

**Projeto de Pesquisa de Iniciação Científica**  
**Insper - Engenharia da Computação**

**Aluno:** Vinícius Grando Eller

**Orientador:** Fábio José Ayres

**Título:** Construção de modelo de geração de música com Deep Learning

**Palavras Chave:** *deep learning, autoencoders, music generation, MIDI*

## 1. Descrição do Problema e Revisão de Literatura

A humanidade sempre desenvolveu parte de suas tecnologias por meio de inspirações que podiam ser observadas na natureza. Em busca de melhoras nos métodos de uso de inteligência artificial, uma análise sobre o funcionamento cerebral inspirou a criação de um novo modelo de *Machine Learning*, o qual foi chamado de redes neurais artificiais (*artificial neural networks* - ANNs) (Géron, 2019). As ANNs são o coração de modelos de redes neurais profundas (*Deep Learning*), que hoje são utilizados com sucesso em diversas tarefas complexas, como análise de imagens, entendimento de texto, e agentes autônomos (Bengio, Goodfellow, & Courville, 2015).

A análise e geração de músicas usando técnicas de *Deep Learning* é uma área de crescente proeminência. Busca-se, pela aplicação de redes neurais profundas, entender e reproduzir melodias, escalas harmônicas e as associações entre a parte falada e tocada de uma música (Briot, Pachet, & Hadjeres, 2019). Há grupos de pesquisa como Magenta (Eck, s.d.) e MuseNet (Payne, 2019) que já possuem modelos avançados nessa área; no caso do primeiro, a geração não é apenas sonora, mas também visual, pois o modelo consegue criar quadros artísticos. Para o segundo caso, tem-se a capacidade de gerar composições com até quatro minutos de duração, sendo que podem ser tocados com 10 tipos de instrumentos diferentes.

Existem vários modelos de geração de música com redes neurais, sem a dominância clara de um modelo em relação a outros no estado atual da arte (Carnovalini & Rodà, 2020). Por exemplo: *Convolutional Neural Networks* (CNNs), que são bastante populares para a análise e geração de imagens, podem ser aplicadas à geração de músicas. Apresentam, contudo, uma limitação: como baseiam-se na relação entre pixels próximos espacialmente, o modelo erra ao analisar músicas com múltiplas escalas e grande variedade de notas (Briot, Pachet, & Hadjeres, 2019). Um

outro modelo bastante utilizado é a arquitetura de rede neural conhecida como *Long Short-Term Memory* (LSTM); esta consegue criar composições lógicas musicalmente, apresentando melhor desempenho que uma CNN, mas suas criações são breves e pouco complexas (Huang & Wu, 2016).

Escolhido o modelo de síntese de música, deve-se construir uma tarefa de aprendizado de máquina que permita o ajuste dos parâmetros do modelo. Estudos apontam uma melhor performance no uso de *Autoencoders* ou de *Generative Adversarial Networks* (Yang, Chou, & Yang, 2018). São estratégias de modelagem conhecidas como semi-supervisionadas, onde constrói-se uma tarefa de aprendizado supervisionado a partir de dados não-rotulados. Modelos com essas arquiteturas e estratégias de aprendizado tem tido destaque na literatura científica e na mídia recentemente, pois são capazes de criar os chamados *DeepFakes* (Sample, 2020): podem gerar fotos de pessoas e ambientes, bem como falas e músicas que não existem (Wang, 2019).

## 2. Objetivos

Este projeto tem como objetivo realizar uma análise do funcionamento de modelos *autoencoders*, conjuntamente de um entendimento de representações de dados musicais (e.g. formato MIDI), com o objetivo de criar um modelo funcional que gere músicas artificiais similares a composições convencionais.

Objetivos específicos do projeto são:

- Entender funcionamento do sistema MIDI de representação, transmissão e armazenamento de composições musicais;
- Compreender o funcionamento de modelos *autoencoders* para a geração de músicas;
- Criar modelo funcional de geração de músicas artificiais baseado em *autoencoders*;
- Avaliação empírica de aspectos qualitativos das músicas geradas, através de formulários de avaliação a serem distribuídos a participantes do estudo.

Objetivo avançado do projeto:

- Construir uma interface de usuário que permita o uso, por músicos, do modelo desenvolvido.

### 3. Metodologia

O projeto seguirá os seguintes passos:

- a) Estudo de representações digitais de composições musicais, em particular o sistema MIDI;
- b) Estudo de elementos de teoria musical, e sua relação com a representação de composições: compasso, notas, tempo, escalas (Harkleroad, 2006);
- c) Estudo de redes neurais profundas e *autoencoders* (Alain & Bengio, 2014);
- d) Criação de um modelo de geração de músicas a partir de um conjunto de exemplos no formato MIDI;
- e) Avaliação qualitativa das músicas geradas, e estudo da relação destas qualidades com os parâmetros e arquitetura do *autoencoder*.

### 4. Resultados Esperados

- Um modelo treinado de geração de músicas;
- Uma análise da qualidade das músicas geradas e sua relação com parâmetros do modelo;
- Um estudo do estado da arte em geração de composições musicais com redes neurais;

### 5. Referências Bibliográficas

Alain, G., & Bengio, Y. (19 de Aug de 2014). What Regularized Auto-Encoders Learn from the Data. *What Regularized Auto-Encoders Learn from the Data*, pp. 2-7.

Bengio, Y., Goodfellow, I., & Courville, A. (2015). Deep Learning. Em Y. Bengio, I. Goodfellow, & A. Courville, *Deep Learning* (pp. 1-25).

Briot, J. P., Pachet, F. D., & Hadjeres, G. (2019). Deep Learning Techniques for Music Generation – A Survey. Em J. P. Briot, F. D. Pachet, & G. Hadjeres, *Deep Learning Techniques for Music Generation – A Survey* (pp. 19, 198).

Carnovalini, F., & Rodà, A. (03 de April de 2020). Computational Creativity and Music Generation Systems. *Computational Creativity and Music Generation Systems: An Introduction to the State of the Art*, pp. 7-13.

Eck, D. (s.d.). *Magenta*. Fonte: Magenta: <https://magenta.tensorflow.org/get-started>

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*.

Huang, A., & Wu, R. (15 de jun de 2016). Deep Learning for Music. *Deep Learning for Music*, pp. 1-9.

Payne, C. (25 de Apr de 2019). *MuseNet*. Fonte: OpenAI: [openai.com/blog/musenet](https://openai.com/blog/musenet)

Sample, I. (13 de Jan de 2020). *What are deepfakes - and how can you spot them?*

Fonte: The Guardian:

<https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>

Wang, P. (Dec de 2019). *This Person Does Not Exist*. Fonte: This Person Does Not Exist: <https://thispersondoesnotexist.com/>

Yang, L.-C., Chou, S.-Y., & Yang, Y.-H. (18 de jul de 2018). Midnet. *Midnet: A convolutional Generative Adversarial Network for Symbolic-domain Music Generation*.

## 6. Cronograma de Atividades

Atividades	2021					2022						
	ago	set	out	nov	dez	jan	fev	mar	abr	mai	jun	jul
Estudo de representações digitais da musica	■	■	■									
Estudo de elementos da teoria musical			■	■	■							
Estudo de autoencoders			■	■	■	■						
Criação de modelos				■	■	■	■	■	■			
Analise dos Resultados								■	■	■	■	
Atividades de escrita de relatório			■			■				■	■	■