

Proposta Kátia Cardoso

Modelos de Aprendizado de Máquina - DatAí Consultoria

versão 1.0.0 - 09/05/2025

[Sumário executivo](#_65r62cryojr2)

[Visão Geral](#_y9j38669v5x7)

[Objetivo principal](#_gwkoaonopuxs)

[Resultado esperado](#_5lohbleczb2s)

[Escopo do projeto](#_6i3x4kco45yj)

[Inclui:](#_qi2a6i1uzmyh)

[Não inclui:](#_efber2cwq6q9)

[Objetivos e Metodologia](#_9h24ezbw1fdz)

[Etapas e cronogramas](#_bj923hvscfy1)

[Entregáveis](#_j0kg9m5lm8g1)

[Precificação](#_tb2wf92jef9j)

[Considerações finais](#_tlnpizsbm64v)

[Sobre a DatAí Consultoria](#_dqhmb4pn6j7k)

[Contatos](#_j7c2sjx1t6z0)

# Sumário executivo

## Visão Geral

O projeto é uma plataforma para conectar encarregados de dados a organizações públicas e privadas que necessitam de um responsável pelos seus dados. A DatAí entrará com uma consultoria para modelar um algoritmo de classificação.

## Objetivo principal

A DatAí tem o papel de desenvolver algoritmos para classificação de organizações públicas e privadas em relação ao nível da competência do encarregado de dados, e/ou classificação de pessoas em relação ao nível de aprendizado sobre os conhecimentos da LGPD.

## Resultado esperado

Um relatório descrevendo os resultados dos modelos testados para classificar entidades ou pessoas.

# Escopo do projeto

## Inclui:

* Tratamento de dados (caso necessário);
* Relatório de análise exploratória (caso desejado);
* Engenharia de atributos;
* Treinamento e validação de modelos de classificação;
* Relatório final com resultados e insights;

## Não inclui:

* Implementação do modelo com sistemas externos;
* Coleta de dados;
* Manutenção contínua do modelo após a entrega;

# Objetivos e Metodologia

Classificar entidades públicas, assim como pessoas em três categorias (alto, médio, e baixo nível) com base em atributos fornecidos. Os modelos serão treinados utilizando a linguagem de programação ***Python*** utilizando bibliotecas desenvolvidas pela comunidade para o desenvolvimento de modelos de aprendizado de máquina. Para a avaliação dos modelos serão utilizadas as métricas de acurácia, *F1-score*, e matriz de confusão.

Antes de iniciar um projeto de aprendizado de máquina é necessário ter um problema a ser resolvido. O problema já está definido, classificação de entidades ou pessoas no seu nível de necessidade/conhecimento sobre a LGPD. Depois do problema definido é necessário ter dados para modelar, no caso será fornecida pela cliente Kátia os dados. Depois da obtenção dos dados é avaliado a qualidade dos dados, dependendo desta qualidade é necessário fazer transformações, e/ou limpezas de dados. Após a obtenção de dados será avaliado se há a necessidade da etapa de transformação e/ou limpeza.

Para o desenvolvimento de qualquer projeto de aprendizado de máquina é necessário fazer uma etapa de análise exploratória de dados, caso seja desejado, será feito um relatório desta etapa.

Após avaliar a natureza dos dados, como as estatísticas, comportamento das variáveis, é iniciado a modelagem dos algoritmos. Após a modelagem dos algoritmos, que é a definição dos hiperparâmetros de cada algoritmo. Então os dados são divididos em duas ou três amostras, que são denominadas amostras de treino, teste e validação, onde os algoritmos aprenderam com estes dados, após isto serão calculados as métricas dos modelos, para a comparação dos modelos e então eleger o melhor modelo. A amostra de validação é utilizada para comparação dos modelos e será feita somente se houver uma quantidade razoável de dados.

Após todas as etapas descritas será feito o relatório final dos modelos. Avaliando cada métrica, e desempenho dos modelos.

# Etapas e cronogramas

| Etapas | Tempo |
| --- | --- |
| Análise da qualidade de dados | 1 a 3 dias |
| Tratamento/limpeza de dados | 0 a 2 dias |
| Análise exploratória dos dados | 5 a 10 dias |
| Modelagem dos algoritmos | 5 a 7 dias |
| Relatório final | 5 a 7 dias |

# 

# 

# 

# Entregáveis

1. Relatório final dos modelos;
2. Relatório da análise exploratória, caso solicitado;

# Precificação

| Etapas | Preço unitário |
| --- | --- |
| Tratamento/limpeza de dados | R$100,00~400,00 |
| Análise exploratória dos dados | R$200,00~500,00 |
| Modelagem dos algoritmos | R$1.500,00~4.000,00 |
| Total | R$1.800,00~4.900,00 |

# Considerações finais

Desde o início do diagnóstico do problema, deixamos claro que desenvolver um modelo de inteligência artificial pode ser eficaz para a resolução do problema específico. Nós não nos comprometemos a encontrar modelos com desempenho bom, ou alcançar determinadas métricas de avaliação de modelos. Vale lembrar que não será feita implementação do modelo a qualquer sistema externo, para este momento será apenas feito apenas a modelagem.

# Sobre a DatAí Consultoria

Somos uma empresa júnior de consultoria em ciência de dados e inteligência artificial. Nossa missão é solucionar problemas, por meio de dados, tecnologia e inovação. Como visão para a nossa empresa queremos tornar-nos uma empresa de referência no mercado de ciência de dados conquistando os nossos clientes por meio de soluções inovadoras e de alta qualidade.

# Contatos

Em caso de dúvidas, poderá entrar em contato:

* Felipe Toledo Neves, responsável do projeto
* [felipe.neves@iesb.edu.br](mailto:felipe.neves@iesb.edu.br)
* (61) 996.882.384
* [datai.tec.br](http://datai.tec.br)