



Extração Inteligente de Dados de Documentos Fiscais – Versão MVP

Equipe “Challenge Accepted”

30 de outubro de 2025

Sumário

Descrição do Tema Escolhido	2
Público-Alvo	2
Dores e Desafios Atuais	2
Oportunidade e Justificativa de Valor	3
Objetivo	4
Objetivo Principal	4
Objetivos Secundários	5
Resultados Esperados	5
Plano de Trabalho	7
Fases do Projeto	7
Ambiente de Execução com Tecnologias Previstas	9
Arquitetura da Solução	9
Visão Geral da Arquitetura:	9
Tecnologias Previstas/Sugeridas	10
Planejamento Estratégico e Viabilidade de Negócio	11
Análise de Mercado	11
Análise de Viabilidade e Riscos	12
Estratégia de Monetização e Modelo de Negócio	12
Aspectos Legais e de Compliance	13
Plano de Marketing e Go-to-Market	13
Roadmap de Evolução e Funcionalidades Futuras	13
Cronograma	15
GitHub Repositório	15
Equipe de Trabalho	15

Descrição do Tema Escolhido

Desenvolver uma solução digital que permita às empresas em geral acelerar o processo de imputação de dados referentes às compras e vendas documentadas em Notas Fiscais para finalidade de criar histórico de compras e/ou vendas, precificação, regiões de maior demanda, características de consumo e ainda, poder formatar propostas competitivas, tendo como base a base pública de notas fiscais de compras governamentais, uma vez que estas notas estão disponíveis em portais de transparências.

Empresas de todos os portes lidam com um volume crescente de documentos fiscais, como Notas Fiscais (NF-e, NFS-e), Conhecimentos de Transporte (CT-e) e faturas. O processo manual de extração de dados desses documentos é lento, propenso a erros e de alto custo operacional. A inserção incorreta de dados em sistemas de gestão (ERP's) pode levar a problemas fiscais, contábeis e de *compliance*, além de dificultar a análise de dados para tomada de decisão estratégica.

Público-Alvo

Principal: Empresas interessadas em vender produtos e serviços para empresas públicas, órgãos governamentais ou empresas de economia mista, as quais buscam entender o comportamento de consumo destes órgãos, identificar carências em produtos e serviços, com base em padrão de consumo de empresas/órgãos correlatos, podendo tomar decisão de participar em processos de compras / licitações.

Secundário: os próprios órgãos governamentais poderiam utilizar a solução para ter maior inteligência na tomada de decisão sobre compras, realizar análise histórica, identificar seus padrões de consumo, identificar valores pagos nas compras e melhores fornecedores, entre outros critérios.

Dores e Desafios Atuais

Sob a ótica de processos / tecnologia:

- **Baixa Eficiência:** O processo manual de digitação consome um tempo significativo da equipe, que poderia ser alocado em atividades de maior valor agregado.
- **Alta Incidência de Erros:** A digitação manual está sujeita a erros humanos, resultando em dados inconsistentes e retrabalho.
- **Dificuldade de Escalabilidade:** O aumento no volume de documentos exige um aumento proporcional da equipe, tornando o processo caro e difícil de gerenciar.
- **Falta de Visibilidade em Tempo Real:** A demora no processamento impede uma visão ágil e em tempo real das informações financeiras e fiscais.
- **Adaptação a Múltiplos Formatos:** A grande variedade de layouts de notas fiscais, especialmente entre diferentes municípios para NFS-e, torna a automação baseada em templates fixos (leitura por posição) ineficaz.

Sob a ótica de negócios:

- **Carência de Informação:** não há uma base centralizada dentro dos conceitos modernos de Inteligência de Negócios que permitem identificar lacunas, demandas de consumo, padrões de comportamento em compras de produtos e serviços por órgãos governamentais.

Oportunidade e Justificativa de Valor

Oportunidade

A digitalização e a transparência das compras públicas no Brasil criaram um ecossistema riquíssimo de dados abertos, especialmente com a obrigatoriedade da publicação de informações sobre contratações no Portal Nacional de Contratações Públicas (PNCP) e em sistemas como o Compras.gov.br. Empresas fornecedoras de produtos e serviços para órgãos públicos enfrentam, porém, desafios como:

Dificuldade em identificar oportunidades de venda em meio ao grande volume de licitações e contratações publicadas diariamente nos diversos portais.

Necessidade de analisar históricos de compras, preços praticados, sazonalidade de demanda e regiões com maior volume de aquisições, para embasar decisões comerciais e formatar propostas competitivas.

Falta de ferramentas integradas que cruzem dados públicos de notas fiscais eletrônicas, editais, contratos e perfis de consumo dos órgãos, limitando a inteligência comercial e a capacidade de precificação estratégica.

Nesse cenário, há uma oportunidade clara para uma solução digital que automatize a coleta, análise e apresentação dessas informações, facilitando a vida dos fornecedores e potencializando sua participação e competitividade nas compras públicas.

Justificativa de Valor

Uma solução digital com o objetivo descrito entrega valor significativo para as empresas fornecedoras e para o próprio setor público, pelos seguintes motivos:

Análise Inteligente de Histórico e Precificação:

Utiliza dados públicos de notas fiscais eletrônicas e bases de compras anteriores para identificar padrões de consumo, preços praticados, sazonalidade e regiões com maior demanda, permitindo ao fornecedor calibrar suas propostas e estratégias de venda.

Propostas Mais Competitivas:

Com base em dados reais de mercado e histórico de compras governamentais, a solução sugere faixas de preço e diferenciais competitivos, aumentando as chances de sucesso nas licitações e evitando a apresentação de valores fora do padrão de mercado.

Expansão de Mercado e Inclusão

Facilita o acesso de pequenas e médias empresas ao mercado público, tradicionalmente dominado por grandes fornecedores, promovendo maior participação e democratização das oportunidades.

Eficiência, Transparência e Redução de Custos:

Automatiza tarefas repetitivas, reduz erros e retrabalho, e contribui para a transparência e conformidade dos processos, tanto para fornecedores quanto para o setor público, alinhando-se às diretrizes da Nova Lei de Licitações (Lei 14.133/2021) e à estratégia de governo digital.

Geração de Inteligência de Mercado:

Permite ao fornecedor monitorar concorrentes, identificar tendências e antecipar demandas, tornando a atuação no segmento público mais estratégica e menos reativa.

Identificação Ágil de Oportunidades::

Automatiza com o uso de agentes de IA a varredura dos principais portais de compras públicas, notificando o fornecedor sobre editais e demandas compatíveis com seu portfólio, reduzindo o tempo e o esforço de busca manual.

O desenvolvimento de uma solução digital baseada na análise da base pública de notas fiscais de compras governamentais responde a uma demanda real do mercado, promovendo eficiência, competitividade e inclusão. Ela agrega valor ao permitir que fornecedores identifiquem oportunidades, compreendam o comportamento de compra dos órgãos públicos e elaborem propostas mais assertivas, potencializando o sucesso nas vendas ao setor público e contribuindo para a modernização do ambiente de compras governamentais

Objetivo

Objetivo Principal

O objetivo da solução é fornecer às empresas fornecedoras de produtos e serviços para órgãos públicos uma ferramenta que permita identificar oportunidades de venda, analisar o histórico de compras governamentais, precificação, regiões de maior demanda e características de consumo. Ancorando-se na base pública de notas fiscais de compras governamentais, essa solução visa ajudar essas empresas a formatarem propostas competitivas, otimizando sua atuação no mercado de compras públicas. Isso inclui o uso de dados detalhados sobre as transações, valores empenhados, padrões sazonais e perfil dos órgãos compradores para apoiar decisões estratégicas e aumentar a competitividade nas licitações e contratos públicos

Para atingir este objetivo, iremos desenvolver e implantar um sistema de software capaz de processar imagens de documentos fiscais (em formatos como CSV, PDF, PNG, JPG), extrair automaticamente dados essenciais e disponibilizar esses dados de forma estruturada. Aplicaremos IA para avaliar os dados estruturados para responder perguntas relacionadas ao padrão de consumo, preços, regiões que mais consomem determinado produto/serviço, baseado nos dados estruturados.

Objetivos Secundários

- Extrair com precisão as seguintes informações:
 - Número da nota, Chave de Acesso
 - Emitente e Destinatário: CNPJ, Razão Social, Inscrição Estadual, Endereço.
 - Itens da Nota: Código, Descrição, quantidade, valor unitário e valor total.
 - Impostos Principais: Base de cálculo e valor de ICMS, IPI, PIS e COFINS.
 - Códigos Fiscais: CFOP (Código Fiscal de Operações e Prestações) e CST (Código de Situação Tributária).
 - Número do pedido
- Garantir que a solução seja adaptável a diferentes layouts de documentos fiscais, minimizando a necessidade de configuração manual para cada novo formato.
- Criar uma arquitetura de software que permita a atualização e o retreinamento dos modelos de NLP para se adaptar a futuras mudanças na legislação fiscal, como a introdução do IVA.
- Fornece uma estrutura bem documentada para a consulta dos dados extraídos.

Resultados Esperados

Identificação Otimizada de Oportunidades de Venda: A solução deve permitir que as empresas encontrem de forma eficiente e rápida novas oportunidades de negócios com órgãos públicos, filtrando por tipo de produto/serviço, região, valor etc.

Análise Abrangente do Histórico de Compras: Fornecer uma visão detalhada das compras passadas dos órgãos públicos, incluindo volumes, frequência, fornecedores anteriores e condições de aquisição.

Inteligência de Precificação: Oferecer insights sobre os preços praticados em licitações e compras anteriores, auxiliando as empresas a definirem preços competitivos e lucrativos para suas propostas.

Mapeamento de Regiões de Maior Demanda: Identificar geograficamente as áreas com maior necessidade de determinados produtos ou serviços, permitindo que as empresas direcionem seus esforços de venda de forma mais estratégica.

Compreensão das Características de Consumo: Analisar padrões de consumo dos órgãos públicos, como sazonalidade, especificações técnicas recorrentes e preferências, para que as empresas possam adaptar suas ofertas.

Formatação de Propostas Competitivas: Ajudar as empresas a elaborarem propostas mais alinhadas às necessidades dos órgãos públicos, aumentando suas chances de sucesso em processos licitatórios.

Redução de Custos e Tempo: Automatizar a pesquisa e análise de dados, diminuindo o tempo e os recursos que as empresas gastariam manualmente para identificar e qualificar oportunidades.

Aumento da Taxa de Sucesso em Licitações: Com informações mais precisas e análises aprofundadas, as empresas estarão mais preparadas para competir e vencer licitações governamentais.

Melhora na Tomada de Decisão Estratégica: Fornecer dados e análises que suportem decisões estratégicas sobre quais mercados atuar, quais produtos/serviços focar e como posicionar a empresa.

Acesso Facilitado a Dados Públicos: Transformar a complexa base de dados de notas fiscais em informações acessíveis e utilizáveis para as empresas, democratizando o acesso a esses dados.

Ao final do projeto, será implementado um sistema automatizado que transforma documentos fiscais não estruturados em dados estruturados e acionáveis, prontos para serem integrados em seus fluxos de trabalho, aumentando a eficiência e a confiabilidade de suas operações fiscais.

O principal entregável será uma plataforma de integração robusta que atuará como um *middleware* inteligente: ela irá consumir os dados já processados, validados e estruturados pelo provedor DaaS via API, para então adaptá-los, enriquecê-los com regras de negócio específicas do cliente e distribuí-los de forma segura para os sistemas internos, como ERP's e plataformas de BI. Com isso, o resultado esperado é um fluxo de dados contínuo e confiável que elimina a necessidade de processamento interno de documentos e garante o acesso imediato a informações fiscais precisas, alavancando a infraestrutura e a especialização de um fornecedor de dados já consolidado no mercado.

Plano de Trabalho

A execução do projeto seguirá uma metodologia híbrida, combinando fases bem definidas (cascata) para o planejamento e design da arquitetura, com uma abordagem Ágil (Scrum) para o desenvolvimento e iteração da solução, permitindo flexibilidade e entregas de valor contínuas.

Fases do Projeto

Fase 1: Coleta de Dados Iniciais

- **Atividades:**
 - **Coleta das Notas Fiscais:** Coletar um volume inicial de documentos para servirem como base para treinamento e testes.
 - **Formatos:** Recebimento de imagens e arquivos (fotos de celular, arquivos de scanner, CSV) e documentos em formato PDF.
 - **Volume:** Definição de um lote inicial de, no mínimo, 200 a 300 documentos variados para garantir uma amostragem representativa.
 - **Entregáveis:** Documentos explicando os formatos, respectivas tecnologias de coleta.

Fase 2: Arquitetura e Prova de Conceito (PoC) do Pipeline de Extração

- **Atividades:**
 - Desenhar a arquitetura geral da solução completa.
 - Desenvolver uma Prova de Conceito focada em validar o fluxo de extração de ponta a ponta em um subconjunto dos documentos coletados. Este PoC incluirá:
 - **Pré-processamento da Imagem:** Implementação de rotinas iniciais para ajustes automáticos, como correção de inclinação (deskew), remoção de ruídos e melhoria de contraste. **Ferramentas:** Bibliotecas Python como **OpenCV** e **Pillow (PIL)**.
 - **OCR (Reconhecimento Óptico de Caracteres):** Teste comparativo entre um modelo open-source (**Tesseract**) e uma solução comercial (**AWS Textract** ou **Google Vision**) para avaliar a acurácia, custo e velocidade.
 - **Extração Estruturada dos Dados:** Desenvolvimento de scripts preliminares para extrair os campos-chave solicitados (CNAE, Tipo de Produto, CEP) a partir do texto bruto gerado pelo OCR.
 - **Entregáveis:** Diagrama de Arquitetura, Relatório da PoC com análise de ferramentas juntamente com um MVP do módulo de extração e processamento dos documentos em base de dados.

Fase 3: Desenvolvimento e Customização do Pipeline

- **Atividades:** Com base nos aprendizados da PoC, o desenvolvimento da plataforma completa será realizado em sprints, focando em construir, testar e refinar cada módulo do pipeline:
 - **Módulo de Pré-processamento:** Construção de um serviço robusto e configurável para o tratamento das imagens, aplicando as técnicas validadas na PoC.
 - **Módulo de OCR e Customização:** Implementação da solução de OCR escolhida. Se necessário, será iniciado o processo de **customização (fine-tuning)** do modelo, treinando-o com documentos específicos para melhorar o reconhecimento de fontes, caracteres e layouts fiscais.
 - **Módulo de Extração Estruturada:** Desenvolvimento da lógica avançada de extração utilizando uma combinação de Processamento de Linguagem Natural (NLP), expressões regulares e regras de negócio para identificar e extrair com alta precisão:
 - **CNAE:** Localização do código numérico e da descrição da atividade.
 - **Tipo de Produto:** Classificação dos itens com base em palavras-chave, contexto e, se necessário, um dicionário de categorias pré-definido.
 - **CEP:** Extração e validação do formato do CEP do estabelecimento.
 - **Módulo de Classificação e Validação dos Dados:** Implementação de um motor de **regras de negócio para verificar a consistência** dos dados extraídos (ex.: validação da estrutura do CNAE, verificação do formato do CEP em base de dados externa).
 - Criação de lógicas de **correção automática** para erros comuns e um sistema de **sinlização (flagging) para revisão humana** quando a confiança do sistema for baixa.
 - **Módulo de Armazenamento e Integração:**
 - Configuração do **banco de dados** (ex.: PostgreSQL para metadados e **NoSQL** com MongoDB ou DynamoDB para os dados extraídos em JSON). Alternativamente, geração de saídas em arquivos **CSV**.
 - **Módulo de Inteligência e Pesquisa de Informações:**
 - Interface que permita ao usuário realizar pesquisas sobre produtos e serviços vendidos por região, seus preços para órgãos públicos
 - **Entregáveis:** Módulos de software funcionais, API documentada, Banco de Dados modelado, Dashboard de gerenciamento (desenvolvido em paralelo).

Fase 4: Testes de ponta-a-ponta e Validação

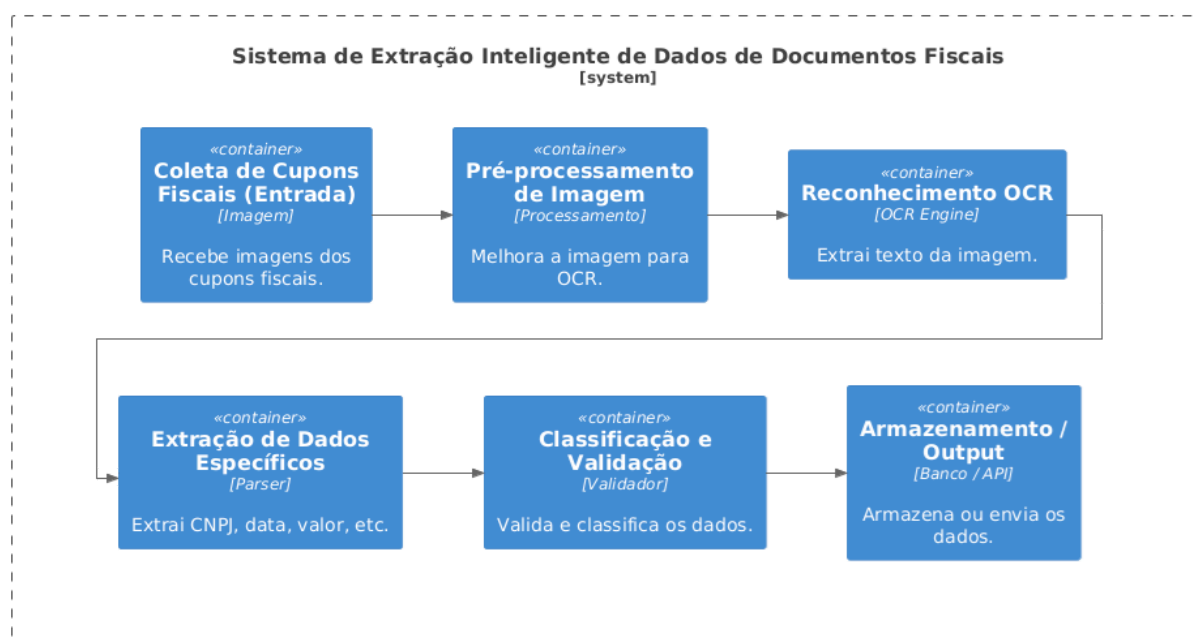
- **Atividades:** Executar testes integrados em todo o pipeline, desde o upload da imagem até a consulta do dado estruturado.
- **Entregáveis:** Relatórios de Qualidade e Acurácia.

Ambiente de Execução com Tecnologias Previstas

Arquitetura da Solução

- **Arquitetura de Microsserviços Orientada a Eventos:** Esta escolha é fundamental para atacar os desafios de adaptabilidade e escalabilidade. Cada etapa do processo (ingestão, pré-processamento, OCR, NLP, validação) será um serviço independente. Isso permite:
 - a. **Flexibilidade:** Para suportar um novo layout de documento, apenas o micro serviço de extração precisa ser atualizado ou um novo ser criado, sem impactar o resto do sistema.
 - b. **Escalabilidade:** Se houver um pico no recebimento de documentos, podemos escalar horizontalmente apenas os serviços de processamento, otimizando custos.
 - c. **Resiliência:** A comunicação via fila de mensagens garante que, se um serviço falhar, a requisição não será perdida e poderá ser reprocessada.

Visão Geral da Arquitetura:



Tecnologias Previstas/Sugeridas

Pré-processamento da Imagem

- **Tecnologias:** **OpenCV** e **Pillow (PIL)**.
- **Justificativa:** Essas bibliotecas são o padrão da indústria em Python para manipulação de imagens. Serão utilizadas para criar rotinas automatizadas que corrigem a inclinação dos documentos, melhoram o contraste, remover ruídos e binarizar a imagem, preparando-a para que o processo de OCR tenha o melhor desempenho possível.

OCR (Reconhecimento Óptico de Caracteres)

- **Tecnologias:** **Tesseract** (open-source) e **Google Vision API** (comercial).
- **Justificativa:** A abordagem dual nos permite flexibilidade. O **Tesseract** pode ser usado para cenários de menor complexidade ou como uma opção de baixo custo, com a possibilidade de ser treinado (fine-tuning) para os layouts específicos dos cupons. A **Google Vision API** será usada para garantir a mais alta precisão em documentos complexos ou de baixa qualidade, aproveitando os modelos de IA do Google.

NLP para Extração Estruturada

- **Tecnologias:** **SpaCy** e **Hugging Face Transformers**.
- **Justificativa:** Após o OCR converter a imagem em texto, utilizaremos modelos de Processamento de Linguagem Natural (NLP) para "entender" o conteúdo. O **SpaCy** é ideal para o Reconhecimento de Entidades Nomeadas (NER) de forma rápida, como localizar CEPs, valores e códigos. Para tarefas que exigem maior entendimento de contexto, como classificar o tipo de produto, os modelos da biblioteca **Hugging Face Transformers** oferecem o estado da arte.

Validação e Regras de Negócio

- **Tecnologia:** Scripts customizados em **Python**.
- **Justificativa:** A validação dos dados extraídos (ex.: verificar se um CNAE é válido ou se um CEP existe) será implementada em Python. Essa abordagem nos dá total controle para criar regras de negócio complexas e específicas para o sistema, além de permitir a fácil integração com fontes de dados externas para validação.

Armazenamento de Dados

- **Tecnologias:** **PostgreSQL** (SQL) e **MongoDB** (NoSQL).
- **Justificativa:** Adotaremos uma estratégia de persistência híbrida. O **PostgreSQL** será usado para armazenar dados estruturados e relacionais, como informações de usuários, metadados do processamento e logs. O **MongoDB** é a escolha ideal para armazenar os dados extraídos dos cupons, pois sua natureza baseada em documentos (JSON) acomoda perfeitamente a variedade e a flexibilidade dos campos encontrados em diferentes tipos de notas.

Planejamento Estratégico e Viabilidade de Negócio

Análise de Mercado

Esta análise visa fornecer uma estimativa do Tamanho Total de Mercado (TAM), Mercado Endereçável Disponível (SAM) e Mercado Obtível (SOM) para o Ai-Agents, o Agente Inteligente de Análise Fiscal, com base em dados do mercado brasileiro de automação e tecnologia.

1. Premissas de Mercado

A análise se baseia nas seguintes premissas extraídas da pesquisa:

- O Brasil é um dos países que mais gasta tempo com impostos no mundo (1.501 horas/ano), o que torna a automação fiscal uma necessidade crítica.
- Apesar da necessidade, 98% das grandes e médias empresas brasileiras ainda não utilizam automação completa na área financeira.
- O mercado de TI no Brasil é robusto, com investimentos que ultrapassaram R\$ 280 bilhões em 2024 (ABES).
- O foco do Ai-Agents é a automação cognitiva da entrada de dados de NFS-e (Notas Fiscais de Serviço Eletrônicas) para PMEs e Grandes Empresas.

2. Estimativa de Mercado (TAM/SAM/SOM)

2.1. TAM (Total Addressable Market)

O TAM representa o valor total de mercado se todos os potenciais clientes comprassem o produto.

- Definição: Mercado global de software e serviços de Automação de Processos Contábeis e Fiscais.
- Estimativa de Base: O mercado global de automação de processos (RPA, AI, etc.) é projetado para atingir centenas de bilhões de dólares. No entanto, focaremos no mercado de TI brasileiro como base para uma estimativa mais conservadora e relevante.
 - Mercado de TI no Brasil (2024): R\$ 280 Bilhões (ABES).
 - Fatia de Mercado (TAM): Considerando que 10% do mercado de TI está relacionado a software de gestão empresarial (ERP, Contábil, Fiscal), e que a automação fiscal é um componente essencial, podemos estimar o TAM no Brasil.
- TAM Estimado: R\$ 28 Bilhões (10% de R\$ 280 Bi) - Representa o valor total de mercado de software de gestão e automação fiscal no Brasil.

2.2. SAM (Serviceable Available Market)

O SAM representa a parcela do TAM que pode ser alcançada com o modelo de negócio atual (SaaS para extração de NFS-e).

- Definição: Empresas no Brasil (PMEs e Grandes) que emitem ou recebem alto volume de NFS-e e que buscam soluções de automação para entrada de dados.
- Base de Clientes: O Brasil possui cerca de 1,5 milhão de PMEs e 30 mil Grandes Empresas.
- Foco do Produto: O Ai-Agents é ideal para empresas que processam mais de 500 NFS-e por mês (PMEs de médio porte e Grandes Empresas). Estimamos que 100.000 empresas se encaixem neste perfil.
- Modelo de Receita (Estimativa de ARPU - Average Revenue Per User):

- Custo médio por nota processada: R\$ 0,50 (Plano Starter) a R\$ 1,50 (Plano Enterprise).
- Volume médio de notas/mês/cliente: 3.000.
- ARPU Anual Estimado: R\$ 1,00 (custo médio por nota) * 3.000 notas/mês * 12 meses = R\$ 36.000/ano.
- SAM Estimado: 100.000 empresas * R\$ 36.000/ano = R\$ 3,6 Bilhões.

2.3. SOM (Serviceable Obtainable Market)

O SOM representa a parcela do SAM que o Ai-Agents realisticamente pode capturar nos primeiros 3 anos.

- Definição: Nossa meta de clientes e receita nos primeiros 36 meses.
- Meta de Clientes (3 anos): Capturar 5% do mercado-alvo (SAM).
 - 5% de 100.000 empresas = 5.000 clientes.
- SOM Estimado (Receita Recorrente Anual - ARR):
 - 5.000 clientes * R\$ 36.000 (ARPU) = R\$ 180 Milhões em ARR.

Análise de Viabilidade e Riscos

Esta seção detalhará os aspectos de viabilidade técnica, econômica e operacional do projeto, bem como os riscos associados e as estratégias de mitigação. Abrange:

- Viabilidade Técnica: Avaliação da maturidade das tecnologias propostas (OCR, NLP, microsserviços), disponibilidade de expertise na equipe e desafios de integração.
- Viabilidade Econômica: Análise de custo-benefício, incluindo custos desenvolvimento, infraestrutura (nuvem, licenças de DaaS/APIs comerciais), manutenção e o retorno sobre o investimento (ROI) esperado para as empresas usuárias (ex.: economia de tempo, redução de erros, aumento da qualidade das informações fiscais).
- Viabilidade Operacional: Capacidade da organização de operar e manter a solução, necessidade de treinamento, e impacto nos processos de negócio existentes.
- Análise de Riscos: Identificação de riscos potenciais (ex.: baixa acurácia do OCR/NLP, mudanças regulatórias, dependência de provedores externos, segurança de dados, escalabilidade, aceitação do usuário) e proposição de planos de mitigação para cada um.

Estratégia de Monetização e Modelo de Negócio

Embora o documento foque na solução técnica, uma seção sobre a estratégia de monetização é crucial para um trabalho completo, especialmente se a intenção é transformar a solução em um produto ou serviço comercializável. Esta seção explora:

- Modelos de Assinatura: SaaS (Software as a Service) com diferentes níveis de serviço (ex.: por volume de documentos processados, por número de usuários, por funcionalidades).
- Precificação por Uso: Pagamento por transação ou por volume de dados extraídos/consultados.
- Serviços de Valor Agregado: Consultoria, customização, treinamento, suporte premium
- Parcerias Estratégicas: Com provedores de DaaS, ERPs, ou associações comerciais.
- Análise de Concorrentes: Como soluções similares (se existirem) são monetizadas e como a proposta se diferencia.

Aspectos Legais e de Compliance

Dado que a solução lida com dados fiscais e de compras públicas, os aspectos legais e de compliance são de suma importância. Esta seção aborda:

- Lei Geral de Proteção de Dados (LGPD): Como a solução garantirá a privacidade e a segurança dos dados pessoais e empresariais, especialmente no tratamento de informações de fornecedores e órgãos públicos.
- Legislação Fiscal: Conformidade com as normas fiscais brasileiras, como a Lei 14.133/2021 (Nova Lei de Licitações e Contratos), e outras regulamentações específicas de notas fiscais eletrônicas.
- Segurança da Informação: Medidas de segurança (criptografia, controle de acesso, auditoria) para proteger os dados em trânsito e em repouso.
- Auditoria e Rastreabilidade: Como a solução permitirá a auditoria dos processos de extração e o rastreamento da origem dos dados.

Plano de Marketing e Go-to-Market

Para que a solução alcance seu público-alvo e gere impacto, um plano de marketing e go-to-market é essencial. Esta seção, detalha:

- Segmentação de Mercado: Identificação de nichos específicos dentro do público-alvo (ex.: pequenas e médias empresas, grandes fornecedores, setores específicos). Canais de Aquisição de Clientes: Marketing digital (SEO, SEM, mídias sociais), eventos do setor, vendas diretas, parcerias.
- Proposta de Valor Única (UVP): Como a solução se diferencia da concorrência e quais benefícios exclusivos oferece.
- Estratégia de Lançamento: Fases de lançamento (beta, lançamento oficial), comunicação e engajamento inicial.
- Métricas de Sucesso de Marketing: KPIs para medir o desempenho das ações de marketing (ex.: número de leads, taxa de conversão, custo de aquisição de cliente).

Roadmap de Evolução e Funcionalidades Futuras

Esta seção apresenta uma visão de longo prazo para a solução, demonstrando seu potencial de crescimento e adaptação.

- Possibilidade de upload de diversas notas fiscais em lote que podem ser acessadas diretamente de um notebook ou de um servidor, ou ainda bancos de arquivos na nuvem.
- Integração com Outras Fontes de Dados: Além das notas fiscais, considerar a integração com outros documentos (ex.: contratos, ordens de compra, editais de licitação em formatos não estruturados) ou bases de dados públicas (ex.: dados do IBGE, dados econômicos).
- Funcionalidades Avançadas de IA: Implementação de modelos preditivos para antecipar tendências de compra, recomendação de produtos/serviços, análise de sentimentos em documentos textuais.
- Dashboards e Relatórios Customizáveis: Permitir que os usuários criem seus próprios painéis e relatórios com base nos dados extraídos.
- Alertas e Notificações Inteligentes: Notificações proativas sobre novas oportunidades, mudanças de preços, ou atividades de concorrentes.
- Expansão para Outros Mercados: Adaptação da solução para outros países ou setores que lidam com grandes volumes de documentos fiscais.

Cronograma

Fase	Duração Estimada	Objetivo
Fase 1: Análise e Coleta de Dados	2 Semanas	Analisar e Coletar os Dados Disponibilizados
Fase 2: Arquitetura e PoC	4 Semanas	Arquitetura Definida, PoC Validada com Sucesso
Fase 3: Desenvolvimento	8 Semanas	Pipeline de Desenvolvimento
	3 Semanas	Técnicas de OCR
	3 Semanas	Treinamento de Agente Inteligente
Fase 4: Testes e Validação	2 Semanas	Ciclo de Testes para validação com outros datasets.
DURAÇÃO TOTAL ESTIMADA	22 Semanas	

GitHub Repositório

<https://github.com/challenge-accepted-i2a2/Ai-agents>

Equipe de Trabalho

Carlos Antônio Campos Jorge
Claudio Fagundes Pereira
David de Freitas Neto
Filipe do Rego Barros Luz
Kattson Alves Bastos
Letícia Ferreira Murça Reis
Raphael Vinholes Bichiarov
Renato Azevedo Sant'Anna
Thayron Sabino Alves dos Santos