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 conversation (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)

Packages or websites:

* Early Therapy chatbot, Eliza from 1960s: [link](https://web.njit.edu/~ronkowit/eliza.html)

Some additional datasets:

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

Keywords: Therapy chatbots ([link](https://scholar.google.ca/scholar?q=therapy+chatbot+nlp&hl=en&as_sdt=0&as_vis=1&oi=scholart)), NLP in Healthcare, education, decision making,

Papers:

Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health [link](https://aclanthology.org/Q16-1033/)

Detecting Depression with Audio/Text Sequence Modeling of Interviews [link](https://groups.csail.mit.edu/sls/publications/2018/Alhanai_Interspeech-2018.pdf)

Ideas: medical notes: what Dr. Ghassemi did for analyzing the notes of doctors and nurses for giving instructions or assistance in decision making for doctors.

Research question:

Can we have models that decide for us by looking at training data?

What we interpret when we see outputs of NLP models?

What are some systems that help with decision making in healthcare?

How do people use the datasets to decide?

Who decides if the answers to our questions generated by NLP models are ethical? Who says we should oblige to their command?