



## Documento de Projeto

---

### Sumário

1. Introdução
2. Contexto e Justificativa
3. Visão Geral do Projeto
4. Casos de Uso
5. Arquitetura do Sistema
6. Componentes Técnicos
7. Implementação na Plataforma Databricks
8. Plano de Implementação
9. Requisitos Técnicos
10. Métricas de Sucesso
11. Limitações e Considerações
12. Protótipos
13. Conclusão

---

## 1. Introdução

O Sicredi CeleiroGPT é um assistente virtual inteligente desenvolvido para atender às necessidades específicas do ecossistema Sicredi. Este sistema integra tecnologias de processamento de linguagem natural, busca vetorial, análise de dados e visualização para oferecer uma interface conversacional que permite aos colaboradores acessar, interpretar e utilizar informações distribuídas em diversos sistemas da cooperativa.

O nome "CeleiroGPT" faz alusão a um celeiro - local onde se armazenam e organizam grãos para uso posterior - representando a capacidade do sistema de armazenar, organizar e disponibilizar conhecimentos sobre o Sicredi de forma estruturada e acessível. A tecnologia GPT (Generative Pre-trained Transformer) fornece a base para a compreensão e geração de linguagem natural, permitindo interações fluidas e contextualmente relevantes.

Este documento apresenta uma visão completa do projeto, incluindo sua justificativa, arquitetura, componentes técnicos, plano de implementação e métricas de sucesso.

## 2. Contexto e Justificativa

### 2.1 Cenário Atual

O Sicredi, como instituição financeira cooperativa, gerencia um volume significativo de dados distribuídos em múltiplos sistemas, incluindo:

- Dados cadastrais de associados
- Histórico de relacionamento e transações
- Informações de produtos e serviços
- Indicadores de desempenho de agências
- Políticas de crédito e investimento
- Regulamentos e normas internas
- Modelos de análise de risco e comportamento

Atualmente, para obter uma visão completa sobre um associado ou analisar o desempenho de uma agência, os colaboradores precisam:

1. Identificar os sistemas que contêm as informações necessárias
2. Acessar cada sistema individualmente
3. Executar consultas específicas, muitas vezes exigindo conhecimento técnico
4. Extrair e compilar dados manualmente
5. Interpretar as informações com base em seu conhecimento e experiência
6. Formatar análises e recomendações para apresentação

### 2.2 Problemas e Desafios

Este processo apresenta diversos desafios:

- **Tempo Excessivo:** A coleta e análise manual de dados pode levar horas ou dias
- **Dependência Técnica:** Muitas consultas exigem conhecimento de SQL ou ferramentas específicas
- **Inconsistências:** Diferentes colaboradores podem interpretar os mesmos dados de formas distintas
- **Conhecimento Fragmentado:** Informações importantes estão dispersas em documentos e sistemas
- **Atendimento Limitado:** Impossibilidade de fornecer análises completas durante atendimentos em tempo real
- **Escalabilidade:** Dificuldade em compartilhar conhecimento especializado entre todas as agências
- **Padronização:** Variação na qualidade e formato das análises entre diferentes colaboradores

## 2.3 Oportunidade

O CeleiroGPT surge como solução para estes desafios, oferecendo:

- **Acesso Simplificado:** Interface conversacional que não exige conhecimento técnico
- **Análise Integrada:** Combinação de dados de múltiplas fontes em uma única resposta
- **Padronização:** Aplicação consistente de políticas e melhores práticas
- **Velocidade:** Respostas em segundos ou minutos, em vez de horas ou dias
- **Democratização:** Acesso ao mesmo nível de análise para colaboradores de todas as agências
- **Escalabilidade:** Capacidade de atender múltiplos usuários simultaneamente
- **Aprendizado Contínuo:** Melhoria constante com base em novas informações e feedback

## 3. Visão Geral do Projeto

### 3.1 Objetivo Principal

Desenvolver um assistente virtual inteligente que permita aos colaboradores do Sicredi acessar, analisar e utilizar informações de forma simples, rápida e contextualizada, através de interações em linguagem natural.

### 3.2 Objetivos Específicos

1. Criar uma interface conversacional intuitiva para acesso a dados e análises
2. Integrar múltiplas fontes de dados em um único ponto de acesso

3. Automatizar análises de perfil, risco e oportunidades para associados
4. Fornecer visualizações e relatórios personalizados sob demanda
5. Aplicar consistentemente políticas e melhores práticas nas análises
6. Reduzir o tempo necessário para obtenção de informações e insights
7. Democratizar o acesso a análises avançadas para todos os colaboradores

### 3.3 Público-Alvo

O CeleiroGPT atenderá diferentes perfis de usuários dentro do ecossistema Sicredi:

- **Gerentes de Relacionamento:** Para análises de associados e identificação de oportunidades
- **Analistas de Crédito:** Para avaliação de risco e capacidade de pagamento
- **Gerentes de Agência e carteiras:** Para análise de desempenho e comparativos
- **Consultores de Investimento:** Para recomendações baseadas em perfil
- **Diretores Gerais e Regionais:** Para visões consolidadas de desempenho

### 3.4 Escopo do Projeto

**Incluso no Escopo:**

- Desenvolvimento do agente orquestrador central
- Implementação do sistema de personas especializadas
- Criação de bases de conhecimento vetoriais para domínios principais
- Desenvolvimento de skills técnicas (busca, SQL, visualização, etc.)
- Integração com fontes de dados principais do Sicredi
- Interface conversacional via web para colaboradores
- Documentação e treinamento para usuários

**Fora do Escopo (Fases Futuras):**

- Processamento de documentos e imagens
- Funcionalidades transacionais (aprovações, contratações)

## 4. Casos de Uso

### 4.1 Análise Completa de Associado

**Cenário:** Um gerente de relacionamento precisa fazer uma análise completa de um associado antes de uma reunião.

**Interação:**

Usuário: "Preciso de uma análise completa do associado João Silva, CPF 123.456.789-00, para uma reunião em 30 minutos."

CeleiroGPT: "Vou preparar uma análise completa do associado João Silva. Aqui está:

## PERFIL DO ASSOCIADO

João Silva é pessoa física, 42 anos, engenheiro civil, associado há 8 anos na cooperativa. Cliente da agência Centro (código 123), carteira Premium gerenciada por Maria Oliveira.

## RELACIONAMENTO COM A COOPERATIVA

- Principalidade: Alta (score 85)
- Produtos atuais: conta corrente, cartão de crédito (limite R\$ 15.000), previdência privada, seguro residencial
- Movimentação média mensal: R\$ 12.500
- Última atualização cadastral: há 45 dias

## SITUAÇÃO FINANCEIRA

- Renda mensal declarada: R\$ 15.000
- Patrimônio total: R\$ 920.000
- Capital social: R\$ 45.000
- Saldo atual em conta: R\$ 8.500
- Investimentos no Sicredi: R\$ 65.000 (3 aplicações)
- Comprometimento de renda: 34,6% (R\$ 5.200/mês)

## ANÁLISE DE RISCO

- Faixa de risco atual: Baixo
- Probabilidade de inadimplência: 2,3%
- Endividamento total: R\$ 320.000 (35% do patrimônio)
- Restrições: Nenhuma restrição identificada
- Histórico de crédito: Excelente, sem atrasos nos últimos 36 meses

## OPORTUNIDADES IDENTIFICADAS

1. Potencial para aumento de limite de cartão (capacidade para até R\$ 25.000)
2. Perfil adequado para produtos de investimento de maior rentabilidade

3. Não possui consórcio - oportunidade para consórcio imobiliário
4. Candidato a upgrade para carteira Private (patrimônio próximo ao mínimo requerido)

Deseja que eu gere um relatório PDF desta análise ou aprofunde algum aspecto específico?"

#### **4.2 Análise Comparativa de Agências**

**Cenário:** Um diretor regional precisa avaliar o desempenho das agências sob sua responsabilidade.

**Interação:**

Usuário: "Compare o desempenho das agências da Regional Sul no último trimestre em termos de captação e crescimento da base de associados."

CeleiroGPT: "Aqui está a análise comparativa das agências da Regional Sul para o último trimestre:

#### **CAPTAÇÃO DE RECURSOS**

As 5 agências com melhor desempenho em captação foram:

1. Porto Alegre Centro: R\$ 12,5 milhões (↑15% vs trimestre anterior)
2. Florianópolis: R\$ 8,7 milhões (↑8%)
3. Caxias do Sul: R\$ 7,2 milhões (↑5%)
4. Gramado: R\$ 6,9 milhões (↑22%)
5. Curitiba Norte: R\$ 6,1 milhões (↑3%)

A média de captação da Regional foi de R\$ 5,3 milhões por agência, superando em 8% a meta trimestral estabelecida.

#### **CRESCIMENTO DA BASE DE ASSOCIADOS**

As 5 agências com maior crescimento percentual foram:

1. Balneário Camboriú: 7,2% (142 novos associados)
2. Gramado: 6,8% (98 novos associados)
3. Florianópolis Sul: 5,9% (115 novos associados)
4. Porto Alegre Zona Norte: 5,5% (87 novos associados)
5. Joinville: 5,1% (104 novos associados)

A média de crescimento da Regional foi de 4,2%, ficando 0,8 pontos percentuais acima da meta.

#### DESTAQUES E OBSERVAÇÕES

- Gramado aparece entre as top 5 tanto em captação quanto em crescimento de associados
- Porto Alegre Centro, apesar do excelente resultado em captação, teve crescimento de apenas 2,1% na base de associados
- Curitiba Sul ficou abaixo da meta em ambos os indicadores

Gostaria de visualizar estes dados em um gráfico ou aprofundar a análise de alguma agência específica?"

#### 4.3 Análise Preditiva e Recomendação

**Cenário:** Um analista de crédito precisa avaliar um pedido de financiamento.

**Interação:**

Usuário: "O associado Pedro Oliveira solicitou um financiamento imobiliário de R\$ 350.000. Qual a análise de viabilidade e risco?"

CeleiroGPT: "Analisarei o pedido de financiamento imobiliário do associado Pedro Oliveira:

#### DADOS DO FINANCIAMENTO SOLICITADO

- Valor: R\$ 350.000
- Finalidade: Aquisição de imóvel residencial
- Prazo desejado: 240 meses (informado em simulação prévia)

#### ANÁLISE DO ASSOCIADO

- Renda mensal comprovada: R\$ 12.800
- Comprometimento atual de renda: 22% (R\$ 2.816)
- Score de crédito: 850 (Excelente)
- Tempo de associação: 6 anos e 4 meses
- Histórico de crédito: Sem restrições ou atrasos nos últimos 48 meses

#### ANÁLISE DE VIABILIDADE

- Capacidade de pagamento: R\$ 3.840/mês (30% da renda)
- Prestação estimada: R\$ 3.325/mês (taxa atual de 9,8% a.a.)
- Comprometimento total estimado: 48% da renda (dentro do limite de 50%)

#### ANÁLISE DE RISCO

- Probabilidade de inadimplência: 3,2% (Baixa)
- LTV (Loan to Value): 70% (imóvel avaliado em R\$ 500.000)
- Classificação de risco: B (Baixo risco)

#### RECOMENDAÇÃO

O financiamento é VIÁVEL com as seguintes considerações:

1. Aprovação condicionada à avaliação formal do imóvel
2. Recomendável seguro prestamista (custo já incluído na simulação)
3. Sugestão de entrada mínima de 30% (R\$ 150.000)

Deseja que eu prepare uma simulação detalhada ou analise cenários alternativos de prazo e valor?"

#### 4.4 Consulta sobre Produtos e Recomendações

**Cenário:** Um consultor de investimentos precisa recomendar produtos para um associado.

**Interação:**

Usuário: "Quais produtos de investimento são mais adequados para a associada Ana Martins, considerando seu perfil e objetivos?"

CeleiroGPT: "Analisarei o perfil da associada Ana Martins para recomendar produtos de investimento:

#### PERFIL DA ASSOCIADA

- Idade: 35 anos



- Perfil de investidor: Moderado
- Patrimônio atual: R\$ 180.000
- Investimentos atuais: R\$ 120.000 (70% em renda fixa, 30% em fundos multimercado)
- Objetivo declarado: Aposentadoria (longo prazo) e reserva para educação dos filhos (médio prazo)
- Horizonte de investimento: 5+ anos para educação, 25+ anos para aposentadoria

## RECOMENDAÇÕES PERSONALIZADAS

Para objetivo de EDUCAÇÃO (médio prazo - 5 a 10 anos):

1. LCA/LCI: Isenção de IR, liquidez após carência, rentabilidade atual de 98% do CDI
2. Fundos DI Premium: Liquidez diária, rentabilidade atual de 103% do CDI
3. Tesouro Direto (Tesouro IPCA+): Proteção contra inflação, adequado para prazo definido

Para objetivo de APOSENTADORIA (longo prazo - 25+ anos):

1. Previdência Privada PGBL: Benefício fiscal de dedução no IR (até 12%)
2. Fundos de Ações: Exposição gradual (sugestão de 20% da carteira)
3. Fundos Imobiliários: Diversificação e renda mensal através de dividendos

## ESTRATÉGIA RECOMENDADA

Considerando o perfil moderado e os objetivos, sugiro a seguinte alocação:

- 50% em produtos de renda fixa (LCA/LCI, Tesouro Direto)
- 30% em previdência privada (PGBL)
- 15% em fundos multimercado
- 5% em fundos de ações (com aumento gradual)

Gostaria que eu detalhasse algum destes produtos ou preparasse uma simulação de rentabilidade para esta estratégia?"

## 4.5 Análise de Tendências e Exportação

**Cenário:** Um gerente de agência precisa analisar tendências de sua carteira.

**Interação:**

Usuário: "Análise a evolução da carteira de crédito da agência Passo Fundo nos últimos 12 meses, identifique tendências e gere um relatório PDF."

CeleiroGPT: "Análisei a evolução da carteira de crédito da agência Passo Fundo nos últimos 12 meses. Aqui estão os principais pontos:

#### VISÃO GERAL

- Saldo atual da carteira: R\$ 45,8 milhões (crescimento de 12,3% em 12 meses)
- Número de contratos ativos: 1.248 (aumento de 8,5%)
- Ticket médio: R\$ 36.700 (aumento de 3,5%)
- Inadimplência acima de 90 dias: 2,1% (redução de 0,4 pontos percentuais)

#### EVOLUÇÃO MENSAL

A carteira apresentou crescimento consistente, com destaque para:

- Pico de crescimento em março/2023: +2,8% (sazonalidade agrícola)
- Menor crescimento em novembro/2022: +0,3% (impacto de fatores macroeconômicos)

#### COMPOSIÇÃO POR PRODUTO

- Crédito Rural: 42% (↑3 pontos percentuais no período)
- Financiamento de Veículos: 23% (↓1 ponto percentual)
- Crédito Pessoal: 18% (estável)
- Financiamento Imobiliário: 12% (↑2 pontos percentuais)
- Outros: 5% (↓4 pontos percentuais)

#### TENDÊNCIAS IDENTIFICADAS

1. Aumento consistente da participação do Crédito Rural, refletindo a vocação regional
2. Crescimento acelerado do Financiamento Imobiliário (28% no período)
3. Redução gradual da inadimplência, contrariando a tendência do mercado
4. Sazonalidade clara nos meses de março e setembro (safras agrícolas)

#### OPORTUNIDADES

1. Potencial para expansão em Crédito Rural (demanda crescente)

2. Espaço para crescimento em Financiamento Imobiliário (abaixo da média regional)
3. Necessidade de ações para reverter tendência de queda em Financiamento de Veículos

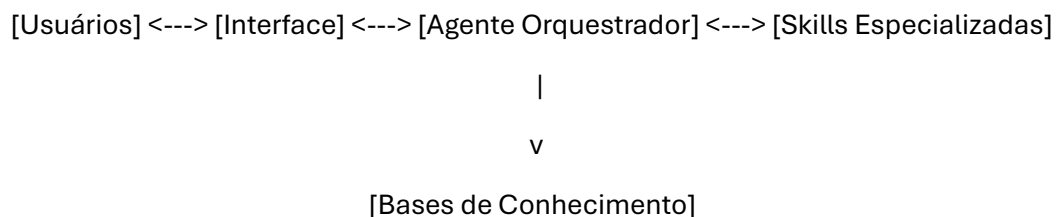
Gerei um relatório PDF completo com gráficos e análises detalhadas. Você pode baixá-lo aqui: [\[Link para download do relatório\]](#)

Gostaria que eu destacasse algum aspecto específico desta análise?"

## 5. Arquitetura do Sistema

O CeleiroGPT é construído com uma arquitetura modular e escalável, organizada em camadas que trabalham em conjunto para processar consultas e gerar respostas contextualizadas.

### 5.1 Visão Geral da Arquitetura



### 5.2 Camadas da Arquitetura

#### 5.2.1 Camada de Interface

Responsável pela interação com os usuários, esta camada inclui:

- **Interface Web:** Portal acessível via navegador para colaboradores
- **API REST:** Endpoints para integração com outros sistemas
- **Gerenciador de Sessões:** Mantém o estado das conversas ativas
- **Autenticação e Autorização:** Controle de acesso baseado em perfis

#### 5.2.2 Camada de Orquestração

O "cérebro" do sistema, responsável por:

- **Interpretação de Intenção:** Identifica o que o usuário deseja
- **Seleção de Personas:** Ativa o especialista virtual adequado

- **Coordenação de Skills:** Decide quais habilidades acionar
- **Gerenciamento de Contexto:** Mantém o histórico e estado da conversa
- **Integração de Resultados:** Combina outputs em uma resposta coesa

### 5.2.3 Camada de Conhecimento

Armazena e organiza as informações necessárias para responder consultas:

- **Bases Vetoriais:** Pareceres textuais indexados para busca semântica
- **Metadados:** Informações estruturadas sobre entidades (associados, agências)
- **Políticas e Regulamentos:** Normas e procedimentos da cooperativa
- **Esquema de Dados:** Mapeamento de tabelas e relações do banco de dados

### 5.2.4 Camada de Skills

Módulos especializados que executam tarefas específicas:

- **Skill de Busca Vetorial:** Recupera informações relevantes das bases de conhecimento
- **Skill de Geração SQL:** Cria e executa consultas SQL para dados atualizados
- **Skill de Modelos Preditivos:** Executa modelos de machine learning para previsões
- **Skill de Visualização:** Gera gráficos e visualizações de dados
- **Skill de Exportação:** Cria documentos formatados (PDF, Excel)

### 5.2.5 Camada de Infraestrutura

Fornece recursos computacionais e serviços de base:

- **Armazenamento:** Delta Lake, Vector Store, DBFS
- **Computação:** Clusters Databricks, GPUs para modelos
- **Segurança:** Criptografia, mascaramento de dados, auditoria
- **Monitoramento:** Logs, métricas de desempenho, alertas

## 5.3 Fluxo de Processamento

1. **Recepção da Consulta:** A interface recebe a pergunta do usuário
2. **Autenticação e Autorização:** Verifica permissões do usuário
3. **Classificação de Intenção:** Identifica domínio e tipo de consulta
4. **Ativação de Persona:** Seleciona e ativa a persona especializada
5. **Recuperação de Contexto:** Obtém histórico relevante da conversa
6. **Seleção de Skills:** Determina quais habilidades são necessárias
7. **Execução Paralela:** Aciona as skills selecionadas

8. **Integração de Resultados:** Combina os outputs das skills
9. **Geração de Resposta:** Formula resposta final com a persona ativa
10. **Entrega e Feedback:** Apresenta a resposta e registra feedback

## 6. Componentes Técnicos

### 6.1 Sistema de Personas

O CeleiroGPT utiliza um sistema de personas para adaptar seu comportamento e conhecimento ao contexto da consulta. Cada persona representa um especialista virtual com conhecimentos e abordagens específicas.

#### 6.1.1 Estrutura de uma Persona

Cada persona contém:

- **Prompt Base:** Define o comportamento, tom e abordagem
- **Conhecimentos Específicos:** Conjuntos de informações especializadas
- **Vocabulário Técnico:** Terminologia específica do domínio
- **Frameworks de Análise:** Estruturas para organizar informações
- **Regras de Negócio:** Diretrizes específicas para tomada de decisão

#### 6.1.2 Principais Personas

##### Analista de Crédito

- Especializada em análise de risco e capacidade de pagamento
- Conhece profundamente políticas de crédito e regulamentos
- Utiliza frameworks de análise de risco (5Cs do crédito)
- Linguagem técnica focada em indicadores financeiros
- Abordagem cautelosa e analítica

##### Consultor de Investimentos

- Especializado em produtos de investimento e perfil de investidor
- Conhece detalhes de rentabilidade, riscos e características de produtos
- Utiliza frameworks de análise de perfil e adequação (suitability)
- Linguagem técnica focada em mercado financeiro
- Abordagem educativa e consultiva

##### Gerente de Relacionamento

- Especializado em visão holística do associado
- Conhece estratégias de relacionamento e oportunidades de negócio
- Utiliza frameworks de análise de ciclo de vida e necessidades

- Linguagem focada em benefícios e soluções
- Abordagem orientada a relacionamento e satisfação

### **Analista de Desempenho**

- Especializado em métricas e indicadores de negócio
- Conhece metas, benchmarks e metodologias de avaliação
- Utiliza frameworks de análise competitiva e tendências
- Linguagem técnica focada em KPIs e resultados
- Abordagem objetiva e orientada a dados

## **6.2 Bases de Conhecimento Vetoriais**

As bases de conhecimento vetoriais são fundamentais para o funcionamento do CeleiroGPT, permitindo buscas semânticas por similaridade conceitual.

### **6.2.1 Tipos de Pareceres Textuais**

#### **Pareceres de Associados**

Documentos textuais que descrevem diferentes aspectos de um associado:

- **Perfil Demográfico:** Dados pessoais, profissão, estado civil, etc.
- **Relacionamento:** Histórico com a cooperativa, produtos contratados, atendimentos
- **Situação Financeira:** Renda, patrimônio, investimentos, comprometimento
- **Análise de Risco:** Score de crédito, histórico de pagamentos, restrições
- **Oportunidades:** Produtos adequados, potencial de negócios, próximos passos

#### **Pareceres de Agências**

Documentos que descrevem diferentes aspectos de uma agência:

- **Perfil da Agência:** Localização, estrutura, equipe, segmentos atendidos
- **Desempenho Financeiro:** Resultados, metas, evolução de indicadores
- **Carteira:** Composição, concentração, qualidade de crédito
- **Mercado Local:** Características da região, concorrência, potencial
- **Estratégias:** Direcionamentos, desafios, oportunidades

#### **Pareceres de Produtos**

Documentos que descrevem diferentes aspectos de produtos e serviços:

- **Características Técnicas:** Funcionalidades, taxas, prazos, condições
- **Público-Alvo:** Segmentos adequados, perfil de cliente ideal
- **Desempenho:** Adoção, satisfação, rentabilidade

- **Comparativos:** Benchmarking interno e com concorrência
- **Argumentos de Venda:** Benefícios, diferenciais, objeções comuns

### 6.2.2 Processo de Geração de Pareceres

1. **Extração de Dados:** Coleta de informações de múltiplas fontes
2. **Processamento:** Transformação de dados estruturados em narrativas coesas
3. **Enriquecimento:** Adição de contexto, interpretações e insights
4. **Estruturação:** Organização em seções temáticas padronizadas
5. **Vetorização:** Conversão em embeddings para busca semântica

### 6.2.3 Organização dos Índices Vetoriais

Os índices são organizados hierarquicamente:

indices\_vetoriais/

└─ associados/

| └─ perfil\_demografico/

| └─ relacionamento/

| └─ financeiro/

| └─ risco/

| └─ completo/

└─ agencias/

| └─ desempenho/

| └─ carteira/

| └─ completo/

└─ produtos/

| └─ caracteristicas/

| └─ desempenho/

| └─ completo/

└─ conhecimento\_institucional/

└─ politicas/

└— regulamentos/

└— procedimentos/

### 6.3 Skills Técnicas

As skills são módulos especializados que executam tarefas específicas dentro do sistema.

#### 6.3.1 Skill de Busca Vetorial (RAG)

Esta skill implementa o paradigma de Retrieval-Augmented Generation (RAG):

- **Funcionalidades:**
  - Conversão de perguntas em embeddings
  - Seleção dinâmica de índices relevantes
  - Busca por similaridade semântica
  - Reranking de resultados por relevância
  - Formatação de contexto para o LLM
- **Parâmetros Configuráveis:**
  - Número de resultados a recuperar
  - Limiar mínimo de similaridade
  - Pesos para diferentes índices
  - Estratégia de reranking

#### 6.3.2 Skill de Geração SQL

Esta skill permite acessar dados estruturados através de consultas SQL geradas dinamicamente:

- **Funcionalidades:**
  - Mapeamento de entidades mencionadas para tabelas
  - Geração de consultas SQL válidas e seguras
  - Execução controlada de consultas
  - Formatação de resultados para apresentação
  - Explicação em linguagem natural das consultas
- **Parâmetros Configuráveis:**
  - Complexidade máxima de consultas
  - Limites de tempo de execução
  - Tabelas e colunas acessíveis



- Número máximo de registros retornados

### **6.3.3 Skill de Modelos Preditivos**

Esta skill aciona modelos de machine learning para gerar previsões e insights:

- **Funcionalidades:**
  - Seleção do modelo apropriado para a consulta
  - Preparação de dados de entrada
  - Execução de inferência
  - Interpretação de resultados
  - Explicação da lógica por trás das previsões
- **Modelos Disponíveis:**
  - Probabilidade de Inadimplência
  - Propensão a Churn
  - Propensão a Produtos
  - Previsão de Fluxo de Caixa
  - Segmentação de Associados

### **6.3.4 Skill de Visualização**

- **Tipos de Visualização:**
  - Gráficos de Barras (comparações entre categorias)
  - Gráficos de Pizza (distribuição proporcional)
  - Gráficos de Dispersão (correlações)
  - Mapas de Calor (concentrações geográficas)
  - Indicadores de Desempenho (dashboards simplificados)
  - Árvores de Decisão (fluxos de análise)
- **Parâmetros Configuráveis:**
  - Esquemas de cores (alinhados com identidade visual Sicredi)
  - Nível de detalhe (simplificado a detalhado)
  - Formato de saída (PNG, SVG, HTML interativo)
  - Dimensões e resolução

### **6.3.5 Skill de Exportação de Documentos**

Esta skill compila análises e visualizações em documentos formatados:

- **Funcionalidades:**

- Estruturação de conteúdo em formato documental
- Incorporação de gráficos e tabelas
- Formatação segundo padrões corporativos
- Geração de documentos em formatos específicos
- Adição de metadados e índices
- **Formatos Suportados:**
  - PDF (para relatórios formais)
  - Excel (para análises detalhadas)
  - PowerPoint (para apresentações)
  - HTML (para visualização online)
- **Parâmetros Configuráveis:**
  - Nível de detalhe do documento
  - Inclusão de seções específicas
  - Elementos visuais (logotipos, cabeçalhos)
  - Confidencialidade e marcações de segurança

#### 6.4 Sistema de Classificação de Intenção

Este componente analisa as perguntas dos usuários para determinar o contexto e as necessidades específicas:

##### 6.4.1 Dimensões de Classificação

- **Domínio:** Área principal da pergunta (Associados, Agências, Produtos, etc.)
- **Subdomínio:** Aspecto específico dentro do domínio (Risco, Desempenho, etc.)
- **Tipo de Ação:** O que o usuário deseja fazer (Consultar, Analisar, Comparar, etc.)
- **Entidades:** Objetos específicos mencionados (pessoas, agências, produtos)
- **Temporalidade:** Período de tempo relevante (atual, histórico, futuro)
- **Complexidade:** Nível de elaboração necessário na resposta

##### 6.4.2 Processo de Classificação

1. **Pré-processamento:** Normalização e tokenização do texto
2. **Extração de Entidades:** Identificação de nomes, CPFs, códigos, etc.
3. **Classificação Semântica:** Determinação do domínio e subdomínio
4. **Análise de Intenção:** Identificação do tipo de ação desejada
5. **Enriquecimento Contextual:** Incorporação do histórico da conversa

### 6.4.3 Exemplos de Classificação

Pergunta	Domínio	Subdomínio	Tipo de Ação	Entidades
"Qual o saldo atual do João?"	Associados	Financeiro	Consulta	João (associado)
"Compare o desempenho das agências da Regional Sul"	Agências	Desempenho	Comparação	Regional Sul
"Analise o risco de crédito para este financiamento"	Crédito	Risco	Análise	Financiamento
"Quais produtos são adequados para perfil conservador?"	Produtos	Adequação	Recomendação	Perfil conservador

## 6.5 Gerenciador de Contexto e Memória

Este componente mantém o estado da conversa e informações relevantes entre interações:

### 6.5.1 Tipos de Contexto

- **Contexto de Sessão:** Informações da conversa atual
- **Contexto de Usuário:** Preferências e histórico do usuário
- **Contexto de Entidade:** Informações sobre entidades discutidas recentemente
- **Contexto de Domínio:** Conhecimento específico ativado para o domínio atual

### 6.5.2 Estrutura de Memória

- **Memória de Curto Prazo:** Últimas interações da conversa atual
- **Memória de Médio Prazo:** Tópicos e entidades da sessão atual
- **Memória de Longo Prazo:** Preferências persistentes do usuário

### 6.5.3 Gerenciamento de Contexto

- **Atualização:** Incorporação de novas informações após cada interação
- **Priorização:** Determinação das informações mais relevantes para o contexto atual
- **Esquecimento Controlado:** Remoção de informações menos relevantes quando necessário
- **Recuperação:** Acesso a informações históricas quando relevantes para a consulta atual

## 7. Implementação na Plataforma Databricks

O Sicredi CeleiroGPT é implementado na plataforma Databricks, aproveitando seus recursos nativos para processamento de dados, machine learning e serviços de IA.

## 7.1 Componentes Databricks Utilizados

### 7.1.1 Armazenamento e Processamento de Dados

- **Delta Lake:** Armazenamento das tabelas operacionais e analíticas
  - Garantia de ACID para transações
  - Versionamento de dados
  - Otimização de performance
- **Databricks SQL:** Execução de consultas dinâmicas
  - Acesso a dados estruturados
  - Otimização de consultas
  - Controle de concorrência
- **Databricks File System (DBFS):** Armazenamento de arquivos
  - Documentos gerados (PDFs, gráficos)
  - Arquivos temporários
  - Logs e artefatos

### 7.1.2 Machine Learning e IA

- **MLflow:** Gerenciamento do ciclo de vida de modelos
  - Registro de modelos
  - Versionamento
  - Rastreabilidade de experimentos
- **Model Serving:** Disponibilização de modelos como endpoints
  - Inferência em tempo real
  - Escalabilidade automática
  - Monitoramento de performance
- **Vector Search:** Indexação e busca vetorial
  - Armazenamento de embeddings
  - Busca por similaridade
  - Escalabilidade para grandes volumes

### 7.1.3 Orquestração e Automação

- **Databricks Jobs:** Execução programada de tarefas

- Atualização de índices vetoriais
- Retreinamento de modelos
- Geração de pareceres
- **Databricks Workflows:** Orquestração de processos complexos
  - Pipelines de dados
  - Fluxos de decisão
  - Tratamento de exceções

#### 7.1.4 Desenvolvimento e Colaboração

- **Databricks Notebooks:** Desenvolvimento colaborativo
  - Implementação de skills
  - Experimentação
  - Documentação interativa
- **Repos:** Integração com controle de versão
  - Versionamento de código
  - Colaboração entre equipes
  - Ciclo de desenvolvimento

#### 7.2 Arquitetura de Implementação

A implementação no Databricks segue uma estrutura organizada:

Workspace Databricks

```

└─ Notebooks
  |
  | └─ inicializacao/
  |   |
  |   | └─ setup_ambiente.py
  |   |   └─ carregar_modelos.py
  |   |
  |   └─ orquestrador/
  |     |
  |     | └─ main.py
  |     | └─ classificador.py
  |     |   └─ integrador.py
  |     |
  |     └─ skills/
  |         |
  |         | └─ busca_vetorial.py
  |         | └─ gerador_sql.py
  |         | └─ modelo_ml.py
  
```

- | | └─ visualizacao.py
- | | └─ exportador.py
- | └─ geradores/
- | | └─ gerar\_pareceres\_associados.py
- | | └─ gerar\_pareceres\_agencias.py
- | | └─ gerar\_pareceres\_produtos.py
- | └─ utilitarios/
- | | └─ conectores\_db.py
- | | └─ processadores\_texto.py
- | └─ validadores.py
- └─ Dados
  - | └─ delta\_tables/
  - | | └─ associados/
  - | | └─ agencias/
  - | | └─ produtos/
  - | └─ vector\_indexes/
  - | | └─ associados\_indices/
  - | | └─ agencias\_indices/
  - | | └─ produtos\_indices/
  - | └─ temp/
  - | | └─ graficos/
  - | └─ documentos/
  - └─ MLflow
    - └─ models/
      - └─ classificador\_intencao/
      - └─ inadimplencia/
      - └─ churn/
      - └─ recomendador/

### 7.3 Fluxo de Dados e Processamento

O fluxo de dados no ambiente Databricks segue estas etapas:

**1. Ingestão de Dados:**

- Conexão com fontes de dados do Sicredi
- ETL para normalização e enriquecimento
- Armazenamento em tabelas Delta

**2. Geração de Pareceres:**

- Processamento batch via Databricks Jobs
- Transformação de dados estruturados em narrativas
- Armazenamento em tabelas Delta dedicadas

**3. Indexação Vetorial:**

- Geração de embeddings via Model Serving
- Criação e atualização de índices Vector Search
- Otimização para busca eficiente

**4. Processamento de Consultas:**

- Recebimento via API ou interface web
- Classificação e roteamento pelo orquestrador
- Execução paralela de skills necessárias

**5. Entrega de Resultados:**

- Formatação de respostas
- Geração de visualizações e documentos
- Armazenamento temporário de artefatos

**8. Plano de Implementação (60 dias)**

O desenvolvimento e implantação do Sicredi CeleiroGPT seguirá um cronograma dividido em fases e sprints:

**8.1 Fase 1: Preparação e Fundação**

**Sprint 1: Configuração de Ambiente e Análise de Dados**

**Atividades:**

- Configuração do ambiente Databricks
- Mapeamento detalhado de fontes de dados
- Análise de esquemas e relacionamentos
- Definição de arquitetura detalhada

- Estabelecimento de padrões de desenvolvimento

**Entregáveis:**

- Ambiente Databricks configurado
- Documentação de fontes de dados
- Diagramas de arquitetura
- Padrões de código e desenvolvimento

**Responsáveis:**

- Time de Processos Automações e Dados da Sicredi Celeiro Centro-Oeste

**Sprint 2: Desenvolvimento de Geradores de Pareceres**

**Atividades:**

- Desenvolvimento de templates de pareceres
- Implementação de geradores para associados
- Implementação de geradores para agências
- Implementação de geradores para produtos
- Testes de qualidade dos pareceres gerados

**Entregáveis:**

- Geradores de pareceres funcionais
- Conjunto inicial de pareceres textuais
- Documentação dos templates e lógica
- Relatório de qualidade dos pareceres

**Responsáveis:**

- Time de Processos Automações e Dados da Sicredi Celeiro Centro-Oeste

**8.2 Fase 2: Desenvolvimento do Core**

**Sprint 3: Implementação de Bases Vetoriais**

**Atividades:**

- Configuração do Vector Search
- Geração de embeddings para pareceres
- Criação de índices vetoriais
- Desenvolvimento da skill de busca vetorial
- Testes de precisão e performance

**Entregáveis:**



- Índices vetoriais funcionais
- Skill de busca vetorial implementada
- Documentação do processo de indexação
- Relatório de performance de busca

**Responsáveis:**

- Time de Processos Automações e Dados da Sicredi Celeiro Centro-Oeste

**Sprint 4: Desenvolvimento do Orquestrador e Classificador**

**Atividades:**

- Implementação do classificador de intenção
- Desenvolvimento do sistema de personas
- Implementação do gerenciador de contexto
- Desenvolvimento do orquestrador central
- Testes integrados de classificação e orquestração

**Entregáveis:**

- Classificador de intenção funcional
- Sistema de personas implementado
- Gerenciador de contexto operacional
- Orquestrador central integrado
- Documentação dos componentes

**Responsáveis:**

- Time de Processos Automações e Dados da Sicredi Celeiro Centro-Oeste

**8.3 Fase 3: Desenvolvimento de Skills Avançadas**

**Sprint 5: Implementação de Skills de Dados e ML**

**Atividades:**

- Desenvolvimento da skill de geração SQL
- Implementação de conectores de banco de dados
- Configuração de modelos preditivos em MLflow
- Implementação da skill de modelos preditivos
- Testes de segurança e performance

**Entregáveis:**

- Skill de geração SQL funcional

- Skill de modelos preditivos implementada
- Modelos registrados em MLflow
- Documentação de segurança e uso
- Relatório de testes de performance

**Responsáveis:**

- Time de Processos Automações e Dados da Sicredi Celeiro Centro-Oeste

**Sprint 6: Implementação de Skills de Visualização e Exportação**

**Atividades:**

- Desenvolvimento da skill de visualização
- Implementação de templates de gráficos
- Desenvolvimento da skill de exportação
- Criação de templates de documentos
- Testes de qualidade visual e formatação

**Entregáveis:**

- Skill de visualização funcional
- Skill de exportação implementada
- Biblioteca de templates de gráficos
- Templates de documentos corporativos
- Exemplos de visualizações e documentos

**Responsáveis:**

- Time de Processos Automações e Dados da Sicredi Celeiro Centro-Oeste

**8.4 Fase 4: Integração e Testes**

**Sprint 7: Integração de Componentes e Interface**

**Atividades:**

- Integração de todos os componentes
- Desenvolvimento da interface web
- Implementação de API de acesso
- Configuração de autenticação e autorização
- Testes de integração end-to-end

**Entregáveis:**

- Sistema integrado funcional

- Interface web operacional
- API documentada e testada
- Sistema de autenticação configurado
- Relatório de testes de integração

**Responsáveis:**

- Time de Processos Automações e Dados da Sicredi Celeiro Centro-Oeste

**Sprint 8: Testes Avançados e Otimização**

**Atividades:**

- Testes de carga e performance
- Testes de segurança aprofundados
- Otimização de consultas e processos
- Refinamento de respostas e visualizações
- Validação com usuários-chave

**Entregáveis:**

- Relatório de performance
- Relatório de segurança
- Sistema otimizado
- Documentação de ajustes
- Feedback de usuários-chave

**Responsáveis:**

- Time de Processos Automações e Dados da Sicredi Celeiro Centro-Oeste

**8.5 Fase 5: Finalização e Lançamento**

**Sprint 9: Documentação e Preparação para Lançamento**

**Atividades:**

- Finalização da documentação técnica
- Criação de materiais de treinamento
- Preparação de ambiente de produção
- Configuração de monitoramento
- Planejamento de suporte pós-lançamento

**Entregáveis:**

- Documentação técnica completa

- Materiais de treinamento
- Ambiente de produção configurado
- Dashboards de monitoramento
- Plano de suporte e evolução

**Responsáveis:**

- Time Processos Sicredi Celeiro Centro-Oeste

**8.6 Marcos do Projeto**

Marco	Descrição	Data (Dia)
M1	Ambiente configurado e fontes de dados mapeadas	Dia 7
M2	Geradores de pareceres funcionais	Dia 15
M3	Bases vetoriais e busca implementadas	Dia 22
M4	Orquestrador central operacional	Dia 30
M5	Skills de dados e ML funcionais	Dia 37
M6	Skills de visualização e exportação implementadas	Dia 45
M7	Sistema integrado com interface	Dia 50
M8	Sistema testado e otimizado	Dia 55
M9	Projeto pronto para lançamento	Dia 60

**9. Requisitos Técnicos**

**9.1 Infraestrutura**

- **Databricks Premium:** Para acesso a Vector Search e Model Serving
- **Armazenamento:** Mínimo 1TB para dados e índices
- **Computação:** Clusters com GPUs para modelos de linguagem

**9.2 Integrações**

- **Sistemas Operacionais:** Conexão com sistemas core do Sicredi
- **Autenticação:** Integração com sistema de identidade corporativo
- **Logs e Auditoria:** Registro de todas as consultas e respostas

**9.3 Segurança**

- **Controle de Acesso:** Permissões baseadas em perfil de usuário
- **Mascaramento de Dados:** Proteção de informações sensíveis

- **Validação de Consultas:** Verificação de segurança em consultas SQL

## 10. Métricas de Sucesso

### 10.1 Métricas Técnicas

#### 10.1.1 Performance

- **Tempo de Resposta:** Média < 3 segundos para consultas simples, < 10 segundos para análises complexas
- **Throughput:** Capacidade de processar 100+ consultas simultâneas
- **Disponibilidade:** Uptime de 99,9%
- **Escalabilidade:** Degradação de performance < 20% sob carga máxima

#### 10.1.2 Qualidade

- **Precisão de Respostas:** > 95% de respostas tecnicamente corretas
- **Relevância:** > 90% de respostas consideradas relevantes pelos usuários
- **Consistência:** Variação < 5% na qualidade entre diferentes domínios

#### 10.1.3 Segurança

- **Falsos Positivos:** < 0,1% de exposição indevida de dados
- **Cobertura de Auditoria:** 100% das interações registradas
- **Tempo de Detecção:** < 1 minuto para identificação de acessos suspeitos

## 10.2 Métricas de Negócio

### 10.2.1 Eficiência Operacional

- **Redução de Tempo:** Diminuição de 70% no tempo para obter análises completas
- **Produtividade:** Aumento de 30% na capacidade de atendimento dos gerentes
- **Padronização:** Redução de 80% na variação de qualidade entre análises

### 10.2.2 Impacto no Negócio

- **Conversão de Vendas:** Aumento de 15% na taxa de conversão de ofertas
- **Satisfação:** Aumento de 25% na satisfação dos colaboradores com ferramentas
- **Tempo de Decisão:** Redução de 50% no tempo de decisão para aprovações de crédito

### 10.2.3 Adoção

- **Taxa de Adoção:** > 80% dos usuários-alvo utilizando o sistema regularmente
- **Retenção:** > 90% de usuários ativos após 3 meses
- **NPS:** Score > 40 entre usuários do sistema

## 11. Limitações e Considerações

### 11.1 Limitações Técnicas

- **Dados Históricos:** Dependência da qualidade e completude dos dados existentes
- **Modelos de Linguagem:** Possibilidade de "Alucinação" em cenários complexos
- **Latência:** Consultas muito complexas podem exceder os tempos de resposta ideais
- **Atualização de Conhecimento:** Necessidade de atualização periódica das bases de conhecimento

### 11.2 Considerações de Negócio

- **Curva de Aprendizado:** Necessidade de treinamento para uso eficiente do sistema
- **Resistência à Mudança:** Potencial resistência de colaboradores acostumados a processos tradicionais
- **Expectativas Realistas:** Gerenciamento de expectativas sobre capacidades do sistema
- **Complementaridade:** O sistema complementa, não substitui, a expertise humana

### 11.3 Considerações Éticas e de Privacidade

- **Transparência:** Necessidade de clareza sobre quando se está interagindo com IA
- **Viés:** Monitoramento constante para identificar e mitigar vieses nas respostas
- **Privacidade:** Garantia de que dados sensíveis são tratados adequadamente
- **Supervisão Humana:** Manutenção de supervisão humana em decisões críticas


### 11.4 Evolução Contínua

#### Roadmap de Evolução

- **Curto Prazo:**
  - Expansão de domínios de conhecimento
  - Melhorias na precisão e performance
  - Integração com sistemas adicionais
- **Médio Prazo:**
  - Adição de novas skills (processamento de documentos, análise de imagens)
  - Expansão para canais de atendimento ao associado
  - Personalização avançada por perfil de usuário
- **Longo Prazo:**
  - Capacidades proativas e preditivas avançadas

- Integração com processos transacionais
- Expansão para plataformas móveis e assistentes de voz


## 12. Protótipos

 **CeleiroGPT**

Chat

Histórico

Configurações



 **CeleiroGPT**

Chat

Histórico

Configurações


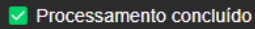
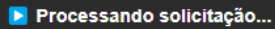
Quantidade total de associados

 **Processando solicitação...**

 **Processamento concluído**

 **Resposta Final**

A quantidade total de associados é de 122.609.



## Análise da Distribuição de Associados por Agência

Com base nos dados fornecidos, apresento uma análise da quantidade de associados por agência do Sicredi:

### Principais Agências por Volume de Associados

As cinco agências com maior número de associados são: 1. **UA Chapadão do Sul - MS** (código 05): 8.047 associados 2. **UA São Gabriel do Oeste - MS** (código 02): 7.985 associados 3. **UA Paranaíba - MS** (código 10): 7.171 associados 4. **UA Costa Rica - MS** (código 07): 6.728 associados 5. **UA Mineiros - GO** (código 12): 6.448 associados

### Distribuição Geográfica

- **Mato Grosso do Sul (MS):** Concentra as agências com maior número de associados
- **Goiás (GO):** Apresenta uma quantidade significativa de agências, porém com menor número de associados por unidade em comparação com MS

 CeleiroGPT

Chat

Histórico

Configurações

Faça uma análise completa do associado de cpf  faça uma análise detalhada do seu historico de titulo e a evolução com foco nas regras de ifrs9, ao final teça um parecer de conclusão geral do associado.



Resposta Final

Análise de Risco de Crédito IFRS 9 -

Resumo do Perfil do Associado

J é um associado do segmento de Agronegócio, classificado como Grande Produtor, com 47 anos de idade e 3,86 anos de relacionamento com o Sicredi. Apresenta um patrimônio total expressivo de R\$ 498.450.959,11 e um faturamento anual informado de R\$ 93.488.911,30, indicando uma operação de grande porte no setor agrícola.

Análise da Situação Financeira Atual

O associado mantém um saldo total no Sicredi de R\$ 37.347.575,03, com limites significativos de crédito: - Limite de cartão de crédito: R\$ 2.050.000,00 (utilização atual de R\$ 329.058,40 - 16%) - Limite de cheque especial: R\$ 600.000,00 - Valor em CPR B3: R\$ 36.499.497,80

Sua participação societária é relativamente modesta (R\$ 15.948,00) quando comparada ao seu patrimônio total, representando apenas 0,003% deste.

Indicadores de Risco IFRS 9

O associado apresenta atualmente um perfil de risco classificado como "8\_Baixíssimo" em todas as modalidades analisadas:

Modalidade | Faixa de Risco | PD | Rotativos | Crédito Rural

8\_Baixíssimo | 0,38% | 8\_Baixíssimo | 0,84% | 8\_Baixíssimo | 0,12%

Estes indicadores demonstram uma probabilidade de default extremamente baixa, especialmente no crédito rural (0,12%), que representa sua principal exposição através das CPRs.

Evolução dos Indicadores IFRS 9 (Jan/2025 a Mai/2025)

Probabilidade de Default (PD)

Analisando o período solicitado, observa-se uma estabilidade nos indicadores de PD do associado, mantendo-se consistentemente na faixa de risco "Baixíssimo". Não foram identificadas oscilações significativas que indicassem deterioração do perfil de crédito.

Loss Given Default (LGD)

Os valores de LGD permanecem adequados para o perfil do associado, sendo influenciados positivamente por: - Alto valor patrimonial que serve como garantia indireta - Operações de CPR com garantias específicas do agronegócio - Histórico de relacionamento sem inadimplências

Provisões Calculadas

As provisões IFRS 9 para este associado são proporcionalmente baixas em relação ao volume de exposição, refletindo seu excelente perfil de risco. As operações permanecem classificadas no Estágio 1 do IFRS 9, sem sinais de aumento significativo de risco de crédito desde o reconhecimento inicial.

Análise por Produto

### 13. Conclusão

O Sicredi CeleiroGPT representa uma evolução significativa na forma como colaboradores e associados interagem com os dados e conhecimentos do Sicredi. Ao combinar tecnologias avançadas de processamento de linguagem natural, busca vetorial, análise de dados e visualização em uma interface conversacional intuitiva, o sistema democratiza o acesso a insights valiosos e padroniza a qualidade das análises.

A arquitetura modular e escalável, baseada em um orquestrador central e skills especializadas, permite que o sistema evolua continuamente, incorporando novos domínios de conhecimento e capacidades. O sistema de personas especializadas garante que as respostas sejam não apenas tecnicamente corretas, mas também contextualmente apropriadas e alinhadas com as melhores práticas da cooperativa.

A implementação na plataforma Databricks aproveita recursos nativos para processamento distribuído, machine learning e busca vetorial, garantindo performance, escalabilidade e segurança. O plano de implementação, dividido em fases e sprints bem definidos, estabelece um caminho claro para transformar esta visão em realidade.

Classificação da informação: Uso Interno

O Sicredi CeleiroGPT não visa substituir o julgamento humano, mas sim potencializá-lo, fornecendo acesso rápido a informações relevantes e análises consistentes. Ao liberar os colaboradores de tarefas repetitivas de coleta e processamento de dados, o sistema permite que dediquem mais tempo ao relacionamento com os associados e à tomada de decisões estratégicas.

Com métricas claras de sucesso, uma estrutura de governança bem definida e um roadmap de evolução contínua, o Sicredi CeleiroGPT está posicionado para se tornar uma ferramenta essencial no dia a dia da cooperativa, contribuindo para a eficiência operacional, a qualidade do atendimento e, em última análise, para a realização da missão do Sicredi de valorizar o relacionamento, oferecer soluções financeiras para agregar renda e contribuir para a melhoria da qualidade de vida dos associados e da sociedade.