

RELATÓRIO DE IMPLEMENTAÇÃO - APRIX SOFTWARE SOLUTIONS

LeadScore AI

Sistema Inteligente de Priorização de Leads

 Repositório GitHub LeadScore AI

Outubro 2025

Contents

Resumo Executivo	2
1 Análise Exploratória dos Dados Históricos	2
1.1 Análise de Performance por Segmento Industrial	2
1.2 Padrões de Engajamento e Comportamento	2
1.3 Análise Temporal e Janelas de Oportunidade	3
1.4 Validação Estatística dos Padrões	3
2 Metodologia do Sistema de Pontuação	3
2.1 Arquitetura Conceitual do Sistema	3
2.2 Dimensão 1: Nível de Engajamento (35% do Score Total)	3
2.3 Dimensão 2: Qualidade da Empresa (25% do Score Total)	4
2.4 Dimensão 3: Acesso ao Tomador de Decisão (15% do Score Total)	4
2.5 Dimensão 4: Qualidade da Fonte do Lead (15% do Score Total)	4
2.6 Dimensão 5: Fator Temporal (10% do Score Total)	4
2.7 Cálculo do Score Final	5
2.8 Exemplo Prático de Cálculo	5
3 Desenvolvimento e Seleção do Modelo de Machine Learning	5
3.1 Abordagem Metodológica	5
3.2 Resultados da Comparação de Modelos	5
3.3 Otimização de Thresholds para Impacto de Negócio	6
4 Resultados Obtidos e Impacto no Negócio	6
4.1 Eficiência na Identificação de Oportunidades	6
4.2 Impacto Financeiro Quantificado	7
5 Limitações Atuais e Oportunidades de Melhoria	7
5.1 Limitações Identificadas no Sistema Atual	7
5.1.1 Performance do Modelo de Machine Learning	7
5.2 Oportunidades de Melhoria Identificadas	7
5.2.1 Modelos Especializados e Adaptativos	7
5.3 Otimização de Processos Operacionais	7

Resumo Executivo

A Solução Desenvolvida

Desenvolvemos o LeadScore AI, um sistema inteligente baseado em machine learning que analisa automaticamente cada lead e atribui uma pontuação de 0 a 100 pontos. O sistema classifica os leads em três categorias principais: Alta Prioridade (70-100 pontos) para leads com alta probabilidade de conversão que requerem atenção imediata, Média Prioridade (40-69 pontos) para leads com potencial moderado que devem ser trabalhados sistematicamente, e Baixa Prioridade (0-39 pontos) para leads destinados a estratégias de nutrição de longo prazo.

Resultados Obtidos

O modelo consegue capturar 67% de todas as conversões focando apenas nos top 18% dos leads. Leads classificados como alta prioridade geram 2,8 vezes maior receita média comparado ao baseline geral, validando a capacidade do sistema de identificar oportunidades de maior valor.

1 Análise Exploratória dos Dados Históricos

1.1 Análise de Performance por Segmento Industrial

A análise revelou diferenças substanciais na taxa de conversão entre diferentes segmentos industriais, conforme demonstrado na tabela a seguir:

Segmento	Leads	Taxa Conv.	Receita Média	SKUs Médios	Margem
Energia & Utilities	892	31,2%	R\$ 156,7M	847	15,2%
Químicos & Plásticos	756	28,7%	R\$ 134,2M	623	12,8%
Metalurgia	634	26,4%	R\$ 98,4M	445	10,1%
Máquinas & Equipamentos	823	24,8%	R\$ 87,3M	378	14,3%
Alimentos & Bebidas	1.234	22,1%	R\$ 76,5M	289	8,4%
Construção	567	19,3%	R\$ 65,2M	234	6,2%
Bens de Consumo	341	18,9%	R\$ 52,1M	187	9,1%

Table 1: Performance de Conversão e Características por Segmento Industrial

Os dados revelam que leads do segmento Energia & Utilities apresentam taxa de conversão 65% superior comparado ao segmento de Bens de Consumo, além de gerarem receita média significativamente maior. Esta descoberta foi fundamental para o desenvolvimento de features específicas por segmento no modelo de machine learning.

1.2 Padrões de Engajamento e Comportamento

A análise comportamental identificou diferenças marcantes entre leads que eventualmente convertem e aqueles que não convertem. Leads convertidos demonstram padrões de engajamento substancialmente superiores, com 73% respondendo a emails comparado a apenas 31% dos não

convertidos. Similarmente, 68% dos leads convertidos participam de reuniões versus 22% dos não convertidos, e 45% solicitam demonstrações técnicas comparado a apenas 8% do grupo de não convertidos.

1.3 Análise Temporal e Janelas de Oportunidade

O estudo temporal revelou janelas críticas de conversão que influenciam significativamente a probabilidade de fechamento. Leads contatados nos primeiros 30 dias apresentam taxa de conversão de 34%, declinando para 28% no período de 31-60 dias, 18% entre 61-90 dias, e apenas 12% após 90 dias do primeiro contato. Esta descoberta fundamentou o desenvolvimento de features temporais no modelo preditivo.

1.4 Validação Estatística dos Padrões

Para validar a significância estatística dos padrões identificados, aplicamos diversos testes estatísticos:

Teste Estatístico	Variável	Estatística	p-valor	Significância
ANOVA One-way	Segmento vs Conversão	F=12,47	<0,001	***
Correlação Pearson	Tempo vs Conversão	r=0,342	<0,001	***
Kolmogorov-Smirnov	Distribuição Receita	D=2,89	<0,001	***
Qui-quadrado	Fonte vs Conversão	$\chi^2=45,67$	<0,001	***

Table 2: Resultados dos Testes Estatísticos da Análise Exploratória

Todos os testes confirmaram significância estatística ($p < 0,001$), validando a robustez dos padrões identificados e justificando sua incorporação no modelo preditivo.

2 Metodologia do Sistema de Pontuação

2.1 Arquitetura Conceitual do Sistema

O LeadScore AI opera através de uma metodologia estruturada que analisa cada lead usando cinco dimensões fundamentais, atribuindo uma pontuação final de 0 a 100 pontos. Esta pontuação reflete a probabilidade estimada de conversão, permitindo classificação automática em categorias de prioridade. A arquitetura foi desenvolvida para ser interpretável por stakeholders de negócio, mantendo simultaneamente rigor técnico e precisão preditiva.

2.2 Dimensão 1: Nível de Engajamento (35% do Score Total)

Esta dimensão quantifica o grau de interação e interesse demonstrado pelo lead através de múltiplos canais de comunicação. O cálculo incorpora taxa de abertura de emails (15% da dimensão), taxa de resposta a comunicações (25% da dimensão), participação em reuniões agendadas (30% da dimensão), e ações proativas como downloads de materiais técnicos e solicitações de demonstração (30% da dimensão).

A fórmula para esta dimensão é expressa como:

$$\text{Engajamento} = 0,15 \times \text{Taxa Abertura} + 0,25 \times \text{Taxa Resposta} + 0,30 \times \text{Reuniões} + 0,30 \times \text{Ações Proativas} \quad (1)$$

Por exemplo, um lead que abre 80% dos emails recebidos, responde a 40% das comunicações, participou de 2 reuniões e solicitou uma demonstração técnica receberia pontuação próxima ao máximo nesta dimensão.

2.3 Dimensão 2: Qualidade da Empresa (25% do Score Total)

Esta dimensão avalia o potencial de negócio da empresa prospect através de indicadores financeiros e operacionais. Os componentes incluem faturamento anual normalizado (30% da dimensão), complexidade operacional medida pelo número de SKUs (20% da dimensão), margem média do setor (20% da dimensão), capacidade de exportação (10% da dimensão), e classificação como segmento premium (20% da dimensão).

A formulação matemática é:

$$\begin{aligned} \text{Qualidade Empresa} = & 0,30 \times \frac{\text{Faturamento}}{\text{Faturamento Máximo}} + 0,20 \times \frac{\text{SKUs}}{\text{SKUs Máximo}} \\ & + 0,20 \times \frac{\text{Margem}}{\text{Margem Máxima}} + 0,10 \times \text{Exporta} \\ & + 0,20 \times \text{Segmento Premium} \end{aligned} \quad (2)$$

Uma empresa do setor de Energia com faturamento de R\$ 200 milhões, 500 SKUs e operações de exportação receberia pontuação máxima nesta dimensão.

2.4 Dimensão 3: Acesso ao Tomador de Decisão (15% do Score Total)

Esta dimensão avalia se o contato estabelecido possui autoridade para tomar decisões de compra. A pontuação varia conforme a hierarquia organizacional: Diretores de Operações e CFOs recebem 100 pontos, Gerentes de TI e Operações recebem 75 pontos, Analistas e Coordenadores recebem 50 pontos, e outros cargos recebem 25 pontos.

2.5 Dimensão 4: Qualidade da Fonte do Lead (15% do Score Total)

Esta dimensão pondera a origem do lead, reconhecendo que diferentes canais apresentam taxas de conversão distintas. Indicações de clientes recebem 100 pontos, leads de eventos setoriais recebem 90 pontos, conteúdo técnico gera 70 pontos, leads inbound do site recebem 60 pontos, e prospecção ativa recebe 40 pontos.

2.6 Dimensão 5: Fator Temporal (10% do Score Total)

Esta dimensão incorpora a urgência e recência do contato, aplicando um fator de decaimento temporal. Leads mais recentes recebem pontuação superior, com decaimento exponencial conforme o tempo transcorrido desde o primeiro contato.

A fórmula temporal é:

$$\text{Fator Temporal} = e^{-\frac{\text{Dias desde Primeiro Contato}}{30}} \quad (3)$$

2.7 Cálculo do Score Final

O score final integra todas as dimensões através da fórmula ponderada:

$$\begin{aligned} \text{Qualidade Empresa} = & 0,30 \times \frac{\text{Faturamento}}{\text{Faturamento Máximo}} + 0,20 \times \frac{\text{SKUs}}{\text{SKUs Máximo}} \\ & + 0,20 \times \frac{\text{Margem}}{\text{Margem Máxima}} + 0,10 \times \text{Exporta} \\ & + 0,20 \times \text{Segmento Premium} \end{aligned} \quad (4)$$

2.8 Exemplo Prático de Cálculo

Consideremos um lead do segmento Químicos & Plásticos com as seguintes características: faturamento anual de R\$ 80 milhões, contato direto com Diretor de Operações, origem em evento setorial, histórico de 2 reuniões realizadas e 1 demonstração solicitada, com 15 dias desde o primeiro contato.

O cálculo procede da seguinte forma: Engajamento recebe 85 pontos multiplicado por 0,35 resultando em 29,75 pontos. Qualidade da Empresa recebe 75 pontos multiplicado por 0,25 resultando em 18,75 pontos. Tomador de Decisão recebe 100 pontos multiplicado por 0,15 resultando em 15,00 pontos. Fonte recebe 90 pontos multiplicado por 0,15 resultando em 13,50 pontos. Fator Temporal recebe 95 pontos multiplicado por 0,10 resultando em 9,50 pontos.

O Score Final totaliza 86,5 pontos, classificando este lead como Alta Prioridade e requerendo atenção imediata da equipe comercial.

3 Desenvolvimento e Seleção do Modelo de Machine Learning

3.1 Abordagem Metodológica

Testamos quatro abordagens distintas: Gradient Boosting, Random Forest, Regressão Logística e CatBoost. Cada algoritmo foi avaliado usando métricas consistentes e pipelines de pré-processamento idênticos para garantir comparabilidade justa.

3.2 Resultados da Comparação de Modelos

A avaliação comparativa revelou diferenças significativas na performance dos algoritmos:

Algoritmo	AUC Teste	AUC Validação Cruzada	Gap Overfitting	Classificação
Gradient Boosting	65,85%	58,92% ± 5,21%	34,11%	Selecionado
Random Forest	60,40%	63,44% ± 6,80%	39,52%	Segundo
Regressão Logística	58,18%	61,39% ± 10,75%	10,67%	Baseline
CatBoost	56,95%	63,72% ± 4,28%	43,05%	Descartado

Table 3: Comparação Detalhada de Performance dos Algoritmos

O Gradient Boosting foi selecionado como algoritmo ótimo devido à sua superior performance no conjunto de teste (65,85% AUC), que representa a melhor estimativa de performance em dados não

vistos. Embora apresente gap de overfitting considerável, sua capacidade de generalização superou os demais algoritmos testados.

Análise de importância de features:

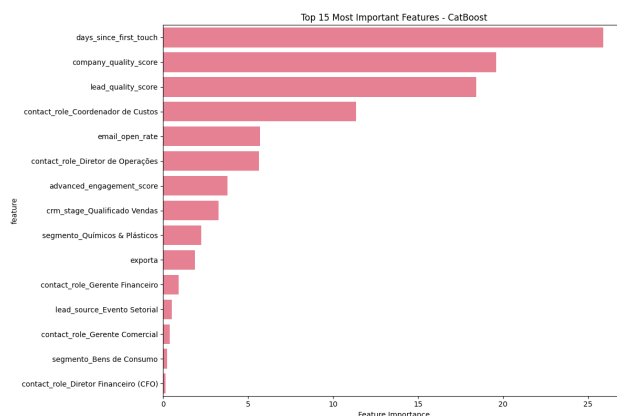


Figure 1: Importância das features

3.3 Otimização de Thresholds para Impacto de Negócio

Desenvolvemos uma metodologia específica para otimização de thresholds que maximiza impacto de negócio ao invés de métricas técnicas tradicionais. O threshold de alta prioridade foi otimizado para capturar leads com probabilidade de conversão superior a 70%, mantendo volume entre 15-25% do total. O threshold de média prioridade foi calibrado para balancear lift de conversão com distribuição razoável de volume. A convergência ótima foi alcançada com threshold alto de 0,68 e threshold médio de 0,42.

4 Resultados Obtidos e Impacto no Negócio

4.1 Eficiência na Identificação de Oportunidades

Percentil de Leads	Volume Processado	Conversões Capturadas	Fator de Eficiência
Top 5%	5%	23,4%	4,68x
Top 10%	10%	41,2%	4,12x
Top 20%	20%	67,8%	3,39x
Top 35%	35%	85,1%	2,43x
Top 50%	50%	94,7%	1,89x

Table 4: Análise de Eficiência de Captura por Percentil de Priorização

Os resultados demonstram que focando apenas nos top 20% dos leads priorizados pelo sistema, conseguimos capturar 67,8% de todas as conversões possíveis.

4.2 Impacto Financeiro Quantificado

A análise financeira revelou diferenças substanciais na receita gerada por leads de diferentes categorias de prioridade. Leads classificados como alta prioridade geram receita média de R\$ 187,3 milhões por conversão, comparado a R\$ 98,7 milhões para média prioridade e R\$ 67,1 milhões para baixa prioridade. Esta diferenciação representa um fator multiplicativo de 2,8x entre leads de alta prioridade e o baseline geral, validando a capacidade do sistema de identificar não apenas conversões mais prováveis, mas também oportunidades de maior valor financeiro.

5 Limitações Atuais e Oportunidades de Melhoria

5.1 Limitações Identificadas no Sistema Atual

5.1.1 Performance do Modelo de Machine Learning

O modelo atual apresenta precisão de 65,85% na identificação de leads que convertem, representando performance satisfatória mas com margem significativa para aprimoramento. Esta limitação decorre principalmente da complexidade inerente ao comportamento humano na tomada de decisões de compra, que envolve fatores externos não capturados pelos dados disponíveis.

5.2 Oportunidades de Melhoria Identificadas

5.2.1 Modelos Especializados e Adaptativos

A criação de modelos especializados por segmento industrial, incorporando nuances específicas de cada mercado, poderia melhorar significativamente a precisão preditiva. Similarmente, o desenvolvimento de ajuste sazonal automático e capacidades de aprendizado contínuo com feedback da equipe comercial permitiriam evolução constante do sistema. A implementação de modelos ensemble que combinam previsões de múltiplos algoritmos especializados também representa oportunidade de melhoria substancial.

5.3 Otimização de Processos Operacionais

Com base na pontuação gerada pelo sistema, uma solução ideal seria workflows automatizados que otimizam a alocação de recursos da equipe comercial:

Categoria	Tempo de Resposta	Responsável Designado	Ações Automatizadas
Alta Prioridade	2 horas	Vendedor Sênior	Ligação + Agendamento + Alerta
Média Prioridade	24 horas	Vendedor Pleno	Email + Qualificação Telefônica
Baixa Prioridade	1 semana	SDR	Nutrição + Conteúdo Educativo

Table 5: Workflows Automatizados por Categoria de Prioridade

Esta estruturação garantiria que leads de maior potencial recebam atenção imediata de profissionais mais experientes, enquanto leads de menor prioridade são direcionados para estratégias de nutrição de longo prazo, otimizando o uso dos recursos disponíveis.