

---

# PREVISÃO DE CONTRATAÇÃO BASEADA EM REDE NEURAL

## Grupo 5

Luiz Tokuhara

Rachel Reuters da Silva

Marcelo Rizzardo Lima

Alan Seigneur Alonso

Caio Cesar Moura Barroso

## PLANO DE NEGÓCIO



## Introdução

No cenário atual de recrutamento e seleção, a eficiência na triagem de candidatos é um fator decisivo para o sucesso de uma empresa. A Paschoalotto, uma das líderes no setor de soluções de crédito e atendimento ao cliente, busca continuamente aprimorar seus processos para garantir a contratação dos melhores talentos, alinhados aos valores e objetivos da empresa.

Neste contexto, a aplicação de tecnologias de ponta, como o Machine Learning (ML), surge como uma solução inovadora para otimizar a previsão de contratações bem-sucedidas. Utilizando dados imputados pelos candidatos, como histórico educacional, experiências profissionais, habilidades específicas e respostas a testes comportamentais, o ML pode oferecer uma análise preditiva que auxilia na tomada de decisão de forma mais assertiva e eficiente.



Este plano de negócios propõe o desenvolvimento e implementação de um sistema de previsão de contratações baseado em Machine Learning, especificamente projetado para atender às necessidades da Paschoalotto. Com essa abordagem, a empresa poderá não apenas reduzir o tempo e os custos associados ao processo seletivo, mas também aumentar a precisão na identificação de candidatos com maior potencial de desempenho e adequação à cultura organizacional.

## OBJETIVO

### Objetivo

Desenvolver um modelo de machine learning para prever a decisão de contratação com base em características dos candidatos.

### Persona

Recrutadores que buscam automatizar a seleção de candidatos para aumentar a eficiência do processo de contratação.

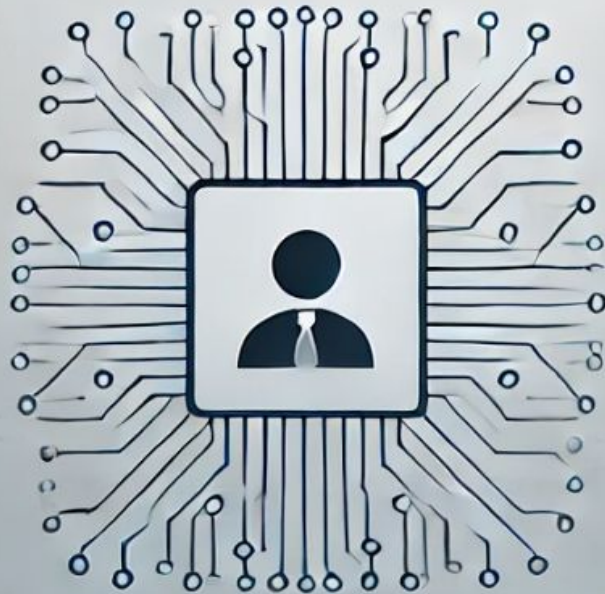
### Modelo

A proposta é utilizar uma Rede Neural Artificial para classificar candidatos como aprovados ou não aprovados com base nas características fornecidas no dataset.

**Fonte:** A base de dados foi obtida de uma fonte fictícia no Kaggle, contendo informações sobre 1500 candidatos com características como idade, nível de educação, anos de experiência, e decisão de contratação.

## MACHINE LEARNING PROJECT: CANDIDATE PATTERN RECOGNITION

AI



OBJECTIVE

PERSONA

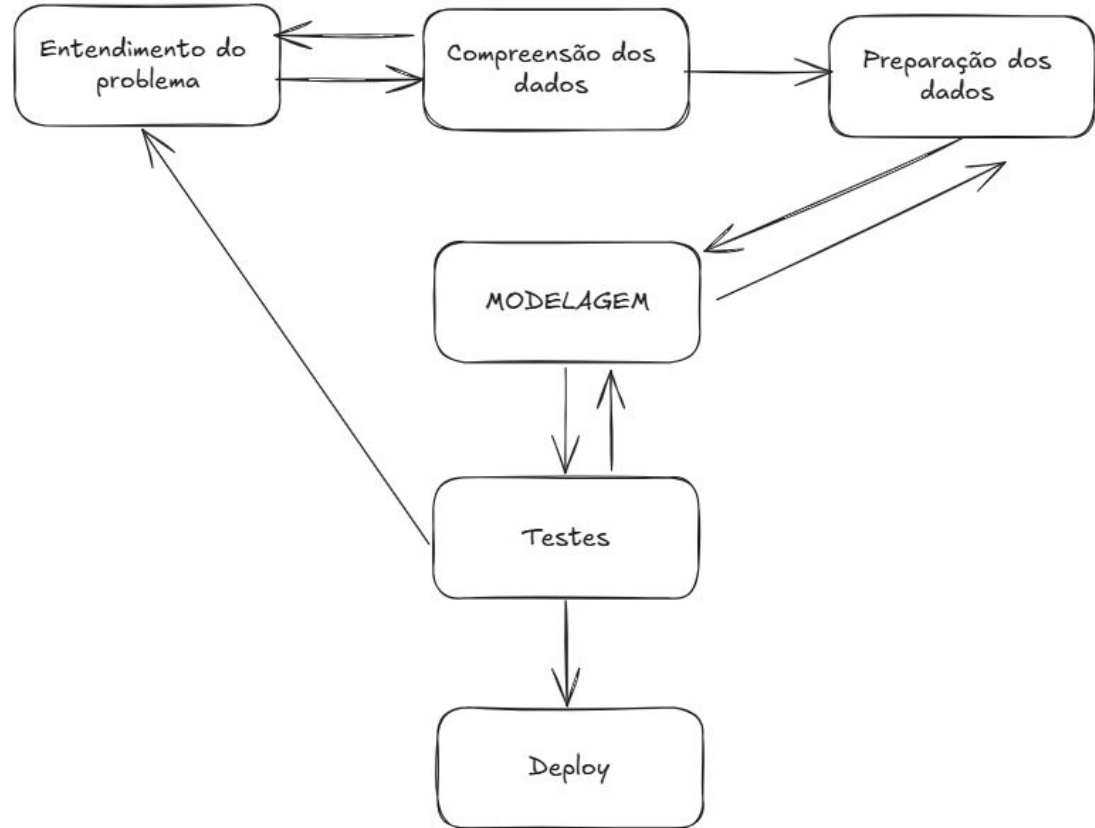


Condições de uso: Este projeto é uma demonstração de machine learning e não deve ser usado para fins comerciais ou legais. Os dados são fictícios e não representam informações reais.

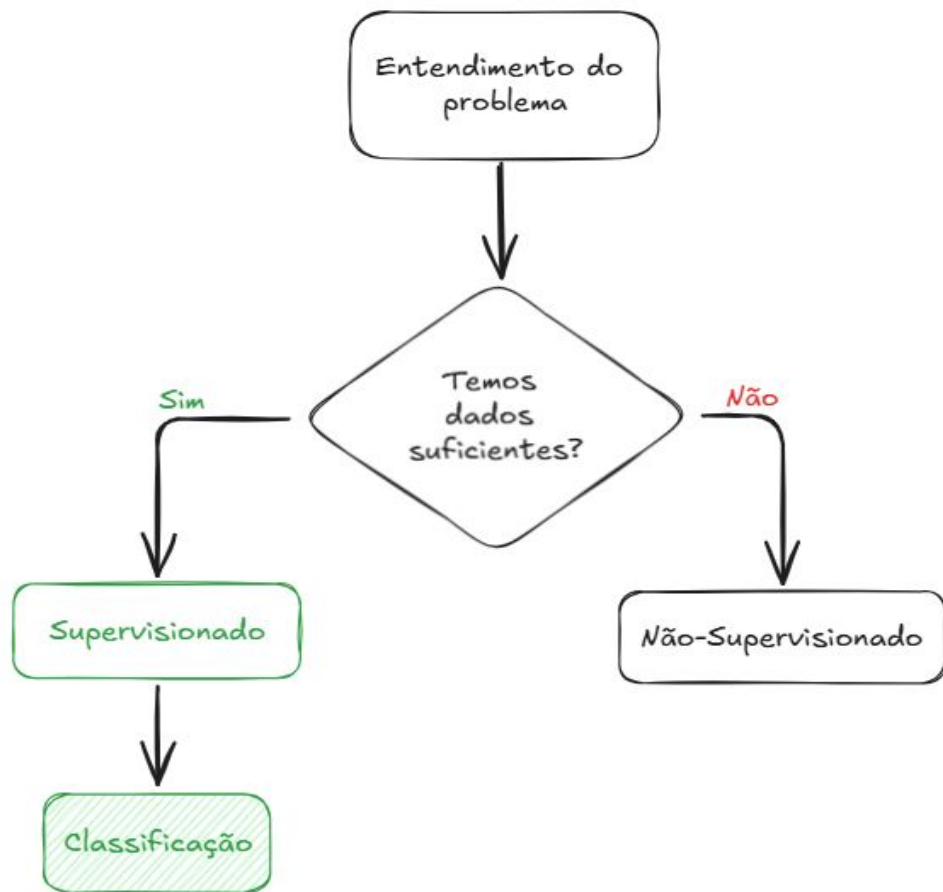
## CRISP DM

**C**Ross  
**I**ndustry  
**S**tandard  
**P**rocess for  
**D**ata  
**M**ining

Modelo de processo de mineração de dados que descreve abordagens comumente usadas por especialistas em mineração de dados para atacar problemas.



## ENTENDIMENTO DO PROBLEMA



### Tipo de Problema

#### Classificação binária

0 - Não aprovado

1 - Aprovado

Y (Target)

Decisão de contratação.

## COMPREENSÃO DOS DADOS

<b>Formato dos dados</b>	CSV
<b>Linhas</b>	1500
<b>Colunas total</b>	11
<b>Colunas de output</b>	1
<b>Colunas de input</b>	10
<b>Dados faltantes?</b>	Não
<b>Dados estão na mesma escala?</b>	Não



# COMPREENSÃO DOS DADOS

## Idade

**Descrição:** Idade do candidato

**Escala:** 20 a 50 anos

**Tipo:** Número Inteiro

## Sexo

**Descrição:** Sexo do candidato

**Escala:** 0 = H ou 1 = M

**Tipo:** Binário

## Nível de Educação

**Descrição:** Maior grau de educação

**Escala:**

**1** = Bacharel tipo 1 | **2** = Bacharel tipo 2

**3** = Mestrado | **4** = PHD

**Tipo:** Categórico

## Pontuação na Entrevista

**Descrição:** Pontuação na entrevista

**Escala:** 0 a 100

**Tipo:** Número Inteiro

## Anos de Experiência

**Descrição:** Anos de experiência

**Escala:** 0 a 15 anos

**Tipo:** Número Inteiro

## Empresas Anteriores

**Descrição:** Quantidade de empresas anteriores que trabalhou

**Escala:** 1 a 5 empresas

**Tipo:** Número Inteiro

## Distância da Empresa

**Descrição:** Distância da residência para a empresa em km

**Escala:** 1 a 50 km

**Tipo:** Número de Ponto Flutuante (decimal)

## Pontuação na Personalidade

**Descrição:** Pontuação na personalidade

**Escala:** 0 a 100

**Tipo:** Número Inteiro

## Pontuação nas Habilidades

**Descrição:** Pontuação nas habilidades técnicas

**Escala:** 0 a 100

**Tipo:** Número Inteiro

## Estratégia de Recrutamento

**Descrição:** Estratégia de recrutamento

**Escala:**

**1** = Agressiva | **2** = Moderada

**3** = Conservadora

**Tipo:** Categórico



## PREPARAÇÃO DOS DADOS

Verificamos que o input "**Recruitment Strategy**" é um dado categórico e não existe uma ordem entre as classes, portanto decidimos aplicar a técnica do **One Hot Encoding**.



Recruitment Strategy
1 = Agressiva
2 = Moderada
3 = Conservadora



Recruitment Strategy 1
1
0
0

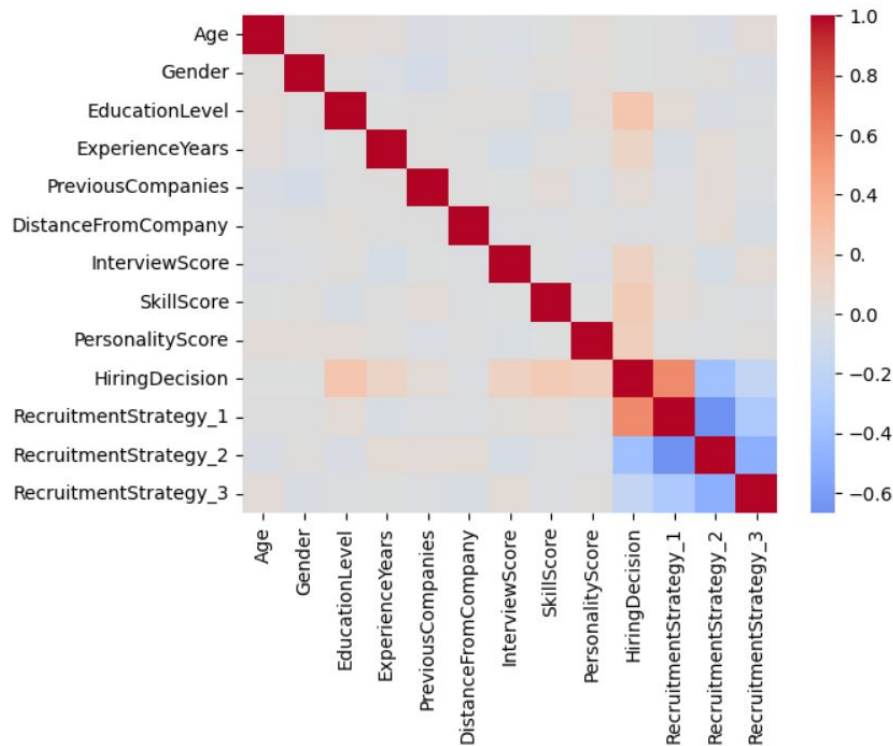
Recruitment Strategy 2
0
1
0

Recruitment Strategy 3
0
0
1



## PREPARAÇÃO DOS DADOS

A fim de escolher as *features* que mais trazem valor ao problema, verificamos a correlação dos dados



### Features Selecionadas

Education Level

Skill Score

Personality Score

Interview Score

Experience Years

Recruitment Strategy

# PREPARAÇÃO DOS DADOS

## Tipo de Problema

Classificação binária

- 0 - Não aprovado
- 1 - Aprovado

## Y (Target)

- Decisão de contratação

## X (Features)

- Nível de Educação
- Nota de Habilidade
- Nota de Personalidade
- Nota da Entrevista
- Anos de Experiência
- Estratégia de Recrutamento

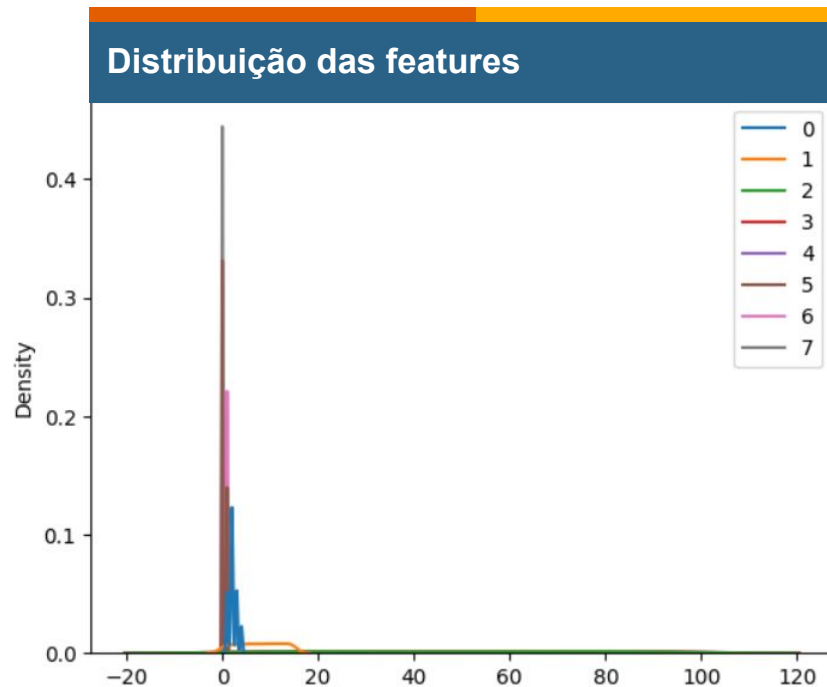
## Correlação utilizando 3 métodos diferentes

	kendal	pearson	spearman
RecruitmentStrategy_1	0.571330	0.571330	0.571330
EducationLevel	0.214417	0.236710	0.230218
SkillScore	0.167029	0.203668	0.203484
PersonalityScore	0.137927	0.169177	0.168043
InterviewScore	0.123372	0.146064	0.150309
ExperienceYears	0.102202	0.122494	0.121406
PreviousCompanies	0.039412	0.044025	0.044063
Age	0.002075	0.001850	0.002500
Gender	-0.002249	-0.002249	-0.002249
DistanceFromCompany	-0.013621	-0.016791	-0.016676
RecruitmentStrategy_3	-0.173982	-0.173982	-0.173982
RecruitmentStrategy_2	-0.385584	-0.385584	-0.385584

# PREPARAÇÃO DOS DADOS

## Configuração das Features Escolhidas

- Nível de Educação: 1 a 4. Categórico. Não aplicamos o OHE pois existe uma ordem intrínseca entre os níveis.
- Nota de Habilidade: 0 a 100
- Nota de Personalidade: 0 a 100
- Nota da Entrevista: 0 a 100
- Anos de Experiência: 0 a 15
- Estratégia de Recrutamento: 1 a 3. Categórico. Foi aplicado o OHE.

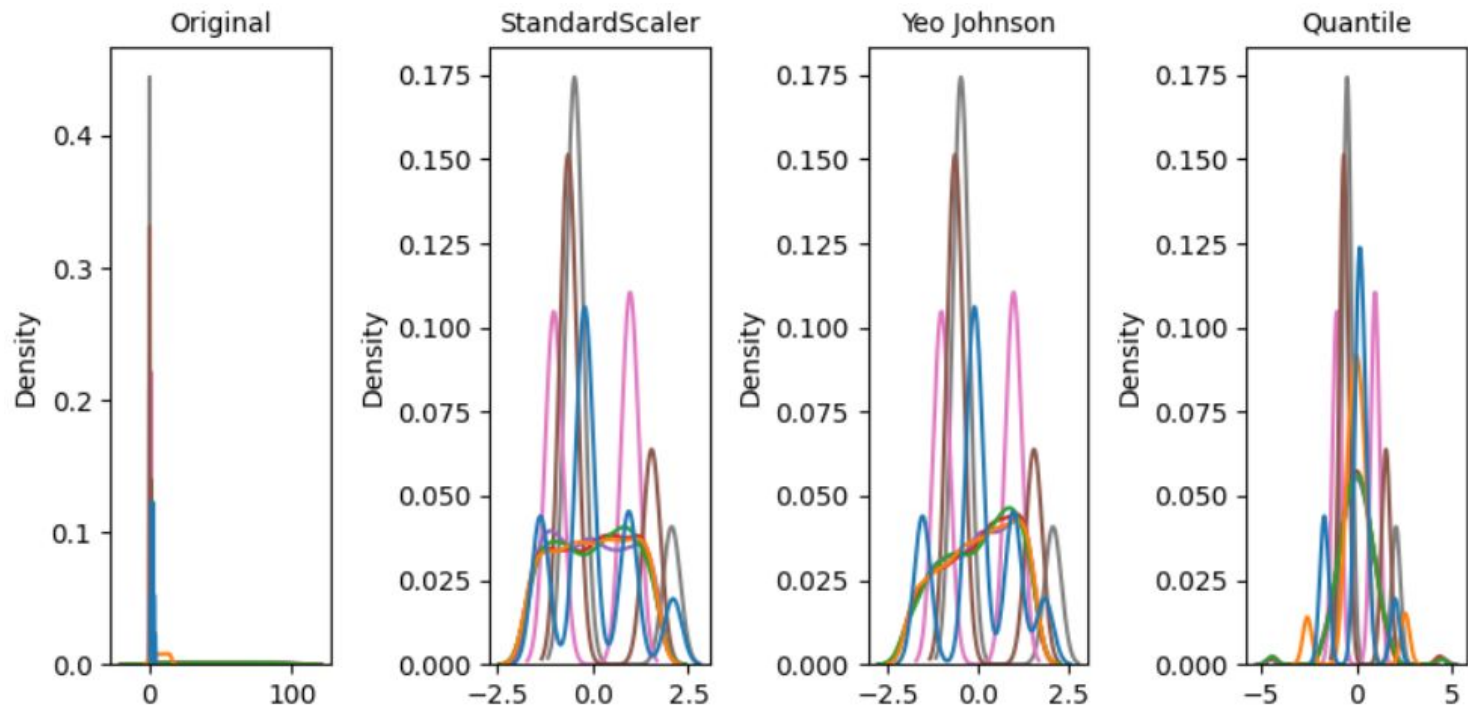


Dados estão normalizados?

Resposta: Não

# PREPARAÇÃO DOS DADOS

## NORMALIZAÇÃO DOS DADOS



## Ativação na Camada Oculta

### Função Tanh (Tangente Hiperbólica).

A ativação Tanh mapeia os valores de entrada para o intervalo de -1 e 1.

## Ativação na Camada de Saída

### Função Sigmóide

Apropriada para problemas de classificação binária, isso permite que a saída seja interpretada como uma probabilidade.

## Quantificação das Variáveis

Todas as variáveis foram codificadas como valores numéricos inteiros.

## Normalização

Os dados foram normalizados para a faixa [-5, 5].

## Camadas Ocultas

Uma camada Tanh, testado com GridSearch, com 8 e 16 neurônios, o que foi descoberto que 16 performa melhor para acurácia.

## Camadas de Saída

Um neurônio, utilizando a ativação Sigmóide.

### Erro de Generalização

Avaliado usando K-Folds Cross Validation (5 folds).

### Seleção de Hiperparâmetros

Grid Search

1500 épocas

Paciência em 150

Validação cruzada para ajustar a taxa de aprendizado, número de neurônios de 8 e 16, número de camadas e 5-Fold.

Adam como otimizador

### Critério de Parada Prematura

Monitoramento do `val_loss` durante o treinamento para evitar overfitting.

### Figura de Mérito

Acurácia x Sensibilidade x Precisão?

### Acurácia

- Busca otimizar a quantidade de acertos dentre as previsões realizadas, seja contratando ou não contratando.
- Ideal para mercados nos quais o custo de contratar um funcionário inadequado seja equivalente ao valor de um bom funcionário.

### Sensibilidade (recall)

- Busca otimizar a contratação de todos bons funcionários, assumindo a contratação de alguns inadequados.
- Ideal para mercados nos quais é muito importante garantir que bons funcionários sejam contratados, pois seu valor agregado compensa algumas contratações inadequadas.

### Precisão

- Busca otimizar a quantidade de bons funcionários entre os contratados, minimizando os inadequados.
- Ideal para mercados nos quais é muito importante garantir que os funcionários contratados sejam bons, pois o valor de um funcionário inadequado é alto.

## DETERMINAÇÃO DA FIGURA DE MÉRITO

Com base nas definições discutidas anteriormente, torna-se claro que a análise da figura de mérito fica intrinsecamente ligado ao mercado para o qual estamos contratando.

Dessa forma, para atender melhor às necessidades de cada cliente devemos considerar qual figura de mérito se encaixa melhor em qual tipo de mercado.

Além disso, outros fatores como senioridade dos cargos a serem preenchidos também devem ser levados em conta.

1

Caso o cliente seja um hospital novo que deve contratar uma alta quantidade de profissionais da saúde (como médicos e enfermeiros), o custo associado à um funcionário inadequado que terá mais chance de cometer um erro grave é muito alto, logo nosso modelo priorizaria uma alta precisão.

2

Em um outro caso onde nosso cliente seja uma investidora que tenha bons mecanismos de parada em caso de maus investimentos por parte dos seus funcionários, o valor agregado de um ótimo funcionário pode compensar com sobras o custo de outros inadequados, dessa forma nosso modelo priorizaria uma alta sensibilidade.

3

Por outro lado, uma para a contratação de cargos CLT em massa, como em telemarketing, onde há uma variação menor entre o custo de um mau funcionário e o valor de um bom funcionário, buscamos apenas minimizar os erros, priorizando uma alta acurácia.



## DETERMINAÇÃO DA FIGURA DE MÉRITO

Com base nas definições discutidas anteriormente, torna-se claro que a análise da figura de mérito fica intrinsecamente ligado ao mercado para o qual estamos contratando.

Dessa forma, para atender melhor às necessidades de cada cliente devemos considerar qual figura de mérito se encaixa melhor em qual tipo de mercado.

Além disso, outros fatores como senioridade dos cargos a serem preenchidos também devem ser levados em conta.

1

Caso o cliente seja um hospital novo que deve contratar uma alta quantidade de profissionais da saúde (como médicos e enfermeiros), o custo associado à um funcionário inadequado que terá mais chance de cometer um erro grave é muito alto, logo nosso modelo priorizaria uma alta precisão.

2

Em um outro caso onde nosso cliente seja uma investidora que tenha bons mecanismos de parada em caso de maus investimentos por parte dos seus funcionários, o valor agregado de um ótimo funcionário pode compensar com sobras o custo de outros inadequados, dessa forma nosso modelo priorizaria uma alta sensibilidade.

3

Por outro lado, uma para a contratação de cargos CLT em massa, como em telemarketing, onde há uma variação menor entre o custo de um mau funcionário e o valor de um bom funcionário, buscamos apenas minimizar os erros, priorizando uma alta acurácia.

## RESULTADOS INICIAIS

### GridSearch

- 1 camada oculta
- 8 ou 16 neurônios
- 1500 épocas
- Ativação = [tanh, relu]

VS

### RandomForest

- 100 Estimadores
- 30% de grupo de teste 70% treino

Acurácia: 0.9177777777777778

Precisão: 0.9076923076923077

Recall: 0.8251748251748252

Especificidade: 0.9609120521172638

F1 Score: 0.8644688644688644

ROC AUC Score: 0.91945513769618

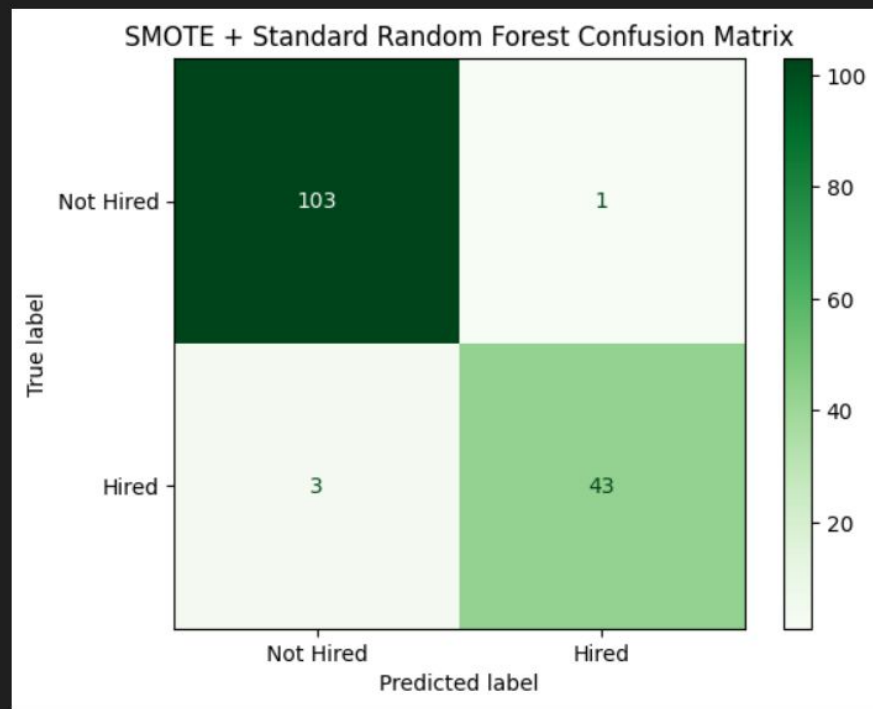
```
Best: 0.924667 using {'epochs': 1500, 'model_activation': 'tanh', 'model_neurons': 16}
0.922000 (0.066953) with: {'epochs': 1500, 'model_activation': 'tanh', 'model_neurons': 8}
0.924667 (0.072498) with: {'epochs': 1500, 'model_activation': 'tanh', 'model_neurons': 16}
0.856667 (0.065727) with: {'epochs': 1500, 'model_activation': 'relu', 'model_neurons': 8}
0.869333 (0.059007) with: {'epochs': 1500, 'model_activation': 'relu', 'model_neurons': 16}
```

### Random Forest com técnica SMOTE

Mean accuracy: 0.942

Mean recall: 0.916

Mean precision: 0.967



## CONCLUSÃO

Este plano de negócios propõe o desenvolvimento e implementação de um sistema de previsão de contratações baseado em Machine Learning, especificamente projetado para atender às necessidades da Paschoalotto. Com essa abordagem, a empresa poderá não apenas reduzir o tempo e os custos associados ao processo seletivo, mas também aumentar a precisão na identificação de candidatos com maior potencial de desempenho e adequação à cultura organizacional.



# FONTES

1. [http://www.nce.ufrj.br/labic/downloads/dicas\\_cfg\\_rna.pdf](http://www.nce.ufrj.br/labic/downloads/dicas_cfg_rna.pdf)
2. <https://medium.com/@reinapeh/16-data-feature-normalization-methods-using-python-with-examples-part-2-of-6-4224c9699253>
3. <https://machinelearningmastery.com/grid-search-hyperparameters-deep-learning-models-python-keras/>
4. <https://medium.com/@gustavocamara/classifica%C3%A7%C3%A3o-com-random-forest-6c508a046537>
5. <https://meridian.allenpress.com/jim/article/47/1/31/131479/Random-Forest>
6. <https://medium.com/sfu-csmp/surviving-in-a-random-forest-with-imbalanced-datasets-b98b963d52eb>

Link para o git : [posGraduacaoIA/RedesNeurais/TRABALHO\\_FINAL/recruitment\\_modeling.ipynb](https://github.com/rachelreuters/posGraduacaoIA) at main · [rachelreuters/posGraduacaoIA \(github.com\)](https://github.com/rachelreuters/posGraduacaoIA)