**FACULTAD DE CIENCIAS Y TECNOLOGÍA**

INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

**GENERACIÓN DE MÚSICA USANDO DCGAN**



Materia: Desarrollo de Aplicaciones Inteligentes (SIS330)

Estudiante: Oropeza Acosta Aylin Pola

C.U.: 111-310

Sucre – Bolivia

2021

ÍNDICE

[1. RESUMEN 3](#_Toc76735705)

[2. ANTECEDENTES 3](#_Toc76735706)

[3. PROBLEMÁTICA 3](#_Toc76735707)

[4. PROPUESTA 3](#_Toc76735708)

[5. OBJETIVO GENERAL 3](#_Toc76735709)

[6. OBJETIVOS ESPECÍFICOS 3](#_Toc76735710)

[7. FUNDAMENTOS TEÓRICOS 4](#_Toc76735711)

[8. DATASET 6](#_Toc76735712)

[9. MÉTRICA DE RENDIMIENTO (FRÉCHET INCEPTION DISTANCE - FID) 6](#_Toc76735713)

[10. DESCRIPCIÓN DEL TRABAJO REALIZADO EN EL MODELO (MATERIALES Y MÉTODOS) 7](#_Toc76735714)

[11. SOFTWARE DE APLICACIÓN DESARROLLADO (MÉTODOS Y MATERIALES) 8](#_Toc76735715)

[12. RESULTADOS FINALES 8](#_Toc76735716)

[13. CONCLUSIONES 9](#_Toc76735717)

[14. RECOMENDACIONES 9](#_Toc76735718)

[15. REFERENCIAS BIBLIOGRÁFICAS 9](#_Toc76735719)

# RESUMEN

Este proyecto busca demostrar los alcances de la Inteligencia Artificial aplicándola a un área artística, la música. Para ello se usa la red DCGAN un modelo de las redes generativas adversarias que recibe como entrada datos pre procesados de un conjunto de audios en formato midi y devuelve su composición en datos numéricos que son transformados a audio en un post procesamiento. La música de entrada y salida son piezas para piano, al ser uno de los instrumentos que maneja la mayor cantidad de información musical.

# ANTECEDENTES

Las matemáticas siempre estuvieron presentes en la música, dando explicación a los sistemas de afinación, ayudando a comprender el sonido. Con la aparición de las computadoras se comenzó a utilizarlas como instrumentos, brindando herramientas que revolucionaron la música, creando nuevos géneros como la música electrónica, el lo-fi y muchos más, estos géneros pasaron a tener una riqueza tímbrica a diferencia de la música acústica, cuyo fuerte es la armonía.

Los algoritmos estuvieron para facilitar el trabajo de composición de los músicos, llegando a ser utilizados también como instrumentos al poder crear inimaginables sonidos que pueden ser utilizados en la música. Pero ahora la Inteligencia Artificial es capaz de componer por su cuenta y no ser solamente una herramienta.

# PROBLEMÁTICA

Se cree que composición está limitada únicamente a los humanos por su contenido emotivo, pero con composiciones de la IA que logren engañar a los melómanos, ésta se encargaría de la industrialización de la música, mientras los artistas pueden dedicarse más a experimentar y jugar con antiguos y nuevos elementos musicales.

# PROPUESTA

Existen varios algoritmos capaces de componer música, se propone utilizar uno para obtener una pequeña obra musical.

# OBJETIVO GENERAL

Utilizar la Inteligencia Artificial para generar audios con forma musical, con una precisión mayor al 90% como evaluación cuantitativa y aceptación de los oyentes como evaluación cualitativa.

# OBJETIVOS ESPECÍFICOS

* Identificar elementos matemáticos, digitales del estado del arte.
* Constituir un dataset de pistas de audio.
* Implementar y ajustar el modelo inteligente.
* Validar el modelo.

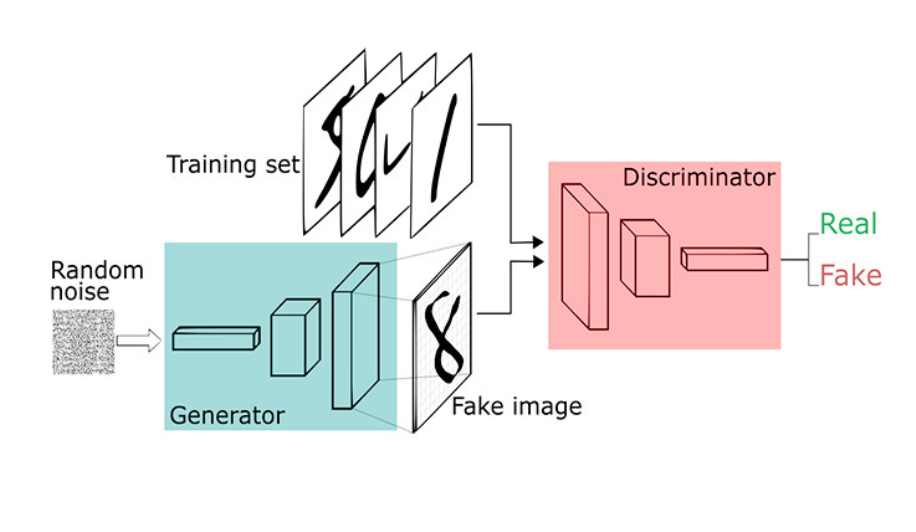
# FUNDAMENTOS TEÓRICOS

## ÁMBITO DE LA INTELIGENCIA ARTIFICIAL

Las redes GAN (Redes generativas adversas) consisten en enfrentar dos redes neuronales que compiten en un constante juego de suma cero (si una gana, la otra pierde).

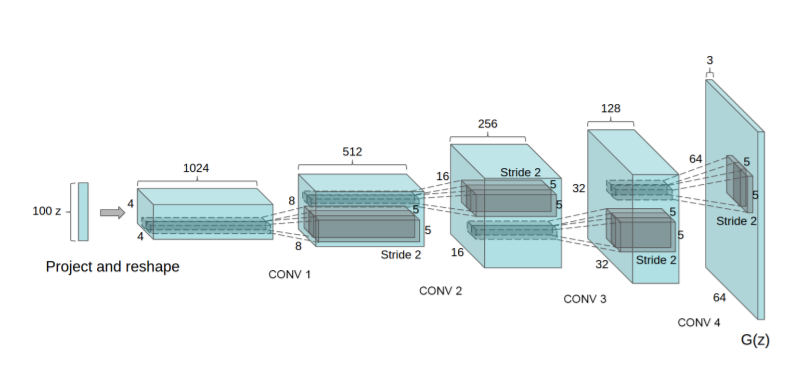
Una es la red generativa, que produce muestras de los que queremos crear (imágenes, textos, sonidos), al comienzo genera ruido y luego va mejorando en cada intento. Mientras la segunda red, la discriminadora, entrenada en la identificación analizará el material producido por la red generativa y determina si se ajusta a lo que buscamos. Entonces, la red generativa trata de hacer pasar sus creaciones como reales, mientras la red discriminadora no deja pasar contenido que considera falso.

Así, puede haber cientos, miles o millones de intentos antes de que la red discriminadora acepte el resultado ofrecido por su rival. Mientras tanto, rechazo tras rechazo, la red generativa habrá ido 'aprendiendo' qué es lo que busca la discriminadora, quien a su vez le habrá ido guiando con la información que aportaba con sus porcentajes de acierto.



En este proyecto se utiliza la arquitectura DCGAN (Deep Convolutional Generative Adversarial Network), que consiste en una representación del aprendizaje no supervisado. Esta arquitectura tiene algunos cambios en las arquitecturas CNN:

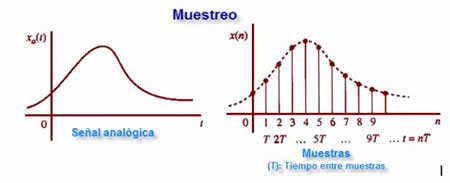
1. La primera red del Generador es totalmente convolucional, reemplazando funciones de agrupación como el maxpooling con convoluciones escalonadas.
2. Se eliminan capas completamente conectadas, la GAN recibe el ruido y devuelve un tensor de 4 dimensiones.



1. Normalización por lotes, se utiliza batch normalization a todas las capas, menos a la capa de salida del Generador y la capa de entrada del Discriminador para mantener la estabilidad de la arquitectura. El Generador la función de activación ReLU excepto en la capa de salida, la cual utiliza la función Tanh y el Discriminador utiliza la función de activación LeakyReLU en todas las capas.

## ÁMBITO AL SE APLICA LA INTELIGENCIA ARTIFICIAL

La música es un conjunto de sonidos en una conexión directa con las emociones humanas. Las matemáticas y la física facilitan el entendimiento de cada sonido y han permitido tener una representación digital de estas señales analógicas tomando millones de muestras (una grabación de alta calidad consiste en 44 000 muestras por segundo), esta señal digital consiste en matrices de números.



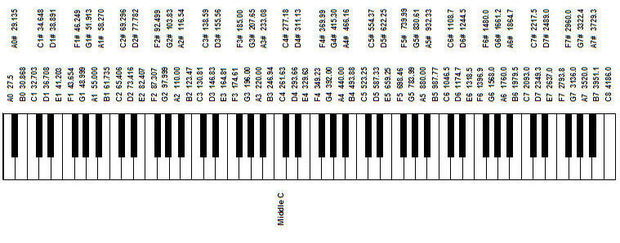
Los archivos MIDI son los que contienen la mayor cantidad de información musical.

Las siglas MIDI corresponden a la abreviatura anglosajona de “Musical Instrument Digital Interface”, una breve explicación de este protocolo creado para la comunicación y mezcla de señales que provienen de varios instrumentos y dispositivos​ de creación musical. Estos podrán enviarse a una computadora digital, concretamente a un software que se encargará de la producción musical de nuestra obra y se obtiene de la grabación de interpretaciones en instrumentos netamente digitales o desde un software simulador, de esta forma los archivos midi contienen información como el nombre de las notas, tono, frecuencia, tempo, velocidad, instrumento, etc

El resultado final de todo sonido que pasa por el midi es un archivo de gran calidad y de un tamaño no demasiado grande, que además permite modificaciones y ajustes.

# DATASET

Se tienen 331 obras para piano de compositores clásicos como Bach, Mozart y Chopin, cuidando que no sean obras demasiado densas musicalmente y evitando las disonancias (notas que no pertenecen a la misma familia, consiguiendo un sonido menos agradable). Con la librería music21 se procesará los archivos midi del dataset y se tendrá la información de cada nota por su nombre, por ejemplo C#5, lo que corresponde a la notación americana de la nota do, el “#” indica que es sostenida (un semitono más arriba) o “-“ para bemol (un semitono más abajo) y el 5 es la octava en la que está ubicada la nota. Una descripción gráfica se encuentra en la figura:



Los acordes son la ejecución de dos o más notas simultáneamente, se accederá a éstos con la concatenación de las notas que lo componen.

Una vez obtenidas las notas y acordes de las pistas se las pasa a un diccionario asignándoles un número.

# MÉTRICA DE RENDIMIENTO (FRÉCHET INCEPTION DISTANCE - FID)

Fréchet Inception Score es una de las mejores métricas aplicable a las redes GAN, que nació para suplir las carencias de su antecesora Inception Score al comparar las estadísticas de las muestras reales y las muestras generadas, en lugar de sólo evaluar las muestras generadas. Desde la propuesta de Heusel, Ramsauer, Unterthiner, Nessler y Hochreiter el año 2017 se usa la distancia Fréchet entre dos matrices gaussianas multivariables.

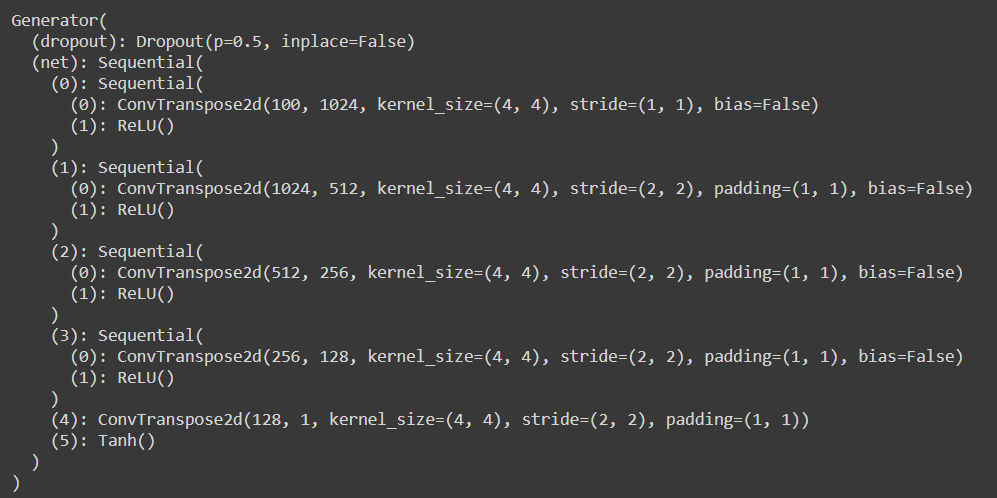
FID=||μr−μg||2+Tr(Σr+Σg−2(ΣrΣg)1/2)

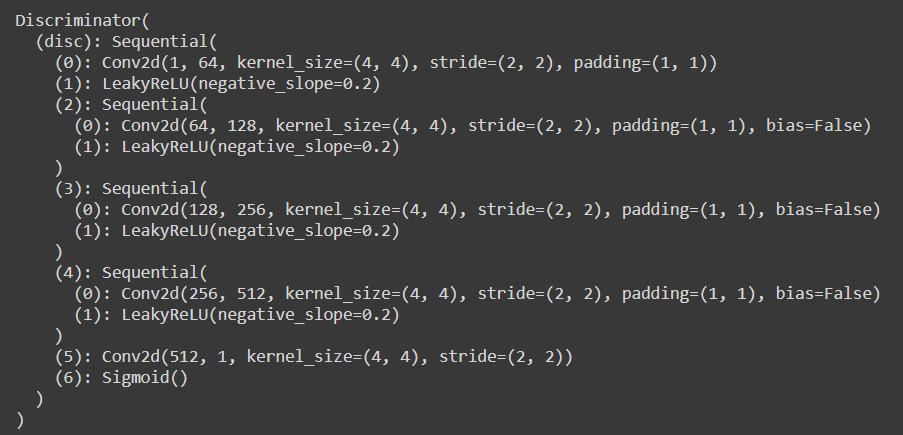
Donde y son las activaciones de la capa pool3 del modelo pre entrenado Inception-v3 para muestras reales y generadas, respectivamente.

# DESCRIPCIÓN DEL TRABAJO REALIZADO EN EL MODELO (MATERIALES Y MÉTODOS)

Google Colab es la herramienta utilizada para la creación, entrenamiento y pruebas del modelo siguiendo la arquitectura DCGAN. El software ha sido desarrollado con PyTorch, para el preprocesamiento de archivos midi se ha usado la librería music21 y el modelo pre entrenado Inception-v3 para las métricas.

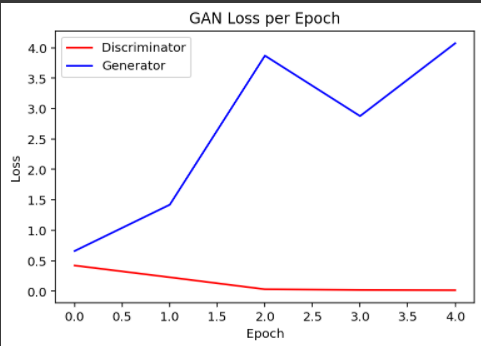
La arquitectura de las redes generadora y discriminadora, se pueden apreciar respectivamente en las imágenes.





Para el entrenamiento del modelo se utilizaron los siguientes hiperparámetros un learning rate de 0.0002 para el generador y el discriminador, batchsize de 128, la dimensión de entrada es 64, al ser este el valor con el que trabaja la DCGAN y 5 epochs, siendo éste el valor con el que se evitan fluctuaciones.

Con estos valores en el entrenamiento se obtuvo la gráfica de los valores de la función de pérdida:



# SOFTWARE DE APLICACIÓN DESARROLLADO (MÉTODOS Y MATERIALES)

El modelo puede ser usado desde una página web desarrollada con el framework minimalista para Python, Flask.

Este micro framework puede ser aplicado en cualquier patrón de diseño, en este proyecto se aplicará el Modelo Vista Controlador (MVC), por su manejo de rutas.

La exportación e importación del modelo entrenado se realizó con “traicing”, pero los valores que se obtienen son netamente numéricos, necesitan una post procesamiento, para ello se exportó el diccionario generado en el preprocesamiento transformándolo en un archivo JSON para consumirlo desde la página web, permitiendo la conversión a audio.

# RESULTADOS FINALES

El audio en formato midi obtenido consiste en una melodía, que dependiendo los intentos puede ser más o menos agradable al oído.

# CONCLUSIONES

La Inteligencia Artificial crece cada vez más, nos ayuda en tareas que años atrás solo se veía en la ficción e incluso, sin darnos cuenta, se vuelve parte de nuestro día a día. Se tienen proyectos científicos y sociales muy lindos en los cuales incursionar, de manera que se vuelve una experiencia enriquecedora y se tiene un acercamiento a esta área que cada vez abarca más.

# RECOMENDACIONES

Para poder llevar el proyecto es necesario trabajo constante, al profundizar en algo que aparentemente ya conocíamos surgen dudas que se habían pasado por alto y es en este punto donde se llega a un aparente estancamiento, donde se investiga mucho y en una de esas lecturas se encuentra la respuesta, mejor aún con una guía. Entonces se avanza hasta dar por concluido el proyecto, porque siempre se encuentran cosas para mejorar, como en este proyecto, aún hay muchas cosas que se pueden ajustar y mejorar, como incluir acordes en el post procesamiento para no tener sólo una melodía, incluir la duración de cada nota o la intensidad de la misma ,se podrían tomar en cuenta más instrumentos, aplicar una clasificación de temas en tonalidades mayores y menores en la generación, así se puede continuar indefinidamente hasta quedar satisfechos con los resultados.

# REFERENCIAS BIBLIOGRÁFICAS

* Radford, A., Metz, L., & Chintala, S. (2015). UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS. *Cornell University*, *2*, 1–4. https://arxiv.org/abs/1511.06434
* Yean, N. (2018, 15 julio). *Fréchet Inception Distance*. Fréchet Inception Distance. https://nealjean.com/ml/frechet-inception-distance/
* Ahmed, I. (2020, septiembre). *GAN in Pytorch with FID*. Kagle. https://www.kaggle.com/ibtesama/gan-in-pytorch-with-fid