Project Plan

Due 3/10/2025

1. **Group:**
   1. Brian Davis – U00556990
2. **Project Dataset:**
   1. **Link to dataset:** https://www.kaggle.com/datasets/carlosgdcj/genius-song-lyrics-with-language-information
   2. **Target (if supervised):** Song Genre (listed in csv as “tag”)
   3. **Features:** title, tag, artist, year, views, features, lyrics, id, language\_cld3, language\_ft, language
   4. **Brief description:** The dataset is quite large, so I will likely use a smaller, balanced subset (or multiple subsets) to train a model with song lyrics to predict genre. I want to try combinations of models like SBERT and RoBERTa with SVM, Logistic Regression, Random Forest, and Multi-Layer Perceptron for these predictions. I might also try to fine-tune on a couple of Google’s T5 models to make predictions as well.
3. **Timeline**

|  |  |  |
| --- | --- | --- |
| Date | Student | Task |
| 17 March | Brian Davis | Data Preprocessing |
| 24 March | Brian Davis | Write code for training models |
| 1 April | Brian Davis | Train model using SVM, LR, RF, and MLP |
| 5 April | Brian Davis | Write Final Report |
| 9 April | Brian Davis | Record Presentation |

Note: While the timeline may change, this will provide a rough idea of the tasks that may need more (or less time). Please leave room for the final project report writing. That will be the outcome/ evaluation for the final project points for this class.