**A Twitter Music Recommendation Engine**

Sean Quigley

General Assembly: Fall 2013 Data science course

**Abstract**

This paper describes and evaluates a system that leverages text data from Twitter to determine artist similarities and make recommendations. Features are extracted through TF-IDF vectorization of artist tweets. These features are then mapped into a vector space and used to make artist recommendations using a nearest neighbors approach. The effectiveness of this unsupervised learning process is then evaluated qualitatively.

**1. Introduction**

As Internet use continues to grow, more and more product discoveries and purchases are occurring online. This has created a huge market for recommendation engines that use vast troves of web data to intelligently place the most relevant products and media in front of a specific user. Music has become a particular area of focus with companies like Pandora, Spotify, The Echo Nest, and Apple all developing recommender systems.

This paper overviews the development and evaluation of a simple music recommendation engine that uses Twitter data specifically. The final product is a tool that can take as input any Twitter handle and return a list of artists whose Twitter feeds contain text data is most similar to that of the input handle.

**2. Set up and Feature Extraction**

The first step in this process was developing a library of artist Tweet data. The Echo Nest API was used to get a list of the top 1000 artists and their corresponding Twitter handles. Top artists were determined with The Echo Nest’s “hotttnesss” metric which uses web mentions, music reviews, play counts, and other data points to measure how much “buzz” an artist is currently generating. Of these 1000 artists, 533 had active Twitter handles. The 200 most recent tweets for each artist were obtained with Tweepy, a Python wrapper library for the Twitter API. Artists with fewer than 10 Tweets were removed from the library in order to ensure a reasonable amount of text data in the feature space. 523 artists are included in the final recommendation engine.