Class Augmentation

Třída ma za účel upravovat data v datasetu, tak že si bude pohrávat s daty a upravovat jak otázky tak i odpovědi nebo celé taxy, důležité je aby samotný význam věty se nezměnil ale jen se věta podala v jiném kontextu a model měl z čeho se učit , bude se k tomu využívat externí modely které specializují se na daný ukony jako je Bert př pro synonymous replacement… , můžeme i kombinovat více modelu

Funkce třídy

1. Def \_\_init\_\_  
   - Tato funkce bude mít nastarost přijmout data na kterých se bude pracovat

* Zavola funkci augment\_setting pro aby sebrala od uživatele jak bude chtít upravi dataset,

1. Def synonymous\_rep

* Za pomocí Bert modelu bude nahrazovat synonyma

1. Def random\_insrt

* Nejspíš také za pomoci bert modelu

1. Def random\_swap

* Nejspíš také za pomoci bert modelu

1. Def random\_deletion

* Můžeme zkusit za pomoci bermodelu

1. Def back\_translation

* Zde to zkusíme za pomocí MarianMT

1. Def terminal\_UI

* Zeptame se uživatele jak bude chtít trénovat model (resp jakou funkci použijeme
* Posléze se zepta jaký model k tomu použijeme nabídka bude Bert\_base\_uncased, GPT-3,5 a LLAMA 3 7b (po celou dobu upravy datasetu bude k dispozici pouze jeden model) - Def start\_model
* Po doběhnutí jedné úpravy zeptáme se uživatele jestli ještě budeme nějak upravovat nebo konec a můžeme se vrátit zpět do train class

Ke GPT budu používat běh přes servery openai

Ke LLama 3 budu používat běh přes servery anyscale

Oboje poběží přes api OpenAI