



**UNIVERSIDADE NOVE DE JULHO
DIRETORIA DOS CURSOS DE INFORMÁTICA**

Kelvin Felipe Dos Santos1 - RA 923200158

Extrator Inteligente de NFS-e com OCR e LLM

**SÃO PAULO
2025**

Kelvin Felipe Dos Santos - RA 923200158

Extrator Inteligente de NFS-e com OCR e LLM

Trabalho apresentado à Universidade Nove de Julho, UNINOVE, em cumprimento parcial às exigências da disciplina Projeto em Computação Aplicada Extrator-NFSe-Ollama: Automação Inteligente de Processamento de Notas Fiscais, sob orientação do Prof. Luis Carlos Dos Santos Junior.

SÃO PAULO 2025 SUMÁRIO

1. RESUMO	4
2. OBJETIVOS	4
3. DESCRIÇÃO DA EMPRESA	4
4. MISSÃO, VISÃO E VALORES DA EMPRESA	5
4.1. MISSÃO	5
4.2. VISÃO	5
4.3. VALORES	5
5. DEFINIÇÃO DA EQUIPE, DIVISÃO DE PAPÉIS E TAREFAS, CRONOGRAMA DE DESENVOLVIMENTO	5

6.	DESENVOLVIMENTO (METODOLOGIA)	6
7.	CONCLUSÃO	6
8.	REFERÊNCIAS	7

1. RESUMO

Este projeto documenta o desenvolvimento do "Extrator Inteligente de NFS-e", uma solução de software destinada a automatizar o processamento de Notas Fiscais de Serviço Eletrônicas para a empresa fictícia Soluções IA Ltda. O contexto do problema reside na ineficiência e no alto custo de erro do processamento manual de documentos fiscais. O objetivo geral do trabalho foi criar um sistema que utiliza Reconhecimento Óptico de Caracteres (OCR) e um Modelo de Linguagem de Larga Escala (LLM) local para extrair e estruturar dados de PDFs e imagens de notas fiscais. A metodologia empregada foi a prototipagem evolutiva, utilizando Python, Streamlit para a interface, Azure OCR para leitura, e Ollama (com phi3:mini) para a extração de dados, com um banco de dados MySQL para armazenamento. Como resultado, o projeto entregou um protótipo funcional capaz de processar lotes de notas, validando a eficácia dos LLMs locais para tarefas de automação financeira. Conclui-se que a abordagem é viável e oferece um potencial significativo de redução de custos e erros para PMEs.

Palavras-chave: Processamento de Notas Fiscais, Inteligência Artificial, OCR, LLM, Automação Financeira.

2. OBJETIVOS

Desenvolver uma solução de software robusta, denominada "Extrator Inteligente de NFS-e", capaz de automatizar o processo de leitura, extração e armazenamento de dados de Notas Fiscais de Serviço Eletrônicas (NFS-e) utilizando tecnologias de Reconhecimento Óptico de Caracteres (OCR) e Modelos de Linguagem de Larga Escala (LLM)

3. DESCRIÇÃO DA EMPRESA

A Soluções IA Ltda. é uma startup de tecnologia (fintech) focada em automação inteligente de processos (RPA) para o setor financeiro. Criada em 2025, a empresa nasceu para satisfazer a demanda crescente de pequenas e médias empresas (PMEs) por eficiência no processamento de documentos fiscais, uma área tradicionalmente manual e suscetível a erros.

De acordo com Dornelas (2015), a descrição da empresa deve conter seu status atual e projeções futuras. As projeções para o mercado de automação fiscal indicam uma demanda

crescente por soluções que integrem IA para lidar com a variedade de layouts de NFS-e no Brasil.

A Soluções IA Ltda. busca manter uma margem competitiva através da entrega de um software (SaaS) de alta precisão, com excelente suporte ao cliente e capacidade de se adequar rapidamente a novos layouts de notas fiscais. O produto principal, o "Extrator Inteligente de NFS-e", beneficia o cliente ao reduzir drasticamente o tempo de digitação, mitigar erros humanos e fornecer dados estruturados para análise financeira imediata, ressaltando as características únicas do serviço

4. MISSÃO, VISÃO E VALORES DA EMPRESA

4.1. MISSÃO

Fornecer soluções de automação inteligente que libertem as equipes financeiras de tarefas manuais repetitivas, permitindo que foquem em análises estratégicas e no crescimento do negócio

4.2. VISÃO

Ser a plataforma líder no Brasil para extração e gerenciamento de dados fiscais para PMEs, reconhecida pela precisão da nossa IA e pela simplicidade da nossa interface.

4.3. VALORES

- **Inovação:** Buscar continuamente as melhores tecnologias (OCR, LLM) para resolver problemas reais.
- **Precisão:** Comprometimento com a acurácia dos dados extraídos.
- **Eficiência:** Criar soluções que economizem tempo e recursos para os nossos clientes.
- **Segurança:** Garantir a confidencialidade e integridade dos dados fiscais.

5. DEFINIÇÃO DA EQUIPE, DIVISÃO DE PAPÉIS E TAREFAS

A equipe de desenvolvimento do projeto é composta pelos seguintes membros, com seus respectivos papéis e tarefas:

Kelvin Felipe Dos Santos: Gerente de Projeto e Desenvolvedor Full-Stack (Backend, Frontend e Banco de Dados).

6. DESENVOLVIMENTO (METODOLOGIA)

Este projeto utiliza uma abordagem de prototipagem evolutiva, combinando elementos de metodologias ágeis. A natureza do projeto, que depende da precisão de modelos de IA (OCR e LLM), exige ciclos rápidos de teste e refinamento. O estudo é de natureza experimental e descritiva, pois aplica tecnologias existentes (LLMs) a um novo problema (layouts de NFS-e específicos) e descreve os resultados de precisão obtidos."

Ferramentas e Tecnologias: As ferramentas e tecnologias utilizadas neste projeto incluem :

- Linguagem de Programação: Python 3.11+
- Framework de Interface (Frontend): Streamlit
- Inteligência Artificial (LLM): Ollama (servidor local) com o modelo phi3:mini.
- Reconhecimento Óptico (OCR): Azure Cognitive Services (Microsoft).
- Banco de Dados (Backend): MySQL.
- Bibliotecas Python: pandas, streamlit-authenticator, mysql-connector-python, PyMuPDF.
- Ambiente de Desenvolvimento: Visual Studio Code.

7. CONCLUSÃO

A conclusão é a síntese para a qual caminha o trabalho. O desenvolvimento do "Extrator Inteligente de NFS-e" atingiu com sucesso os objetivos propostos. Foi implementado um sistema funcional que integra Azure OCR e o LLM local phi3:mini, capaz de processar e extrair dados de notas fiscais com alta precisão. A interface em Streamlit provou ser uma ferramenta eficaz para a validação humana dos dados, e o banco de dados MySQL estruturou as informações para futuras análises. Os resultados obtidos [aqui você manifesta seu ponto de vista, ex: "demonstraram que, mesmo com modelos de IA locais e mais leves, é possível alcançar uma automação robusta..."]. Conclui-se que a solução é viável e apresenta um potencial significativo para a otimização de processos financeiros na empresa Soluções IA Ltda.

8. REFERÊNCIAS

<http://docs.uninove.br/arte/pdfs/Manual-Elaboracao-de-Trabalhos-ABNT.pdf>

DORNELAS, José. Empreendedorismo: transformando ideias em negócios. 5. ed. Rio de Janeiro: LTC, 2015.

SEVERINO, Antonio Joaquim. Metodologia do trabalho científico. 22. ed. rev. ampl. São Paulo: Cortez, 2002.

MICROSOFT. Documentação Oficial do Azure Cognitive Services for Vision. Disponível em: <https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/overview-ocr>. Acesso em: 10 out. 2025.

OLLAMA. Ollama: Run large language models locally. Disponível em: <https://ollama.com>. Acesso em: 10 out. 2025.

PYTHON SOFTWARE FOUNDATION. Python Language Reference. Disponível em: <https://www.python.org>. Acesso em: 05 out. 2025.

STREAMLIT. Streamlit Documentation. Disponível em: <https://docs.streamlit.io>. Acesso em: 15 out. 2025.