# Machine Learning and Neural Networks CM3015 Mid-Term Coursework

**Introduction**

Source for the dataset: <https://www.kaggle.com/datasets/jarredpriester/taylor-swift-spotify-dataset/data?select=taylor_swift_spotify.csv>

Taylor Swift is undeniably one of the most successful and influential women in music today, as stated by *Forbes* magazine. (Dailey, 2023) Her 2023 *Eras* tour was the first ever music tour to gross over 1 billion dollars (Associated Press, 2023). According to *BusinessInsider,* it even had a profound impact on the US economy last summer (Gaines, 2023). The entire marketing premise behind Swift’s *Eras* tour was to visually and narratively promote a different “feel” for each of the singer’s albums, accompanied by a complete change of set and costume whenever she sang the songs from that particular record – and this tactic was incredibly successful and lucrative, as described above,.

The main aim of this project is to either confirm or disprove whether, from a *musical* perspective, it is indeed the case that each Taylor Swift album has a distinctive sound and character. The dataset used here (available on Kaggle, but originally compiled through the Spotify API) represents each Taylor Swift song as a row/sample. There are columns representing musical attributes/features such as “acousticness”, danceability, energy and loudness. Furthermore, each row/song also has an album attribute to designate which record it was on – these will constitute our “labels”. To achieve this aim, several classification algorithms, including K-Nearest Neighbour, Decision Trees and Naïve-Bayes Classification will be implemented in order to predict which album a song belongs to, based on these metrics. The errors produced by each classifier will be compared in order to assess the accuracy and recall of each classifier. Furthermore, k-means clustering with *k* being equal to the number of albums/classes included in the sample will be used to see if these albums are distinguishable based solely on their musical dimensions. Additionally, I will look at whether these musical features (e.g. energy) have any predictive power in determining the popularity of the song using a Linear Regressor and a Decision Tree.