NLP CHATBOT-
Fictional Character Conversational ChatBot with huggingface

PF26	33	1032180786	Umang Sakuja	1032180786@mitwpu.edu.in
PF13	14	1032180387	Rutuj Tatle	1032180387@mitepu.edu.in
PE64	77	1032182014	Yash Rathore	1032182014@mitwpu.edu.in
PA21	25	1032180379	Aditya Rao	1032180379@mitwpu.edu.in

Problem Statement

- We have used dialogues of Harry Potter book which has 1600 lines which will be preprocessed and then used for training the model

- After training the model we hypertune the parameters for best results

- Finally we deploy the model to huggingface api where we get the up and running chatbot.

DataSet

- https://www.kaggle.com/gulsahdemiryurek/harry-potter-dataset

269.08 kB

▥ Characters.csv
▥ Harry Potter 1.csv
▥ Harry Potter 2.csv
▥ Harry Potter 3.csv
▥ Potions.csv
▥ Spells.csv
▥ shortversioncharacters.csv

< **Harry Potter 1.csv** (67.44 kB)

Detail   Compact   Column

About this file

Harry Potter and the Sorcerer's Stone Movie Script

▲ Character		▲ Sentence	
1529 unique values		[null]	65%
		Harry.	1%
		Other (545)	34%
Dumbledore		I should've known that you would be here, Professor McGonagall.	
McGonagall		Good evening, Professor Dumbledore.	
McGonagall		Are the rumors true, Albus?	
Dumbledore		I'm afraid so, professor.	
Dumbledore		The good and the bad.	

Preprocessing

- Loading data into dataframe
- Using front and back of the dialogue to make sense and create reference

	response	context	context/0	context/1	context/2	context/3	context/4	context/5
137	Go on!	Dumbledore is particularly famous for his defe...	I found him!	Besides, if anyone cared to notice my eyebrows...	I don't appreciate the insinuation, Longbottom.	You to set my bloody kneecups on fire!	No, that's all I need!	I'll do the countercurse!
109	Believed to be the work of Dark wizards or wit...	Listen.	Hey Ron! Somebody broke into Gringotts.	The only problem is, I can't remember what I'v...	When the smoke turns red, it means you've forg...	I've read about those.	Hey, look! Neville's got a Remembrall	Can I borrow this? Thanks.
128	Nicholas Flamel...where are you?	15th Century Fiends...Flamel...	Famous fire eaters...	It just said, "Use it well."	There was no name.	I wonder who gave it to you.	They're really rare.	I'm invisible??
59	What's in there, Hagrid?	Vault 713.	Didn't think your mum and dad will leave you w...	Key, please.	Lamp, please.	Vault 687.	Very well.	It's about You-Know-What in vault you-know-which.
29	Thank you.	Baked it myself, words and all.	Afraid I might have sat on it at somepoint, bu...	Got something for you.	Well, of course you are.	I am.	I'm not Harry.	Particularly 'round the middle.

Models Used

- After train test splitting of 70-30

- We used [Microsoft DialoGPT conversational model](#) to fine-tuned on the game transcript

- DialoGPT is a SOTA large-scale pretrained dialogue response generation model for multiturn conversations. The [human evaluation results](#) indicate that the response generated from DialoGPT is comparable to human response quality under a single-turn conversation Turing test. The model is trained on 147M multi-turn dialogue from Reddit discussion thread.

Training and Evaluation

- Large epochs caused underfitting of data

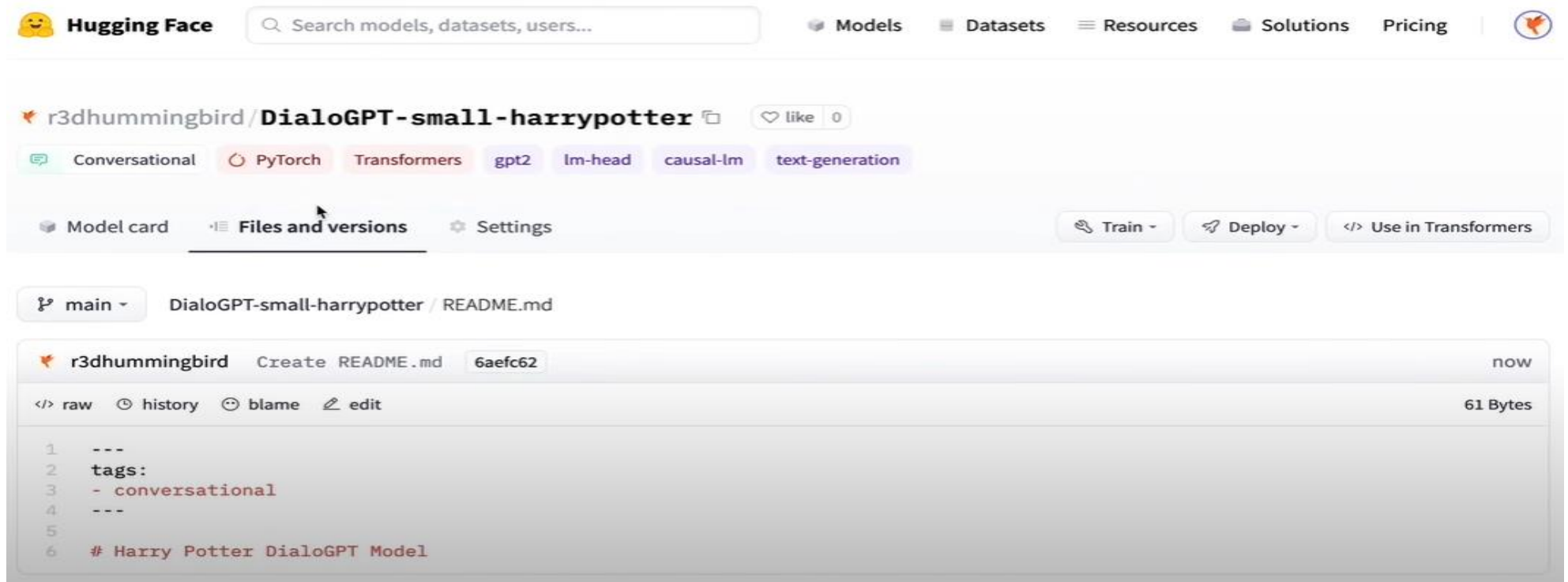- We used small sized dialogGPT with 12 epochs

```
08/08/2021 17:51:16 - INFO - __main__ -      Batch size = 4
Evaluating: 100%  [████████████]  4/4 [00:00<00:00, 17.33it/s]
08/08/2021 17:51:16 - INFO - __main__ -    ***** Eval results *****
08/08/2021 17:51:16 - INFO - __main__ -      perplexity = tensor(11.2926)

{'perplexity_': tensor(11.2926)}
```

- Lower the perplexity lesser the confused is the model

Deploying the Model

- We used git large file storage for push and pull of our model

- We used huggingface free api to deploy our model

- After setting permissions of send messages and reply or model got active.

Final ChatBot Outputs

ThankYou.

.