

**Universidad Nacional de Colombia - sede Bogotá**

**Facultad de Ingeniería**

**Maestría en Ingeniería de Sistemas y Computación**

**Curso: Machine Learning 2020-I**

**Estudiantes:** Lina Fernanda Prado. Maestría matemática aplicada. Cod. 1032478936  
 Duvan Alberto Andrade Cuenca. Pregrado Ing. de Sist. y Comp.

1110583261   
 Javier Jiménez. Maestría en Estadística. Cod. 1018483021

**Propuesta proyecto ML**

**Introducción y Problema Principal:**

La recuperación de información musical (MIR) es un área interdisciplinar en la que diferentes ramas del conocimiento extraen información de la música y trabajan en herramientas de análisis de audio. Entre las aplicaciones encontramos clasificadores de género, reconocimiento de canciones, identificación de acordes, detección de eventos de sonido, detección de humor y extracción de características, entre otros **[1]**. Una de las aplicaciones de MIR es la transcripción automática de la música, procesar el audio y generar su respectiva notación simbólica, para lograr este objetivo es necesario llevar a cabo diferentes tareas de análisis de audio: identificación de instrumentos, estimación de duración, extracción de la armonía, de la melodía y del ritmo, detección de tonos, entre otros, es importante resaltar que entre mayor sea el número de instrumentos mayor es la complejidad de esta tarea.

En el presente documento se propone la construcción de un modelo de *Deep learning* de MIR para la transcripción automática de la notas musicales de un audio, es decir, un modelo que procesa el audio y genera la secuencia de notas correspondientes.

La intención de este proyecto es que más adelante haga parte de una aplicación móvil por medio de la cual sea posible que estudiantes, profesionales o personas que disfrutan mucho de una canción, y que desean interpretarla pero no tengan la partitura ni el entrenamiento musical para transcribirla, consigan una transcripción automática con solo ingresar el audio.

**Objetivo general:**

* Identificar la secuencias de notas presentes en un audio.

**Objetivos específicos:**

* + Generar una representación numérica del audio
  + Generar una representación visual del audio
  + Construir un modelo que procese la representación numérica
  + Construir un modelo que procese la representación visual
  + Combinar los modelos anteriores y crear una opción multimodal

**Metodología**

Se propone utilizar un conjunto de Datasets de archivos de audio de un (1) segundo de duración o menos, donde cada audio será una nota musical interpretada por un instrumento. Se tendrán en cuenta audios de todas las notas musicales: Do, Re, Mi, Fa, Sol, La, Si, (natural, bemol, y sostenido), interpretadas por los siguientes instrumentos: piano, flauta y violín.

Luego, utilizando embedding de audio, obtendremos una representación numérica y una visual, un vector numérico (por ejemplo con la Transformada de Fourier) y imagen (el espectrograma). Usaremos diferentes modelos, una CNN-MLP, una RNN-MLP y un modelo multimodal, para identificar en un audio corto la nota musical que se está interpretando. Se planea utilizar los Datasets FMA, NSynth, APL, etc..., (presentes en el gist de /alexanderlerch **[2]**).

Una vez el modelo sea capaz de identificar la nota musical de un fragmento de audio, lo aplicaremos para toda una canción (interpretada por un solo instrumento). Se dividirán las ondas de las señales en intervalos usando el Teorema del Muestreo y se identificará la nota para cada intervalo de sonido, para así conseguir la secuencia de notas de la canción. Con el conjunto de notas encontradas procederemos a usar algún software musical (por ejemplo un piano online) para tocar las notas y mostrar que tanto del sonido original se logró extraer.

**Bibliografía**

**[1]** Nasrullah, Z., & Zhao, Y. (2019, July). Music Artist Classification with Convolutional Recurrent Neural Networks. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

**[2]** [alexanderlerch](https://gist.github.com/alexanderlerch), 2019 Gist Music Datasets:

<https://gist.github.com/alexanderlerch/e3516bffc08ea77b429c419051ab793a>

**Referencias**

**[3]** Music information retrieval. (2020). En Wikipedia. Recuperado de: [<https://en.wikipedia.org/wiki/Music_information_retrieval>]