25/04/22 (Tuesday)