# Machine Learning Project Proposal

Kyle Hoffpauir

Dr. Juola

COSC532

Generative AI is, in my opinion, more interesting to create and play with than analytics. While less useful, they provide more fun. This is what led me to look for rap music datasets. While I myself am not a huge rap fan, I think the genre is specifically unique and interesting in it’s themes and delivery. I found the following datasets of rap lyrics: [here](https://www.kaggle.com/zhongtr0n/mumble-rap) and [here](https://www.kaggle.com/rikdifos/rap-lyrics). Inspired by AI dungeon: <https://play.aidungeon.io/main/home>, my application would generate a chunk of a rap given a user’s input may be an interesting challenge. The exact amount of text I generate may vary based on the results once I begin experimenting. I plan to begin with generating a single line, and if that goes well I may attempt to generate a bar or two. If the application can create bars then I may attempt to make it create entire songs.

I plan to generate the text using recurrent neural networks and python’s TensorFlow building off of tutorials [here](https://stackabuse.com/text-generation-with-python-and-tensorflow-keras/) and [here](https://www.tensorflow.org/text/tutorials/text_generation). Generating similar text to user input based on a dataset of text is similar to what we did with the 20 newsgroup data and the project could be extended to have a separate function to take in a user’s input and tell them which rapper their input most resembles based on the classification.

Different models for different rappers may be an interesting avenue to explore as well, allowing the user to select which rappers the model will be trained on to allow for their styles to be more apparent in the generated text. This could come with challenges relating to too small of a sample size and may not be implemented.

The end result would be an application that could generate rap lyrics and classify them based on what rappers have created combined with what the user has generated. While not particularly useful, this application of machine learning would be both fun and cool as it allows users to explore their creativity while bouncing rhymes off of my model, getting ideas from all of their favorite rappers (given their rappers exist in the dataset).

If text generation and classification are not sufficiently interesting, I have a parallel idea using a dataset of [philosopher’s quotes](https://www.kaggle.com/kouroshalizadeh/history-of-philosophy) or [poetry](https://www.kaggle.com/datasets/tgdivy/poetry-foundation-poems) in which I could identify if a given quote is more like poetry, philosophy, or rap and which author out of those groups the quote most resembles. This would result in an interesting classification problem given large amounts of diverse text.

## Works Cited

<https://www.kaggle.com/zhongtr0n/mumble-rap>

<https://www.kaggle.com/rikdifos/rap-lyrics>

<https://www.tensorflow.org/text/tutorials/text_generation>

<https://stackabuse.com/text-generation-with-python-and-tensorflow-keras/>

<https://www.kaggle.com/datasets/tgdivy/poetry-foundation-poems>

<https://www.kaggle.com/kouroshalizadeh/history-of-philosophy>