Idea 1:

The Multi-Domain Sentiment Dataset contains product reviews taken from Amazon.com from many product types (domains). Some domains (books and dvds) have hundreds of thousands of reviews. Others (musical instruments) have only a few hundred. Reviews contain star ratings (1 to 5 stars) that can be converted into binary labels if needed. This page contains some descriptions about the data. If you have questions, please email Mark Dredze or John Blitzer.

Data: <https://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

Solution: Can be a sentiment extraction use case

Idea 2:

MCTest is a freely available set of 660 stories and associated questions intended for research on the machine comprehension of text. The data was gathered using Mechanical Turk

Data: <https://mattr1.github.io/mctest/>

Solution: Can be used to train a model that can answer questions based on paragraph given - data is less