

# INTRODUÇÃO AO MACHINE LEARNING

22/04/2024

MsC Hanna Carla  
Email: [hannacarla47@gmail.com](mailto:hannacarla47@gmail.com)





# O que é Machine Learning?



Machine Learning, ou aprendizado de máquina, é um ramo da inteligência artificial que se concentra no desenvolvimento de algoritmos e técnicas que permitem aos computadores aprenderem a partir de dados e experiências passadas, sem serem explicitamente programados para isso.



# Tipos de Aprendizado

Aprendizado Supervisionado -  
Dados Rotulados

- Classificação
- Previsão/Regressão

Aprendizado Não Supervisionado -  
Dados Não Rotulados

- Clustering
- Redução de dimensionalidade

Aprendizado Semi-Supervisionado -  
Dados Semi-Rotulados

- Algoritmos de AS+ANS

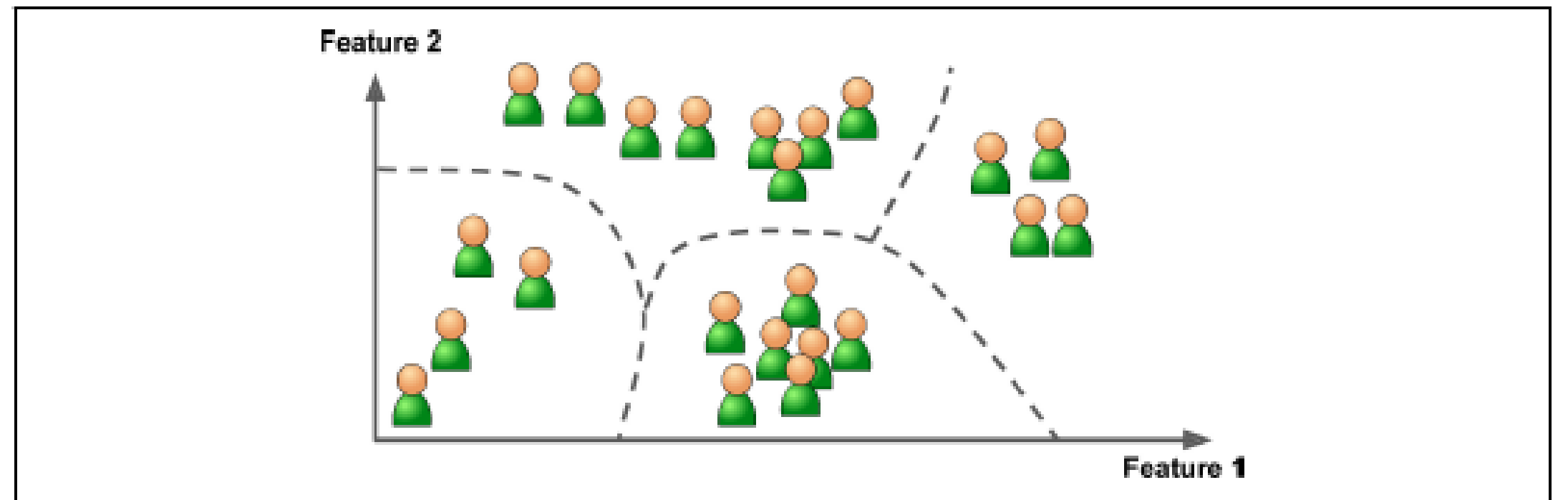
# Algoritmos de Aprendizado Supervisionado

- Regressão Linear
- Regressão Logística
- K-Nearest Neighbours
- Support Vector Machines
- Decision Trees
- Random Forest
- Neural Networks



# Algoritmos de Aprendizado Não Supervisionado

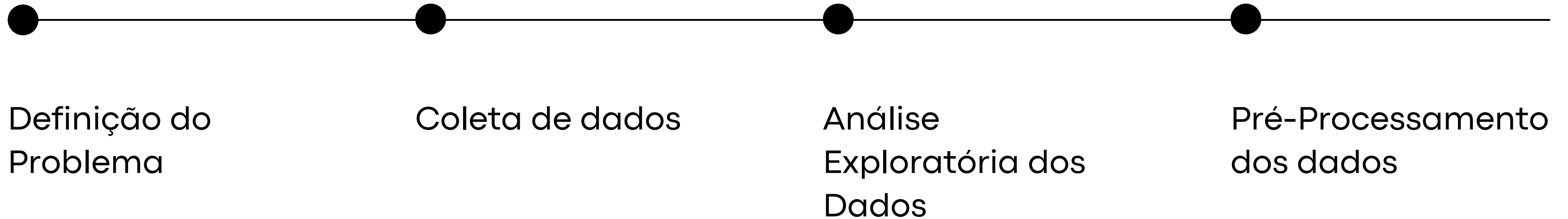
- K-Means
- DBSCAN
- HCA
- PCA



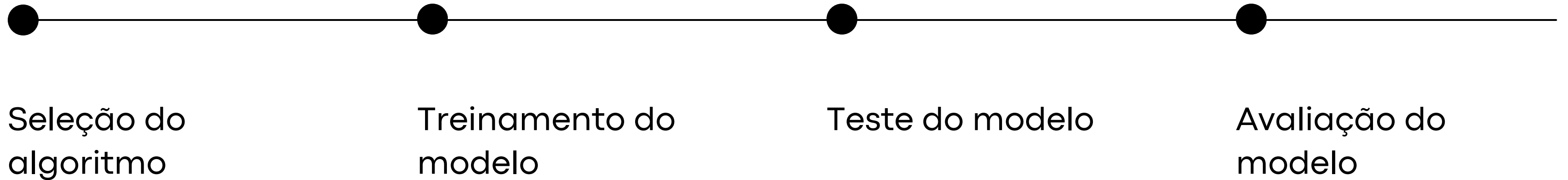
*Figure 1-8. Clustering*

O que deve conter um Projeto de  
ML?

# Linha do tempo



# Linha do tempo





# Análise Exploratória dos Dados

Conheça os seus dados



Visualizar os dados

Estatística descritiva

Identificar valores ausentes ou nulos

Verificar correlações

# Pré-Processamento

Trate os seus dados

- Transformação de variáveis categóricas em variáveis numéricas
- Normalização dos dados
- Detecção de outliers
- Redução de dimensionalidade
- Balanceamento dos dados
- Feature Engineering
- Pipelines

# Desafios do ML



Quantidade Insuficiente de Dados

Dados de Treinamento Não Representativos

Dados de Baixa Qualidade

Overfitting e Underfitting

# Avaliação de Modelos de ML



# Métricas de Performance

Algoritmos Supervisionados

## Classificação

Matriz de Confusão

Acurácia

Precisão

Recall

F1-Score

Curva ROC

Curva Precisão-Recall

## Regressão

Erro Absoluto Médio (MAE)

Erro Quadrático Médio (MSE)

Raiz do MSE (RMSE)

Coeficiente de Determinação  
( $R^2$ )



# Matriz de Confusão

Classificação

		Classe Predita	
		Positiva	Negativa
Classe Real	Positiva	Verdadeiro Positivo	Falso Negativo
	Negativa	Falso Positivo	Verdadeiro Negativo

# Métricas de Performance

## Classificação

### 1. Acurácia (Accuracy):

$$\text{Acurácia} = \frac{\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos} + \text{Verdadeiros Negativos} + \text{Falsos Negativos}}$$

### 2. Precisão (Precision):

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos}}$$

### 3. Revocação (Recall ou Sensibilidade):

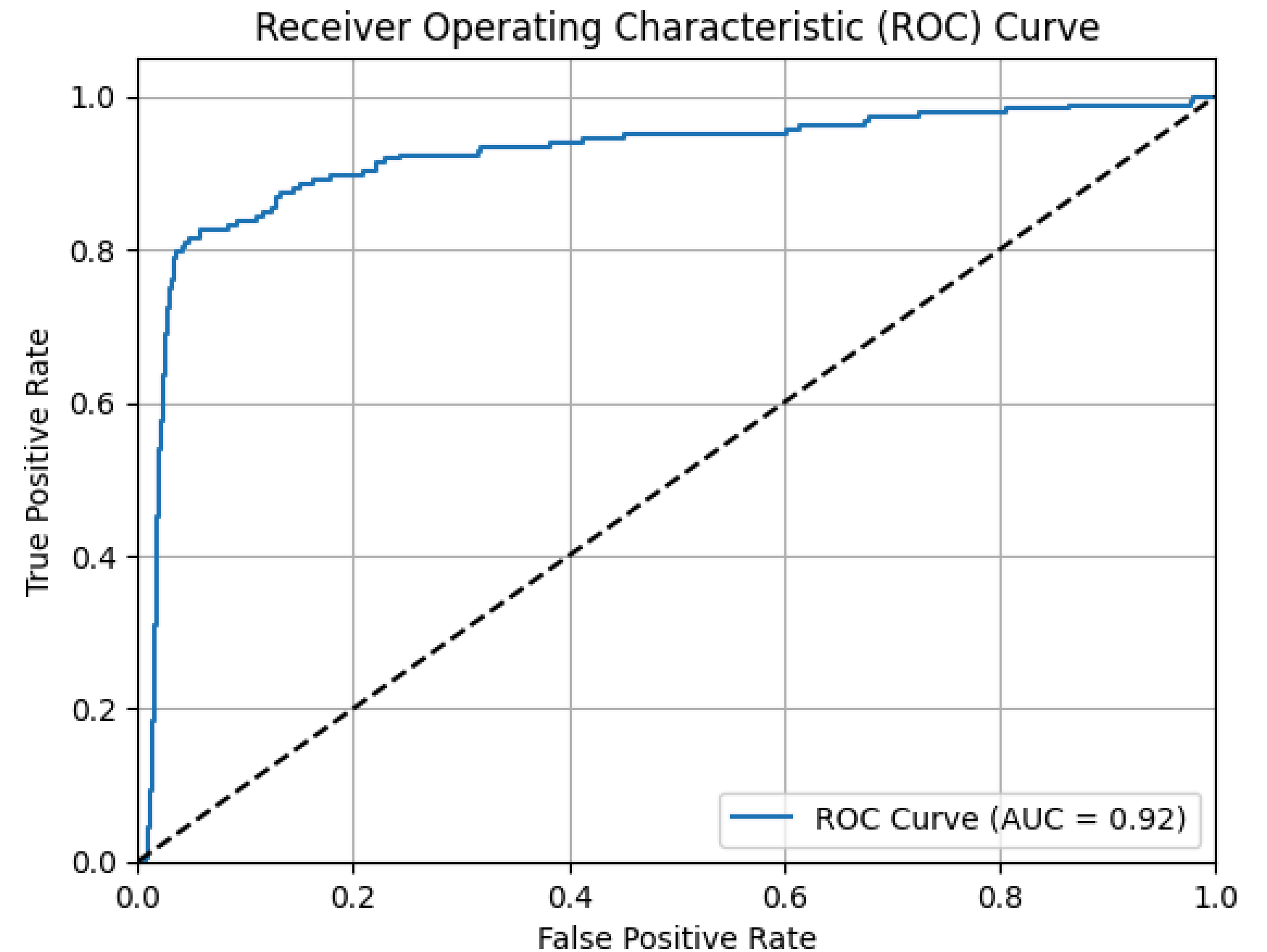
$$\text{Revocação} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}}$$

### 4. F1-Score:

$$\text{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

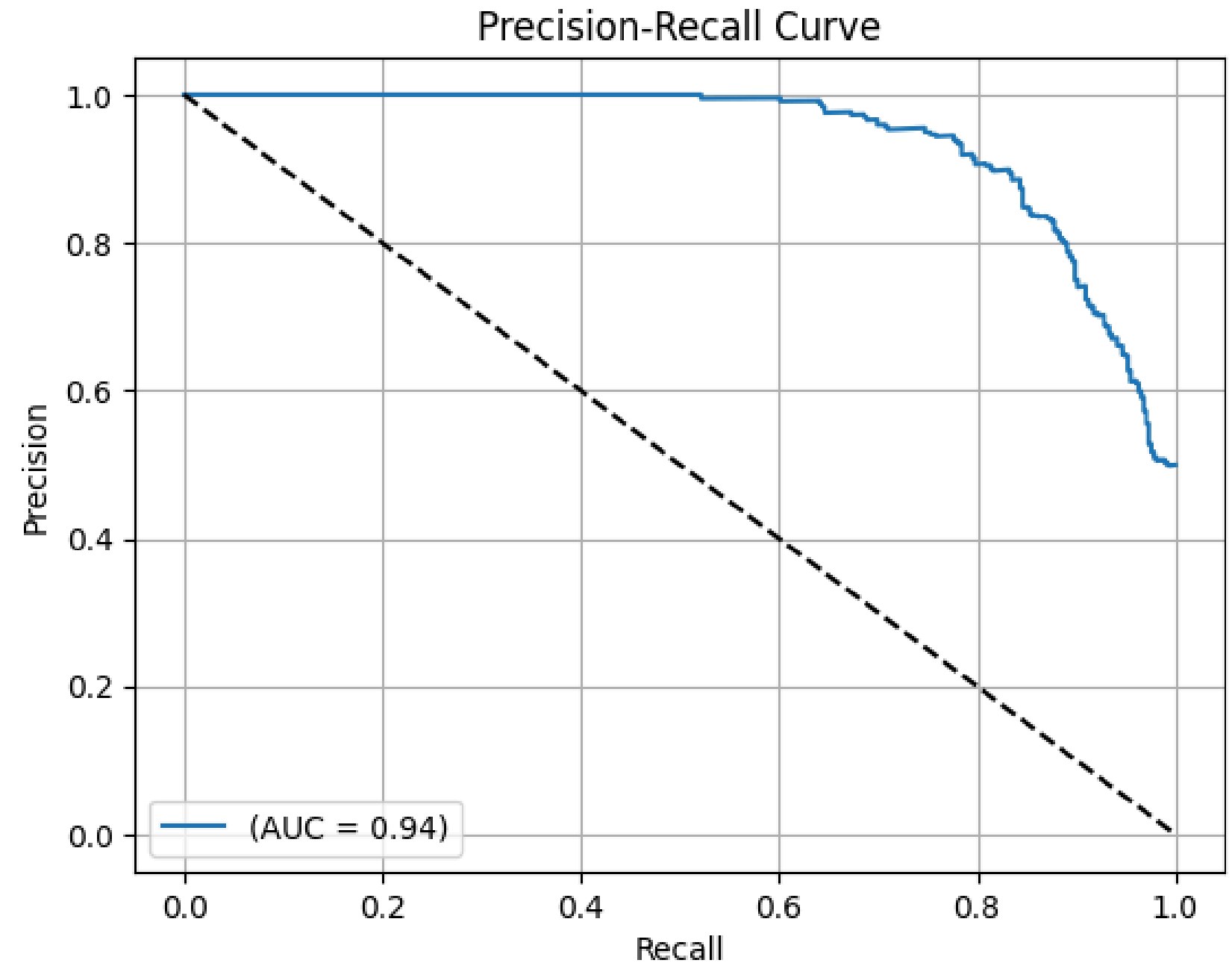
# Curva ROC

A curva ROC plota a taxa de verdadeiros positivos (TPR, do inglês) versus a taxa de falsos positivos (FPR, do inglês)



# Curva Precisão- Recall

A curva precisão-recall é uma representação gráfica da relação entre a precisão e o recall para diferentes valores de limiar de decisão em um modelo de classificação binária.



# Mãos à Obra

<https://github.com/HannaCarla/machine-learning-classes>

HannaCarla/**machine-learning-classes**

Machine learning basics



 1  
Contributor

 0  
Issues

 1  
Star

 0  
Forks



---

**HannaCarla/machine-learning-classes: Machine learning basics**

Machine learning basics . Contribute to HannaCarla/machine-learning-classes development by creating an account on GitHub.

 GitHub