LMF VISTORIAS

Donizete Marcos Gomes

Mariana Borges Curvêlo

Nátalia Silveira Toledo

Pedro Valadares Junior

MODELO PREDITIVO DE RISCO E PRIORIZAÇÃO DE VISTORIA (MPRPV): UTILIZAÇÃO DE DATA SCIENCE PARA OTIMIZAÇÃO DE VISTORIAS IMOBILIÁRIAS

Sumário

[1 INTRODUÇÃO 3](#_Toc212274954)

[1.1 Contextualização 3](#_Toc212274955)

[1.2 Problema de Pesquisa 3](#_Toc212274956)

[1.3. Problema Central 3](#_Toc212274957)

[1.4. Detalhamento do Problema 4](#_Toc212274958)

[1.4.1. Eixo 1: Risco e Qualidade da Avaliação (Inconsistência) 4](#_Toc212274959)

[1.4.2. Eixo 2: Eficiência e Produtividade (Ineficiência Operacional) 4](#_Toc212274960)

[1.5. Solução Proposta pelo Projeto (MPRPV) 4](#_Toc212274961)

[1.5.1 Prever o Risco (Preditivo): O modelo MPRPV calcula a probabilidade de dano antes da visita. 4](#_Toc212274962)

[1.3 Objetivos 5](#_Toc212274963)

[1.3.1 Objetivo Geral 5](#_Toc212274964)

[1.3.2 Objetivos Específicos 5](#_Toc212274965)

[1.4 Justificativa 6](#_Toc212274966)

[2 REFERENCIAL TEÓRICO 7](#_Toc212274967)

[2.1 A Vistoria Imobiliária e o Rigor Técnico 7](#_Toc212274968)

[2.2 Aplicações de Machine Learning no Setor Imobiliário (*Proptech*) 7](#_Toc212274969)

[2.3 Visão Computacional (*Computer Vision*) na Construção Civil 7](#_Toc212274970)

[3 METODOLOGIA 8](#_Toc212274971)

[3.1 Tipo e Abordagem 8](#_Toc212274972)

[3.2 Coleta e Tratamento de Dados 8](#_Toc212274973)

[3.3 Desenvolvimento do Modelo Preditivo (MPRPV) 8](#_Toc212274974)

[3.3.1 Modelo de Classificação de Risco 8](#_Toc212274975)

[3.3.2 Modelo de Visão Computacional (Suporte ao Laudo) 9](#_Toc212274976)

[3.4 Avaliação e Validação 9](#_Toc212274977)

[4 RECURSOS E CRONOGRAMA 10](#_Toc212274978)

[4.1 Recursos Necessários 10](#_Toc212274979)

[4.2 Cronograma (Plano de Ação - 6 Meses) 11](#_Toc212274980)

[APÊNDICE A - DICIONÁRIO DE DADOS PRELIMINAR 12](#_Toc212274981)

[ANEXO A - DOCUMENTAÇÃO TÉCNICA 13](#_Toc212274982)

# 1 INTRODUÇÃO

## 1.1 Contextualização

O setor de vistorias imobiliárias, vital para a mitigação de riscos em transações de locação e compra e venda, é tradicionalmente caracterizado por processos manuais e alta dependência da percepção humana. Na LMF Vistorias, uma empresa que busca se estabelecer no mercado por meio da excelência e da inovação, a necessidade de padronização e otimização do processo é crítica. O volume crescente de dados gerados em cada laudo e foto representa uma oportunidade inexplorada.

A Ciência de Dados (Data Science) permite transformar este acervo histórico de informações em inteligência estratégica, reduzindo a subjetividade, aumentando a precisão dos laudos e, crucialmente, elevando a eficiência operacional da equipe, um fator de diferenciação importante para uma empresa em fase inicial.

## 1.2 Problema de Pesquisa

Considerando o desafio de uma empresa nova em otimizar recursos e garantir alta precisão técnica, como a aplicação de técnicas avançadas de *Machine Learning* e *Visão Computacional* pode criar um Sistema Preditivo de Risco e Anomalias, utilizando dados históricos da LMF Vistorias para otimizar a qualidade e a eficiência de cada inspeção?

## 1.3. Problema Central

A LMF Vistorias tem uma alta variabilidade na qualidade dos laudos e um processo de vistoria ineficiente, porque não possui um método padronizado e preditivo para direcionar a atenção do vistoriador aos danos mais prováveis ou graves.

## 1.4. Detalhamento do Problema

Este problema se manifesta em dois eixos principais:

### 1.4.1. Eixo 1: Risco e Qualidade da Avaliação (Inconsistência)

* + Situação Atual: O tempo e o foco do vistoriador são distribuídos igualmente em todos os imóveis, independentemente do seu risco real de apresentar patologias graves (como vazamentos ou rachaduras estruturais).
  + Consequência: A avaliação de danos é subjetiva e depende da experiência individual do técnico. Isso leva a inconsistência nos laudos, aumentando o risco de *falsos negativos* (defeitos não identificados) e abrindo margem para futuras disputas e prejuízos.

### 1.4.2. Eixo 2: Eficiência e Produtividade (Ineficiência Operacional)

* + Situação Atual: Sendo uma empresa nova, a LMF precisa maximizar a produtividade. O processo atual exige que o vistoriador gaste tempo analisando minuciosamente áreas de baixo risco.
  + Consequência: Baixa produtividade da equipe, tempo de laudo prolongado e alto custo operacional. Não há um sistema de priorização que permita que o vistoriador se concentre nos 20% do imóvel que geram 80% dos problemas (*Princípio de Pareto*).

## 1.5. Solução Proposta pelo Projeto (MPRPV)

O projeto de Ciência de Dados resolve este problema ao:

### 1.5.1 Prever o Risco (Preditivo): O modelo MPRPV calcula a probabilidade de dano antes da visita.

1. Otimizar a Ação (Diretivo): O vistoriador recebe um "Checklist Focado" (o resultado da predição), sabendo exatamente quais pontos inspecionar com maior rigor, reduzindo o tempo de vistoria sem sacrificar a qualidade nos pontos críticos.
2. Padronizar o Laudo (Objetivo): O modelo de Visão Computacional oferece uma avaliação objetiva (máquina) do dano, reduzindo a subjetividade do texto no laudo final.

## 1.3 Objetivos

### 1.3.1 Objetivo Geral

Desenvolver, validar e implementar um Modelo Preditivo de Risco e Priorização de Vistoria (MPRPV), com o intuito de *estimar a probabilidade e o tipo de danos em imóveis antes da visita*, qualificando a alocação de tempo e recursos da LMF Vistorias.

### 1.3.2 Objetivos Específicos

a) Coleta e Tratamento: Consolidar, limpar e harmonizar os dados estruturados (características do imóvel e localização) e não estruturados (fotos e textos de laudos) do acervo da empresa.

b) Modelagem de Risco: Construir um modelo de classificação capaz de gerar uma Pontuação de Risco para cada novo imóvel, indicando a probabilidade de ocorrência de danos graves ou específicos.

c) Automação da Análise de Imagem (Visão Computacional): Treinar algoritmos de *Deep Learning* para a identificação e categorização automática de anomalias visuais (e.g., rachaduras, infiltrações, mofo) nas fotos tiradas em campo.

d) Integração Operacional: Desenvolver um Módulo de *Business Intelligence* (BI) para visualizar os resultados preditivos e integrá-los como um *checklist* customizado para a equipe de vistoria.

## 1.4 Justificativa

O projeto justifica-se pela busca por excelência e eficiência operacional, pilares essenciais para o crescimento da LMF Vistorias:

* Posicionamento de Mercado: Diferenciação por meio da tecnologia e entrega de laudos com maior nível de detalhe e menor subjetividade, em conformidade com as exigências de rigor técnico (ABNT NBR 14653).
* Otimização de Custos (Empresa Nova): A alocação inteligente do tempo do vistoriador em áreas de risco real minimiza o retrabalho e otimiza a produtividade da equipe, impactando diretamente o *bottom line* da empresa.
* Previsão e Mitigação de Riscos: A identificação precoce de imóveis com alto risco de danos graves auxilia a empresa e seus clientes na proteção contra responsabilidades futuras.
* Inteligência de Dados: O projeto cria a base para futuras análises estratégicas, como a avaliação de desgaste médio por tipo de material ou construtora.

# 2 REFERENCIAL TEÓRICO

## 2.1 A Vistoria Imobiliária e o Rigor Técnico

Conforme a ABNT NBR 14653-1, a vistoria é um procedimento que visa o conhecimento e a caracterização do bem a ser avaliado. O Projeto Integrador de Ciências de Dados atua no aprimoramento desta etapa, fornecendo ao profissional um subsídio técnico-científico para orientar sua inspeção, transformando a experiência subjetiva em dado quantificável e rastreável.

## 2.2 Aplicações de Machine Learning no Setor Imobiliário (*Proptech*)

O uso de modelos preditivos e de classificação (*Machine Learning*) é o motor das *Proptechs* no mundo. Serão explorados modelos como o *Random Forest* e a Regressão Logística para lidar com a classificação de risco (variável dependente binária ou multiclasse) e algoritmos de *clustering* para identificar padrões geográficos de danos (segmentação de risco por bairro).

## 2.3 Visão Computacional (*Computer Vision*) na Construção Civil

A utilização de Redes Neurais Convolucionais (CNNs) é o estado da arte para a identificação de patologias em imagens. O referencial teórico abordará casos de sucesso na detecção automática de:

* Rachaduras e fissuras em paredes e pisos.
* Manchas de umidade, mofo e infiltrações.
* Desgaste de materiais e acabamentos.

# 3 METODOLOGIA

## 3.1 Tipo e Abordagem

O projeto será conduzido como uma Pesquisa Aplicada, visando a resolução de um problema prático da LMF Vistorias. A abordagem será predominantemente quantitativa, com foco na modelagem estatística e no treinamento de algoritmos de Machine Learning.

## 3.2 Coleta e Tratamento de Dados

| Fonte de Dados | Variáveis Chave | Técnica de Tratamento |
| --- | --- | --- |
| Estruturados | Idade do Imóvel, Área, Tipo, Localização (CEP), Histórico de Reparos | Limpeza (tratamento de *missings* e *outliers*), Normalização, *Feature Engineering*. |
| Não Estruturados (Texto) | Descrições dos danos nos laudos (*free text*). | Processamento de Linguagem Natural (NLP) para vetorização e extração de sentimentos de dano. |
| Não Estruturados (Imagens) | Fotos das vistorias (antes e depois do uso). | Rotulagem manual das patologias e redimensionamento para treinamento de CNNs. |

## 3.3 Desenvolvimento do Modelo Preditivo (MPRPV)

A modelagem será dividida em duas frentes:

### 3.3.1 Modelo de Classificação de Risco

1. Objetivo: Prever a probabilidade de um novo imóvel apresentar Dano Grave ($\text{P}(Dano Grave) > x\%$) nos primeiros 12 meses de locação/uso.
2. Métodos: Aplicação de modelos como Regressão Logística e Random Forest, utilizando características do imóvel, dados geográficos e o histórico de patologias de imóveis vizinhos ou semelhantes.
3. Saída: Uma "Pontuação de Risco" de 0 a 100 para cada agendamento.

### Modelo de Visão Computacional (Suporte ao Laudo)

1. Objetivo: Classificar e localizar automaticamente o tipo de patologia (ex: rachadura fina, mofo) em uma imagem de campo.
2. Métodos: Treinamento de uma Rede Neural Convolucional (CNN), usando bibliotecas como TensorFlow ou PyTorch, para identificar as patologias rotuladas.
3. Saída: Etiquetagem e localização precisa de danos na foto, conferindo objetividade ao laudo.

## 3.4 Avaliação e Validação

O desempenho dos modelos será avaliado utilizando métricas como Acurácia, Precisão e Recall, buscando um equilíbrio que minimize a taxa de falsos negativos (imóvel de alto risco classificado como baixo).

# 4 RECURSOS E CRONOGRAMA

## 4.1 Recursos Necessários

| Categoria | Descrição |
| --- | --- |
| Recursos Humanos | Cientista de Dados (Desenvolvimento e Modelagem); Especialista de Vistoria da LMF (Validação e Rotulagem). |
| Infraestrutura Tecnológica | Servidor/Ambiente Cloud (para treinamento dos modelos de Visão Computacional); *Pipeline* de dados (ETL); Ambiente de desenvolvimento Python. |
| Licenças de Software | Licença(s) para plataforma de *Business Intelligence* (BI) para a visualização dos resultados. |

## 4.2 Cronograma (Plano de Ação - 3 Meses)

| **Mês** | **Etapa Principal** | **Atividades Focais** | **Entregáveis e Validações Chave** |
| --- | --- | --- | --- |
| **Mês 1** | **PLANEJAMENTO E PREPARAÇÃO DOS DADOS** | **1. Planejamento:** Definição de papéis, escopo final e KPIs estratégicos com a LMF. **2. Aquisição de Dados:** Coleta dos dados brutos e configuração do ambiente (Jupyter/Colab). **3. Tratamento Inicial:** Limpeza, remoção de *outliers* e pré-processamento essencial dos dados (Pandas). | **Entregável:** Dados tratados e prontos para uso. **Validação:** Evidências de reuniões e aprovação do escopo pela LMF. |
| **Mês 2** | **DESENVOLVIMENTO E PROTÓTIPO DE BI** | **1. Análise Exploratória (EDA):** Identificação dos padrões de risco e correlações (Matplotlib/Seaborn). **2. Engenharia de Variáveis:** Criação de colunas de risco e cálculo dos KPIs (Taxa de Retrabalho, Eficiência, etc.). **3. Desenvolvimento do Dashboard:** Criação do protótipo funcional no Power BI/Tableau, focado na visualização dos KPIs e no diagnóstico de risco. | **Entregável:** Protótipo funcional do Dashboard de BI. **Validação:** Feedback da empresa (Validação Intermediária) sobre a usabilidade e clareza do painel. |
| **Mês 3** | **VALIDAÇÃO, DOCUMENTAÇÃO E ENTREGA** | **1. Refinamento:** Ajustes finais no Dashboard e nos cálculos dos KPIs com base no *feedback* da LMF. **2. Documentação ABNT:** Redação final do Relatório Técnico (Introdução, Metodologia, Resultados e Conclusões). **3. Preparação para Entrega:** Criação da Apresentação Final e gravação do Vídeo Pitch. **4. Apresentação:** Entrega final do projeto à banca e à LMF Vistorias. | **Entregável Final:** Relatório ABNT, Dashboard completo e pronto para uso. **Validação:** Aprovação final do projeto pelo professor/banca. |

# APÊNDICE A - DICIONÁRIO DE DADOS PRELIMINAR

* [Listar e descrever as colunas (variáveis) do dataset que será utilizado no projeto.]

# ANEXO A - DOCUMENTAÇÃO TÉCNICA

* [Incluir documentação técnica relevante, como a ABNT NBR 14653-1 ou outros padrões da construção civil.]