

Técnicas Tradicionais de Classificação de Imagens

Classificação: Regressão
Logística e Optuna



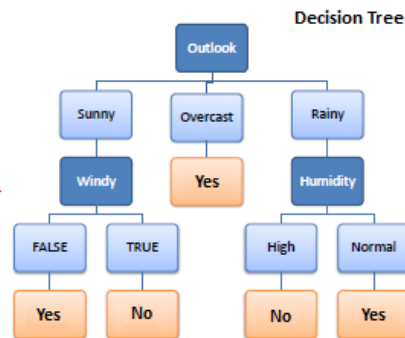
Recapitulação

Árvores de Decisão

Classificação

Árvore de Decisão

Predictors				Target
Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



1. Calcula-se a entropia para classe (0 = dados homogêneos; 1 = dados igualmente distribuídos).

Play Golf	
Yes	No
9	5

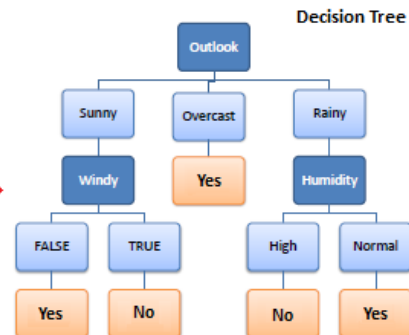
$$\begin{aligned}
 \text{Entropy(PlayGolf)} &= \text{Entropy}(5,9) \\
 &= \text{Entropy}(0.36, 0.64) \\
 &= -(0.36 \log_2 0.36) - (0.64 \log_2 0.64) \\
 &= 0.94
 \end{aligned}$$

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Classificação

Árvore de Decisão

Predictors				Target
Outlook	Temp.	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



R_1 : IF (Outlook=Sunny) AND (Windy=FALSE) THEN Play=Yes

R_2 : IF (Outlook=Sunny) AND (Windy=TRUE) THEN Play=No

R_3 : IF (Outlook=Overcast) THEN Play=Yes

R_4 : IF (Outlook=Rainy) AND (Humidity=High) THEN Play=No

R_5 : IF (Outlook=Rain) AND (Humidity=Normal) THEN Play=Yes

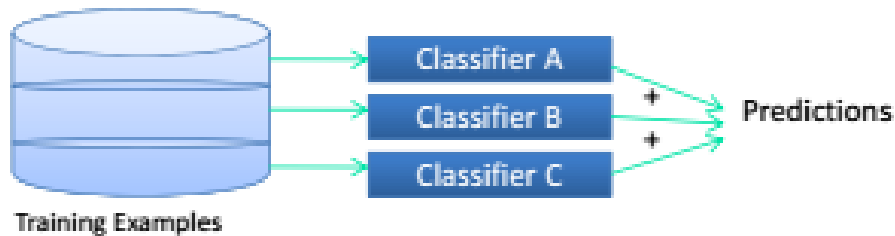
Regras de decisão

Comitê

Classificação

Comitê

- Agregar múltiplos modelos treinados com o objetivo de melhorar a desempenho do modelo conjunto.
- Intuição: simula o que fazemos quando combinamos conhecimento de especialistas em um processo de tomada de decisão.



Comitê: Técnicas

Classificação

Comitê - Técnicas

- Bagging vs. Boosting

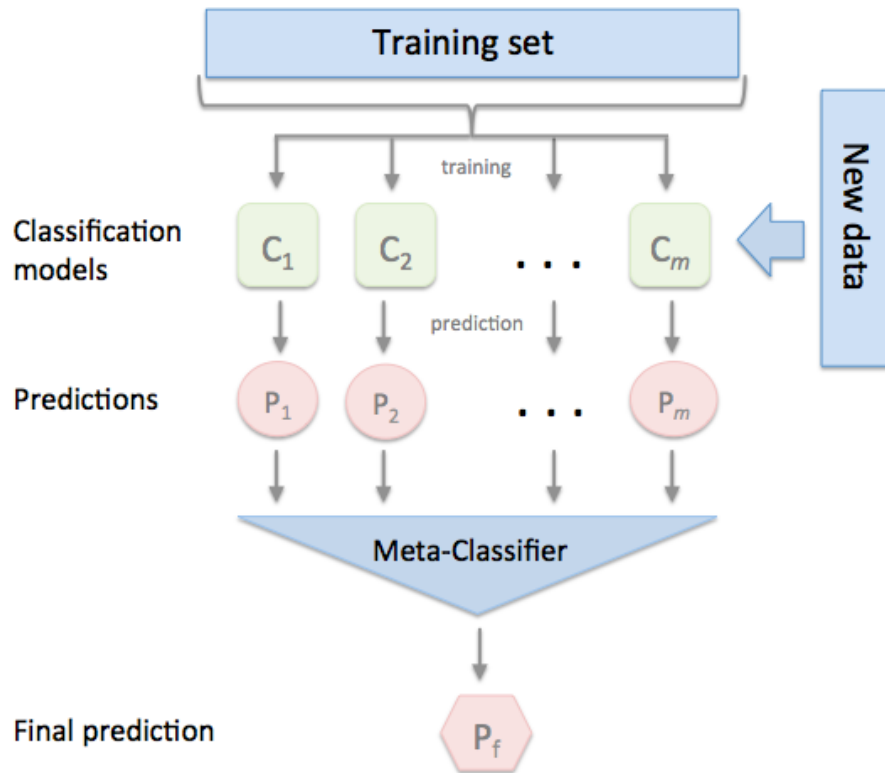


Queridinho do Kaggle: XGBoost

Classificação

Comitê - Técnicas

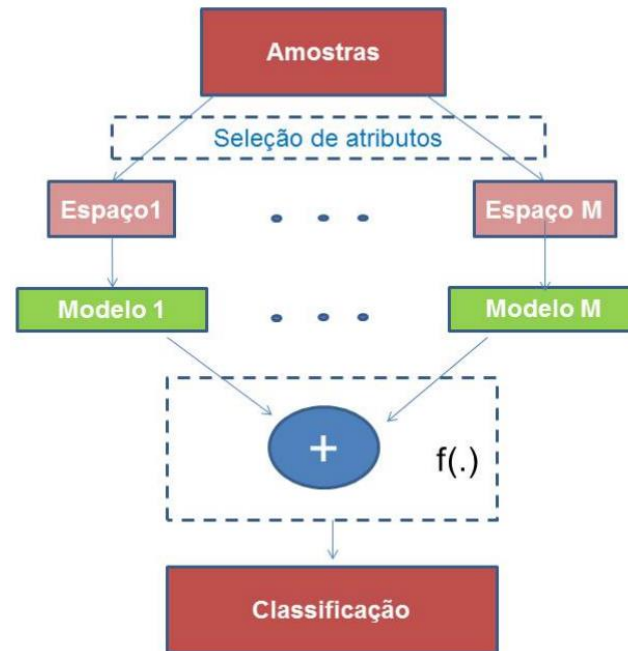
- Stacking



Classificação

Comitê - Técnicas

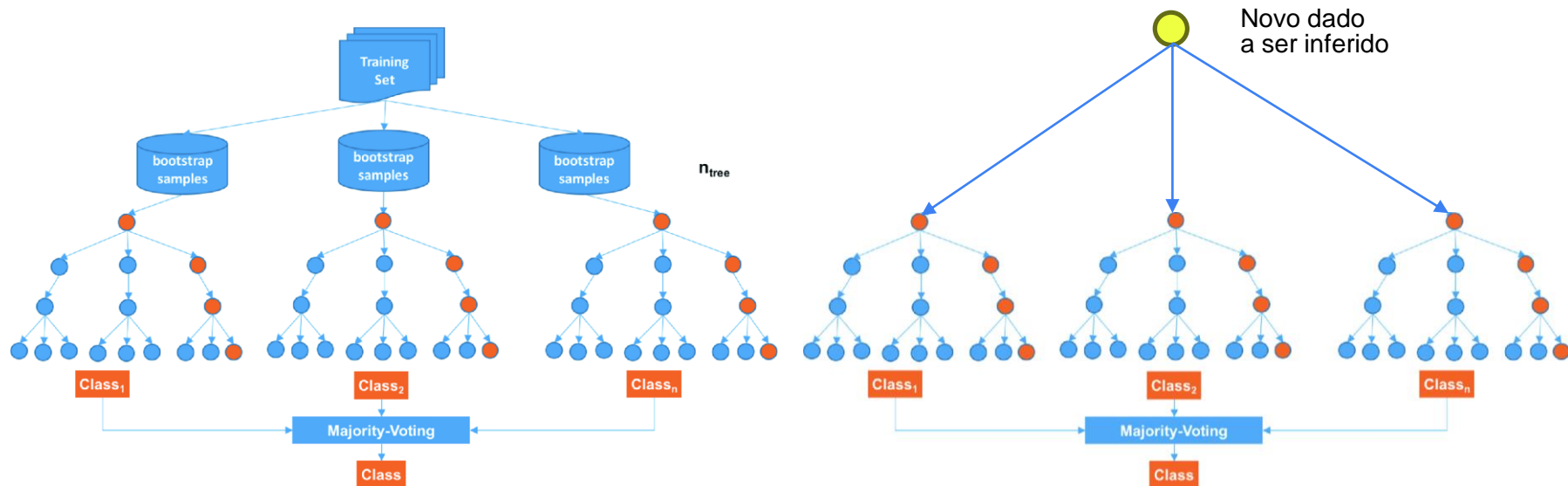
- Random Subspace Method (RSM)
- Similar ao Bagging, mas com aleatorização sobre os atributos.
- Classificadores-base aprendem nos subespaços S de mesma dimensão.
- Decisão final é por votação.



Random Forest

Classificação

Random Forest



A proporção de votos diferentes da classe target em relação ao total de votos é o erro OOB (Out-Of-Bag estimate)

Estudo de Caso

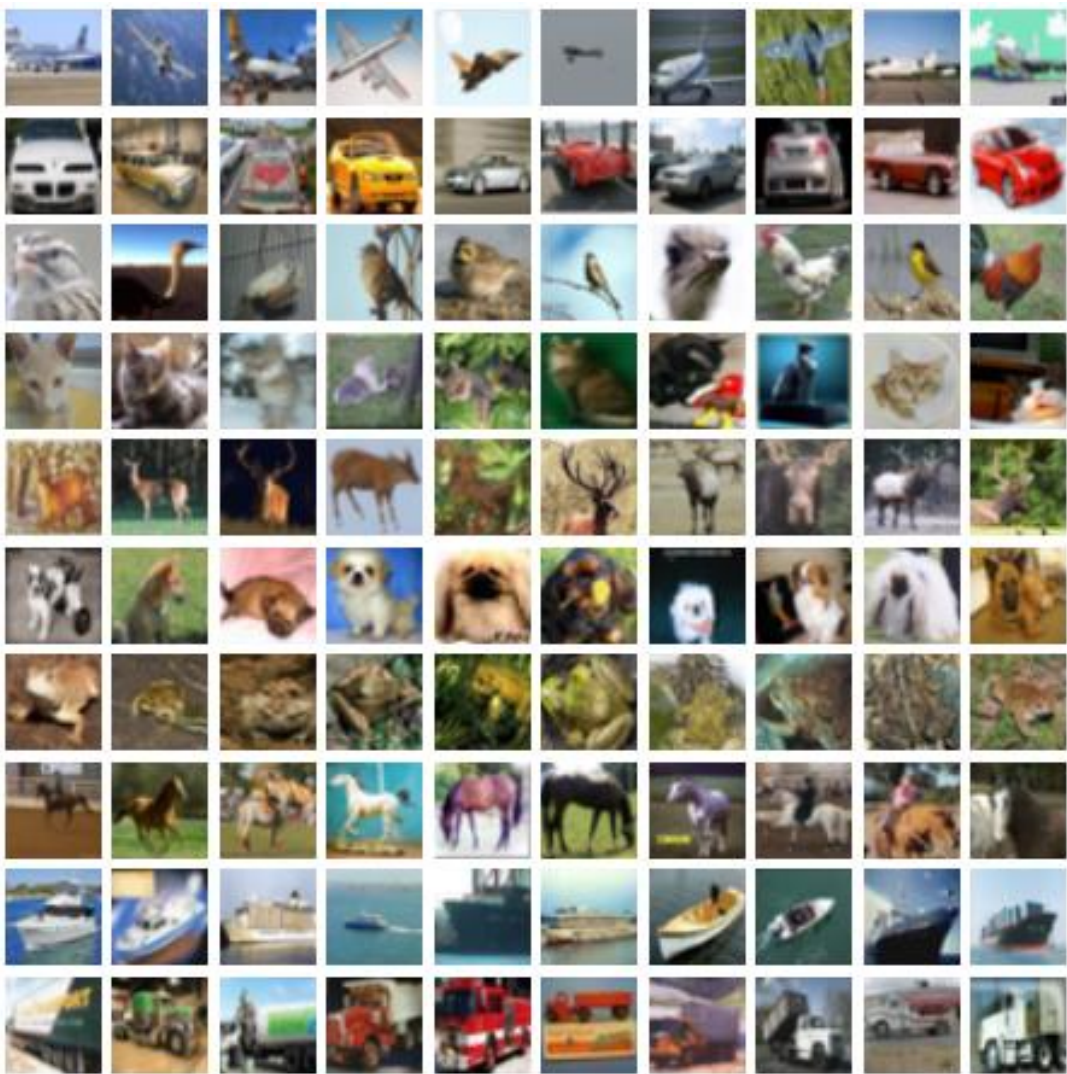
Classificação – Estudo de Caso

KNN

- CIFAR-10
 - 50k imagens de treino
 - 10k imagens de teste
 - 10 classes
 - 32x32x3
 - 8 bits
 - RGB

Curiosidades:

- Geoffrey Hinton foi um dos criadores do dataset. Não só o CIFAR-10, mas também o CIFAR-100 (combinando um total de 80M de imagem rotuladas).



Trabalho

Trabalho

Avaliação

- Análise exploratória;
- Pré-processamento;
- Teste de diferentes modelos;
- Análise de diferentes métricas.

Trabalho

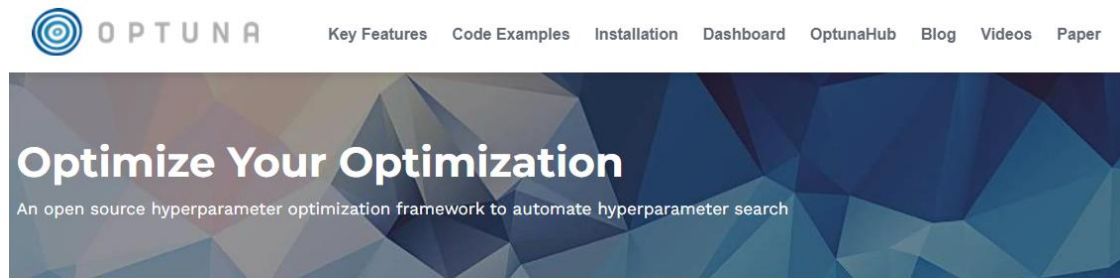
Avaliação

- O trabalho pode ser feito individualmente ou em grupo de até 5 pessoas.
- Sintam-se livres para trabalhar em qualquer base de interesse.
- Vocês podem utilizar R, Python, ou qualquer outro software ou linguagem que prefiram.
- Ao terminar o trabalho, enviem-me um relatório (pdf, doc, ppt ou comentários no próprio notebook) com o relato dos experimentos e resultados pelo **CLASSROOM**.
- Dúvidas: *classroom* ou prof.manoela@ica.ele.puc-rio.br

Optuna

Tuning de Hiperparâmetros

Optuna



Key Features

Eager search spaces



Automated search for optimal hyperparameters using Python conditionals, loops, and syntax

State-of-the-art algorithms



Efficiently search large spaces and prune unpromising trials for faster results

Easy parallelization



Parallelize hyperparameter searches over multiple threads or processes without modifying code



Github



Twitter

LinkedIn

<> Tutorials



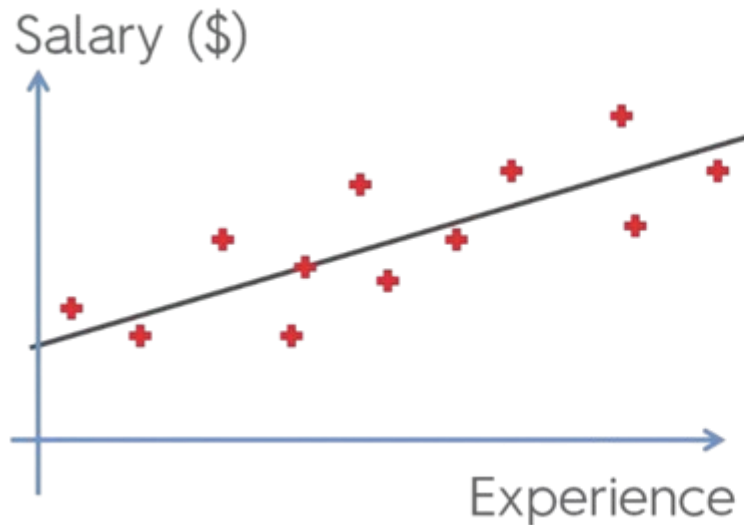
Docs

Regressão Logística

Classificação

Regressão Logística

- Regressão Linear Simples



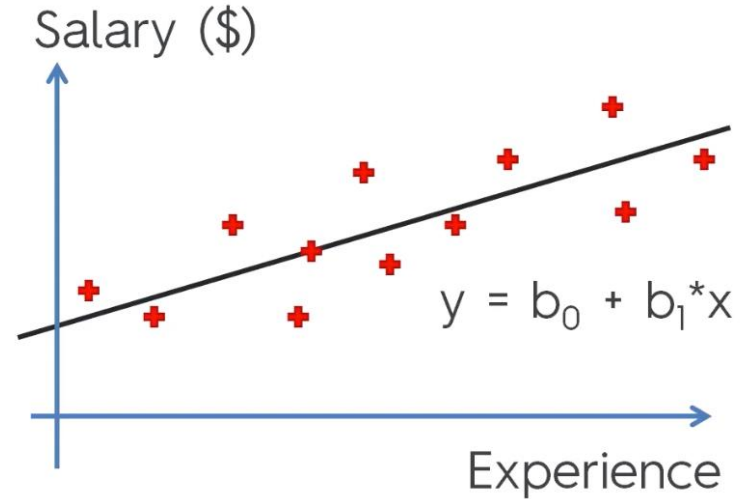
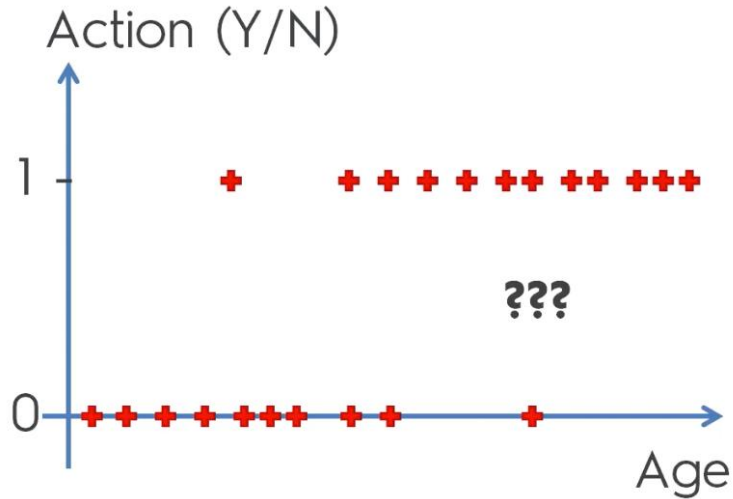
$$y = b_0 + b_1 * x$$



$$\text{Salary} = b_0 + b_1 * \text{Experience}$$

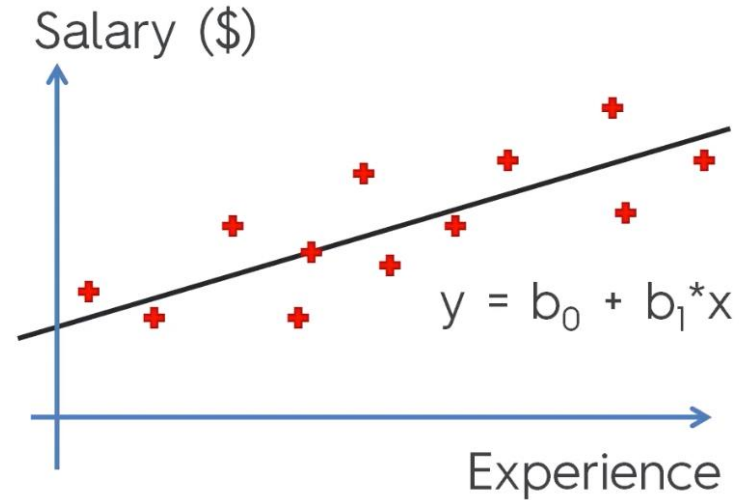
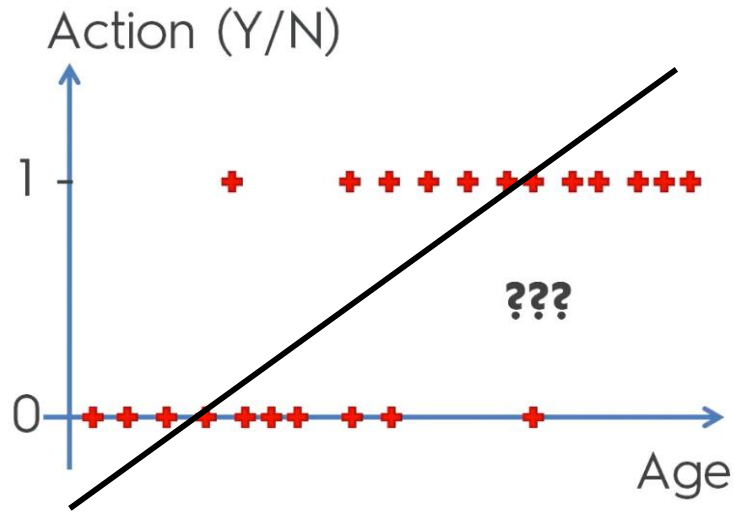
Classificação

Regressão Logística



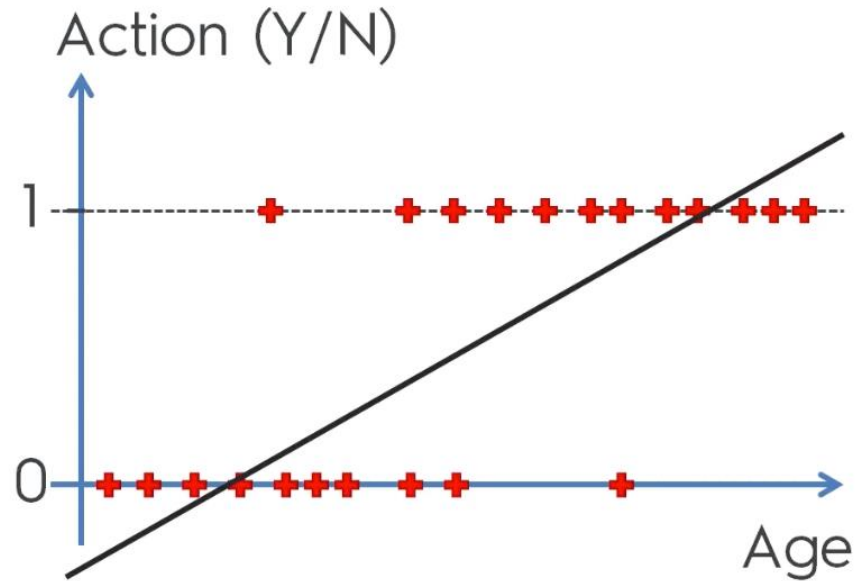
Classificação

Regressão Logística



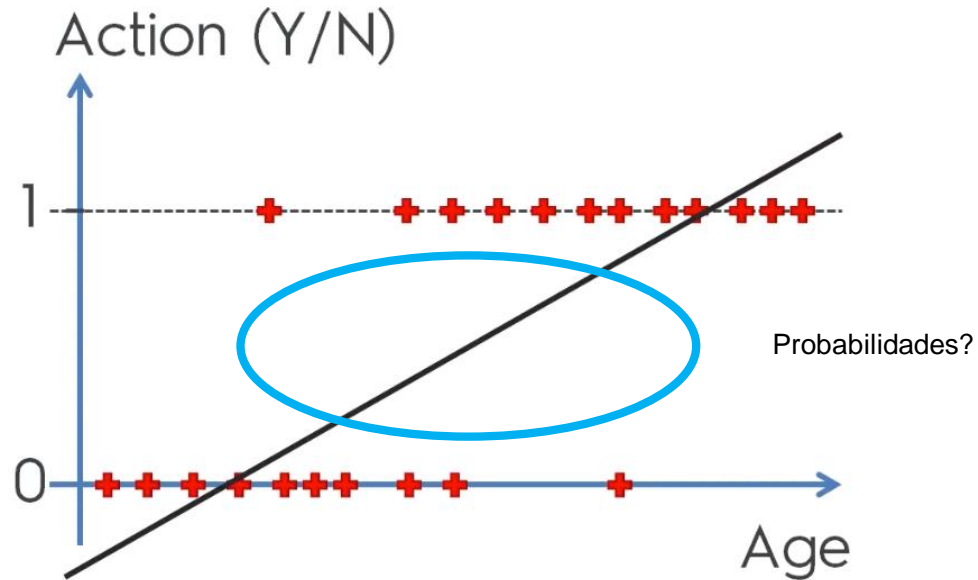
Classificação

Regressão Logística



Classificação

Regressão Logística

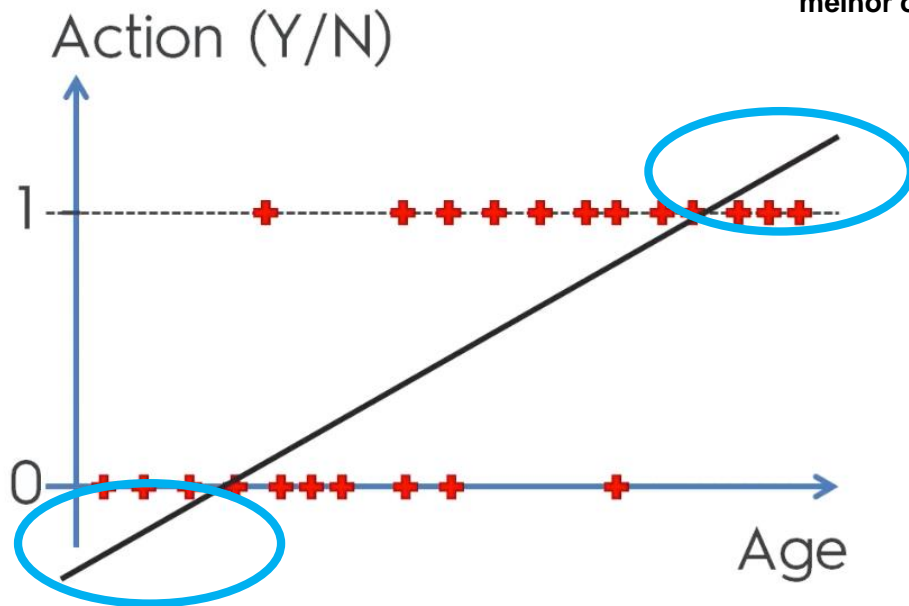


Classificação

Regressão Logística

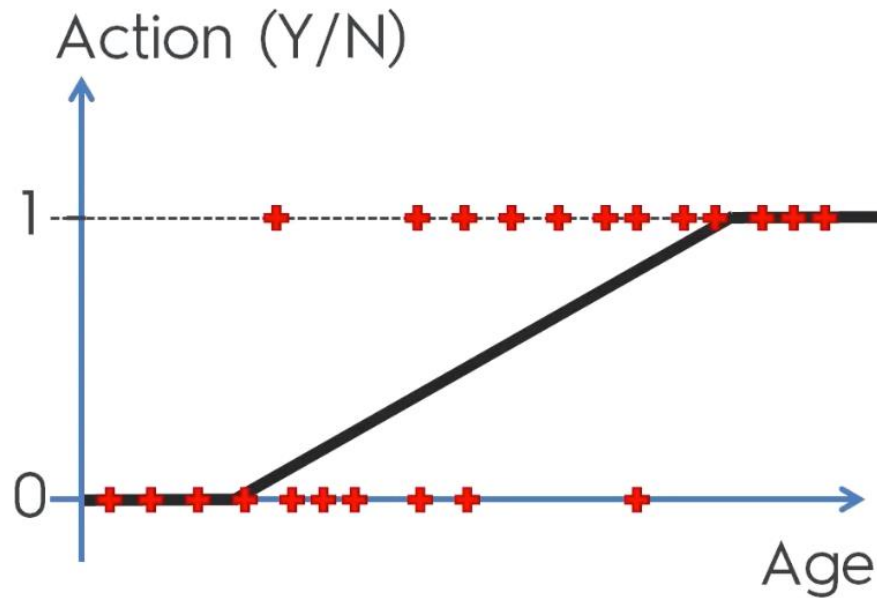
Pensando em probabilidades, esses 'pedaços' não fazem mais sentido.

Poderíamos alterar para modelar melhor o problema.



Classificação

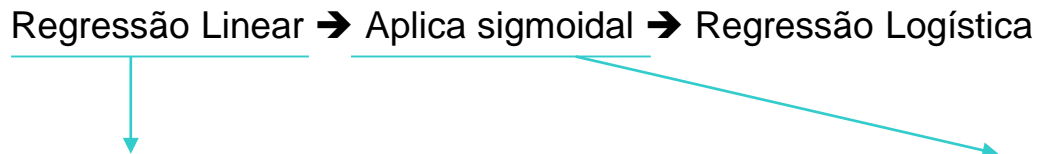
Regressão Logística



Classificação

Regressão Logística

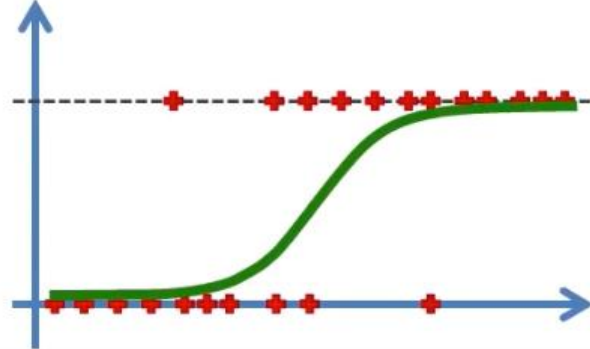
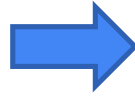
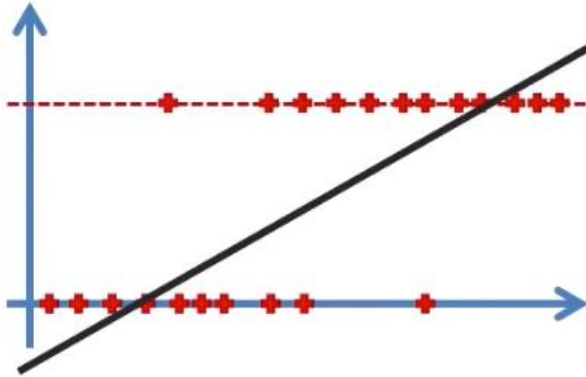
Regressão Linear → Aplica sigmoidal → Regressão Logística


$$y = b_0 + b_1 x_1 \quad \Rightarrow \quad f(a) = \frac{1}{1 + e^{-a}}$$

$$y' = f(y) = \frac{1}{1 + e^{-(b_0 + b_1 x_1)}}$$

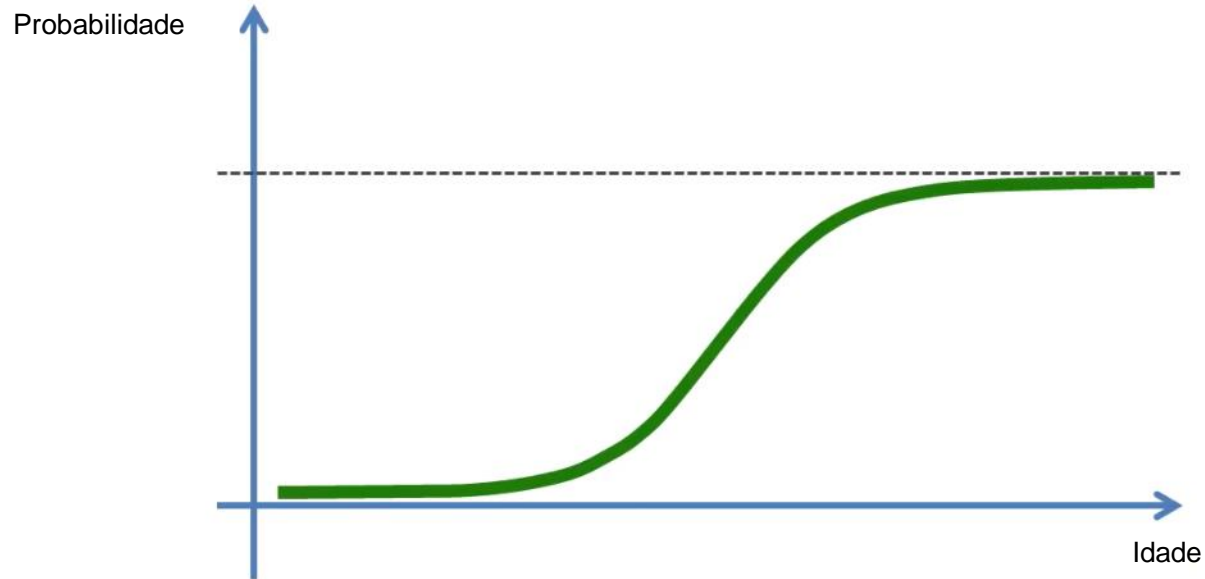
Classificação

Regressão Logística



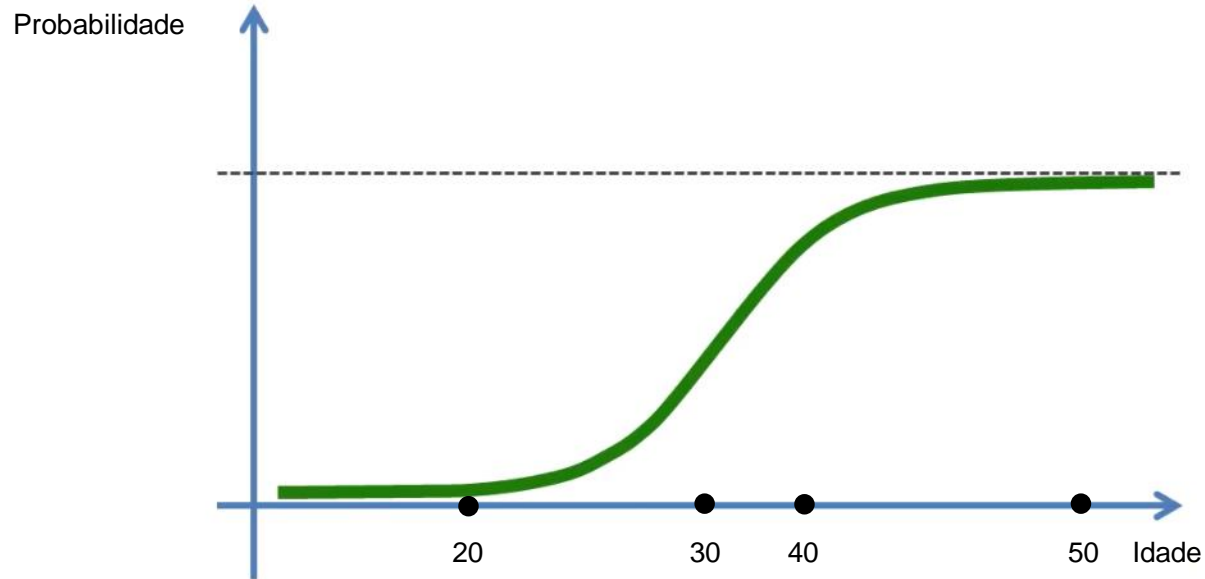
Classificação

Regressão Logística



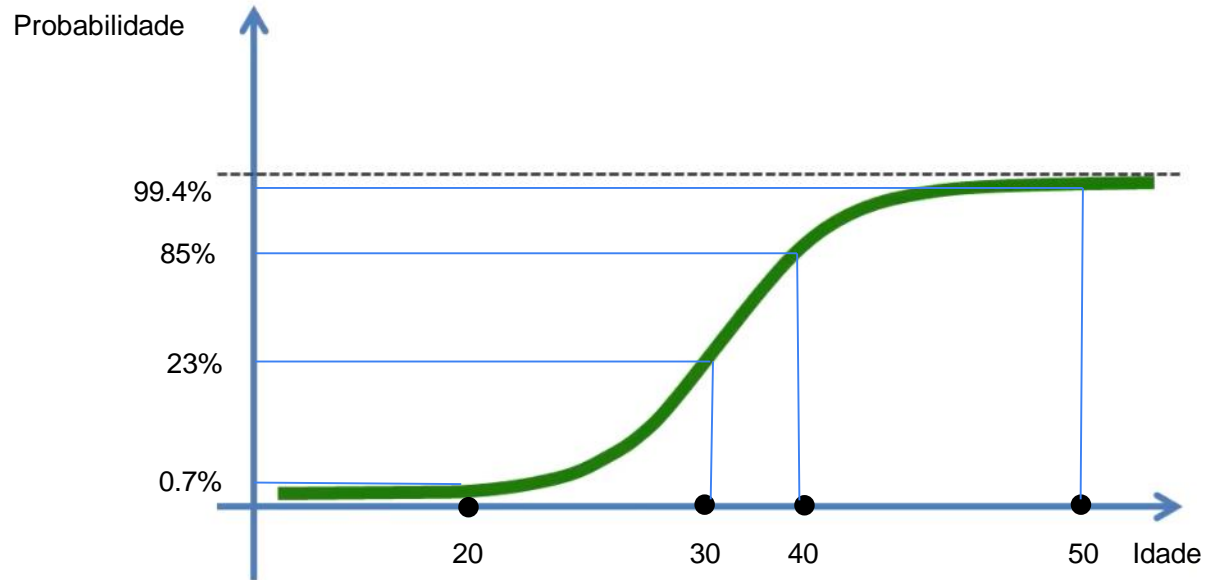
Classificação

Regressão Logística



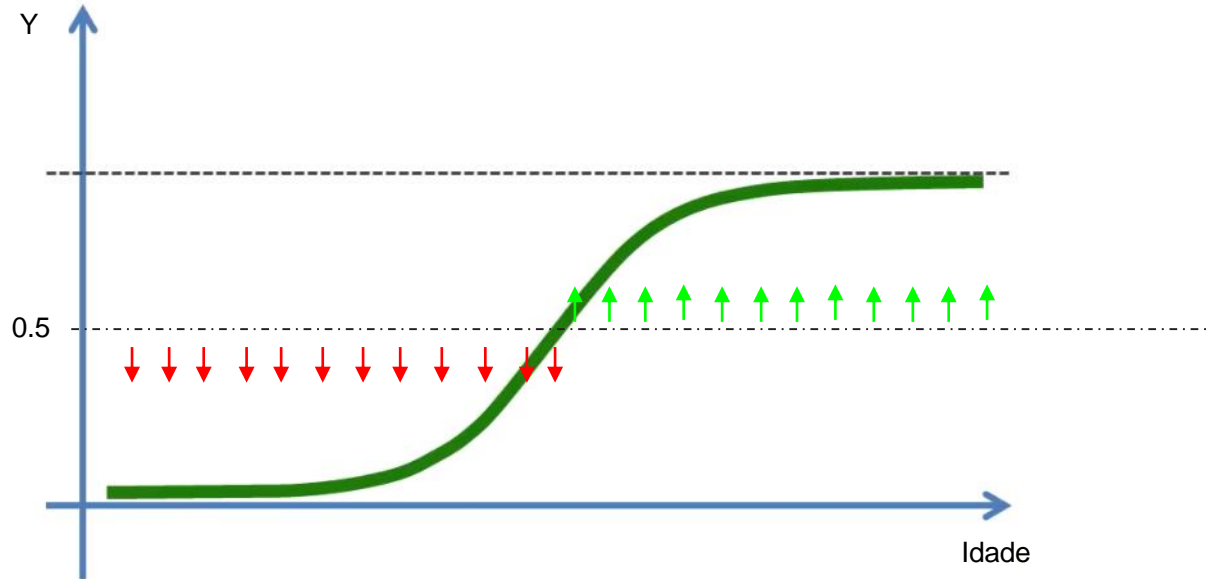
Classificação

Regressão Logística



Classificação

Regressão Logística



AI Kaggle Competitions

The Home of Data Science & Machine Learning

Kaggle helps you learn, work, and play

Create an account

or

Host a competition



Competitions ›

Climb the world's most elite machine learning leaderboards

[Want to host a competition?](#)

Datasets ›

Explore and analyze a collection of high quality public datasets

Kernels ›

Run code in the cloud and receive community feedback on your work

Kaggle Competitions

Por quê?

- Ótima plataforma para aprender e praticar
 - Python ou R
 - EDA
 - Machine Learning
 - Problemas reais
 - Problemas sintéticos
 - Comunidade muito ativa e colaborativa

Kaggle Competitions

Passo a Passo

1. Escolha a competição.

<https://www.kaggle.com/competitions>

Competitions

Grow your data science skills by competing in our exciting competitions. Find help in the [documentation](#) or learn about Community Competitions.

Host a Competition

Your Work



Search competitions

Filters

All Competitions

Everything, past & present

Featured

Premier challenges with prizes

Getting Started

Approachable ML fundamentals

Research

Scientific and scholarly challenges

Community

Created by fellow Kagglers

Playground

Fun practice problems

Active Competitions

Hotness

NFL Big Data Bowl 2023

Help evaluate linemen on pass plays

Analytics

\$100,000 2 months to go

Feedback Prize - English Language Learning

Evaluating language knowledge of ELL st...

Featured

Code Competition - 1809 Teams

\$55,000 a month to go

Big Data Derby 2022

Analyze horse racing data to improve the ...

Analytics

\$50,000 16 days to go

RSNA 2022 Cervical Spine Fracture Detection

Identify cervical fractures from scans

Featured

Code Competition - 871 Teams

\$30,000 2 days to go

2022 Kaggle Machine Learning

Open Problems - Multimodal

Novozymes Enzyme Stability

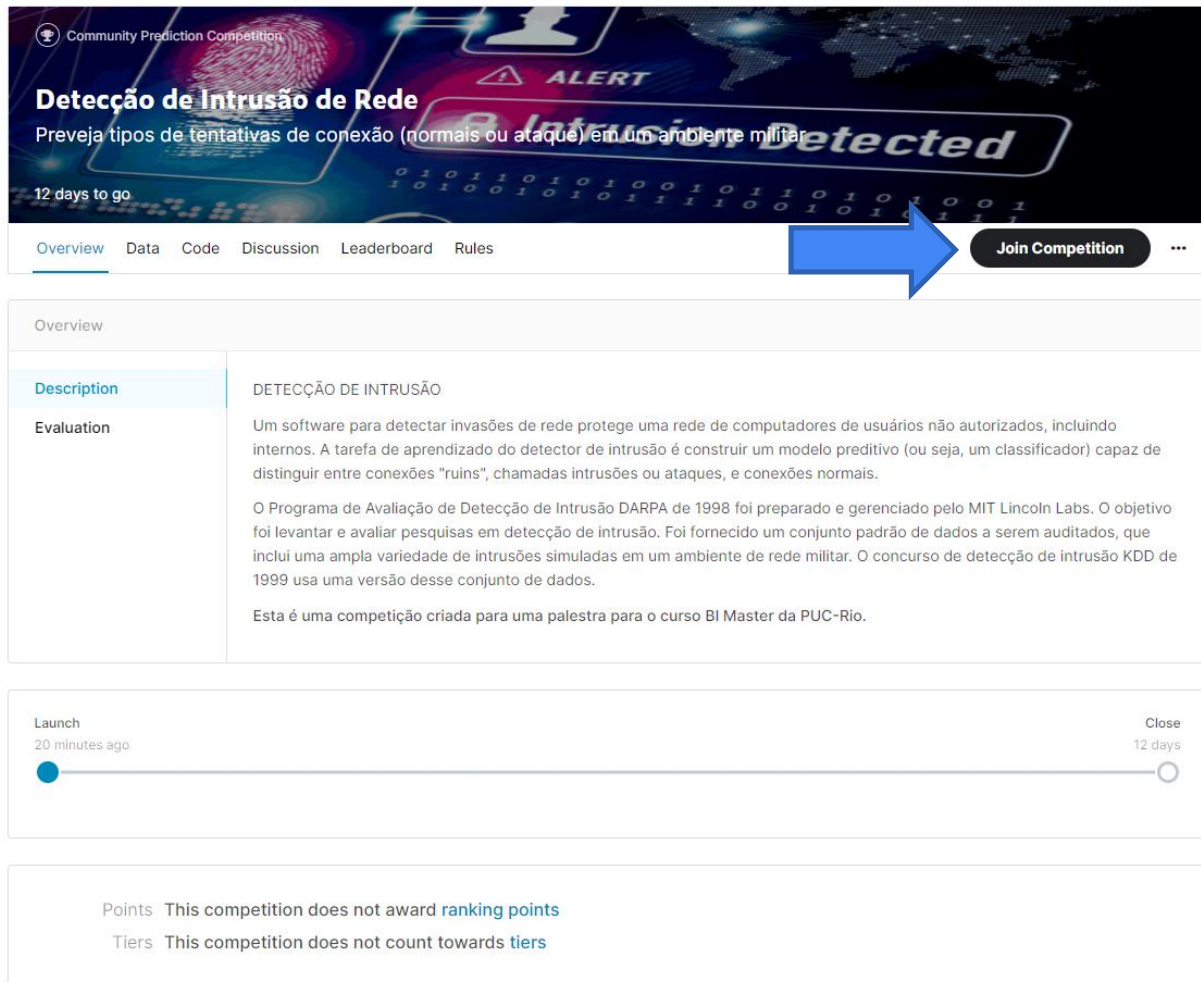
G2Net Detecting Continuous

Kaggle Competitions

Passo a Passo

2. Junte-se à competição.

➔ Leia a descrição da competição, aceite as regras e ganhe acesso à base de dados.



The screenshot shows the Kaggle competition page for "Detecção de Intrusão de Rede". The header features a dark blue banner with a world map, a fingerprint, and a red "ALERT" sign. The title "Detecção de Intrusão de Rede" is prominently displayed, followed by the subtitle "Preveja tipos de tentativas de conexão (normais ou ataque) em um ambiente militar". A countdown timer indicates "12 days to go". Below the banner, there are tabs for "Overview", "Data", "Code", "Discussion", "Leaderboard", and "Rules". A large blue arrow points to the "Join Competition" button. The "Overview" section is expanded, showing a table with "Description" and "Evaluation" rows. The "Description" row contains the title "DETECÇÃO DE INTRUSÃO". The "Evaluation" row contains a detailed description of the competition, including the goal of building a model to distinguish between normal connections and attacks, and the historical context of the DARPA 1998 program. A "Launch" section at the bottom shows a progress bar that is 20 minutes into a 12-day period. The "Points" and "Tiers" sections indicate that this competition does not award ranking points or count towards tiers.

Community Prediction Competition

Detecção de Intrusão de Rede

Preveja tipos de tentativas de conexão (normais ou ataque) em um ambiente militar

12 days to go

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#)

[Join Competition](#)

Overview

Description	DETECÇÃO DE INTRUSÃO
Evaluation	<p>Um software para detectar invasões de rede protege uma rede de computadores de usuários não autorizados, incluindo internos. A tarefa de aprendizado do detector de intrusão é construir um modelo preditivo (ou seja, um classificador) capaz de distinguir entre conexões "ruins", chamadas intrusões ou ataques, e conexões normais.</p> <p>O Programa de Avaliação de Detecção de Intrusão DARPA de 1998 foi preparado e gerenciado pelo MIT Lincoln Labs. O objetivo foi levantar e avaliar pesquisas em detecção de intrusão. Foi fornecido um conjunto padrão de dados a serem auditados, que inclui uma ampla variedade de intrusões simuladas em um ambiente de rede militar. O concurso de detecção de intrusão KDD de 1999 usa uma versão desse conjunto de dados.</p> <p>Esta é uma competição criada para uma palestra para o curso BI Master da PUC-Rio.</p>

Launch
20 minutes ago

Close
12 days

Points This competition does not award [ranking points](#)

Tiers This competition does not count towards [tiers](#)

Kaggle Competitions

Passo a Passo

3. Baixe a base de dados.

→ Pode não ser um arquivo único!

Detecção de Intrusão de Rede
Preveja tipos de tentativas de conexão (normais ou ataque) em um ambiente militar.

12 days to go

View Data Code Discussion Leaderboard Rules Team Host Submissions Submit Predictions

Dataset Description

Este conjunto de dados é composto por uma parte do dataset original do KDD Cup 99.

Arquivos

- treino.csv - base para treino e teste
- teste_sem_rotulo.csv - conjunto a ser inferido e submetido
- exemplo_submissao.csv - um arquivo exemplo de como as inferências devem ser submetidas

Files

3 files

Size: 3.63 MB

Type: csv

Data Explorer

3.63 MB

- exemplo_submissao.csv
- teste_sem_rotulo.csv
- treino.csv

Summary

3 files

37 columns

Download All

New Version

treino.csv (3.62 MB)

Detail Compact Column 10 of 18 columns

About this file

This file does not have a description yet.

id	protocol_type	flag	src_bytes	logged_in
1	icmp	SF	77%	0
2	tcp	S0	18%	1
3	Other (1936)	Other (2512)	5%	0
4	icmp	SF	1832	0
5	tcp	S0	284	0
6	icmp	SF	528	0
7	icmp	SF	1832	0
8	udp	SF	145	0
9	icmp	SF	1832	0
10	icmp	SF	1832	0
11	icmp	SF	1832	0
12	icmp	SF	1832	0
13	icmp	SF	1832	0
14	icmp	SF	1832	0

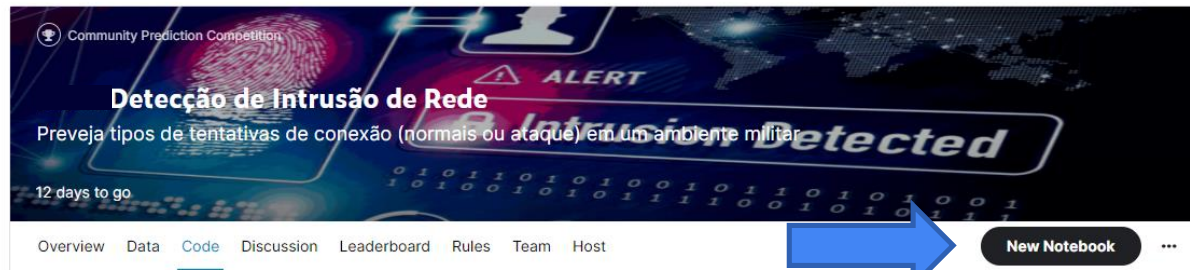
> kaggle competitions download -c detecao-de-intrusao-de-rede

Kaggle Competitions

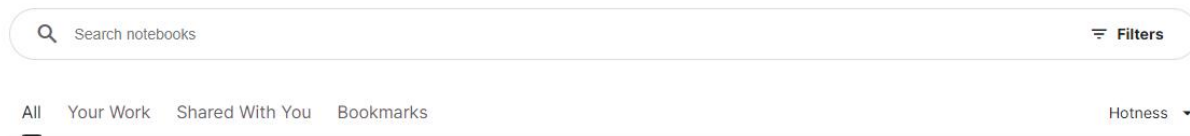
Passo a Passo

3. Trabalhe na solução!

➔ Localmente, no colab, ou usando sua ferramenta de preferência, ou usando o próprio kaggle!



Notebooks



No notebooks found

No notebooks to show.

Kaggle Competitions

Passo a Passo

3. Trabalhe na solução!

➔ Localmente, no colab, usando sua ferramenta de preferência, ou usando o próprio kaggle!

The screenshot shows a Kaggle notebook titled 'notebookb77e27b0bc'. The code is in Python and includes comments about the environment and data sources. It imports numpy and pandas, and uses os to list files in the input directory. The code then reads a CSV file and displays the first few rows of the data.

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 2GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session

/kaggle/input/detecao-de-intrusao-de-rede/teste_sem_rotulo.csv
/kaggle/input/detecao-de-intrusao-de-rede/treino.csv
/kaggle/input/detecao-de-intrusao-de-rede/exemplo_submissao.csv
```

The code is executed, and the output shows the first few rows of the data:

	id	protocol_type	flag	src_bytes	logged_in	count	src_count	sensor_rate	srv_sensor_rate	error_rate	srv_error_rate	diff_srv_rate	dst_host_diff_srv_rate	dst_host_diff_srv_rate	dst_host_diff_srv_rate	dst_host_diff_srv_rate	dst_host_diff_srv_rate
0	1	icmp	SF	1032	0	511	511	0.0	0.0	0.0	0.0	0.00	0.00	0.00	1.00	0.0	
1	2	tcp	SF	204	1	17	17	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.02	0.0	
2	3	tcp	SF	0	0	250	8	1.0	1.0	0.0	0.0	0.06	0.06	0.00	0.00	1.0	
3	4	icmp	SF	520	0	511	511	0.0	0.0	0.0	0.0	0.00	0.00	0.00	1.00	0.0	
4	5	icmp	SF	1032	0	511	511	0.0	0.0	0.0	0.0	0.00	0.00	0.00	1.00	0.0	

Kaggle Competitions

Passo a Passo

4. Faça sua submissão!

➔ Leia atentamente a descrição do formato e informações que devem estar presentes no arquivo de submissão. Algumas competições disponibilizam um exemplo de um arquivo de submissão.

The screenshot shows the Kaggle competition interface for 'Detecção de Intrusão de Rede'. The header features a banner with a fingerprint and the text 'Community Prediction Competition', 'Detecção de Intrusão de Rede', 'Preveja tipos de tentativas de conexão (normais ou ataque) em um ambiente militar', and '9 days to go'. Below the banner is a navigation bar with links: Overview, Data, Code, Discussion, Leaderboard, Rules, Team, Host, Submissions, and a 'Submit Predictions' button. The main content area is titled 'Overview' and contains a 'Description' section. The description explains that the software is for detecting network intrusions and mentions the DARPA 1998 program. At the bottom, there is a 'Launch' section with a progress bar indicating '3 days ago' and a 'Close' button with '9 days' remaining.

Community Prediction Competition

Detecção de Intrusão de Rede

Preveja tipos de tentativas de conexão (normais ou ataque) em um ambiente militar

9 days to go

Overview Data Code Discussion Leaderboard Rules Team Host Submissions **Submit Predictions**

Overview

Edit

Description

Evaluation

+ Add Page

DETECÇÃO DE INTRUSÃO

Um software para detectar invasões de rede protege uma rede de computadores de usuários não autorizados, incluindo internos. A tarefa de aprendizado do detector de intrusão é construir um modelo preditivo (ou seja, um classificador) capaz de distinguir entre conexões "ruins", chamadas intrusões ou ataques, e conexões normais.

O Programa de Avaliação de Detecção de Intrusão DARPA de 1998 foi preparado e gerenciado pelo MIT Lincoln Labs. O objetivo foi levantar e avaliar pesquisas em detecção de intrusão. Foi fornecido um conjunto padrão de dados a serem auditados, que inclui uma ampla variedade de intrusões simuladas em um ambiente de rede militar. O concurso de detecção de intrusão KDD de 1999 usa uma versão desse conjunto de dados.

Esta é uma competição criada para uma palestra para o curso BI Master da PUC-Rio.

Launch 3 days ago

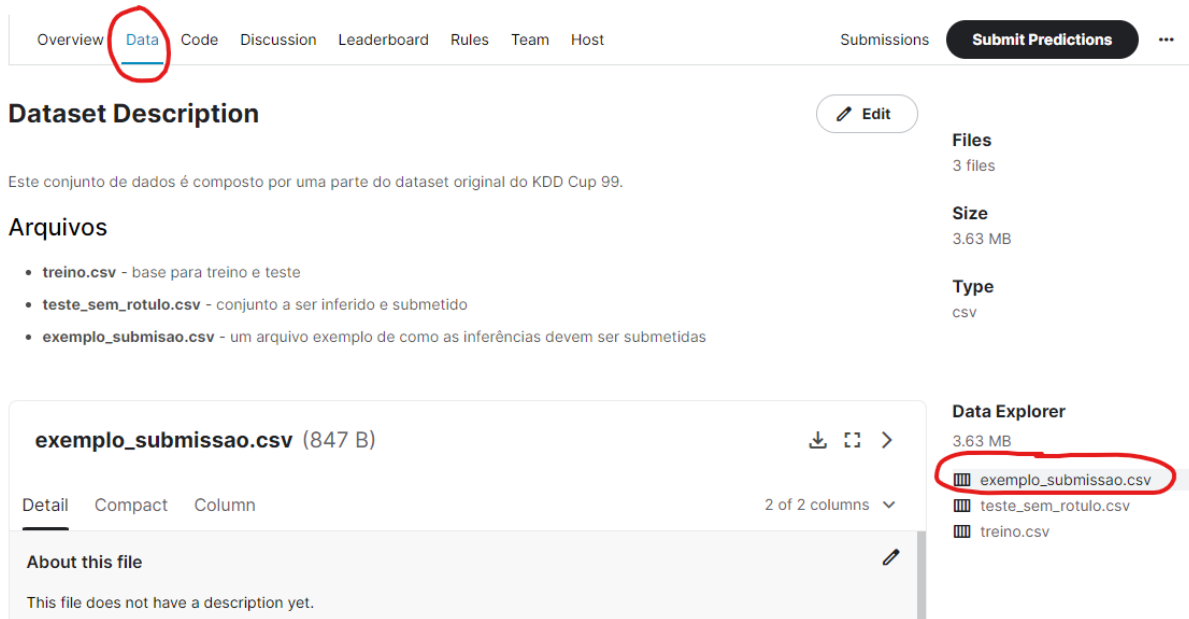
Close 9 days

Kaggle Competitions

Passo a Passo

4. Faça sua submissão!

→ Leia atentamente a descrição do formato e informações que devem estar presentes no arquivo de submissão. Algumas competições disponibilizam um exemplo de um arquivo de submissão.



The screenshot shows the 'Data' tab of a Kaggle competition. The 'Data' tab is highlighted with a red circle. The 'Dataset Description' section states: 'Este conjunto de dados é composto por uma parte do dataset original do KDD Cup 99.' The 'Arquivos' section lists three files: 'treino.csv' (base for training and testing), 'teste_sem_rotulo.csv' (set to be inferred and submitted), and 'exemplo_submissao.csv' (example of how submissions should be made). The 'Data Explorer' section shows three files: 'exemplo_submissao.csv' (3.63 MB), 'teste_sem_rotulo.csv', and 'treino.csv'. The 'exemplo_submissao.csv' file is highlighted with a red circle. The 'About this file' section states: 'This file does not have a description yet.'

Overview **Data** Code Discussion Leaderboard Rules Team Host Submissions **Submit Predictions** ...

Dataset Description

Este conjunto de dados é composto por uma parte do dataset original do KDD Cup 99.

Arquivos

- **treino.csv** - base para treino e teste
- **teste_sem_rotulo.csv** - conjunto a ser inferido e submetido
- **exemplo_submissao.csv** - um arquivo exemplo de como as inferências devem ser submetidas

Data Explorer

3.63 MB

- exemplo_submissao.csv
- teste_sem_rotulo.csv
- treino.csv

exemplo_submissao.csv (847 B)

Detail Compact Column 2 of 2 columns

About this file

This file does not have a description yet.

Kaggle Competitions

Passo a Passo

4. Faça sua submissão!

➔ Leia atentamente a descrição do formato e informações que devem estar presentes no arquivo de submissão. Algumas competições disponibilizam um exemplo de um arquivo de submissão.

✕ Submit to Competition

File Upload Notebook



Detecção de Intrusão de Rede

You have 20 submissions remaining today. This resets in 12 hours.



Drag and drop file to upload

(e.g., .csv, .zip, .gz, .7z)

or

Browse Files

Your submission should be a CSV file with 60 rows and a header. You can upload a zip/gz/7z archive.

DESCRIPTION

Enter a description

0 / 500

>_ kaggle competitions submit -c detecao-de-intrusao-de-rede -f su_



Cancel

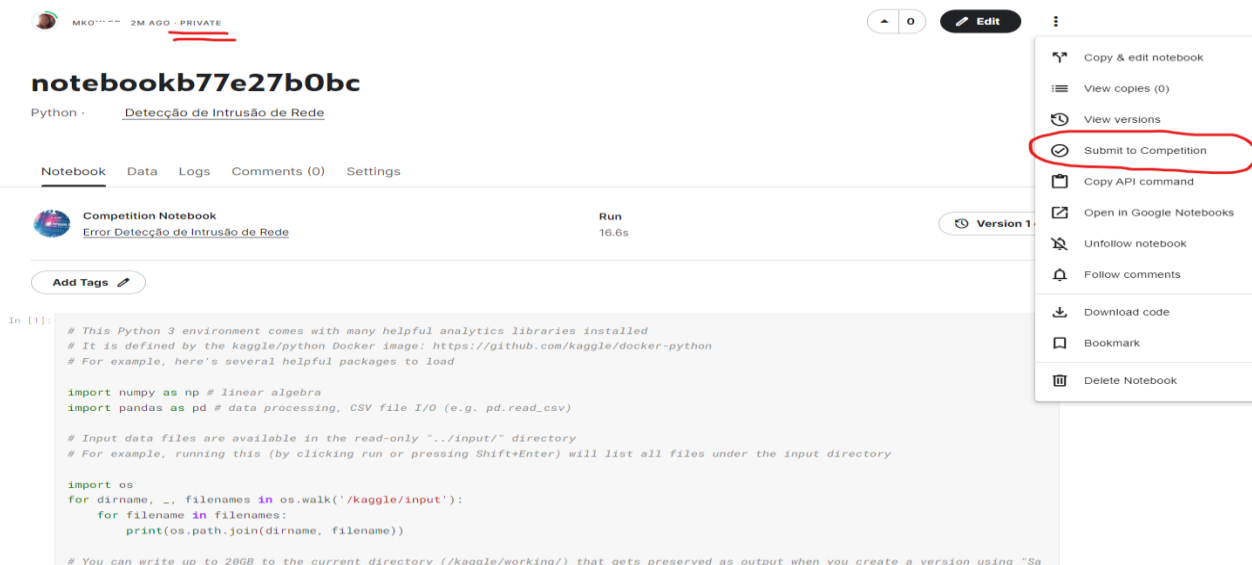
Submit

Kaggle Competitions

Passo a Passo

Utilizando o kaggle, crie o notebook na aba 'Code', desenvolva a solução normalmente. O único requisito é que o próprio script gere um csv no formato exigido para submissão.

4.1. Quando pronto, clique em 'Submit to Competition'



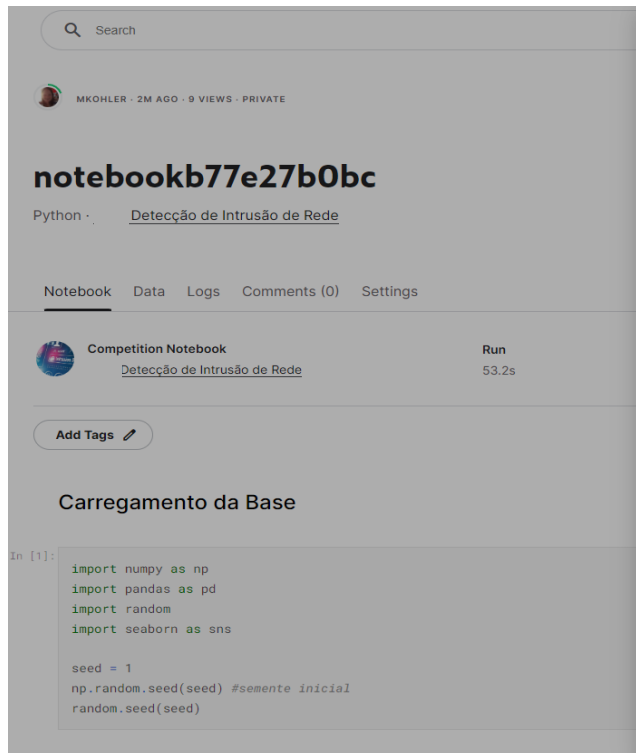
The screenshot shows a Kaggle notebook titled 'notebookb77e27b0bc' with the subtitle 'Detecção de Intrusão de Rede'. The notebook is in the 'Code' tab. A dropdown menu is open on the right side, showing various actions. The 'Submit to Competition' option is highlighted with a red circle. Below the notebook title, there is a table with columns 'Competition Notebook' and 'Run'. The 'Competition Notebook' column shows 'Error Detecção de Intrusão de Rede' and the 'Run' column shows '16.6s'. At the bottom, there is a code editor with Python code for file I/O and directory listing.

```
In [1]:  
  
# This Python 3 environment comes with many helpful analytics libraries installed  
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python  
# For example, here's several helpful packages to load  
  
import numpy as np # linear algebra  
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)  
  
# Input data files are available in the read-only "../input/" directory  
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory  
  
import os  
for dirname, _, filenames in os.walk('/kaggle/input'):  
    for filename in filenames:  
        print(os.path.join(dirname, filename))  
  
# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using "Sa
```

Kaggle Competitions

Passo a Passo

4.2. Escolha o notebook, a versão no notebook e o arquivo a ser submetido.



The screenshot shows a Kaggle Competition Notebook page. At the top, there is a search bar and a header for the competition 'notebookb77e27b0bc' by MKOHLER, 2M AGO, 9 VIEWS, PRIVATE. The notebook is titled 'Detecção de Intrusão de Rede' and is written in Python. Below the title, there are tabs for 'Notebook', 'Data', 'Logs', 'Comments (0)', and 'Settings'. The 'Notebook' tab is selected, showing a 'Competition Notebook' section with the title 'Detecção de Intrusão de Rede' and a 'Run' button. Below this, there is an 'Add Tags' button. The main content area shows the code for 'Carregamento da Base'.

```
In [1]:  
import numpy as np  
import pandas as pd  
import random  
import seaborn as sns  
  
seed = 1  
np.random.seed(seed) #semente inicial  
random.seed(seed)
```

Submit to Competition

File Upload Notebook



Detecção de Intrusão de Rede

You have 19 submissions remaining today. This resets in 2 hours.

SELECT NOTEBOOK

notebookb77e27b0bc

NOTEBOOK VERSION

Version 8

OUTPUT FILE

previsoes.csv

DESCRIPTION

Enter a description

0 / 500

Cancel

Submit

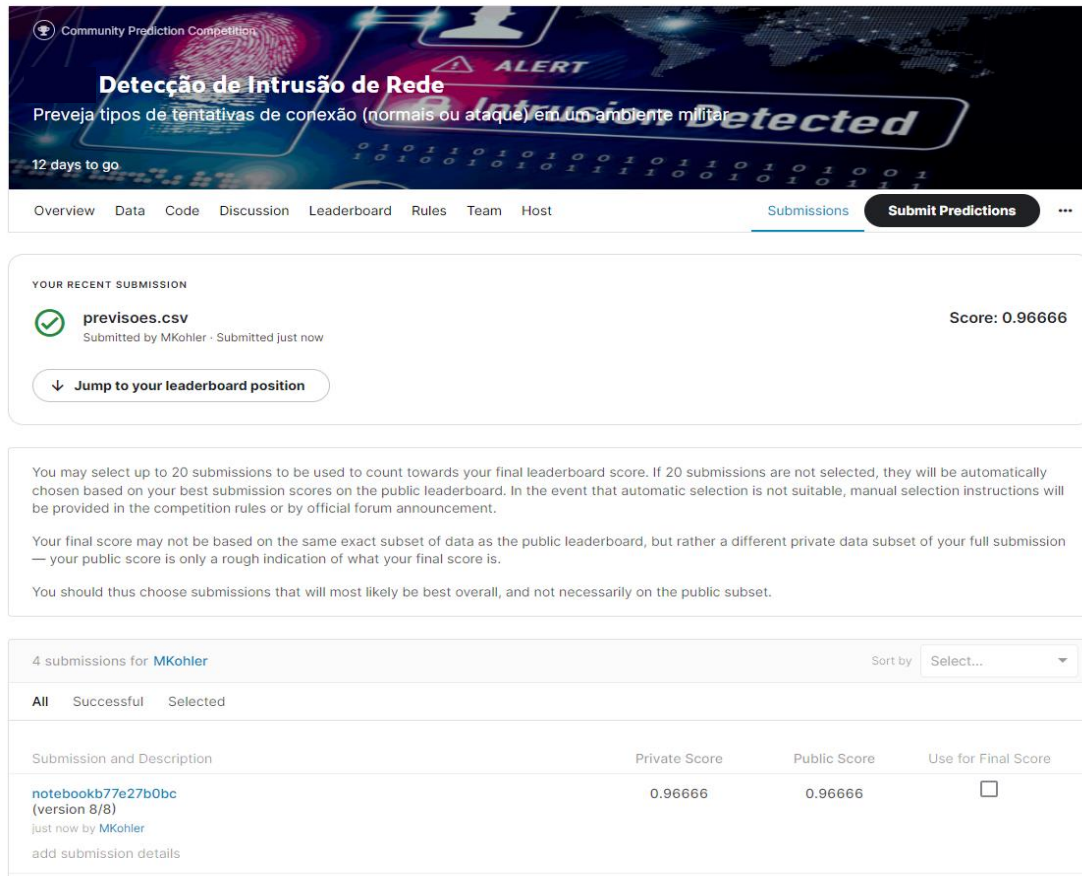
Kaggle Competitions

Passo a Passo

4.3. Clique em 'Submit'

Notebook submetido, o csv escolhido para submissão já é automaticamente validado, a métrica da competição calculada e sua submissão entra para o *leaderboard*.

****As competições podem ter *leaderboards* privado e público (para evitar tentativa e erro ou algum tipo de trapaça). O score privado só é divulgado ao final da competição (em geral). As competições também definem a quantidade de submissões permitidas por dia e o tamanho máximo de uma equipe.**



The screenshot shows the Kaggle competition interface for 'Detecção de Intrusão de Rede'. The header features a dark blue background with a world map and binary code. The main title is 'Detecção de Intrusão de Rede' with a subtitle 'Preveja tipos de tentativas de conexão (normais ou ataque) em um ambiente militar'. A countdown timer shows '12 days to go'. The navigation bar includes 'Overview', 'Data', 'Code', 'Discussion', 'Leaderboard', 'Rules', 'Team', 'Host', 'Submissions', and a 'Submit Predictions' button. Below the navigation bar, the 'YOUR RECENT SUBMISSION' section shows a green checkmark for 'previsoes.csv' submitted by 'MKohler' just now, with a score of 0.96666. A button 'Jump to your leaderboard position' is visible. The 'Submissions' section explains the selection process for the final leaderboard score, noting that up to 20 submissions can be chosen, or the system will automatically select the best ones. It also states that the public score is only a rough indication of the final score. At the bottom, a table shows 4 submissions for 'MKohler', with one submission 'notebookb77e27b0bc (version 8/8)' having a private score of 0.96666 and a public score of 0.96666, with a checkbox for 'Use for Final Score'.

Submission and Description	Private Score	Public Score	Use for Final Score
notebookb77e27b0bc (version 8/8) just now by MKohler add submission details	0.96666	0.96666	<input type="checkbox"/>

Kaggle Competitions

Passo a Passo


5. Acompanhe o *leaderboard*!

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [Submissions](#) [Submit Predictions](#) [...](#)

Leaderboard

[Raw Data](#) [Refresh](#)


YOUR RECENT SUBMISSION


 **primeira_submissao.csv** Score: 1.00000

Submitted by profManoela · Submitted a few seconds ago

[Jump to your leaderboard position](#)

This leaderboard is calculated with all of the test data.

#	Team	Members	Score	Entries	Last	Code
1	profManoela		1.00000	1	1s	

 Your First Entry!
Welcome to the leaderboard!

Kaggle Competitions

Passo a Passo

6. Lembre-se de selecionar quais submissões irão para o leaderboard (limite definido na pela competição)!

Successfully selected submission.

You may select up to 20 submissions to be used to count towards your final leaderboard score. If 20 submissions are not selected, they will be automatically chosen based on your best submission scores on the public leaderboard. In the event that automatic selection is not suitable, manual selection instructions will be provided in the competition rules or by official forum announcement.

Your final score may not be based on the same exact subset of data as the public leaderboard, but rather a different private data subset of your full submission — your public score is only a rough indication of what your final score is.

You should thus choose submissions that will most likely be best overall, and not necessarily on the public subset.









4 submissions for **profManoela** Sort by Select...

All Successful Selected

Submission and Description	Public Score	Use for Final Score
previsoes.csv a minute ago by profManoela add submission details	0.96666	<input type="checkbox"/>
exemplo_submissao.csv 4 minutes ago by profManoela add submission details	0.21666	<input type="checkbox"/>
resposta.csv 5 minutes ago by profManoela add submission details	1.00000	<input type="checkbox"/>
primeira_submissao.csv 8 minutes ago by profManoela add submission details	1.00000	<input checked="" type="checkbox"/>








Kaggle Competitions

Para praticar

	House Prices - Advanced Regression Techniques Predict sales prices and practice feature engineering, RFs, and gradient boosting Getting Started · 3990 Teams · Ongoing	Knowledge ...
	Bag of Words Meets Bags of Popcorn Use Google's Word2Vec for movie reviews Getting Started · 577 Teams · 7 years ago	Knowledge ...
	First Steps With Julia Use Julia to identify characters from Google Street View images Getting Started · 56 Teams · 6 years ago	Knowledge ...
	Facial Keypoints Detection Detect the location of keypoints on face images Getting Started · 175 Teams · 6 years ago	Knowledge ...
	Data Science London + Scikit-learn Scikit-learn is an open-source machine learning library for Python. Give it a try here! Getting Started · 190 Teams · 8 years ago	Knowledge ...
	Just the Basics - Strata 2013 After-party Live from Santa Clara, CA Getting Started · 48 Teams · 10 years ago	Knowledge ...
	Just the Basics - Strata 2013 Live from Santa Clara, CA - Core Data Science Skills with Kaggle's Top Competitors Getting Started · 49 Teams · 10 years ago	Knowledge ...
	Titanic - Machine Learning from Disaster Start here! Predict survival on the Titanic and get familiar with ML basics Getting Started · 14642 Teams · Ongoing	Knowledge ...

Kaggle Competitions

Para praticar

	NFL Big Data Bowl 2023 Help evaluate linemen on pass plays Analytics · 0 Teams · 2 months to go	\$100,000 ...
	Feedback Prize - English Language Learning Evaluating language knowledge of ELL students from grades 8-12 Featured · Code Competition · 1881 Teams · a month to go	\$55,000 ...
	Big Data Derby 2022 Analyze horse racing data to improve the health of the horse and strategy of competition Analytics · 0 Teams · 13 days to go	\$50,000 ...
	2022 Kaggle Machine Learning & Data Science Survey The most comprehensive dataset available on the state of ML and data science Analytics · 0 Teams · a month to go	\$30,000 ...
	Open Problems - Multimodal Single-Cell Integration Predict how DNA, RNA & protein measurements co-vary in single cells Featured · 1042 Teams · 18 days to go	\$25,000 ...
	Novozymes Enzyme Stability Prediction Help identify the thermostable mutations in enzymes Featured · 967 Teams · 2 months to go	\$25,000 ...
	G2Net Detecting Continuous Gravitational Waves Help us detect long-lasting gravitational-wave signals! Research · 328 Teams · 2 months to go	\$25,000 ...

Estudo de Caso

Estudo de Caso

Dogs vs. Cats

- Dogs vs. Cats
 - 25k imagens rotuladas
 - 12.5k imagens não rotuladas
 - 2 classes
 - 64x64x3
 - 8 bits

Curiosidades:

- 1st place - \$126 donation to the ASPCA or an animal charity of your choosing
- 2nd place - A stuffed Boo



Obrigada!

Prof. Manoela Kohler



prof.Manoela@ica.ele.puc-rio.br



www.linkedin.com/in/manoelakohler