

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

VOTORANTIM / SP

2025

Sumário

1 INTRODUÇÃO	3
1.1 Contextualização.....	3
1.2 Problema de Pesquisa	3
1.3 Objetivos.....	3
1.3.1 Objetivo Geral	3
1.3.2 Objetivos Específicos	4
1.4 Justificativa	4
2 REFERENCIAL TEÓRICO	6
2.1 A História Imobiliária e o Rigor Técnico	6
2.2 Aplicações de Machine Learning no Setor Imobiliário (<i>Proptech</i>)	6
2.3 Visão Computacional (<i>Computer Vision</i>) na Construção Civil	6
3 METODOLOGIA.....	7
3.1 Tipo e Abordagem	7
3.2 Coleta e Tratamento de Dados	7
3.3 Desenvolvimento do Modelo Preditivo (MPRPV)	7
3.3.1 Modelo de Classificação de Risco.....	7
3.3.2 Modelo de Visão Computacional (Suporte ao Laudo)	8
3.4 Avaliação e Validação	8
4 RECURSOS E CRONOGRAMA.....	9
4.1 Recursos Necessários	9
4.2 Cronograma (Plano de Ação - 6 Meses)	10
APÊNDICE A - DICIONÁRIO DE DADOS PRELIMINAR.....	11
ANEXO A - DOCUMENTAÇÃO TÉCNICA	12

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 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 <i>missings</i> e <i>outliers</i>), Normalização, <i>Feature Engineering</i> .
Não Estruturados (Texto)	Descrições dos danos nos laudos (<i>free text</i>).	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

- Objetivo:** Prever a probabilidade de um novo imóvel apresentar Dano Grave ($P(\text{Dano Grave}) > x\%$) nos primeiros 12 meses de locação/uso.

- b. *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.
- c. *Saída*: Uma "Pontuação de Risco" de 0 a 100 para cada agendamento.

3.3.2 Modelo de Visão Computacional (Suporte ao Laudo)

- a. *Objetivo*: Classificar e localizar automaticamente o tipo de patologia (ex: rachadura fina, mofo) em uma imagem de campo.
- b. *Métodos*: Treinamento de uma Rede Neural Convolucional (CNN), usando bibliotecas como TensorFlow ou PyTorch, para identificar as patologias rotuladas.
- c. *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); <i>Pipeline</i> de dados (ETL); Ambiente de desenvolvimento Python.
Licenças de Software	Licença(s) para plataforma de <i>Business Intelligence</i> (BI) para a visualização dos resultados.

4.2 Cronograma (Plano de Ação - 6 Meses)

Mês	Atividade Principal	Responsável (Ex.)
1	Planejamento Detalhado e Aquisição/Limpeza Inicial dos Dados.	Equipe de Dados
2	Pré-processamento e Engenharia de Recursos (Incluindo NLP no texto).	Equipe de Dados
3	Treinamento e Validação do Modelo de Classificação de Risco (MPRPV).	Cientista de Dados
4	Rotulagem das Imagens e Treinamento do Modelo de Visão Computacional.	Vistoriador / Cientista de Dados
5	Integração dos Modelos e Desenvolvimento do <i>Dashboard</i> de Risco.	Cientista de Dados
6	Testes de Aceitação (<i>UAT</i>), Documentação Final e Implantação do MVP.	Equipe de Dados / Gestão da LMF

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.]