

# Relatório AutoML - AutoKeras

Rafael Romualdo e Marco Túlio

## O que é

AutoKeras é uma ferramenta de aprendizado AutoML que simplifica o processo de criação de modelos de aprendizado, deixando todo o processo, desde o pré-processamento até a escolha do modelo mais automatizado. Ele foi desenvolvido open-source e integrado ao Keras, o AutoKeras busca facilitar a utilização de aprendizado de máquina para usuários sem conhecimentos em machine learning.

## Funcionamento

O AutoKeras atua na seleção de arquitetura, ajuste de hiperparâmetros e pré-processamento de dados. Ele utiliza técnicas de busca de arquitetura neural (NAS - Neural Architecture Search) para encontrar a melhor estrutura de rede para um dado conjunto de dados e tarefas.

## Resumo

**Modelos:** A ferramenta tem modelos para tarefas comuns como classificação de imagens, regressão, classificação de texto e previsão de séries temporais.

**Busca de Arquitetura Neural:** O AutoKeras utiliza métodos avançados de otimização para explorar diferentes arquiteturas de redes neurais e encontrar a mais adequada, e faz a otimização automática dos hiperparâmetros do modelo.

**Treinamento e Avaliação:** Ele também automatiza o processo de treinamento e avaliação do modelo e usa técnicas de validação cruzada.

## Etapas de Pré-processamento

Ao fazer o input dos dados, ele consegue suportar dados em csv, txt, imagens e dados temporais.

Ele consegue transformar variáveis categóricas em numéricas, que são adequadas para redes neurais. Além disso, escala os dados numéricos para que estejam dentro de uma faixa binária).

O Auto Keras também tokeniza as imagens para uso de linguagem natural(LLM), transformando textos em sequências de tokens.

## **Processo de Busca e Otimização**

- A única coisa que o usuário precisa definir é a definição do problema, ou seja, qual tipo de machine learning vai ser usado no problema, que são entre:
  - Classificação de imagem
  - Regressão de imagem
  - Classificação de texto
  - Regressão de texto
  - Classificação tabular
  - Regressão tabular
  - Classificação de séries temporais
- O AutoKeras faz uma busca em hiperparâmetros e arquiteturas de rede neural para encontrar o melhor modelo para o problema, utilizando técnicas de busca como Random Search, Bayesian Optimization, e Hyperband para encontrar a combinação de hiperparâmetros e arquitetura. Porém, o usuário também consegue configurar alguns parâmetros, entre eles:
  - Epochs: Número de épocas pra treinar o modelo.
  - Batch size: Tamanho do lote de dados pra cada iteração de treinamento.
  - Tuning: Número de tentativas de ajuste de hiperparâmetros que o AutoKeras vai fazer..
  - Max trials: Número máximo de experimentos pra encontrar o melhor modelo.

## **Avaliação**

- O AutoKeras avalia o desempenho dos modelos gerados usando técnicas de validação cruzada ou separando um conjunto de validação dos dados de treinamento.
- Ele seleciona o modelo com o melhor desempenho e o retorna ao usuário.

## **Implementação**

- O modelo final pode ser exportado pra ser usado em escala.
- Pode-se salvar e carregar modelos treinados para reutilização.

## **VANTAGENS**

- Tem integração fácil com outras ferramentas do TensorFlow.
- Pode ser automatizado do início ao fim de todo o processo
- Oferece integração com Deep Learning (algo que não é oferecido por poucas ferramentas).
- Por ser open-source, ele sempre está se atualizando e recebendo melhorias.
- Ele pode ser personalizado para cada usuário ao automatizar um processo.

## **DESVANTAGENS**

- Apesar de ser fácil de gerar um modelo com ele, pode ser difícil de aprender com ele, por ser difícil ter conhecimento de todas as suas funcionalidades.
- Ele exige muito esforço computacional, precisando de GPUs largas para aguentar ele.

O modelo foi testado no Google Colab, disponível no link abaixo:

<https://github.com/mtulioddev/autokeras>