



Modelo Preditivo de Grupos de risco de Imóveis

Produto financeiro: Consórcio



lab.data

24/10/2023

Pós-Graduação em Analytics e Inteligência Artificial – Data Science

**Nome do Aluno:**

Roberto Paulo Sianciulis Júnior

Coordenadores:

Profª Drª Alessandra de Ávila Montini

Profª Dr. Adolpho Walter Pimazoni Canton



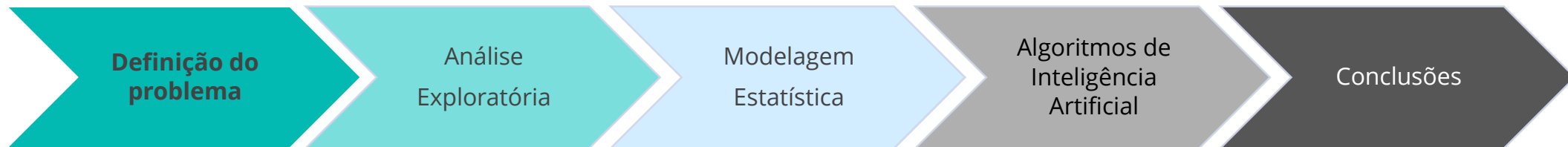
Agenda

- 1. Objetivo do Trabalho
- 2. Contextualização do Problema
- 3. Base de Dados
 - i. Bases originais
 - ii. Filtros
 - iii. Principais variáveis
 - iv. Processo de redução de variáveis
- 4. Análise Exploratória de Dados
- 5. Modelagem com Estatística Tradicional
- 6. Modelagem com Inteligência Artificial
- 7. Conclusões
- 8. Sugestão para Trabalhos Futuros



Metodologia de Análise de dados

4



Planejamento

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

Análise Preliminar

- Medidas resumo
- Análise bivariada
- Correlações
- Redução de variáveis

Avaliação das técnicas

- Regressão Logística
- Árvore de Decisão

Avaliação das técnicas

- Random Forest
- SVM
- Gradient Boosting Tree
- XGBoosting
- LightGBM
- CatBoost

Definição da técnica

- Validação dos resultados com negócios
- Escolha da técnica que melhor se adequa ao uso e estratégias da área de negócio



O que é **Consórcio**?

5

Juntar em grupos pessoas que desejam um mesmo bem, sendo que um sorteio define quem pode pegar o valor para comprar o bem a cada mês.

Etapas principais do **Consórcio**:



Consórcio não é como um empréstimo ou financiamento.
Nesses casos, o cliente recebe o valor do bem na hora da contratação.

Com o Consórcio, o cliente se planeja para o futuro, pagando parcelas menores e recebe o crédito após a contemplação!





Planejamento

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

Análise Preliminar

- Medidas resumo
- Análise bivariada
- Correlações
- Redução de variáveis

Avaliação das técnicas

- Regressão Logística
- Árvore de Decisão

Avaliação das técnicas

- Random Forest
- SVM
- Gradient Boosting Tree
- XGBoosting
- LightGBM
- CatBoost

Definição da técnica

- Validação dos resultados com negócios
- Escolha da técnica que melhor se adequa ao uso e estratégias da área de negócio



O consórcio é um **produto planejado** e uma simples alteração nos indicadores pode comprometer a perenidade de toda a operação.

As Administradoras tem verificado um **aumento expressivo em cancelamento** e em inadimplência das cotas nos últimos anos e a necessidade de **construção de um modelo preditivo** de grupos de risco de imóveis.

→ Cenário de alta competitividade no segmento de imóveis + dinâmica peculiar do produto

→ Modelo: **grande diferencial competitivo que** auxiliará a administradora a trabalhar de forma otimizada com seus grupos de imóveis.



Objetivo

Predizer grupos de risco de consórcio, especificamente da categoria de **imóveis**.

Como

Utilizando **dados históricos**, **modelos estatísticos** e **algoritmos de Inteligência Artificial**, que selecionarão as **características mais relevantes** que explicam o perfil de alto ou baixo risco de um grupo de consórcio de imóveis.

Para

Traçar **estratégias de relacionamento**, desenvolver **régua de comunicação** e **ações preventivas**,
Minimizar perdas financeiras futuras





Base Original

100.261 observações

Base de partida

Base com os registros mensais de grupos de Imóveis de Consórcio, divulgadas em jan/18 até ago/22 (56 meses).



Filtro 1 – Exclusões de observações de grupos formados antes de 2018

19.139 observações

Redução da base 1

Para criação do indicador perc_churn, foi fundamental selecionar somente grupos criados de 2018 em diante.



Filtro 2– Exclusões: erros de preenchimento

15.971 observações

Redução da base 2

Optou-se pela remoção dos registros dos grupos das administradoras da base (registros em que os valores medios de bem venda estavam zerados, ou a variação de cotas excluídas entre assembleias estava negativa ou o perc_churn estava maior que 1).



Base Final

15.971 observações – 31 Variáveis

Modelagem

Consideraremos usar 80% da base para treino (12.776 registros) e 20% para teste (3.195 registros).



Tratamentos



Detalhamento dos tratamentos



base original



base final

- 1 **Remover** espaços duplos em todos os campos
- 2 **Excluir** a última linha (se não tiver conteúdo válido)
- 3 **Mudar** o campo "#Nome_da_Administradora" para "Nome_da_Administradora"
- 4 **Criar** variável chave (nome_da_Administradora + "_" + Código_do_grupo)
- 5 F1 - **filtrar** base deixando somente grupos de administradoras que começaram em 2018 pra frente (isto é, tem Número_da_assembleia_geral_ordinária=1)
- 6 **Classificar** a base por chave (nome administradora + número do grupo) e Número_da_assembleia_geral_ordinária em ordem crescente
- 7 **Criar** o campo Variação Cotas excluídas --> fórmula: se o Número_da_assembleia_geral_ordinária for igual a 1 do campo "chave" então a Var_cotas_Excluidas é igual a 0, senão sempre fazer a linha abaixo menos a linha de cima pelo campo Quantidade_de_cotas_excluídas
- 8 **Criar** o campo parcela --> fórmula: (Valor_médio_do_bem + (Valor_médio_do_bem*Taxa_de_administração))/Prazo_do_grupo_em_meses
- 9 **Criar** o campo carteira --> fórmula: Quantidade_de_cotas_ativas_em_dia + Quantidade_de_cotas_ativas_contempladas_inadimplentes + Quantidade_de_cotas_ativas_não_contempladas_inadimplentes
- 10 **Adicionar** variável externa a base INCC_DI (utilizando o campo Data_base como chave)
- 11 **Criar** o campo percentual de Churn sobre carteira ativa --> fórmula: Var_cotas_Excluidas / Carteira
- 12 F2 - erros de preenchimento de informação: **tirar** da base grupos de administradoras com valores medios de bem venda zerados, var_cotas_excluidas negativa ou perc_churn_s_carteira >1
- 13 F3 - **criar** o campo condição (grupo "novo" criado em 2021 e 2022 e grupo "antigo" criado anterior a 2021, basta filtrar o Número_da_assembleia_geral_ordinária igual 1 e aplicar as regras segundo a data base)
- 14 **Criar** o campo Faixa_parcelas
- 15 **Criar** o campo percentual de inadimplência sobre carteira ativa --> fórmula: (Quantidade_de_cotas_ativas_contempladas_inadimplentes + Quantidade_de_cotas_ativas_não_contempladas_inadimplentes) / carteira
- 16 **Excluir** o codigo_do_segmento pois avaliaremos um segmento apenas no estudo (código segmento =1 - bens imóveis)
- 17 **Criar** o campo idade --> fórmula: se o campo Número_da_assembleia_geral_ordinária for maior do que Prazo_do_grupo_em_meses então 1, senão Número_da_assembleia_geral_ordinária / Prazo_do_grupo_em_meses
- 18 **Criar** o campo faixa_idade
- 19 **Criar** o campo faixa_prazo
- 20 **Criar** o campo Receita_potencial --> fórmula: carteira * Taxa_de_administração * Valor_médio_do_bem
- 21 **Criar** o campo Risco --> fórmula: se o Perc_churn_s_carteira <=3%, então 0 (risco baixo) senão 1 (risco alto)



Principais variáveis (31)



Variáveis cadastrais (5)

- Chave_nome_adm_e_cod_grupo
- Nome_da_Administradora
- CNPJ_da_Administradora
- Codigo_do_grupo
- Prazo_do_grupo_em_meses



Variáveis transacionais (12)

- Data_base
- Numero_da_assembleia_geral_ordinária
- Valor_medio_do_bem
- Taxa_de_administracao
- Quantidade_de_cotas_ativas_em_dia
- Quantidade_de_cotas_ativas_contempladas_inadimplentes
- Quantidade_de_cotas_ativas_nao_contempladas_inadimplentes
- Quantidade_de_cotas_ativas_contempladas_no_mes
- Quantidade_de_cotas_excluidas
- Quantidade_de_cotas_ativas_quitadas
- Quantidade_de_cotas_ativas_com_credito_pendente_de_utilizacao
- Índice de correção



Variáveis criadas (12)

- Carteira_ativa
- Carteira_inadimplente
- Parcela
- Idade
- Var_cotas_Excluidas
- Perc_churn
- Perc_inadimplencia
- Condição
- Fx_parcela
- Fx_prazo
- Fx_idade
- Receita_potencial



Variável resposta (1)

Risco:
1 = grupo de alto risco
0 = grupo de baixo risco



Variável externa (1)

- INCC_DI

Legenda:

Variáveis de identificação e controle

Variáveis Categóricas

Variáveis Numéricas

