Topic: NLP and decision making

List of datasets I think can be helpful:

Dataset .1:

Name: Deep-NLP

Description:

* a dataset consisting of people’s response to a question (“*Describe a time when you have acted as a resource for someone else*”)
* It has been manually annotated as “flagged” or “not flagged”.

Task:

* When it is flagged, the person can continue talking to the chat bot.
* If not flagged, the user will be stopped.

Link: [here](https://www.kaggle.com/datasets/samdeeplearning/deepnlp?datasetId=897&sortBy=voteCount)

My notebook on it: [link](https://www.kaggle.com/code/mitramir5/text-eda-rnn-lstm-gru-and-many-more)

Additional notes: This feature can be used in many fields. It can for example, tell the customer service rep to continue talking about one specific topic, or not. It can sense whether the information he/she is providing is useful or not. Then sense if the customer is satisfied.

Dataset .2:

Name:

Description:

* A set of scraped tweets with their manually annotated sentiments.
* In each tweet, one span of text is mostly expressing the sentiment of that tweet.

Task:

* Find the span that is mostly expressing that emotion by finding the start and end index of that span.

Link: [here](https://www.kaggle.com/competitions/tweet-sentiment-extraction/data)

Chatbots telling people what they have to look for or what they have to do. These models are either question answering models or text generative models.

For generative models we can use many datasets.

Dataset .3:

Name: transcripts of The Big Bang Theory Show

Description:

* I scraped this dataset a while ago to create a talking model with characteristics of the characters I like. (I eventually created a GPT model that can talk like characters when prompted.)
* It can be used to create another dataset: a dataset with every record having a conversations (two records/dialogues as one input) so it can learn to respond.

Link: [here](https://www.kaggle.com/datasets/mitramir5/the-big-bang-theory-series-transcript)

Some additional datasets:

Dataset: Extractive question answering sample: SQUAD ([link](https://rajpurkar.github.io/SQuAD-explorer/explore/v2.0/dev/Force.html))