Utilizing Neural Networks To Categorize Songs

Max Bermont

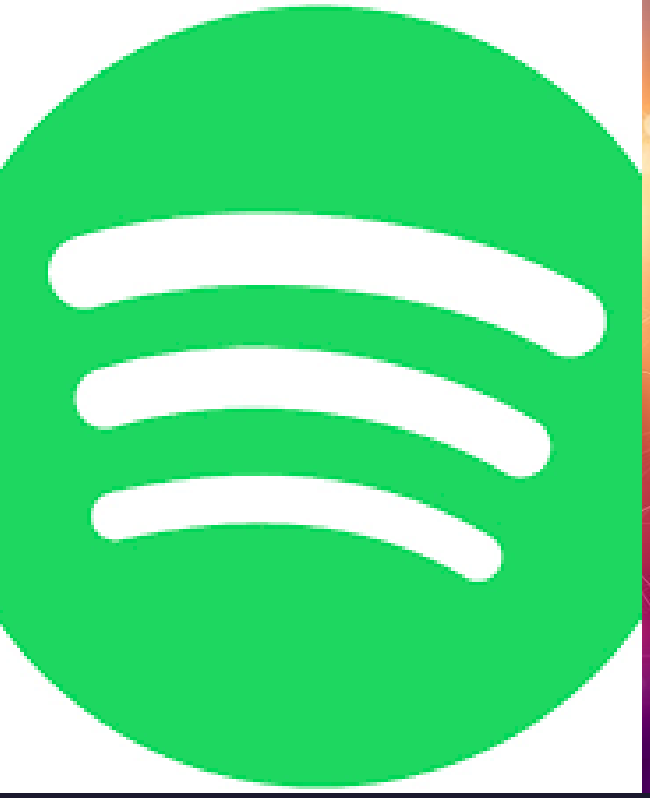
Agenda

Define the Problem

The Data

The Model

The Results



Introduction

- Spotify is one of the largest audio streaming services in the world. They have 345 million active users and over 70 million songs.
- Competitors SoundCloud and Apple Music

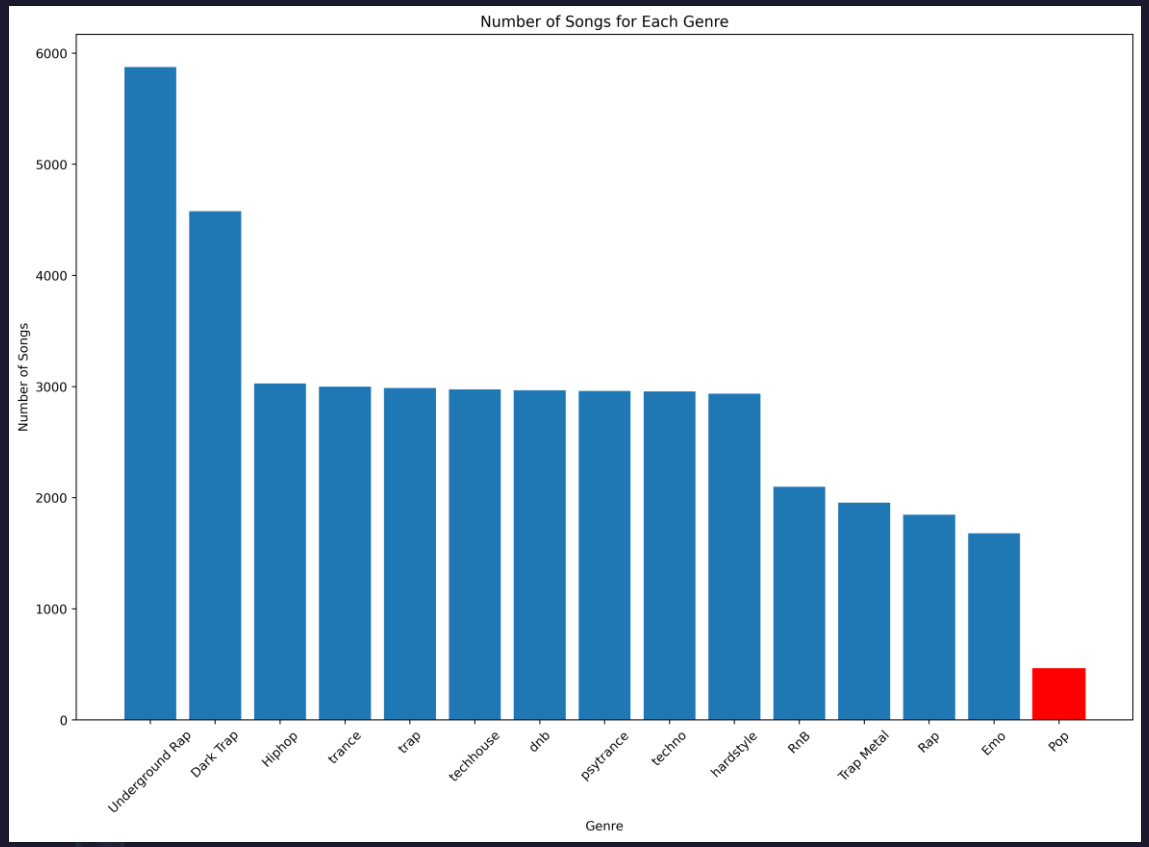
Defining the Problem

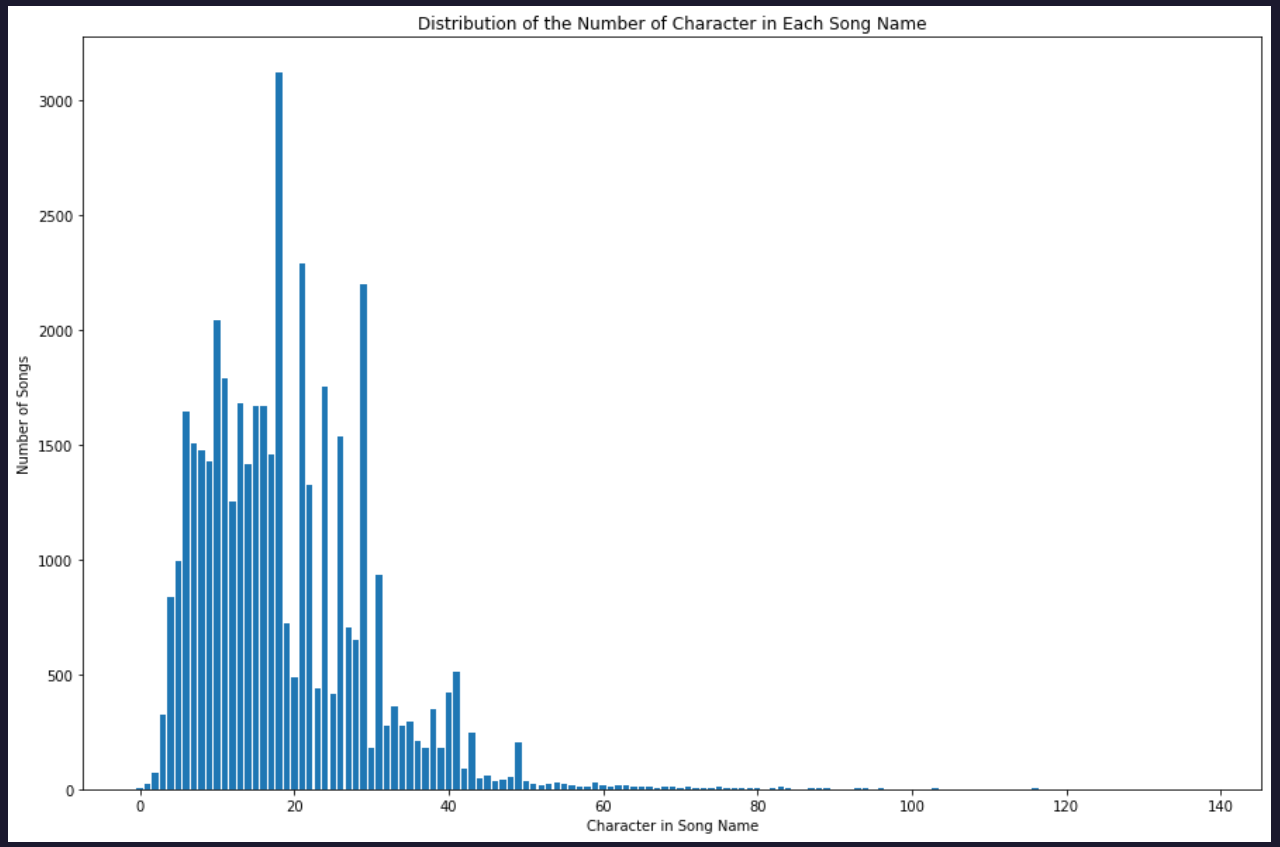
I have been tasked by Spotify to create a model that will assign genres to songs that are being added to the platform.

The Data

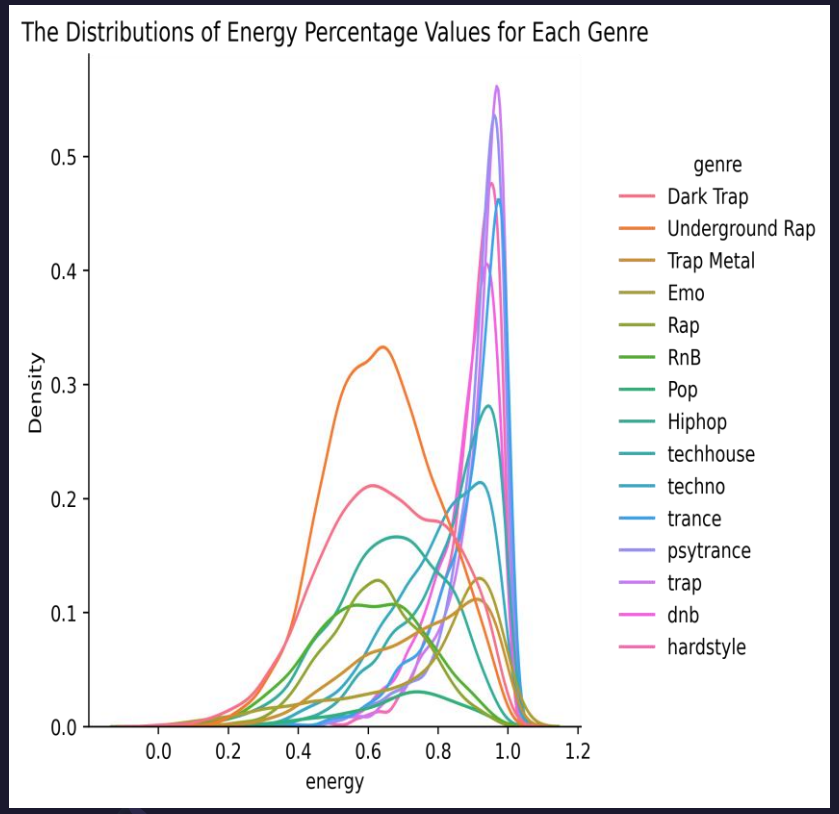
- From Kaggle
- Data Process
- Feature Validation
- Target Values

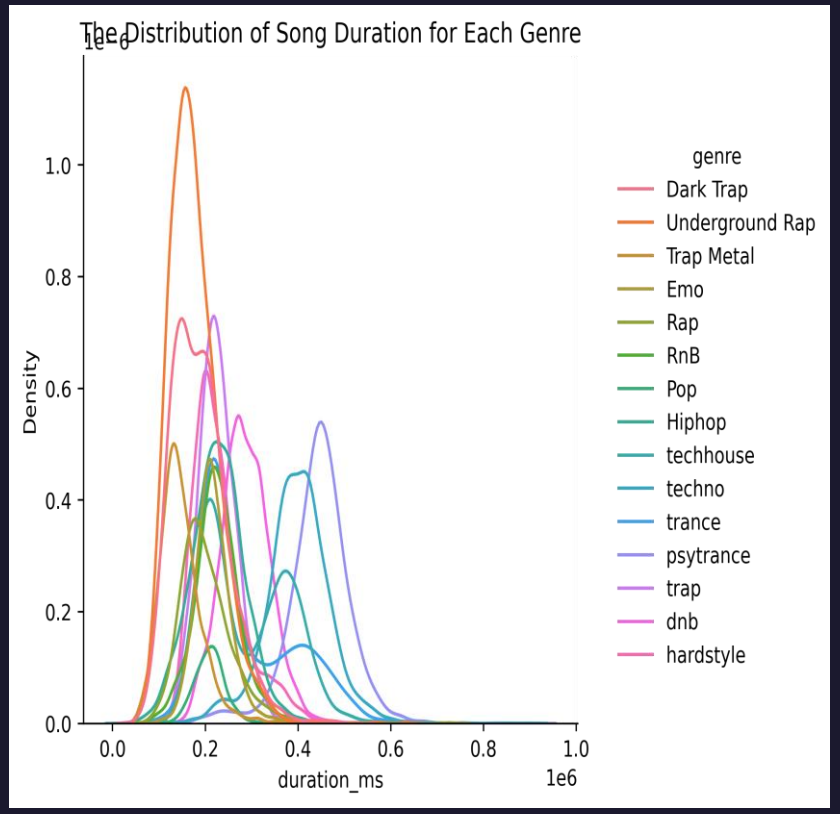
Charts

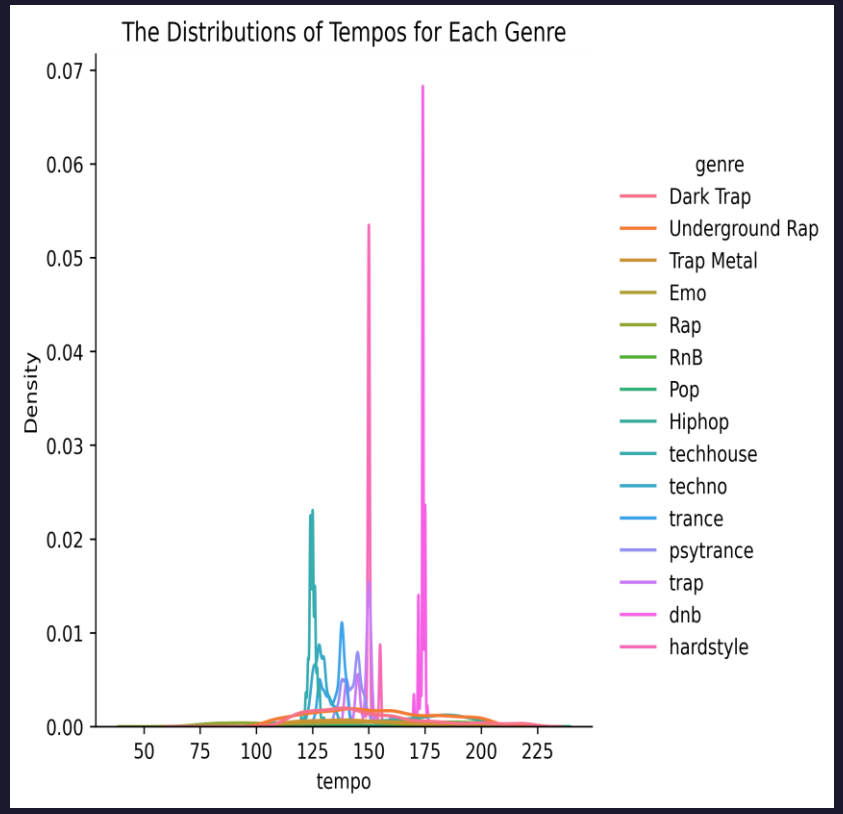

Number of Songs for Each Genre
Number of Songs
Genre
0
1000
2000
3000
4000
5000
6000
Underground Rap
Dark Trap
Hiphop
trance
trap
techhouse
dnb
psytrance
techno
hardstyle
RnB
Trap Metal
Rap
Emo
Pop


Distribution of the Number of Character in Each Song Name
Number of Songs
Character in Song Name
0
500
1000
1500
2000
2500
3000
0
20
40
60
80
100
120
140

Charts (contd.)

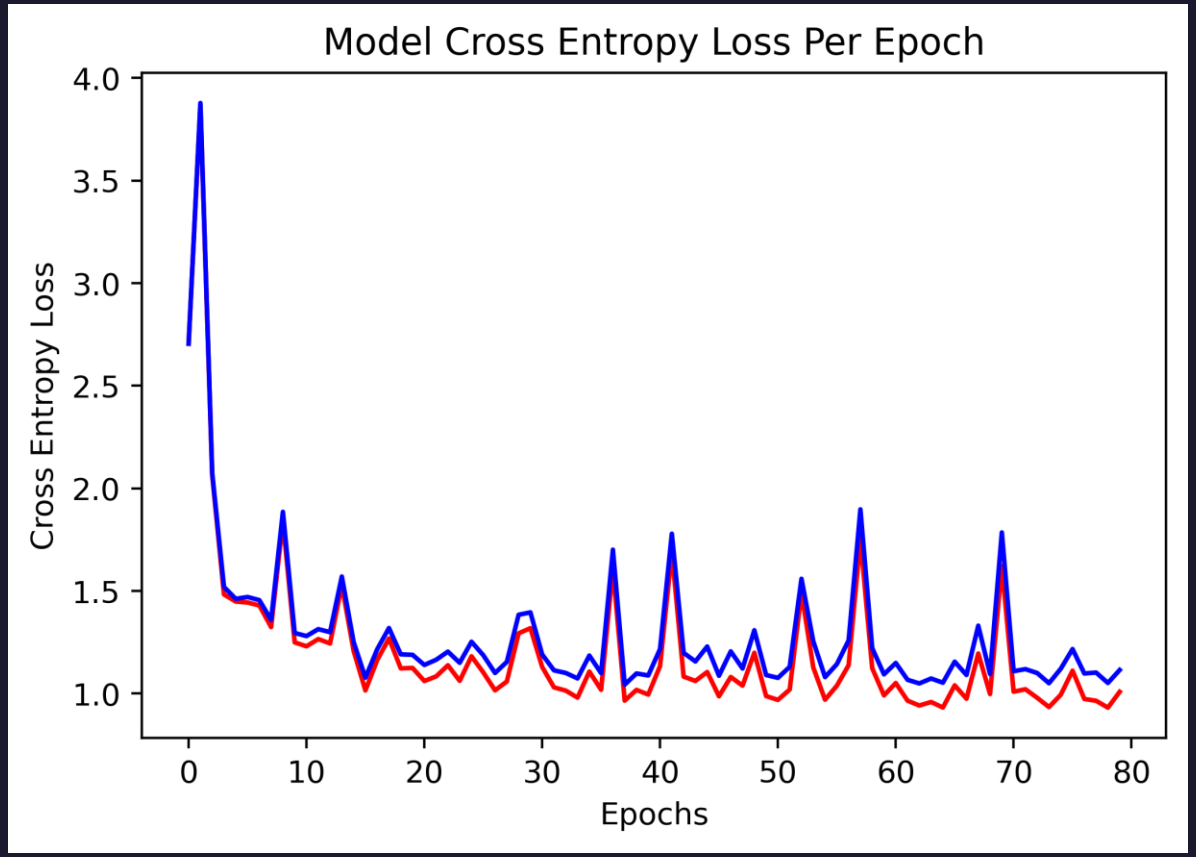




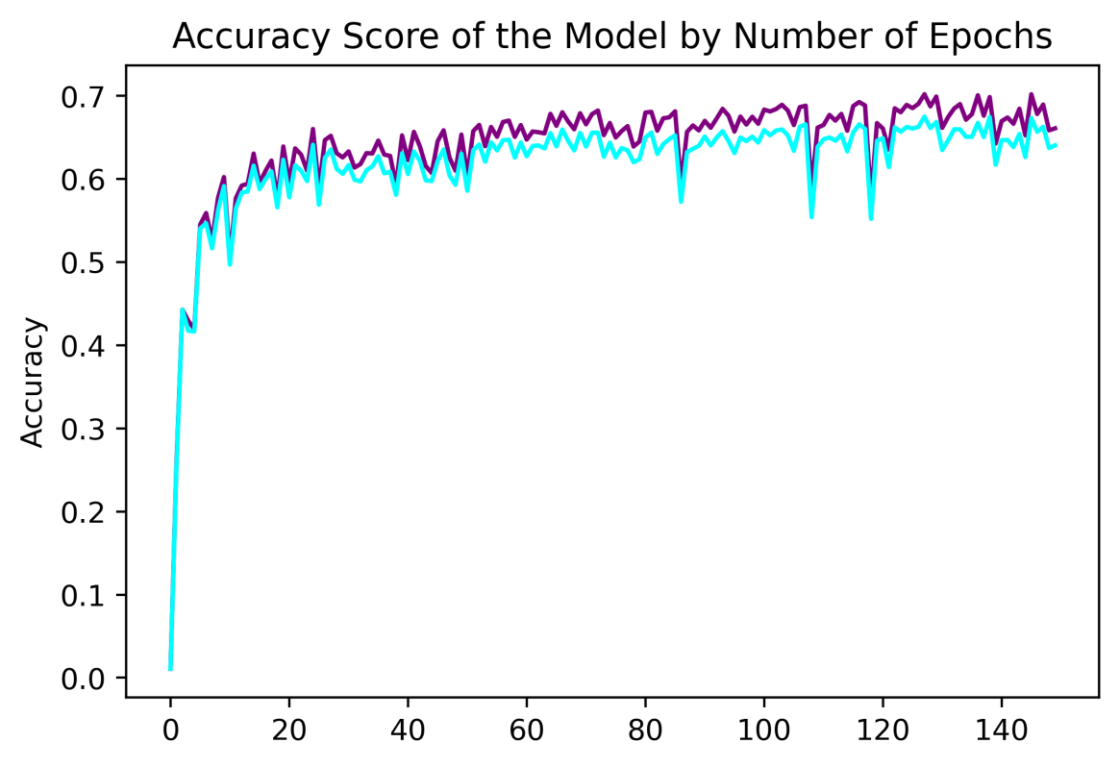


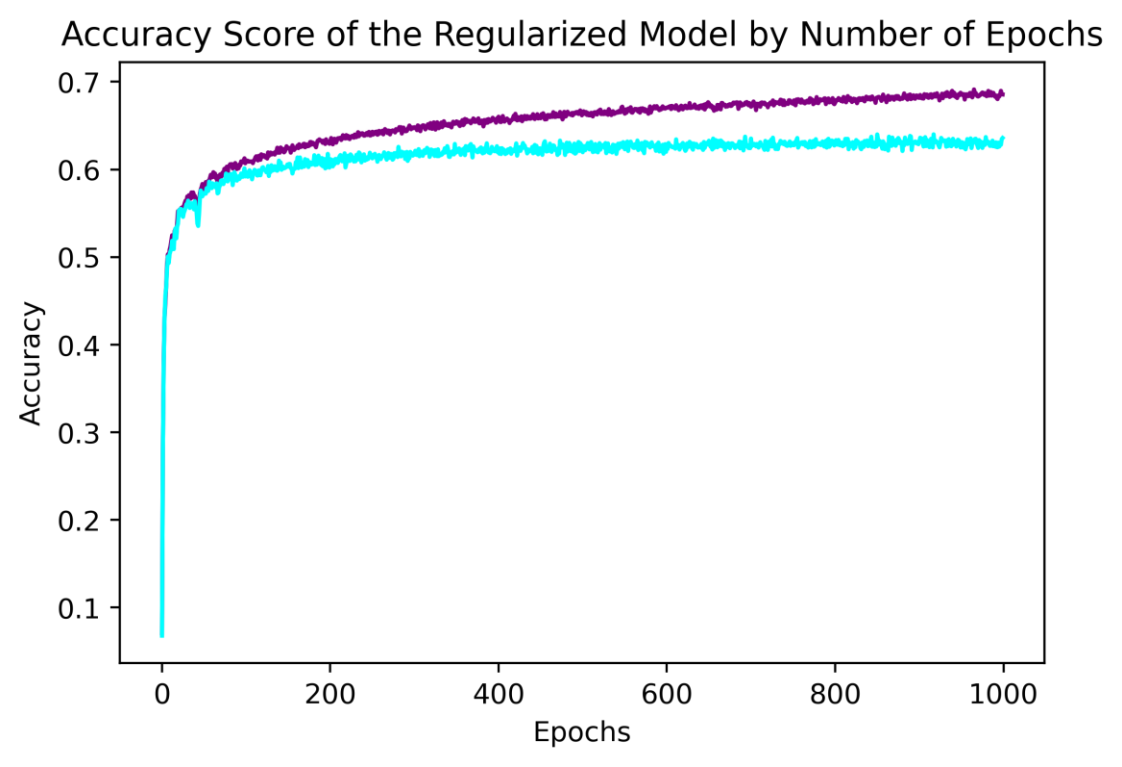
The Model:

- Multiclassification Neural Network (using PyTorch)
- Baseline Accuracy: ~6.59%
- Best Testing Accuracy of Model: 67.50%
- 6 Layers
  - 4 Hidden Layers
  - 1 Input Layer
  - 1 Output Layer



To Regularize?





Conclusion & Recommendations

- I created a neural network model to predict the genre of a song given attribute data about the song.
- This model without regularization is over ten-fold better than the baseline model.
- Utilizing Dropout Regularization did not improve the model in terms of accuracy, due to the data deficiency. It did improve the model by making the model less over fit and far less erratic. However, regardless of the model set up, I have never been able to achieve an accuracy score above 70%

Thank You

Max Bermont

mjbermont@gmail.com

https://github.com/mbermont/