



# Projeto Final - Serviços de Software para Inteligência Artificial

## Nome dos integrante

Gabriel Matheus da Silva Santos, 20.84144-2

Fernanda Tanajura Piva, 20.83990-0

Rodrigo Franciozi R. da Silva, 20.83984-7

Nome dos integrante

1.0 - Introdução

2.0 - Dados

3.0 - Resultados

## 1.0 - Introdução

O intuito desse projeto é aplicar um modelo de machine learning para predição que diz qual seria a idade de uma pessoa, baseada em características de seu animal de estimação (no caso o animal em estudo é o cachorro). O processo foi realizado com o auxilio do AutoML

disponibilizado pela plataforma Azure e um Front foi desenvolvido para o consumo dos resultados do modelo.

## 2.0 - Dados

Os dados utilizados para o processo foram retirados da plataforma Kaggle e se encontram no seguinte link:

Dogs of Zurich  
Data about Dog Owners in Zurich, Switzerland  
 [https://www.kaggle.com/kmader/dogs-of-zurich?select=20160307\\_hundehalter.csv](https://www.kaggle.com/kmader/dogs-of-zurich?select=20160307_hundehalter.csv)



Esse conjuntos de dados, contém informações sobre cachorros de Zurich (na Suíça) que tem como uma de suas linguagens oficiais o Alemão, sendo assim os dados se encontram todos nessa lingua. Uma tradução de algumas colunas da tabela sera descrito abaixo:

1. ALTER -> Idade do indivíduo (Humano)
2. GESCHLECHT -> Gênero do indivíduo (Humano)
3. STADTKREIS -> Bairro ou distrito da cidade
4. RASSE1 -> Raça primaria do Cachorro
5. RASSE2 -> Raça secundaria do Cachorro
6. GEBURTSJAHR\_HUND -> Data de nascimento do Cachorro
7. GESCHLECHT\_HUND -> Gênero do Cachorro
8. HUNDEFARBE -> Cor do Cachorro

O conjunto de dados, possui informações entre os anos de 2015 e 2017.

### 2.1 - Análise Exploratória

Temos 3 tabelas como a seguinte quantidade de registros em cada uma delas:

Ano 2015: 6980

Ano 2016: 6930

Ano 2017: 7155

Juntamos ambas as 3, fazendo com que a tabela principal possua 21065 registros, sendo que 3 colunas tem vários valores nulos. Como isso pode atrapalhar na predição e criação do modelo, as mesmas foram desconsideradas antes do processo de treinamento.

Decidimos plotar 2 gráficos simples para verificar a distribuição de alguns dados. O primeiro deles ilustrado na figura 1, mostra a distribuição dos registros de acordo com a faixa de idade da tabela, podendo assim perceber que estaremos trabalhando com tipo de classificação Multi-classes. Já a figura 2, mostra somente a distribuição de registros em relação ao gênero (humano).

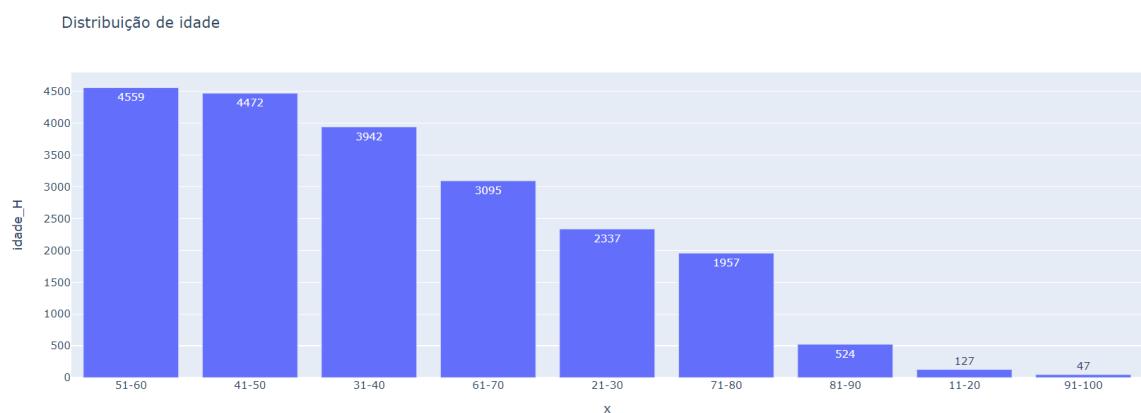


Figura 1 - Distribuição de idade dos dados

Distribuição dos dados por gênero

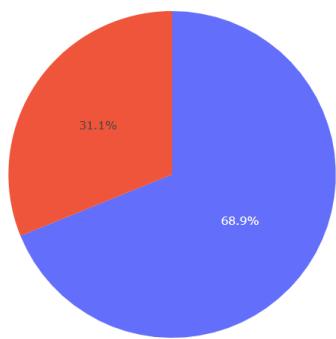


Figura 2 - Distribuição de dados por sexo (Masculino ou Feminino) do dono. 6% dos registros são homens, e apenas 31% são mulheres.

## 2.2 - Processo de treinamento

O modelo foi treinado com o auxilio do AutoML como mencionado anteriormente e os dados foram separados em 70% para treino e 30% para teste.

## 3.0 - Resultados

O melhor resultado alcançado foi de 58%. Caso mais dados tivessem sido usados, e um processo de tratamento ou normalização tivesse sido realizado, poderíamos ter obtido uma melhora (singela ou não) no resultado final do modelo.

Um front foi criado para que o consumo desses resultados fosse mais palatável para o usuário final, e o mesmo se encontra no link do github a seguir:

[https://github.com/FeTPiva/the\\_last](https://github.com/FeTPiva/the_last)