Project Goal: The goal of the project is to create an NLP language model that will generate text in Shakespeare style. As a result, the user will have a chance to provide the desired length of the text to generate Shakespeare-style text by the trained model.

Dataset: I used *Shakespeare_data.csv* for model training. The dataset is available online on the website Kaggle (Dataset link). It includes multiple columns; however, for this project, I used one of the columns, PlayerLine, which included the necessary data for the model, texts of Shakespeare plays.

Research Methods: To complete the task, I searched for different papers that describe the problem and suggest solutions to it. I explored different approaches, however, tried to focus more on PyTorch models as it was preferable. I will share the result of the model that I fine-tuned for the task, which provided possible best results compared to other models.

Project Guidelines: I created *NLP problem.zip* where you can find all necessary files. After unzipping the folder, there you can find:

- *Shakespeare_data.csv,* the downloaded dataset for the model
- *Model.ipynb,* the jupyter notebook file, where I included the PyTorch model code. I divided the code into some logical parts, gave them titles and added comments to explain how each part is working. There are detailed notes and comments how everything is working.
- *model_10epoch.pth,* the saved model after training 10 epochs
- *model_dict_50epoch.pth,* the saved model after training 50 epochs
- *Model.py*, the python file, which includes the same code as in *Model.ipynb,* however I created this py file in order to make model testing process for users easier. So, to test the model, just open one terminal window, change the path where the *NLP problem* folder is, type *python Model.py,* after running it input window will be opened with text *Enter the desired text length:* just provide number of desired text length, for example 600, and see the Shakespeare style generated text! 🙂