

Modelo preditivo para análise de *customer churn*  
nos serviços de telecomunicações  
usando algoritmos de Machine learning.

*Projeto final de da disciplina de TP 555 – IA e Machine Learning*

- **Docente:** Felipe de Figueiredo.
- **Estudante:** António Luciano. - Mat. 823.
- **Estudante:** Daniel Quiteque. - Mat. 823.

# agenda

- 1 Contextualização.
- 2 Conceitos iniciais. Factores que implicam no *churn*.
- 3 Data minning. Algumas features/Variáveis
- 4 Machine Learning e Modelos preditivos
- 5 Código em python para análise de churn
- 6 Resultados. Trabalhos futuros. Conclusão



# 1.CONTEXTUALIZAÇÃO

Em teoria, o sucesso de qualquer negócio É maiores quando se trabalha a retenção dos clientes, pois geralmente, é mais caro conquistar novos clientes do que manter os já existentes.

Em muitos setores, o custo de aquisição de novos clientes pode ser cinco vezes superior ao seu custo de retenção (KURTZ e CLOW, 1998), mas porém, no dia a dia, é comum vermos as empresas dando pouca atenção a esta análise da retenção de clientes.



# 1.CONTEXTUALIZAÇÃO

---

Clientes podem abandonar uma organização ainda que altos investimentos em prospecção e retenção sejam realizados, o que requer diagnóstico e compreensão.

Em setores altamente competitivos, como no caso do setor de telecomunicações, clientes estão cada vez mais exigentes. E a satisfação dos mesmos desempenha papel crítico para o seu relacionamento e consequentemente, para sua manutenção.

Este é o problema que o presente trabalho deseja analisar.

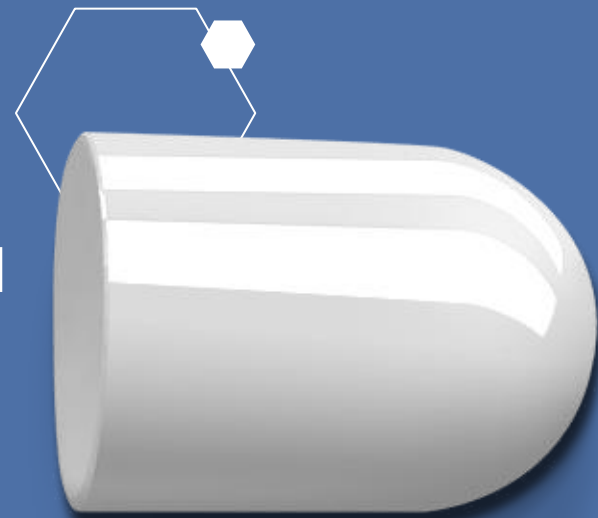
## 2. *Conceitos* INICIAIS

A perda de clientes sofrida por uma empresa, ao longo de um período de tempo específico, também é conhecida como **CHURN RATE**. É uma medida de clientes com atrito, relacionada à taxa anual de turnover da base de clientes (LEJEUNE, 2001). Também conhecido como abandono.





## FATORES QUE INFLUENCIAM *O churn*



*A experiência vivida no atendimento ou qualidade do serviço/produto.*

*O relacionamento da empresa na pré e na pós venda. CRM.*

*O oferta de propostas que superem as suas expectativas.*

*Variáveis macro-econômicas.*



# ***ANÁLISE DO COMPORTAMENTO COM DATA MINNING***

As aplicações de **data minning** (mineração de dados) mostram que a decisão do cliente de interromper o relacionamento com a empresa **acontece meses antes da real interrupção**.

Por isto, a detecção antecipada das mudanças no comportamento dos clientes que sinalizam um potencial abandono ou mudança de comportamento (como a redução do *ticket* médio de compra) é fundamental, ao contrário da detecção apenas durante o evento (BOSE e CHEN, 2009).

# Algumas features/Variáveis

- 1 Necessidades do cliente.
- 2 Preferências.
- 3 Hábitos de compra.
- 4 Condição sócio econômica.

Entre outras...





# 1. ALGORITMOS DE PREDIÇÃO PARA MACHINE LEARNING

Aprendizado de Máquina (machine learning) é uma área da inteligência artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado e a construção de sistemas capazes de adquirir conhecimentos de forma automática (MONARD, BARANAUSKAS, 2003). De acordo com Mitchell (1997).



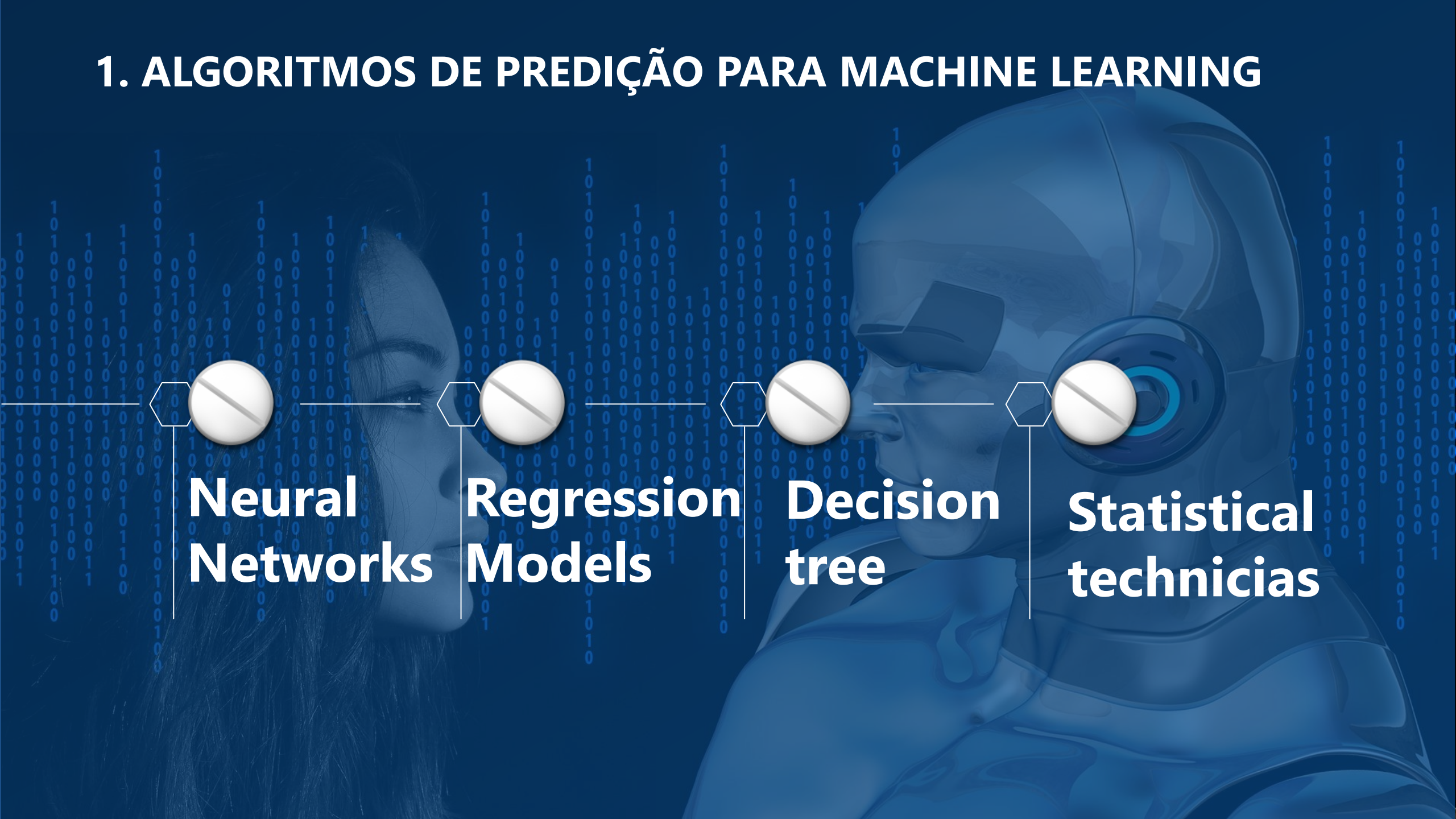


# 1. ALGORITMOS DE PREDIÇÃO EM MACHINE LEARNING

**Algoritmos de machine learning** têm sido de grande valor prático para diversas aplicações, como, problemas de Mineração de Dados (Data Mining), onde grandes banco de dados são analisados automaticamente, na busca de padrões que possam ser úteis para as empresas tomarem decisões.



# 1. ALGORITMOS DE PREDIÇÃO PARA MACHINE LEARNING



**Neural  
Networks**

**Regression  
Models**

**Decision  
tree**

**Statistical  
technicas**





# *Regressão logística*

Um método geralmente usado em data mining para a classificação de clientes, principalmente para a previsão de churn,

Mozer e outros (2000) usaram tal modelo para a previsão da probabilidade de abandono do cliente, usando como variáveis independentes as características individuais de usuários de telefonia celular.

# *Como prever o churn de clientes?*

1. Camada básica para prever o abandono futuro de clientes são os dados do passado.
2. Analisam-se os dados de clientes que já saíram e suas características / comportamento (preditores) antes que o churn acontecesse.
3. Ajustando um modelo estatístico que relaciona dados x resposta, tenta-se prever a resposta para os clientes existentes.

# *Passos para o modelo preditivo*

- 1** Conhecer o negócio/Modelo a a aplicar
- 2** Colecta de dados e Limpeza de dados.
- 3** Seleção das features e Engenharia de dados.
- 4** Machine Learning e Modelo preditivo. Teste.
- 5** Insights e ações a tomar para o negócio

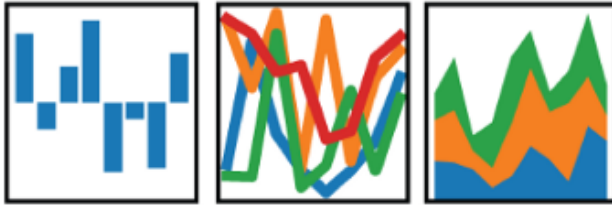




## Ferramentas usadas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



# 5. Código em python para análise de churn

Jupyter Código\_projecto\_Churn\_TP555 Last Checkpoint: uma hora atrás (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

```
In [ ]: # Importante libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
from pylab import rcParams
%matplotlib inline
# Loading the CSV with pandas
data = pd.read_csv('..Customer Churn/Telco-Customer-Churn.csv')
```

# Exploração e seleção de features

Visão geral dos dados de um conjunto de mais de 7043 linhas com 21 colunas

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSup
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	

Fonte de dados: Kaggle.com & IBM.com



# Percentagem de clientes em churn

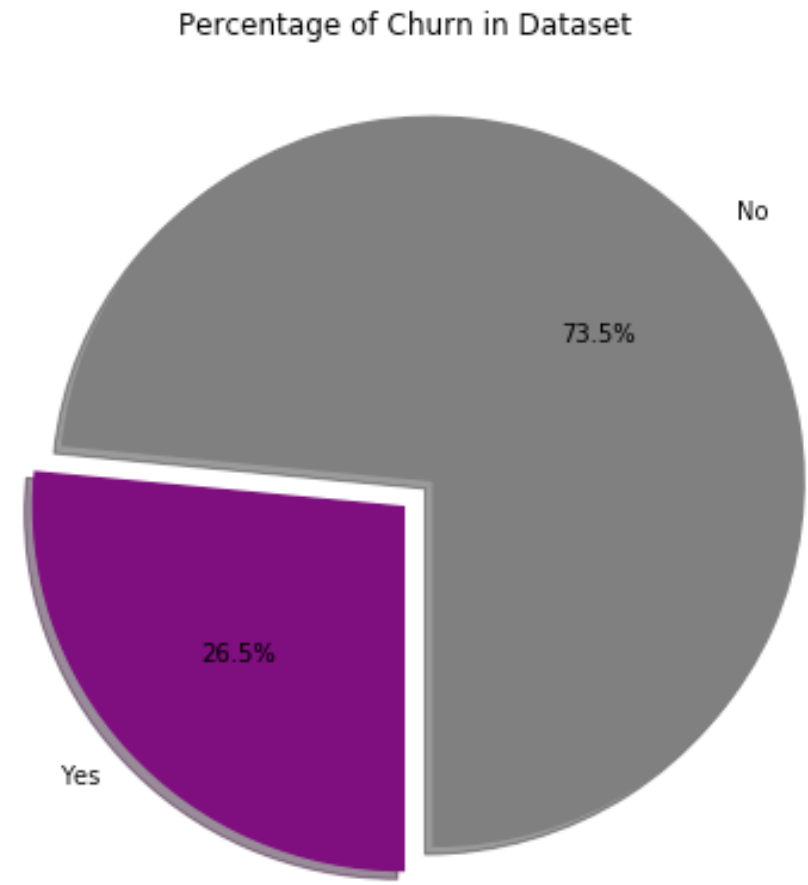
```
jupyter Código_projecto_Churn_TP555 Last Checkpoint: uma hora atrás (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

import warnings
warnings.filterwarnings("ignore")
from pylab import rcParams
%matplotlib inline
# Loading the CSV with pandas
data = pd.read_csv('..Customer Churn/Telco-Customer-Churn.csv')

In [ ]:

In [ ]: # Dados a plotar
sizes = data['Churn'].value_counts(sort = True)
colors = ["grey", "purple"]
rcParams['figure.figsize'] = 5,5
# Plot
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=270,)
plt.title('Percentage of Churn in Dataset')
plt.show()
```



# Trabalho com dados do dataset

Os dois códigos abaixo mostram duas operações:

1-

```
In [ ]: data.drop(['customerID'], axis=1, inplace=True)
```

```
In [3]: data['TotalCharges'] = pd.to_numeric(data['TotalCharges'])
```

# Percentagem de clientes em churn

Rotina para a divisão do dataset em dados de teste e dados de treinamento

```
In [ ]: #Divisão ou separação dos dados em dados de treinamento e dados de teste

data["Churn"] = data["Churn"].astype(int)
Y = data["Churn"].values
X = data.drop(labels = ["Churn"],axis = 1)
# Create Train & Test Data
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=101)
```



# Aplicação da Regressão logística

- **Etapa 1.** Vamos importar o modelo que queremos usar do sci-kit learn
- Etapa 2. Criamos uma instância do modelo
- **Etapa 3.** Está treinando o modelo no conjunto de dados de treinamento e armazenando as informações aprendidas com os dados.

```
In [ ]: # Aplicação da Regressão logística usando sci-kit learn
```

```
from sklearn.linear_model import LogisticRegression  
model = LogisticRegression()  
result = model.fit(X_train, y_train)
```

## 6. Precisão do modelo. Resultado. Conclusão

Vamos importar o modelo que queremos usar do sci-kit learn

- dados.

```
In [ ]: from sklearn import metrics
prediction_test = model.predict(X_test)
# Print the prediction accuracy
print(metrics.accuracy_score(y_test, prediction_test))
```

**Precisão do modelo : 80%**



**Muito  
Obrigado!**

---