BLUEPRINT OF THE MODEL **Date: 18 Oct 2021**

Using NLP Techniques to Predict Song Skips on Spotify based on Sequential User and Acoustic Data

The blueprint follows the following pattern:

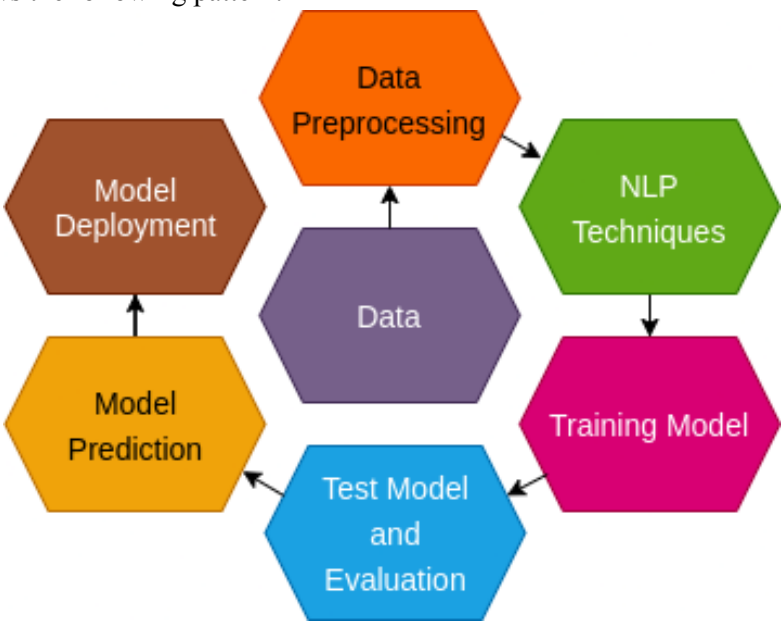

Fig: Dataflow Diagram

Data:	The dataset is consist of two files tf_mini.csv(50705 rows, 30 columns) and log_mini.csv(167881 rows, 21 columns).
	We collected the data from the GitHub repo.
	The File Format of the dataset is CSV.
Data Preprocessing:	In Data Cleaning, we deal with missing values and Outliers.
	In feature engineering, we perform categorical data encoding and merge user behaviour and acoustic features.
NLP Techniques:	Different NLP techniques that we perform are Tokenization, Word2vec, Bag of words and word Embedding for feature engineering.
	Data visualization is done to get an insight into the data.
Training Model:	80% of the dataset will be used for training purposes.
	We will train the model using GBT, LSTM, Bi-LSTM, and Transformers.
	Hyperparameter Tuning will be performed if needed.

Test Model and Evaluation:	20% of the dataset will be used for testing on the trained model.
Model Prediction:	Prediction will be done.
	The accuracy score will be checked using a classification report.
Model Deployment:	We will finally deploy the model using the python frameworks. We then check the functionality of ours model.