



INSTITUTO  
UNIVERSITÁRIO  
DE LISBOA

---

## **Processamento automático de documentos: aplicação à faturação de fornecedores**

João Carlos Almeida Duarte Balaia

Mestrado em Business Analytics

Orientadores:

Professor Doutor Raul Manuel da Silva Laureano, Professor Associado, ISCTE Business School, Departamento de Métodos Quantitativos de Gestão e Economia

Professor Paulo Jorge de Almeida Pereira, Professor Auxiliar convidado, ISCTE Business School, Departamento de Métodos Quantitativos de Gestão e Economia

Setembro, 2023



BUSINESS  
SCHOOL

---

Departamento de Métodos Quantitativos para Gestão e Economia

**Processamento automático de documentos: aplicação à faturação de fornecedores**

João Carlos Almeida Duarte Balaia

Mestrado em Business Analytics

Orientadores:

Professor Doutor Raul Manuel da Silva Laureano, Professor Associado, ISCTE Business School, Departamento de Métodos Quantitativos de Gestão e Economia

Professor Paulo Jorge de Almeida Pereira, Professor Auxiliar convidado, ISCTE Business School, Departamento de Métodos Quantitativos de Gestão e Economia

Setembro, 2023

# Agradecimentos

A conclusão da dissertação de mestrado acrescenta mais um momento memorável à minha vida, pois marca o final do meu ciclo académico.

Ao longo desta jornada, que começou em 2018 no ISCAC e termina agora no ISCTE, conheci inúmeras pessoas, algumas que levo comigo para o resto da vida. Tive a oportunidade de fazer amigos, conhecer lugares incríveis, aprender com profissionais excelentes e crescer tanto a nível profissional como a nível pessoal.

Quero assim agradecer a todos os meus familiares, que me apoiaram sempre ao longo do meu percurso académico, sem o esforço deles tudo isto seria impossível. Agradeço também todo o seu apoio, paciência, motivação e compreensão nos momentos mais difíceis desta etapa.

Quero agradecer aos meus amigos, aos que fiz ao longo do meu percurso e aos que fazem parte da minha vida, por todos os momentos que foram vividos e pelos que ainda estão para vir.

Quero agradecer a todos os professores com quem me cruzei ao longo destes cinco anos, agradecer por todo o conhecimento que de forma direta ou indireta me foi transmitido por eles.

Por fim quero agradecer especialmente aos meus orientadores, o professor Raul Laureano e o professor Paulo Pereira, agradecer por todo o apoio, disponibilidade e conhecimentos transmitidos ao longo deste projeto.



# Resumo

Este projeto procurou entender a importância que o processamento automático de documentos tem para as empresas e entender de que forma estas tecnologias podem ajudar na transformação digital das empresas. Este estudo foi desenvolvido para ajudar a empresa Adriano Luz Duarte Balaia, na forma como a mesma trata a faturação dos fornecedores.

A revisão da literatura teve por objetivo conhecer melhor a tecnologia, perceber as suas vantagens, desvantagens, perceber o que é o processamento automático de documentos, perceber as melhores técnicas e ferramentas disponíveis atualmente. Contou com a leitura e análise de 27 documentos sobre o tema de forma a aprofundar melhor o assunto.

Este projeto seguiu uma adaptação do modelo CRISP-DM, foi realizada a compreensão do negócio e dos dados, de forma a perceber qual o problema a resolver. Contou ainda com duas fases de modelação, a primeira é referente ao fluxo para ler, analisar e extrair informações chave das faturas dos fornecedores, a segunda fase é a modelação de um *dashboard*.

Os resultados obtidos são inseridos num *dashboard*, que permite ao gerente da empresa analisar melhor e de forma mais clara os gastos da empresa.

Palavras-chaves: Processamento automático de documentos, Processamento inteligente de documentos, Extração de informação, análise de dados, *Dashboard*, Processamento automático.

Código JEL: M41, A10



# Abstract

This project aimed to understand the importance of automatic documents processing for companies and how these technologies can contribute to the digital transformation of businesses. The project was developed to assist the company Adriano Luz Duarte Balaia in how it handles supplier invoicing.

The literature review aimed to gain a better understanding of the technology, its advantages, disadvantages, what automatic document processing entails, and to identify the best techniques and tools currently available. It involved reading and analyzing 27 documents on the subject to delve deeper into the topic.

This project followed an adaptation of the CRISP-DM model. It began with understanding the business and data to identify the problem to be addressed. It also included two modeling phases: the first phase focused on the workflow for reading, analyzing, and extracting key information from supplier invoices, and the second phase involved creating a dashboard.

The results obtained in this project are integrated into a dashboard that allows the company's manager to better and more clearly analyze the company's expenses.

Keywords: Automatic document processing, Intelligent document processing, Information extraction, Data analysis, Dashboard, Automatic processing.

JEL code: M41, A10





# **Acrónimos, siglas e abreviaturas**

ANN - Rede Neural Artificial

DL - Deep Learning

DPA - Processamento Automático de Documentos

DT - Árvores de Decisão

et al - e outros

e.g. - em geral (por exemplo)

IA - Inteligência Artificial

IDP - Processamento Inteligente de Documentos

IEEE - Institute of electrical and electronics engineers

K-NN - K-éssimo vizinho mais próximo

ML - Machine learning

NB - Naive Bayes

NPL - Natural Language Processing

OCR - Reconhecimento Ótico de Caracteres

RPA - Automação de Processos Robóticos

RSL - Revisão Sistemática da Literatura

SVM - Máquina de Vetores de Suporte



# Índice

Agradecimentos.....	i
Resumo .....	iii
Abstract .....	v
Acrónimos, siglas e abreviaturas .....	vii
1. Introdução.....	1
1.1. Contexto e problema .....	1
1.2. Questão de investigação e objetivos .....	1
1.3. Abordagem metodológica .....	2
1.4. Estrutura.....	3
2. Revisão Sistemática da Literatura.....	5
2.1. Protocolo da RSL .....	5
2.1.1. Objetivos e questões de pesquisa .....	5
2.1.2. Processo de seleção dos documentos a incluir na RSL.....	5
2.1.3. Análise e avaliação dos estudos da RSL.....	7
2.2. Caracterização dos estudos selecionados.....	8
2.3. Síntese dos artigos .....	12
2.3.1. Conceito e objetivo de DPA e IDP .....	13
2.3.2. Reconhecimento Ótico de Carateres.....	15
2.3.3. Desafios e benefícios da implementação do IDP .....	16
2.3.4. Técnicas de extração de informação .....	17
2.3.5. Ferramentas de automação de processos e de extração de informação.....	19
2.3.6. Contributos, limitações e pistas futuras.....	20
2.4. Avaliação dos estudos da RSL .....	22
3. Metodologia .....	25
3.1. Compreensão do negócio.....	25
3.2. Compreensão e preparação dos dados .....	28

3.2.1. Fluxo.....	28
3.2.2. <i>Dashboard</i> .....	29
3.3. Modelação.....	31
3.3.1. Modelação do fluxo.....	31
3.3.2. Modelação do <i>dashboard</i> .....	34
3.4. Avaliação.....	35
3.4.1. Avaliação do fluxo.....	35
3.4.2. Avaliação do <i>dashboard</i> .....	35
3.5. Implementação.....	36
4. Resultados e Discussão.....	37
4.1. Resultado do fluxo.....	37
4.2. Resultado do <i>dashboard</i> .....	39
4.2.1. Resultado e explicação da página dos fornecedores.....	39
4.2.2. Resultado e explicação da página dos produtos.....	42
4.2.3. Resultado e explicação da página de controlo.....	45
5. Conclusão, limitações e pistas futuras.....	49
5.1. Sumário.....	49
5.2. Contributos.....	49
5.3. Limitações.....	50
5.4. Trabalho futuro.....	51
Referencias.....	53
Anexo.....	57

# Índice de Figuras

Figura 1.1 Metodologia CRISP-DM adotada .....	3
Figura 2.1 Processo de seleção dos estudos a incluir na RSL.....	7
Figura 2.2 Nuvem de palavras dos resumos dos estudos incluídos na RSL .....	11
Figura 2.3 Número de estudos por ano.....	11
Figura 2.4 Número de estudos por autor.....	12
Figura 2.5 Contribuição percentual de cada tipo de estudos.....	12
Figura 2.6 Funcionamento de um IDP .....	15
Figura 2.7 Evolução das técnicas para extração de informação de documentos .....	19
Figura 3.1 Arquitetura do sistema .....	26
Figura 3.2 Modelo de dados .....	29
Figura 3.3 Quando chega um novo email .....	31
Figura 3.4 Variável para definir fornecedor .....	31
Figura 3.5 Modelo customizado para extrair cabeçalhos das faturas .....	32
Figura 3.6 Modelo automático para extrair linhas das faturas .....	32
Figura 3.7 Adicionar dados dos cabeçalhos ao ficheiro Excel.....	33
Figura 3.8 Condição para validar a taxa de precisão .....	33
Figura 3.9 Condição para validar data de pagamento.....	33
Figura 3.10 Fluxo final .....	34
Figura 4.1 Pontuação geral do modelo customizado .....	37
Figura 4.2 Pontuação do modelo automático nos cabeçalhos das faturas.....	37
Figura 4.3 Pontuação do modelo automático nas linhas das faturas.....	38
Figura 4.4 <i>Dashboard</i> página dos fornecedores .....	40
Figura 4.5 Gasto total, total de descontos e nº de fornecedores .....	40
Figura 4.6 Gráficos referentes aos principais fornecedores e principais produtos.....	40
Figura 4.7 Botões para alterar forma do gráfico .....	41
Figura 4.8 Gastos e quantidades por fornecedores e gastos por produto .....	41
Figura 4.9 Total de descontos por ano e por mês .....	41
Figura 4.10 Gasto total por ano e por mês.....	42
Figura 4.11 <i>Dashboard</i> página dos produtos consumidos .....	43
Figura 4.12 Preço, quantidade e contributo percentual no gasto total.....	43
Figura 4.13 Preço unitário médio .....	44
Figura 4.14 Quantidade média.....	44
Figura 4.15 Valor total e quantidade total por mês.....	44
Figura 4.16 Gráfico cascata montante gasto por mês .....	45
Figura 4.17 <i>Dashboard</i> de controlo de fluxo .....	45

Figura 4.18 Média de precisão na extração dos cabeçalhos e das linhas .....	46
Figura 4.19 Número total de extrações por categoria.....	46
Figura 4.20 Tabelas de informações sobre as faturas.....	46

## Índice de Tabelas

Tabela 2.1 Critérios de inclusão e exclusão.....	6
Tabela 2.2 Critérios para avaliar os estudos incluídos na RSL .....	8
Tabela 2.3 Artigos incluídos na RSL.....	8
Tabela 2.4 Outros estudos incluídos na RSL .....	9
Tabela 2.5 Documentos não científicos incluídos na RSL.....	10
Tabela 2.6 Objetivo dos artigos e dos estudos incluídos na RSL.....	13
Tabela 2.7 Conceito e objetivo de DPA e IDP .....	14
Tabela 2.8 Benefícios do IDP .....	17
Tabela 2.9 Desafios do IDP .....	17
Tabela 2.10 Técnicas de extração de informação.....	17
Tabela 2.11 Contributos, limitações e pistas futuras .....	21
Tabela 2.12 Avaliação dos artigos e dos estudos .....	22
Tabela 3.1 Comparação entre as ferramentas existentes .....	27
Tabela 3.2 Comparação entre as ferramentas de <i>dashboard</i> .....	27
Tabela 3.3 Categorias de extração .....	29
Tabela 4.1 Avaliação sobre a satisfação dos utilizadores .....	38
Tabela 4.2 Avaliação da satisfação global dos utilizadores do <i>dashboard</i> .....	39





# 1. Introdução

## 1.1. Contexto e problema

As empresas procuram cada vez mais a transformação digital, na qual os sistemas possam tornar-se inteligentes para se adaptarem à dinâmica dos negócios e às preferências dos compradores (Lievano-Martínez *et al.*, 2022).

As soluções de processamento automático de documentos podem, por exemplo, transformar informações não estruturadas em dados que podem ser utilizados para uma análise futura. Os dados empresariais são de grande importância na transformação digital das mesmas, mas a maioria dos dados são dados não estruturados, em formatos PDF e *emails*, entre outros formatos de documentos. O processamento de documentos vai ser, assim, capaz de extrair e processar esses dados, por meio do uso de tecnologias de Inteligência Artificial (IA) (Automation Anywhere, nd).

Este projeto é desenvolvido numa pequena empresa de construção civil, que pretende proporcionar um crescimento tanto a nível empresarial, como a um nível competitivo na empresa. Como a mesma recebe um grande número de faturas mensalmente, processá-las manualmente pode aborrecer o colaborador e levar ao erro, algo que pode ser eliminado com a adoção de uma tecnologia como o processamento automático de documentos (DPA).

A principal motivação para a escolha deste tema é a importância que o projeto pode trazer para o autor, bem como o facto de unir o trabalho académico com o mundo profissional. Além disso, a empresa tem o desejo de perceber melhor de que forma o seu dinheiro é gasto, quer identificar os principais gastos e identificar os seus principais fornecedores.

## 1.2. Questão de investigação e objetivos

Dentro do problema que a empresa apresentou, a falta de capacidade de perceber onde são gastos os seus fundos, é formulada uma questão de investigação: Como automatizar o processo de registo de faturas e controlo de gestão?

De forma a responder à questão de investigação são definidos os seguintes objetivos:

1. Criar um fluxo para ler as faturas e guardar os dados necessários;
2. Criar um *dashboard* para controlo de faturação dos fornecedores;

Desta forma, este projeto tem dois objetivos principais, o primeiro objetivo é criar um fluxo para recolher, ler e extrair informações de forma automática das faturas dos fornecedores. O segundo objetivo visa a elaboração de um *dashboard*, de forma a auxiliar o gestor da empresa

a perceber melhor os gastos que a mesma tem. A empresa tem inúmeros gastos com fornecedores ao longo do mês, estes gastos podem dar um grande contributo para a empresa caso sejam bem analisados.

A resposta à questão de investigação traz contributos, quer para a empresa alvo do estudo, quer para o conhecimento científico. Para a empresa o principal contributo é a criação de uma ferramenta para automatizar o processo de faturação dos fornecedores e clarificar os gastos da mesma. Para o conhecimento científico, este projeto apresenta uma conjugação de várias ferramentas e incorpora uma abordagem de *low code* para a recolha e extração de dados vitais para as empresas.

### **1.3. Abordagem metodológica**

A metodologia escolhida para a realização deste projeto é baseada no *Cross Industry Standard Process for Data Mining* (CRISP DM). Esta é uma das diversas metodologias disponíveis para guiar projetos deste tipo. Esta metodologia é aplicada em projetos de mineração de dados, mas pode ser aplicada noutras áreas de estudo relacionadas com a análise de dados. A flexibilidade é um dos benefícios do CRISP-DM, pois permite a sua personalização de acordo com a necessidade de cada projeto (Chapman *et al.*, 1999).

Destaca-se que é um estudo de caso em que se automatiza os processos de faturação dos fornecedores da empresa Adriano Luz Duarte Balaia. Este estudo de caso examina 93 faturas de fornecedores de outubro de 2022 a junho de 2023. As ferramentas para automatizar o processo pertencem à Microsoft e, também, se desenvolveu um protótipo de *dashboard* para apresentar os resultados, com recurso à ferramenta Power BI.

Para a realização deste projeto, a metodologia é composta por seis fases, Figura 1.1.

1. Compreensão do negócio
2. Compreensão e preparação dos dados
3. Modelação do fluxo
4. Avaliação do fluxo
5. Modelação do *dashboard*
6. Implementação

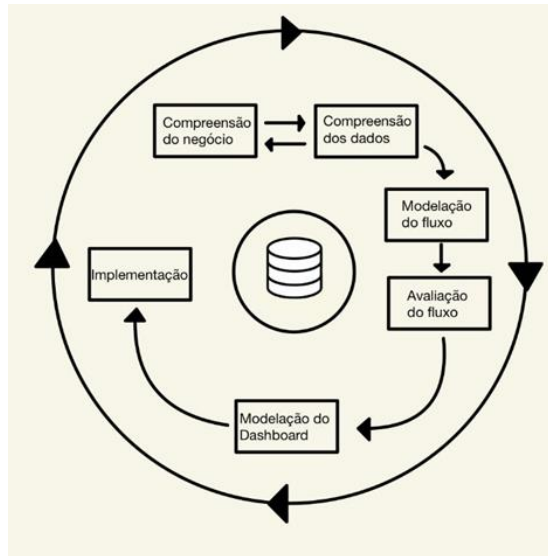


Figura 1.1 Metodologia CRISP-DM adotada

Adaptado de Chapman *et al.* (1999, pág. 13).

#### 1.4. Estrutura

Este estudo está dividido em cinco capítulos, incluindo esta introdução. No capítulo dois apresenta-se uma revisão sistemática da literatura (RSL) sobre as tecnologias DPA e IDP, a RSL apresenta os conceitos, as vantagens e desvantagens desta tecnologia. O capítulo dois serve também para perceber como estas tecnologias trabalham e de que forma podem ajudar as empresas.

O capítulo três é sobre a metodologia seguida para a criação deste projeto, neste caso é usado o CRISP-DM. Este capítulo descreve o processo de extração e preparação dos dados, apresenta também a criação e avaliação do fluxo para extrair e processar as informações dos documentos e apresenta também a criação do *dashboard* usado para visualizar os dados recolhidos.

No capítulo quatro são apresentadas os resultados e explicações do *dashboard* construído, de forma a clarificar o leitor sobre todas as páginas do *dashboard* e sobre todos os elementos visuais do mesmo.

Por último o capítulo cinco apresenta as conclusões do projeto, bem como as limitações que existiram na realização do projeto e apresenta alguns tópicos para guiar um futuro trabalho.



## 2. Revisão Sistemática da Literatura

Este capítulo aborda quais as características que definem um DPA e IDP, de que forma a tecnologia de extração de informação tem evoluído ao longo do tempo e quais as vantagens e desafios da mesma. Neste capítulo também é sintetizado o conhecimento científico sobre o tema.

O protocolo da RSL é um processo metodológico para identificar, selecionar e avaliar de forma criteriosa todos os estudos relevantes sobre um tópico específico (Kitchenham, 2004). Algumas das razões pela qual se deve realizar uma revisão da literatura são fornecer uma visão geral da literatura existente. A estratégia para esta RSL está assente no protocolo proposto por Silva *et al.* (2021).

Assim, inicialmente, são definidos os objetivos e as questões de pesquisa, em seguida é construída uma *query* para efetuar a pesquisa em bases de dados e, finalmente, são definidos critérios de inclusão e de exclusão para filtrar os estudos.

Selecionados e caracterizados os estudos é realizada a sua sistematização e análise crítica de forma a responder às questões de pesquisa formuladas e, por fim, os estudos são avaliados quanto à sua qualidade.

### 2.1. Protocolo da RSL

#### 2.1.1. Objetivos e questões de pesquisa

O principal objetivo desta RSL passa por sintetizar o que já se conhece sobre esta tecnologia, gerando conhecimento que permite perceber de que forma esta tecnologia tem ajudado as empresas no seu dia-a-dia e perceber o seu funcionamento. Para concretizar o objetivo, são definidas as seguintes questões de investigação:

- i. O que é um DPA ou um IDP e seus objetivos?
- ii. Quais são os desafios/benefícios da implementação do IDP?
- iii. Quais e como as técnicas de DPA podem ser usadas para extrair informações?
- iv. Quais as ferramentas de DPA existentes e quais as suas vantagens e desvantagens?
- v. Qual a melhor metodologia para a construção de um projeto de DPA?
- vi. Quais os impactos, contributos, limitações e pistas futuras dos estudos?

#### 2.1.2. Processo de seleção dos documentos a incluir na RSL

Os estudos para este trabalho são identificados em base de dados de publicações científicas, neste estudo a base de dados utilizada foi a Scopus. Os estudos são identificados com base em palavras-chaves, idioma dos documentos e critérios de inclusão e de exclusão. Outros documentos são obtidos em páginas específicas deste tipo de projetos de automação de

documentos, como, por exemplo, a página da Microsoft Power Automate (powerautomate.microsoft.com), IBM (www.ibm.com) e a Automation Anywhere (www.automationanywhere.com). É necessário recorrer a estas páginas para entender melhor a diferença entre o DPA e a automação de processos robóticos (RPA), uma vez que os estudos recolhidos não fazem uma diferenciação entre estas duas tecnologias.

Para os estudos científicos o método de pesquisa tem em atenção os seguintes pontos:

- Idioma, são selecionados os estudos científicos de idioma inglês.
- Livre acesso, todos os documentos recolhidos para a realização deste trabalho são de livre acesso ou permitem efetuar o seu *download* através da assinatura da instituição (ISCTE)
- Palavras-chaves, a seleção dos estudos tem por base palavras-chaves, relacionadas com DPA ou IDP e faturas e organizadas na forma de *query*.

A *query* para encontrar os estudos científicos é ("OCR" OR "optical character recognition") OR ("DPA" OR "process\* automation") OR ("rpa" OR robotic process\* automat\*) OR ("IDP" OR "intelligent automation" OR "intelligent document processing " OR "Intelligent process automation") OR ("low code") OR ("artificial intelligence" or "AI") AND ("billing" OR "invoice").

A *query* é aplicada nos títulos, nos resumos e nas palavras-chaves dos documentos, na base de dados Scopus, sendo a mesma construída em conjunto com um dos orientadores.

Identificou-se 388 documentos com a *query*, que podem ser incluídos ou excluídos de acordo com os critérios da Tabela 2.1. A aplicação destes critérios é feita após uma leitura parcial dos documentos, neste caso o título do documento e o resumo. Quando o resumo e o título não são conclusivos é feita uma leitura mais profunda do documento.

**Tabela 2.1 Critérios de inclusão e exclusão**

<b>Critérios de inclusão:</b>
Artigos que abordam técnicas de processamento automático de documentos como: <ul style="list-style-type: none"> <li>• Reconhecimento ótico de caracteres (OCR)</li> <li>• Análise de imagens</li> <li>• Reconhecimento de caracteres</li> <li>• Extração de informação</li> </ul>
<b>Critérios de exclusão:</b>
<ul style="list-style-type: none"> <li>• Linguagem do artigo não ser inglês</li> <li>• Não ser um artigo científico ou uma revisão ou um artigo de conferência</li> <li>• Não ser de acesso livre (ou não ser possível fazer o download)</li> <li>• Artigos duplicados</li> </ul>

- Artigos sem contributo ou sem aplicação numa área de economia, finanças ou tecnologia.

Notas: OCR- Reconhecimento ótico de caracteres

Ao aplicar os critérios de exclusão, restam 275 estudos em análise. Em seguida aplica-se o critério das palavras-chaves (“*Information Processing*”, “*Image Processing*”, “*Text Processing*” e “*Information extraction*”), este critério foi aplicado após a *query* ter identificado os estudos, através dele foi possível identificar 100 estudos. Em seguida, com base na análise aos títulos e aos resumos, mantêm-se 68 estudos relevantes. Por fim, após leitura mais profunda dos estudos são considerados 27 estudos para entrarem na RSL, dos quais 12 são artigos, 12 são *conference paper*, 1 é uma *review*, 1 um *research paper* e 1 *proceeding paper*. A Figura 2.1 sintetiza todo o processo de seleção dos estudos.

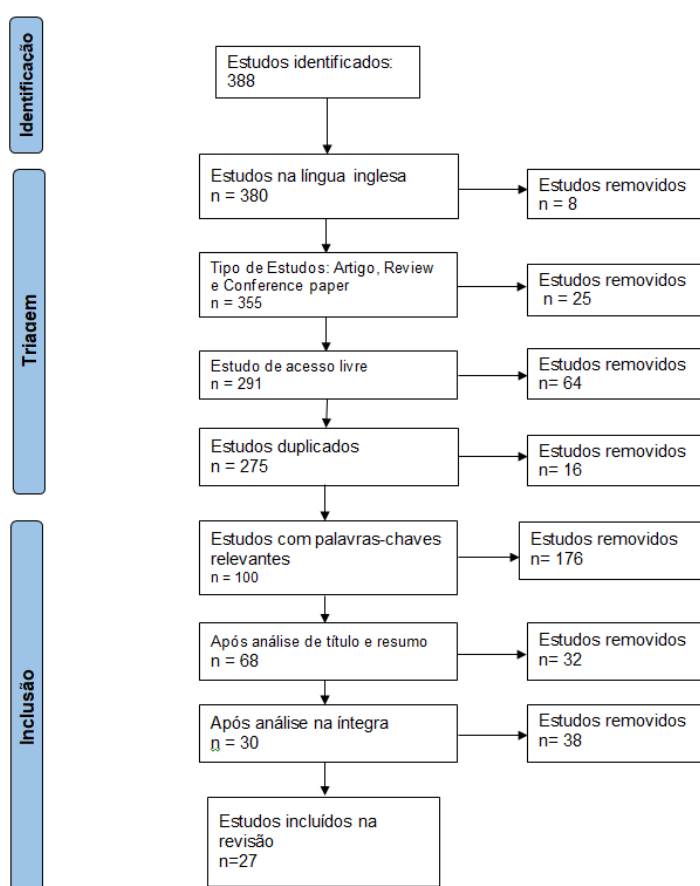


Figura 2.1 Processo de seleção dos estudos a incluir na RSL

### 2.1.3. Análise e avaliação dos estudos da RSL

Para compreender melhor cada um dos 27 estudos incluídos na RSL, é criado um ficheiro no Google docs com o resumo de cada um, de forma a responder às questões de pesquisa.

Por fim, de modo a perceber quais os estudos que melhor respondem às seis questões formuladas são aplicados critérios para avaliar os estudos analisados. Os critérios, operacionalizados em perguntas, encontram-se na Tabela 2.2, e para cada é atribuída uma pontuação: 1 para estudos que respondem totalmente à pergunta; 0,5 para os que respondem parcialmente; e 0 para os que não respondem.

**Tabela 2.2 Critérios para avaliar os estudos incluídos na RSL**

ID	Critério para a avaliação dos estudos
1.1	Define objetivamente o que é um DPA ou um IDP?
1.2	Define claramente o que é RPA?
2.1	Apresenta claramente os benefícios da implementação destes sistemas?
2.2	Apresenta os desafios da implementação destes sistemas?
3.1	Compara adequadamente as diferentes formas de extração de informação?
3.2	Descreve com clareza o processo de extração de informação?
4.1	O estudo apresenta as ferramentas existentes?
4.2	Apresenta adequadamente as vantagens e desvantagens da ferramenta?
5.1	Apresenta adequadamente o tratamento dos dados recolhidos?
5.2	Apresenta adequadamente a modelação do projeto?
6.1	Os resultados da pesquisa estão devidamente apresentados?
6.2	São apresentados os contributos do estudo?
6.3	São apresentadas as pistas futuras para investigação?
6.4	São apresentadas as limitações do estudo?

**Notas:** DPA - Processamento automático de documentos; IDP - Processamento inteligente de documentos; RPA - Automação de processos robóticos

## 2.2. Caracterização dos estudos seleccionados

Os artigos seleccionados são identificados na Tabela 2.3 e na Tabela 2.4 estão identificados os outros estudos, ambas as tabelas são apresentadas por ordem cronológica e são apresentadas algumas características, nomeadamente, o ano, o título, o *journal* ou conferência e ou autores de cada estudo.

**Tabela 2.3 Artigos incluídos na RSL**

(continua)

ID	Ano	Título	Journal	Autores
1	2000	Machine learning for intelligent processing of printed documents	Journal of intelligent information systems	Floriana Esposito; Donato Malerba; Francesca A. Lisi
2	2001	Automatic document classification and indexing in high-volume applications	International journal on document analysis and recognition	E. Appiani; F. Cesarini; A.M. Colla; M. Diligenti; M. Gori; S. Marinai; G. Soda
3	2011	A review of artificial intelligence algorithms in document classification	International journal of electronics and telecommunications	Adrian Bilski
4	2020	Robotic process mining: Vision and challenges	Business & information systems engineering	Volodymyr Leno; Artem Polyvyanyy; Marlon Dumas; Marcello La Rosa; Fabrizio Maria Maggi
5	2022	A framework for implementing robotic process automation projects	Information systems and e-Business management	Lukas-Valentin Herm; Christian Janiesch; Alexander Helm; Florian Imgrund; Adrian Hofmann; Axel Winkelmann



ID	Ano	Título	Journal	Autores
6	2022	Research on artificial intelligence of accounting information processing based on image processing	Mathematical biosciences and engineering	Juanjuan Tian; Li Li
7	2022	Low-code development and model-driven engineering: Two sides of the same coin?	Software and systems modeling	Davide Di Ruscio; Dimitris Kolovos; Juan de Lara; Alfonso Pierantonio; Massimo Tisi; Manuel Wimmer

**Tabela 2.3 Artigos incluídos na RSL**

ID	Ano	Título	Journal	Autores
8	2022	Intelligent automation of invoice parsing using computer vision techniques	Multimedia tools and applications	Anisha Chazhooor; Vergin Raja Sarobin
9	2022	Automated invoice data extraction using image processing	IAES International journal of artificial intelligence	Akanksh Aparna Manjunath; Manjunath Sudhakar Nayak; Santhanam Nishith; Satish Nitin Pandit; Shreyas Sunkad; Pratiba Deenadhayalan; Shobha Gangadhara
10	2022	Intelligent process automation: An application in manufacturing industry	Sustainability	Federico A. Lievano-Martínez; Javier D. Fernández-Ledesma; Daniel Burgos; John W. Branch-Bedoya; Jovani A. Jimenez-Builes
11	2022	Robotic process automation: A review of organizational grey literature	International journal of information systems and project management	Chugh; Ritesha; Macht; Stephanieb; Hossain; Rahat
12	2022	Information extraction from scanned invoice images using text analysis and layout features	Signal processing: image communication	H.T. Ha; A. Horák

**Notas:** IAES - Institute of advanced engineering and science

**Tabela 2.4 Outros estudos incluídos na RSL**

ID	Ano	Título	Publicação	Autores
13	2004	Aligning process automation and Business Intelligence to Support Corporate Performance Management	10th Americas conference on information systems	Melchert; Florian; Klesse; Mario; Winter; Robert
14	2005	Intelligent document processing	8th International conference on document analysis and recognition	F. Esposito; S. Ferilli; T.M.A. Basile; N. Di Mauro
15	2017	A benchmark and evaluation for text extraction from PDF	ACM/IEEE Joint conference on digital libraries	Hannah Bast; Claudius Korzen
16	2019	Alignment of business in robotic process automation	International journal of crowd science	Ning Zhang; Bo Liu
17	2020	Intelligent document processing based on RPA and machine learning	Chinese automation congresso	Xufeng Ling; Ming Gao; Dong Wang
18	2020	Robotic process automation and consequences for knowledge workers; a mixed-method study	19th Conference on e-Business, e-Services, and e-Society	Tom Roar Eikebrokk; Dag Håkon Olsen
19	2020	Data extraction from invoices using computer vision	IEEE 15th International conference on industrial and information systems	Mayur Sudhakar Satav; Tanmay Varade; Dhanashree Kothavale; Snehal Thombare; Pranali Lokhande
20	2020	Analysis of image classification for text extraction from bills and invoices	11th International conference on computing, communication and networking technologies	K.M. Yindumathi; Shilpa Shashikant Chaudhari; R. Aparna

ID	Ano	Título	Publicação	Autores
21	2020	Automatic document classification using deep feature selection and knowledge transfer	1st International conference on innovative research in applied science, engineering and technology	Aissam Jadjli; Mustapha Hain
22	2021	Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions	Institute of Electrical and Electronics Engineers	Dipali Baviskar; Swati Ahirrao; Vidyasagar Potdar; Ketan Kotecha
23	2021	A survey on automatic text summarization techniques	Internacional Conference on Physics and Energy	S Hima Bindu Sri; Sushma Rani Dutta
24	2021	Designing an information system for the electronic document management of a university: Automatic classification of documents	Journal of physics: Conference series	Tkachenko A.L.; Denisova L.A
25	2022	Intelligent document processing with small and relevant training dataset	International conference on intelligent systems and computer vision	Lina Nicolaieff; Mohamed Mehdi Kandi; Younes Zegaoui; Christophe Bortolaso
26	2022	E-Mail assistant - automation of e-mail handling and management using robotic process automation	International conference on decision aid sciences and applications	Arpit Khare; Sudhakar Singh; Richa Mishra; Shiv Prakash; Pratibha Dixit
27	2022	OCR-Free document understanding transformer	European conference on computer vision	Geewook Kim; Teakgyu Hong; Moonbin Yim; Jeongyeon Nam; Jinyoung Park; Jinyeong Yim; Wonseok Hwang; Sangdoon Yun; Dongyoon Han; Seunghyun Park

**Notas:** ACM/IEEE - Association for computing machinery e IEEE; PDF - Portable document format; IEEE - Institute of electrical and electronics engineers; RPA - Automação de processos robóticos; OCR - Reconhecimento ótico de caracteres

**Tabela 2.5 Documentos não científicos incluídos na RSL**

ID	Ano	Título	Publicação	Autor
28	2021	What is IBM automation document processing	What is IBM automation document processing: IBM documentation	IBM
29		Processamento inteligente de documentos	Processamento inteligente de documentos: Como funciona o IDP	Automation Anywhere

**Notas:** IBM - Internacional business machines; IDP - Processamento inteligente de documentos

Para melhor avaliar a pertinência dos documentos em análise a Figura 2.2 mostra uma nuvem de palavras criada a partir dos resumos dos documentos identificados. Assim, podem-se ver que as palavras que mais ocorrem são palavras *document*, *ocr*, *text*, *process*, *recognition* e *system*.

As palavras *document* e *text* podem ser explicadas devido ao facto de muitos documentos utilizarem o processamento de documentos e o processamento de texto, a palavra OCR pode ser explicada devido ao facto de muitos estudos utilizarem esta tecnologia para efetuarem o reconhecimento de texto.



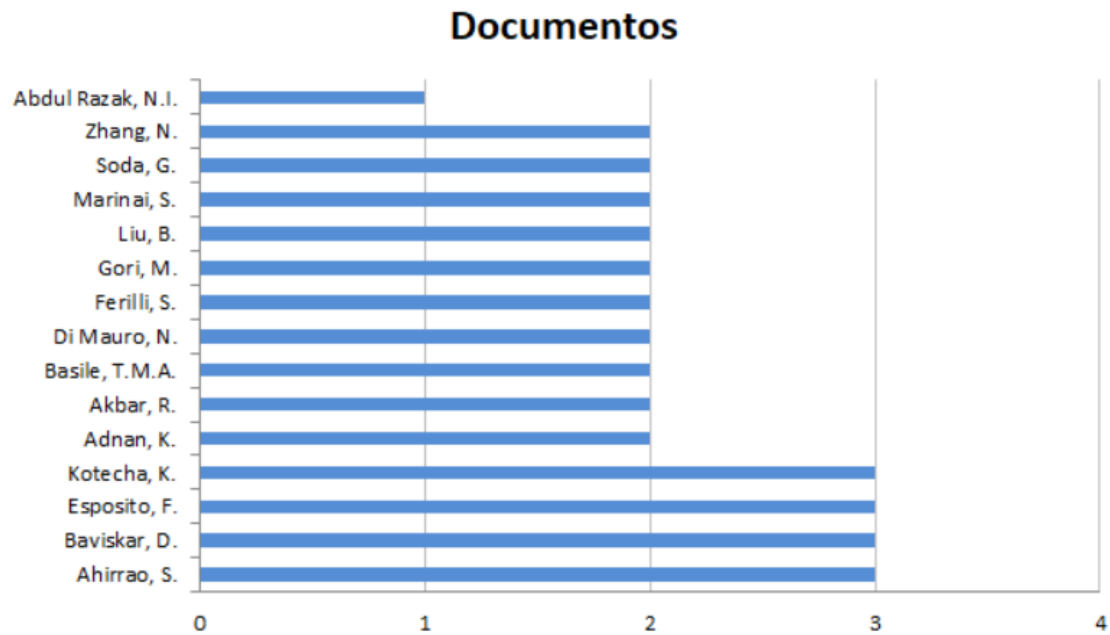


Figura 2.4 Número de estudos por autor

Por último, a Figura 2.5 mostra cada distribuição do tipo de estudo, verificando-se que o tipo de estudo mais usado é o *conference paper* com 53%, seguido dos artigos com 41%.

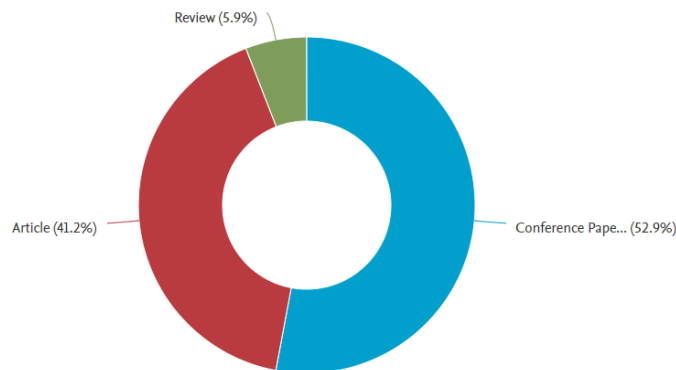


Figura 2.5 Contribuição percentual de cada tipo de estudos

### 2.3. Síntese dos artigos

Na Tabela 2.6 é apresentado o objetivo de cada artigo e de cada estudo selecionado. Destaca-se que a maior parte (9 estudos) tem como objetivo o processamento e automatização de documentos (e.g., Lievano-Martínez *et al.*, 2022; Ling *et al.*, 2020 ; Nicolaieff *et al.*, 2022). Em segundo lugar com seis estudos, são os que visam a extração de informação (e.g., Chazhooor & Sarobin, 2022; Yindumathi *et al.*, 2020; Baviskar *et al.*, 2021). Importa ainda referir que apenas um estudo visa a otimização de *email* (Khare *et al.*, 2022).

Tabela 2.6 Objetivo dos artigos e dos estudos incluídos na RSL

ID	Objetivos dos artigos e dos estudos
1	Técnicas de <i>machine learning</i> (ML) no processamento de documentos
2	Classificação automáticas de documentos
3	Técnicas e metodologias para a classificação de texto
4	Automatização de processos
5	Desenvolvimento e avaliação de projetos de RPA
6	Processamento de informações através de imagens
7	Abordagens de <i>low code</i>
8	Extração de texto a partir de imagens
9	Extrair informação através de técnicas de extração de texto e OCR
10	Automação de processos inteligentes
11	Implementação do RPA
12	Extrair informações através de técnicas de processamento de imagem, OCR, análise de texto
13	Automação de Processos
14	Automatizar a análise de documentos digitais
15	Ferramentas de extração de texto
16	Automação das estratégias de negócios com a tecnologia RPA
17	Automatização de documentos
18	Contribuir com a literatura sobre questões do RPA
19	Extrair informações usando OCR
20	Extração de texto a partir de imagens
21	Classificação automática de documentos
22	Extração automática de documentos
23	Estudo sobre os diferentes métodos, abordagens de resumo automático de textos
24	Classificação de documentos
25	Processamento de documentos
26	Otimização de <i>e-mails</i>
27	Explicar em que consiste o método OCR

Notas: ML - Machine learning; RPA - Automação de processos robóticos, OCR - Reconhecimento ótico de caracteres

### 2.3.1. Conceito e objetivo de DPA e IDP

Na literatura analisada, existem poucas definições e poucos artigos e estudos sobre a automatização de processos recorrendo ao DPA ou ao IDP, uma vez que estes conceitos muitas das vezes são associados ao conceito de RPA.

Assim, a Tabela 2.7 apresenta três artigos (Esposito *et al.*, 2000; Tian *et al.*, 2022; Lievano-Martínez *et al.*, 2022) e dois estudos (Esposito *et al.*, 2005; Nicolaijeff *et al.*, 2022) que definem DPA e IDP, e, também, os respetivos objetivos.

Tabela 2.7 Conceito e objetivo de DPA e IDP

ID	CONCEITO	OBJETIVO
1	Conjunto de técnicas de IA, para analisar, interpretar e extrair informações de documentos.	Usar o OCR, a classificação de documentos para extrair informações.
6	Uso de tecnologias e técnicas de IA para automatizar tarefas repetitivas.	Aumentar a eficiência, reduzir erros e liberar recursos humanos para outras tarefas.
10	Conjunto de tecnologias de IA com técnicas de <i>natural language processing</i> (NPL), ML, que ajudam na previsão, análise e otimização de processos.	Automatizar processos e reduzir tempo e gastos associados a esses processos.
14	Conjunto de técnicas e tecnologias para extrair informações.	Automatizar o processamento de documentos.
25	Conjunto de tecnologias que envolve o processamento de imagem, o ML e o OCR para processar documentos e extrair informações.	Automatizar o processo de extração de informações de documentos.

**Notas:** IA - Inteligência artificial; OCR - Reconhecimento óptico de caracteres; NPL - Natural Language Processing; ML - Machine learning

Assim, pode concluir-se que o IDP é uma área de IA que se concentra em desenvolver técnicas e algoritmos para automatizar a análise, a classificação e a extração de informação de documentos. De outra forma, o IDP pode ser entendido como um conjunto de tecnologias que envolve o processamento de imagem, o OCR, processamento de texto e ML para processar documentos e extrair informações relevantes (Nicolaiëff *et al.*, 2022). Este tipo de IA é amplamente utilizado em áreas da sociedade, como a saúde, finanças, governo, meios de comunicação e área jurídica (Zhang & Liu, 2019).

Em Esposito *et al.* (2005), o conceito de IDP é definido como um conjunto de técnicas e tecnologias destinadas a extrair informações relevantes e categorizar os dados para análise e processamento, com o objetivo de automatizar o processamento de documentos e reduzir o tempo e gastos associados ao trabalho manual de processamento de documentos.

Em consequência da escassez de estudos neste tema, os outros estudos selecionados não diferem muito destas definições. Assim, para ter uma ideia mais abrangente deste tema a IBM define IDP como um processo que fornece os recursos que ajudam a criar uma ferramenta de enriquecimento de dados com tecnologia de IA para processamento e armazenamento de documentos (IBM Documentation, 2021).

Por outro lado, os algoritmos de IDP podem ser divididos em duas categorias principais, o processamento de texto e o processamento de imagem. O processamento de texto inclui técnicas como OCR, análise de sentimentos e sumarização automática de texto (Ha & Horák, 2022). O processamento de imagem inclui técnicas como segmentação de imagem (Tian *et al.*, 2022), reconhecimento de caracteres (Manjunath *et al.*, 2022) e análise de *layout* (Ha & Horák, 2022).



Figura 2.6 Funcionamento de um IDP

Fonte: (Automation Anywhere, nd)

### 2.3.2. Reconhecimento Óptico de Carateres

O OCR vem ganhando uma grande popularidade ao redor do mundo, uma vez que permite às empresas processar os seus documentos, permitindo manter os dados facilmente acessíveis e economizar tempo e espaço (Satav *et al.*, 2020). O OCR permite a digitalização de documentos impressos e a conversão de imagens em texto digital, tornando-os mais fáceis de pesquisar e compartilhar. Além disso, é amplamente utilizado para automatizar o processo de classificação de documentos, o que pode economizar tempo e esforço humano (Satav *et al.*, 2020).

Outra aplicação comum do OCR é a extração automática de informações, como nomes, endereços, números de telefone e data. Isso é especialmente útil em campos como finanças e saúde, onde é importante processar grandes volumes de dados de forma precisa e eficiente (Chazhour & Sarobin, 2022; Satav *et al.*, 2020).

Alguns estudos feitos com a tecnologia OCR revelam resultados interessantes. Ha e Horák (2022) apresentam um método para extrair informações de faturas digitalizadas usando análise de texto e características de *layout*. Os autores apresentam um método baseado em quatro etapas, o pré-processamento de imagem, a segmentação de texto, OCR e extração de informação, concluindo que o método proposto é capaz de extrair informação com alta precisão, mesmo com diferentes tipos de faturas, diferentes tipos de fontes de texto e diferentes idiomas.

Yindumathi *et al.* (2020) analisam a abordagem mais tradicional de extração de texto, abordagem essa que envolve a digitalização da imagem e a utilização de uma só ferramenta de OCR. Eles concluem que esta técnica é pouco precisa para lidar com diferentes *layouts* e texto ilegível. Assim, com o objetivo de analisar a precisão dos diferentes modelos, os autores examinam a utilização de métodos de ML na extração de informação, juntamente com o uso do OCR. Testam três modelos de classificação de imagem e todos mostram um bom desempenho na extração de texto, com uma precisão média de 90%. Concluem, então, que o método proposto supera a abordagem tradicional em termos de precisão e velocidade de extração de texto. Eles também concluem que é uma estratégia útil e precisa para empresas que necessitam de processar grandes quantidades de documentos.

### **2.3.3. Desafios e benefícios da implementação do IDP**

As Tabela 2.8 e Tabela 2.9 apresentam o conjunto dos artigos que falam sobre os benefícios e desafios desta tecnologia. De um modo em geral, o IDP tem o potencial de melhorar significativamente a eficiência e a precisão dos processos de negócios, bem como tornar a informação mais rápida, mais acessível para as pessoas (Lievano-Martínez *et al.*, 2022). Para além destes benefícios, a tecnologia melhora a eficiência no trabalho, ao reduzir a taxa de erro e reduzir os gastos de mão-de-obra (Tian *et al.*, 2022).

Para além dos benefícios anteriores, Chugh *et al.* (2022) identificam benefícios da automação de processos e categoriza-os em monetários, simplicidade, eficiência e produtividade, flexibilidade e escalabilidade, confiabilidade e consistência, conformidade, satisfação do cliente e eficiência dos funcionários.

No entanto, é importante notar que as novas tecnologias também têm desafios e consequências, Tabela 2.9. Por exemplo, Eikebrokk & Olsen (2020) fornecem informações importantes sobre as implicações da automação de processos para os trabalhadores, embora os resultados do seu estudo indiquem que muitos trabalhadores percebem os benefícios desta tecnologia, muitos deles também mostram preocupações em relação à segurança do emprego e à qualidade do emprego, assumindo que pode levar a mudanças significativas das suas funções, exigindo aos trabalhadores novas competências e novas habilidades. Assim os autores concluem que é importante que as organizações implementem medidas para garantir a segurança do emprego e o desenvolvimento de novas habilidades e competências relevantes para os trabalhadores.



Tabela 2.8 Benefícios do IDP

ID	Benefícios
4	Aumento de rapidez e eficiência
5	Múltiplas tarefas, redução de gastos
6	Aumento de eficiência, redução de gastos e redução de erros
8	Aumento de eficiência e redução de gastos
9	Redução de carga de trabalho
10	Automatização de processos e aumento de lucratividade
11	Simplicidade, eficiência, produtividade e confiabilidade
14	Reduzir tempo e gastos
17	Reduzir taxa de erro, gastos e melhorar a eficiência
18	Redução de gastos, flexibilidade e melhorar a eficiência
19	Reduzir erros, aumentar a eficiência e automatizar processos
22	Melhorar a competitividade e produtividade
26	Reduzir erro

Tabela 2.9 Desafios do IDP

ID	Desafios
4	Clareza e dificuldade no tratamento de grande volume de dados
5	Integração de TI fraca é menos robusta
6	Mudança na forma de trabalhar
8	Baixo desempenho de algoritmos em cenários da vida real
10	Infraestrutura preparada
11	Incerteza sobre o futuro e desafio de gestão de mudança
13	Infraestrutura preparada, organização e renovação dos processos de negócio
18	Aprender novas competências
19	Conhecimento técnico especializado

### 2.3.4. Técnicas de extração de informação

A Tabela 2.10 apresenta o conjunto dos artigos que melhor identificam as técnicas de extração de informação.

Tabela 2.10 Técnicas de extração de informação

ID	TÉCNICA
1	Técnicas de ML para adquirir o conhecimento específico exigido por um sistema inteligente de processamento de documentos, WISDOM++.
3	Classificadores Naive Bayes (NB), K-éssimo vizinho mais próximo (K-NN), máquina de vetores de suporte (SVM) e sistemas híbridos.
8	Combinação de algoritmos de detecção de objetos com OCR
14	Sistema baseado em conhecimento com uso de métodos de ML aplicados, para aplicar automaticamente as regras para a realização de várias etapas de processamento.
17	Tecnologia RPA, combinada com tecnologia OCR e tecnologia de <i>deep learning</i> (DL).
20	Combinação do mecanismo Tesseract com técnicas de segmentação de texto.
26	Criação de um RPA através do UiPath, que usa algoritmos de RPA, IA e <i>web scraping</i>

**Notas:** ML - Machine learning; NB - Naive Bayes; K-NN - K-éssimo vizinho mais próximo; SVM - Máquina de vetores de suporte; OCR - Reconhecimento óptico de caracteres; DL - Deep learning; RPA - Automação de processos robóticos; IA - Inteligência artificial

A extração automática de informação é um tema que está em constante evolução, pois vive-se num mundo onde o progresso tecnológico evolui constantemente, e a cada dia que passa está-se mais dependente destas mesmas tecnologias.

Com o evoluir da tecnologia, surgem as primeiras soluções de extração automática de documentos, soluções essas que passam por técnicas de OCR que transforma o conteúdo de documentos digitalizados em formato digital. As versões preliminares do OCR exigem a apresentação de uma imagem de cada caractere e eram limitadas a reconhecer apenas uma fonte por vez (Baviskar *et al.*, 2021).

Como se pode ver pela Figura 2.7, começa aparecer as soluções baseadas em RPA, que é uma tecnologia que permite que as organizações automatizem as suas tarefas repetitivas (Leno *et al.*, 2021), de forma a substituir os processos baseados em regras, estruturados e repetitivos por um robô de *software* (Baviskar *et al.*, 2021).

Nos dias de hoje a automação baseada em inteligência artificial utiliza várias técnicas como NPL, ML e análise de texto (e.g., Baviskar *et al.*, 2021; Lievano-Martínez *et al.*, 2022).

Bilski (2011) apresenta vários algoritmos de IA para a classificação de documentos, onde são incluídos os algoritmos de rede neural artificial (ANN), SVM, KNN, classificador NB, árvores de decisão (DT) e algoritmos de indução de regras.

Ling *et al.* (2020) descrevem uma abordagem para automatizar o processamento de documentos utilizando uma combinação de ML com RPA. O RPA lida com as tarefas do dia-a-dia e as tarefas repetitivas, enquanto os algoritmos de ML extraem as informações dos documentos e melhoraram a precisão e eficiência do processo. No fim do estudo é destacado o potencial que a combinação tem para melhorar o processamento de documentos em empresas e organizações.

Por fim, importa referir Khare *et al.* (2022), que descrevem o uso de um RPA para automatizar a caixa de correio eletrónico de uma empresa, de forma a processar, classificar e responder a correio eletrónico de forma eficiente. O sistema proposto usa técnicas de NPL para entender o correio eletrónico e em seguida tomar uma decisão conforme as regras de negócio definidas.

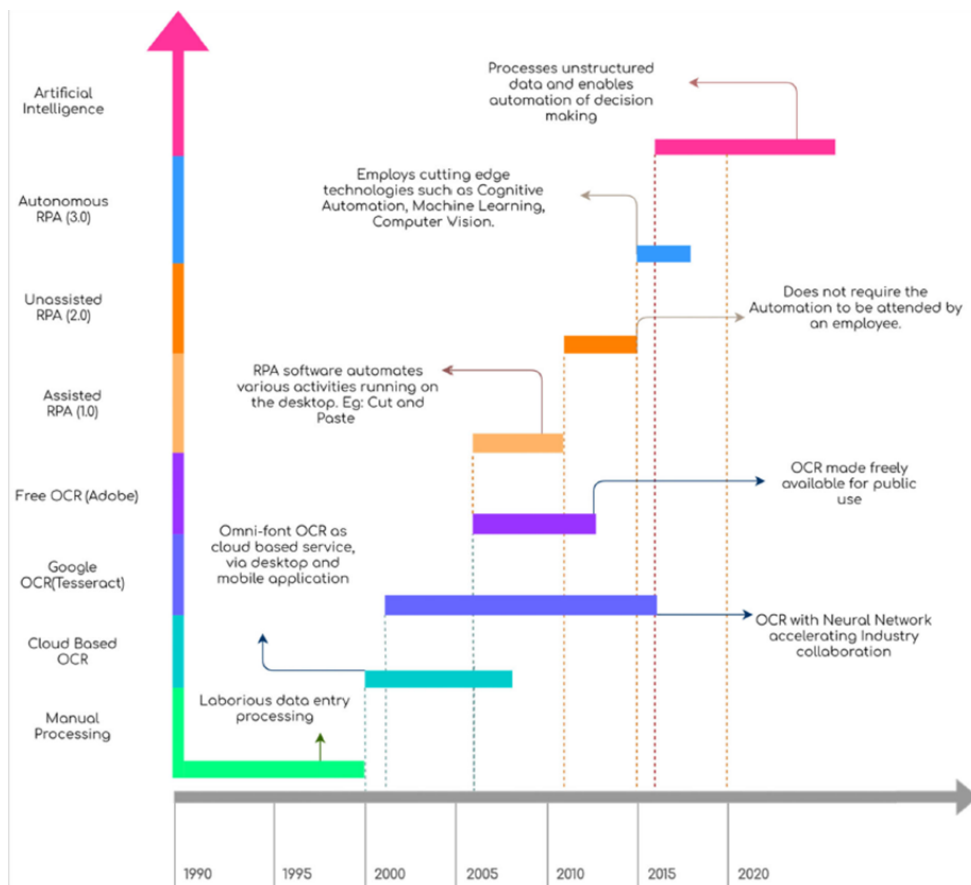


Figura 2.7 Evolução das técnicas para extração de informação de documentos

Fonte: Baviskar *et al.* (2021, pág. 3).

### 2.3.5. Ferramentas de automação de processos e de extração de informação

Com a evolução das tecnologias, novas ferramentas para extrair informação têm um papel cada vez mais importante para as empresas, uma vez que são capazes de extrair informações importantes de dados de forma mais eficiente e mais precisa, ajudando as empresas a tomar as melhores decisões.

Baviskar *et al.* (2021) mostram uma comparação entre outras ferramentas, alguns protótipos, outras ferramentas OCR comerciais. Uma das ferramentas mencionadas no seu estudo é o CloudScan que é um sistema de análise de faturas com uma interface gráfica do usuário de configuração zero e não requer anotação prévia. No entanto, o CloudScan falha no processamento de recursos de imagem, como logotipos e marcas d'água ou qualquer fundo de imagem.

Por outro lado, (Manjunath *et al.*, 2022) propõem o *design* e a implementação de um sistema de processamento de faturas para *web*. Este contém uma ferramenta de anotação e uma ferramenta de extração, os campos de interesse na fatura são anotados pela ferramenta de anotação e a ferramenta de extração usa algoritmos de visão computacional de código aberto, com o objetivo de detetar o texto. O sistema é projetado para fornecer flexibilidade e automatizar o processo de extração de informações importantes das faturas.

### **2.3.6. Contributos, limitações e pistas futuras**

Todos os estudos, principalmente os científicos, devem apresentar contributos, limitações e pistas futuras para investigação futura.

Assim, na Tabela 2.11, são apresentados os contributos, limitações e pistas futuras de cada documento.

No que diz respeito aos contributos, alguns dos estudos mostram uma nova abordagem, (e, g., Leno *et al.*, 2021; Yindumathi *et al.*, 2020; Jadli & Hain, 2020). Por exemplo, Yindumathi *et al.*, (2020) apresenta uma nova proposta, que consiste em uma abordagem alternativa na extração de texto nas faturas. Utiliza técnicas de classificação de imagem antes de aplicar o OCR em vez da abordagem tradicional, concluem que esta nova abordagem é mais eficaz e precisa em relação á abordagem tradicional do OCR.

Outros estudos optam por aprofundar uma abordagem em estudo (e, g., Chazhoor & Sarobin, 2022; Satav *et al.*, 2020), em Satav *et al.*, (2020) os autores apresentam um modelo híbrido que combina tecnologia OCR com técnicas de processamento de imagem e ML para extrair informações das faturas. Concluem que esta abordagem teve uma melhoria significativa na precisão de extração de dados.

Por fim, alguns estudos falam sobre a importância destas tecnologias para as empresas, quais as suas vantagens e desvantagens (e, g., Herm *et al.*, 2022; Chugh *et al.*, 2022; Ling *et al.*, 2020). Melchert *et al.*, (2004) apresenta uma discussão sobre a importância da integração entre a automação de processos e a inteligência empresarial para ter uma gestão eficiente de desempenho corporativo.

No que diz respeito às limitações, a grande parte dos estudos indica que a grande limitação que existe é relativamente aos dados (e, g., Herm *et al.*, 2022; Chugh *et al.*, 2022; Yindumathi *et al.*, 2020), já que quando se trata de faturas, não existe muitos dados disponíveis, e nem sempre é fácil obter esses dados. Outra das limitações é relativamente ao *layout* das faturas (e, g., Ha & Horák, 2022; Esposito *et al.*, 2005), nem sempre as faturas seguem o mesmo *layout*, tornando-se assim difícil o treino do algoritmo. Também é identificada a limitação referente a problemas de reconhecimento de caracteres (e, g., Manjunath *et al.*, 2022; Satav *et al.*, 2020).

Relativamente às pistas futuras, a grande parte dos estudos prevê que como trabalho futuro vão aperfeiçoar o sistema em causa (e, g., Chazhoor & Sarobin, 2022; Lievano-Martínez *et al.*, 2022; Satav *et al.*, 2020), outra pista futura é o desenvolvimento e combinação de algoritmos com técnicas de IA (e, g., Ling *et al.*, 2020; Yindumathi *et al.*, 2020; Baviskar *et al.*, 2021). Apenas um estudo (Chugh *et al.*, 2022), tem como objetivo futuro ver os impactos que estas tecnologias provocam nas empresas e no mundo empresarial.

**Tabela 2.11 Contributos, limitações e pistas futuras**

ID	Contributos	Limitações	Pistas Futuras
1	Técnicas de ML para processamento inteligente de documentos	N. E	Melhorar desempenho a diferentes níveis de <i>layout</i>
2	N. E	N. E	Melhorar o sistema
3	Aprofundamento da abordagem em estudo	N. E	N. E
4	Nova abordagem	Falta de dados	N. E
5	Estado da arte dos projetos de RPA;	Falta de dados	N. E
6	Melhorar a eficiência, reduzir gastos e erros através do processamento de imagem no setor da contabilidade	N. E	Melhorar o sistema
7	Comparação entre abordagem de <i>low-code</i>	N. E	N. E
8	Aprofundamento da abordagem em estudo	Falta de dados	Melhorar o OCR com faturas mais compactas
9	Nova abordagem	Erros de reconhecimento de caracteres	N. E
10	N. E	Falta de dados	Melhorar automação do processo
11	Benefícios e desafios	Falta de estudos	Mais estudos sobre o impacto da RPA nas organizações
12	Nova abordagem	Problemas com <i>layout</i> de faturas	Melhorar desempenho a diferentes níveis de <i>layout</i>
13	Integração entre a automação de processos e a inteligência empresarial.	Integração entre sistemas, treinamento dos funcionários	Uso de novas tecnologias
14	N. E	Problemas com <i>layout</i> de faturas	N. E
15	Avaliar ferramentas de extração	N. E	N. E
16	N. E	Falta de estudos	Implicações da RPA na perspetiva

ID	Contributos	Limitações	Pistas Futuras
			do cliente, dos <i>stakeholders</i> ou do negócio.
17	Benefícios e desafios	N. E	Desenvolver e aplicar algoritmos inteligentes com base em IA
18	Benefícios e desafios	Falta de dados e avanço tecnológico	N. E
19	Aprofundamento da abordagem em estudo	Erros de reconhecimento de caracteres	Melhorar técnicas
20	Nova abordagem	Falta de dados	Desenvolver técnica baseada em IA com OCR
21	Nova abordagem	Falta de dados e de qualidade	Novas técnicas de seleção e mais dados.
22	Comparar técnicas	N. E	Desenvolver técnica baseada em IA com OCR
23	Aprofundamento da abordagem em estudo	N. E	Trabalhar várias combinações de técnicas
24	Aprofundamento da abordagem em estudo	N. E	N. E
25	N. E	Falta de dados	Experimentos com mais dados.
26	N. E	Manutenção do RPA	Melhorar o RPA
27	N. E	N. E	Melhorar o objetivo do pré treino

**Notas:** N. E - não especificou; RPA - Automação de processos robóticos; OCR - Reconhecimento ótico de caracteres; IA - inteligência artificial

## 2.4. Avaliação dos estudos da RSL

Na Tabela 2.12 está sintetizada a avaliação dos 27 estudos que fazem parte desta RSL. As pontuações para cada critério, conforme estabelecido no subcapítulo 2.1.3, são apresentadas e é, também, apresentada a pontuação final de cada documento e a pontuação final de cada pergunta de investigação.

O estudo com melhor classificação (11 pontos em 14 possíveis) é o de Chazhoor & Sarobin (2022) e em segundo lugar Yindumathi *et al.*, (2020) com (10 pontos em 14 possíveis), correspondentes ao ID=8 e ID= 20. Estes foram publicados no Multimedia Tools and Applications e na International Conference on Computing, Communication and Networking Technologies, em 2022 e 2020 respetivamente.

**Tabela 2.12 Avaliação dos artigos e dos estudos**

ID	Q1.1	Q1.2	Q2.1	Q2.2	Q3.1	Q3.2	Q4.1	Q4.2	Q5.1	Q5.2	Q6.1	Q6.2	Q6.3	Q6.4	Total
1	0,5	0	0	0	1	1	0,5	0	0	1	0,5	0	1	0	5,5
2	0	0	0	0	0,5	1	1	0	1	1	1	1	1	1	8,5
3	0	0	0	0	1	0,5	0	0	0	0	1	1	0	0	3,5
4	0	1	1	1	0,5	1	0,5	0	0	0,5	1	1	0	1	8,5
5	0	1	1	1	0	1	0,5	0	1	1	1	1	0	1	9,5
6	0,5	0	1	1	0,5	1	0	0	1	1	0,5	0	1	0	7,5
7	0	0	0	0	0	0	0,5	0	0	0	0	0	0	0	0,5
8	0	0	1	1	1	1	1	0,5	0,5	1	1	1	1	1	11

ID	Q1.1	Q1.2	Q2.1	Q2.2	Q3.1	Q3.2	Q4.1	Q4.2	Q5.1	Q5.2	Q6.1	Q6.2	Q6.3	Q6.4	Total
9	0	0	1	0	0	1	1	0,5	0,5	0,5	1	1	0	1	7,5
10	0,5	1	1	1	0	0,5	0	0	0,5	0,5	1	1	1	1	9
11	0	1	1	1	0	0	0,5	0,5	0	0	1	1	1	1	8
12	0	0	0,5	0	0	1	1	0	0,5	0,5	1	1	1	0,5	7
13	0	0	0,5	1	0	0	1	0	0	0	0,5	1	1	1	6
14	0,5	0	1	0	1	0,5	0	0	0,5	0	0,5	0	0	1	5
15	0	0	0	0	0	0	1	1	0	0	1	1	0	0	4
16	0	1	0,5	0	0	0	1	0	0,5	0	0,5	0	1	1	5,5
17	0	1	1	0,5	1	1	0	0	0	1	1	1	1	0	8,5
18	0	1	1	1	0	0	0	0	0,5	0	1	1	0	1	6,5
19	0	0	1	1	0,5	1	1	0,5	1	0,5	0,5	1	1	1	10
20	0	0	0,5	0,5	1	1	1	0,5	1	1	1	1	1	1	10,5
21	0	0	0,5	0	0,5	1	1	0,5	0,5	1	1	1	1	1	9
22	0	0,5	1	0,5	1	0,5	0	1	0,5	1	1	1	1	1	10
23	0	0	0	0	0,5	0,5	1	0	0	0	1	1	1	0	5
24	0	0	0	0	0,5	0,5	0	0	0	0	0	1	0	0	2
25	1	0	0	0	1	1	1	0	1	1	1	0	1	1	9
26	0,5	0,5	1	0	0	1	1	1	0	0	1	0	1	1	8
27	0	0	0	0	1	0,5	0	0	0	0	1	0	1	0	3,5
Total	3,5	8	15,5	10,5	12,5	17,5	15,5	6	10,5	12,5	22	19	18	17,5	

Através da leitura da Tabela 2.12, percebe-se que os documentos mais relevantes para este estudo são os com ID 5, 8, 10, 19, 20,21, 22 e 25 (Herm *et al.*, 2022; Chazhoor & Sarobin, 2022; Lievano-Martínez *et al.*, 2022; Satav *et al.*, 2020; Yindumathi *et al.*, 2020; Jadli & Hain 2020; Baviskar *et al.*, 2021; Nicolaieff *et al.*, 2022).

Os documentos 22 (Baviskar *et al.*, 2021) e 8 (Chazhoor & Sarobin, 2022) falham na definição de DPA e IDP, que pode ser explicado pelo facto de muitas vezes estes conceitos serem associados ao conceito de RPA. Destes estudos o único que fala sobre IDP é o documento 25 (Nicolaieff *et al.*, 2022). O documento 21 (Jadli & Hain, 2020) apresenta uma nova abordagem na classificação automática de documentos que supera as abordagens tradicionais e alcança uma alta precisão.

Através da Tabela 2.12, também é possível constatar que a questão com maior pontuação, retirando as questões das limitações, pistas futuras e contributos, é a questão 3.2 (Descreve com clareza o processo de extração de informação?), com um total de 17 pontos em 27 possíveis, seguido da 2.1 (Apresenta claramente os benefícios da implementação destes sistemas?), com um total de 15,5 pontos em 27 possíveis. Já por outro lado a questão com menor pontuação é a questão 1.1, relativo ao conceito de DPA e IDP (com 3,5 pontos em 27 possíveis) e em seguida a questão 4.2 relativo às vantagens e desvantagens das ferramentas existentes para a extração de informação (com 6 pontos em 27 possíveis).



### 3. Metodologia

Para abordar a questão desta investigação, que consiste em identificar onde são feitas os gastos da empresa, neste caso uma empresa de construção civil, pretende-se desenvolver um sistema que possa fazê-lo levando em consideração a faturação dos seus fornecedores.

Conforme afirmado no subcapítulo 1.3, a metodologia escolhida é o CRISP-DM, esta metodologia foi modificada e, portanto, consistirá em seis fases, cada uma será explicada ao longo deste capítulo.

Esta foi a metodologia escolhida uma vez que é uma metodologia adequada para trabalhar com a mineração e análise de dados, é simples de entender e por permitir interação com o gerente da empresa (Schröer *et al.*, 2021).

#### 3.1. Compreensão do negócio

O principal objetivo desta investigação é criar um fluxo para automatizar o processamento de documentos, neste caso as faturas dos fornecedores, de forma a realizar esta tarefa automaticamente e assim libertar os colaboradores da empresa para outra atividade.

Como segundo objetivo este projeto passa também por criar um *dashboard* analítico para auxiliar o gestor da empresa na tomada de decisão. Este *dashboard* permite ao gestor ver os seus gastos, ver em qual fornecedor gasta mais, ver qual o produto que gasta mais, entre outras análises.

Para realizar o primeiro objetivo são identificados dois critérios sucesso, de forma a medir o desempenho e a eficácia do fluxo. Assim, como primeiro critério, o fluxo tem de ter uma precisão de recolha superior a 80%. Esta percentagem foi determinada como critério de sucesso, uma vez que, o AI Builder considera que uma pontuação abaixo dos 80% é uma pontuação fraca.

Como segundo critério de sucesso, define-se uma satisfação média global dos potenciais utilizadores com o fluxo superior a 7, numa escala crescente de satisfação de 1 (completamente insatisfeito) a 10 (completamente satisfeito). Para responder a este critério foram feitas três questões aos potenciais utilizadores do fluxo que se encontram no subcapítulo 3.4.2.

Para a concretização deste projeto são usadas ferramentas do ecossistema Microsoft, pois a empresa já utiliza algumas ferramentas deste ecossistema e tenciona no futuro utilizar apenas ferramentas da Microsoft.

A ferramenta Power Automate é utilizada para criar o fluxo, uma vez que permite conectá-lo com todas as ferramentas do ecossistema Microsoft, com o Outlook para receber os *emails* com as faturas, com o Excel para guardar os dados extraídos das faturas, com o OneDrive para guardas as faturas. Esta ferramenta é uma plataforma de automatização inteligente, que permite criar fluxos de trabalho automatizados de forma a simplificar e otimizar os processos de negócio. Pode simplificar o trabalho, uniformizar processos e reduzir erros humanos. Além disso é uma ferramenta baseada em nuvem que permite a integração com varias aplicações da Microsoft (*Power Automate, nd*).

É utilizado o AI Buidler para recolher os dados utilizando o OCR para ler os documentos e extrair os dados. O AI Builder é a ferramenta escolhida uma vez que faz parte do Power Automate, esta é uma funcionalidade do Microsoft Power Platform que fornece modelos de IA para otimizar os processos de negócio. Permite automatizar tarefas e recolher dados de outras aplicações usando capacidades de IA e não necessita de qualquer conhecimento de programação (*V-aangie et al., 2023*).

Por último o Power BI é a ferramenta escolhida para criar o *dashboard*. Esta ferramenta é a escolhida por fazer parte do ecossistema Microsoft e porque o mestrando possui mais conhecimento sobre ela do que sobre o Tableau, sobre a qual o mestrando possui um conhecimento básico, ou até mesmo sobre o Qlik, com a qual o mestrando não possui nenhuma experiência. O Excel é uma alternativa viável, mas é escolhido o Power BI por preferência dos potenciais utilizadores do sistema.

Como critério de sucesso, define-se uma satisfação média global dos potenciais utilizadores com o *dashboard* superior a 7, numa escala crescente de satisfação de 1 (completamente insatisfeito) a 10 (completamente satisfeito). Para obter a satisfação dos potenciais utilizadores do *dashboard*, foi elaborado um serie de perguntas.

Para se perceber melhor o projeto, a Figura 3.1 mostra a arquitetura planeada do sistema proposto.

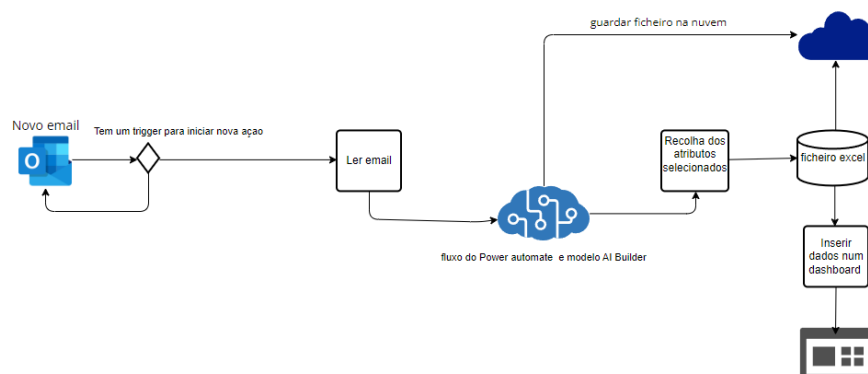


Figura 3.1 Arquitetura do sistema

Para um melhor suporte da seleção das ferramentas a utilizar no projeto, é feita uma comparação entre as diversas ferramentas existentes para realizar um projeto como este. Assim a Tabela 3.1 e a Tabela 3.2 apresentam uma comparação entre as ferramentas utilizadas, Microsoft Power Automate e Power BI, com algumas concorrentes.

**Tabela 3.1 Comparação entre as ferramentas existentes**

Ferramenta	Vantagens	Desvantagens	Preço (USD) /mês	Formatos suportados	Teste grátis
Power automate	Integração com o ecossistema Microsoft Interface amigável e fácil Amplas opções de conectividade	Dependência da plataforma Microsoft Necessidade de conexão à internet	500	PDF, Office 365, XML, JSON, TXT, CSV	Sim
Zapier	Interface simples. Boa integração entre apps	Ações e <i>triggers</i> podem ser um pouco limitadas	30	N. E	SIM
UiPath	Automatização de processos complexos Interface simples e fácil	Dependência de licenças Manutenção contínua Requisitos de infraestruturas	420	PDF, TXT, CSV, JSON, XML, HTML, CSV, Office 365	SIM
Automation Anywhere	Recursos avançados de automação Conectividade e integração abrangentes	Dependência de licenças Requisitos de infraestrutura Período de aprendizagem	750	PDF, Office 365, web scraping, XML, JSON,	SIM
Blue Prism	Robustez e escalabilidade Interface intuitiva Segurança e conformidade	Preço elevado Período de aprendizagem Dependência de infraestrutura	13000 ano	PDF, office 365, web scraping, XML, JSON, CSV, TXT	SIM
ReportMiner	Extração de dados não estruturados Processa grandes volumes de dados	Complexidade inicial Dependência de layout Limitações em documentos complexos	299	PDF, TXT, JSON, PRN, RTF, XLX, XLSX	SIM

Fonte: Capterra (nd), Getapp (nd)

**Tabela 3.2 Comparação entre as ferramentas de dashboard**

Ferramenta	Vantagens	Desvantagens	Preço (USD)/ mês	Teste grátis
Power BI	Fácil de usar Intuitivo Fácil integração com várias fontes de dados	Recursos de análise avançados Elevada curva de aprendizagem	10	SIM
Tableau	Interface intuitiva Ampla gama de visualizações	Preço elevado Elevada curva de aprendizagem	75	SIM
Excel	Simple e prático Fácil aprendizagem	Difícil de trabalhar com ficheiros complexos	7	SIM
Qlik	Facilidade na criação de gráficos Interface amigável	Preço alto Recursos complexos Difícil de aprender	30	SIM

Fonte: Capterra (nd), Getapp (nd)

### 3.2. Compreensão e preparação dos dados

Nesta fase, o objetivo passa por compreender os dados disponíveis para o projeto, bem como a recolha dos mesmos e a descrição das variáveis.

Neste subcapítulo é explicado a compreensão e preparação dos dados utilizados na construção do fluxo e na construção do *dashboard*.

#### 3.2.1. Fluxo

Os dados utilizados neste projeto são obtidos junto da empresa Adriano Luz Duarte Balaia, e o conjunto de dados recolhidos é composto por um conjunto de dados não estruturados com o formato PDF.

Os dados dizem respeito a 93 ficheiros, que correspondem a faturas de nove fornecedores da empresa, pertencendo à janela temporal de outubro de 2022 a junho de 2023.

Os dados das faturas a considerar neste projeto são os seguintes:

- Fornecedor: corresponde ao nome do fornecedor;
- NIF: corresponde ao número de identificação fiscal do fornecedor;
- NºFatura: corresponde ao número da fatura;
- Data Fatura: corresponde à data de emissão da fatura;
- Data pagamento: corresponde à data-limite de pagamento da fatura;
- Valor s/iva: corresponde ao valor líquido da fatura;
- Valor do Iva: corresponde ao montante de iva;
- Valor total: corresponde ao montante final a pagar;
- Desconto: corresponde ao valor de desconto;
- Localização: corresponde à localização do fornecedor;
- Código Produto: corresponde ao código do produto;
- Descrição: identificação do item comprado;
- Quantidade: quantidade do item comprado;
- Unidade: unidade do item comprado;
- Valor unitário: corresponde ao valor unitário do item;
- Montante: corresponde ao montante a pagar por cada item

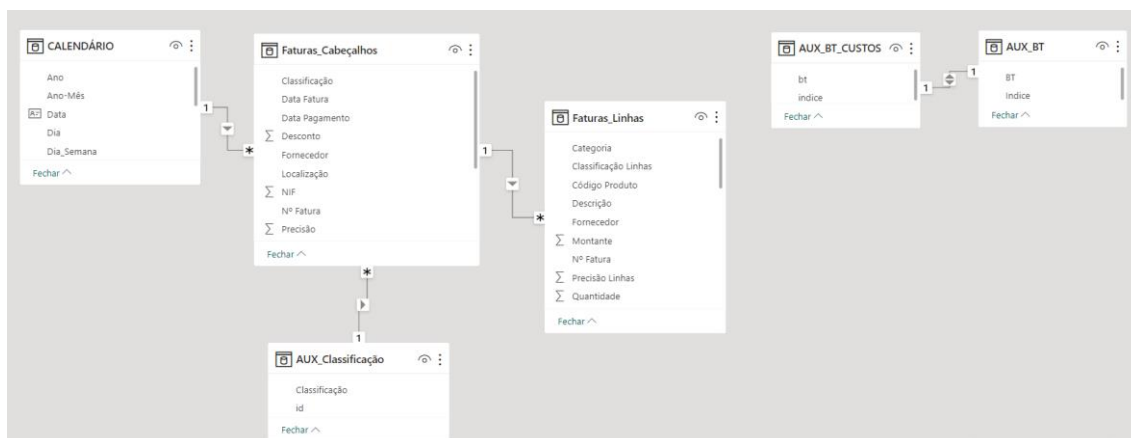
Para além destas variáveis a extrair automaticamente com o fluxo, é criada uma variável categoria, para agrupar os produtos por categorias, de forma a facilitar a visualização no *dashboard*. Também são criadas duas variáveis, Classificação\_cabeçalhos para classificar a extração dos cabeçalhos da fatura e Classificação\_linhas para classificar a extração das linhas da fatura. Estas variáveis vão permitir agrupar as faturas consoante a sua classificação de extração. Na Tabela 3.3 pode ver-se cada uma das categorias<sup>1</sup>.

**Tabela 3.3 Categorias de extração**

Precisão	Categoria
50%-79%	Mau
80%-89%	Razoável
90%-94%	Bom
95%-100%	Excelente

### 3.2.2. Dashboard

Os dados recolhidos são importados para o Microsoft Power BI através de um ficheiro Excel, onde se criaram duas tabelas para auxiliar a criação de filtros (a tabela AUX\_BT\_CUSTOS e a tabela AUX\_BT). O modelo de dados que compõem este projeto está identificado na Figura 3.2.



**Figura 3.2 Modelo de dados**

<sup>1</sup> No nexa A, é apresentado o código de criação das duas variáveis.

Para além das tabelas extraídas pelo fluxo, são criadas três tabelas, a tabela Calendário<sup>2</sup>, a tabela AUX\_BT\_Custo<sup>3</sup> e a tabela AUX\_BT<sup>4</sup>. A tabela AUX\_BT\_Custo é uma tabela auxiliar e o seu objetivo é criar os três botões que servem o gráfico de gastos totais da Figura 4.10. A tabela AUX\_BT é também uma tabela auxiliar e o seu objetivo é também criar três botões para escolher a medida a visualizar nos 2 cartões da página do produto (Figura 4.12). A tabela Calendário<sup>5</sup> é uma tabela auxiliar que contem informações relacionadas com o período escolhido, esta tabela vai ter todas as informações como o ano, o mês, a semana, o dia da semana do período escolhido, neste caso é desde o dia 1/01/2022 até 31/12/2030.

Em seguida, para apresentar algumas visualizações, são definidas algumas fórmulas DAX (métricas), nomeadamente:

- Contar quantas faturas tem uma boa precisão na recolha dos cabeçalhos; % Boa Precisão = `CALCULATE(COUNTROWS('Faturas_Cabeçalhos'), 'Faturas_Cabeçalhos'[Precisão]>=0.8);`
- Contar quantas faturas tem uma boa precisão na recolha dos cabeçalhos ; % Ma Precisão = `CALCULATE(COUNTROWS('Faturas_Cabeçalhos'), 'Faturas_Cabeçalhos'[Precisão]<0.8);`
- Custo a visualizar = `SWITCH(TRUE(), [Escolher custo a visualizar]=1, SUM('Faturas_Cabeçalhos'[Valor s/iva]), [Escolher custo a visualizar]=2, SUM('Faturas_Cabeçalhos'[Valor do Iva]), [Escolher custo a visualizar]=3, SUM('Faturas_Cabeçalhos'[Valor total]));`
- Contar quantas faturas tem uma boa precisão na recolha das linhas; % Precisão Boa TabLinhas = `CALCULATE(COUNTROWS('Faturas_Linhas'), 'Faturas_Linhas'[Precisão Linhas]>=0.8);`
- Qual o percentual de cada produto  
%total=`DIVIDE(SUM('Faturas_Linhas'[Montante]),CALCULATE(SUM('Faturas_Linhas'[Montante]),ALL('Faturas_Linhas')));`
- Max\_preco = `MAX('Faturas_Linhas'[Valor unitário]);`
- Max\_qt = `MAX('Faturas_Linhas'[Quantidade]);`
- Med\_qt = `AVERAGE('Faturas_Linhas'[Quantidade]);`
- MedPreco = `AVERAGE('Faturas_Linhas'[Valor unitário]);`
- Min\_qt = `MIN('Faturas_Linhas'[Quantidade]);`
- MinPreco = `MIN('Faturas_Linhas'[Valor unitário]);`
- MedidaDinamicaQT = `SWITCH(TRUE(), [Escolher medida a visualizar]=1, [Min_qt],[Escolher medida a visualizar]=2,[Med_qt],[Escolher medida a visualizar]=3,[Max_qt]);`
- MedidaDinPreco = `SWITCH(TRUE(), [Escolher medida a visualizar]=1, [MinPreco],[Escolher medida a visualizar]=2,[MedPreco],[Escolher medida a visualizar]=3, [Max_preco], BLANK());`
- Contar quantas faturas tem uma boa precisão na recolha das linhas; % Precisão Ma TabLinhas = `CALCULATE(COUNTROWS('Faturas_Linhas'), 'Faturas_Linhas'[Precisão Linhas]<0.8);`

<sup>2</sup> Ver no anexo B a tabela Calendário.

<sup>3</sup> Ver anexo C, tabela para auxiliar a visualização do gasto a escolher na página dos fornecedores.

<sup>4</sup> Ver anexo D, tabela para auxiliar a visualização da medida a escolher na página do produto.

<sup>5</sup> Ver no anexo E, o código de construção da tabela calendário.

### 3.3. Modelação

#### 3.3.1. Modelação do fluxo

Nesta nova fase são selecionados os modelos e as técnicas a utilizar para atingir o objetivo inicial. Uma vez que este projeto passa pela utilização de ferramentas de inteligência artificial e pela utilização do *low-code*, as técnicas já estão incorporadas no próprio algoritmo da ferramenta. A estrutura do fluxo criado está detalhada ao longo deste subcapítulo.

O fluxo é iniciado quando chega um novo *email* e se detetar o *trigger* que está identificado previamente<sup>6</sup>, nomeadamente, os *triggers* são o endereço de email e o assunto da mensagem (Figura 3.3).

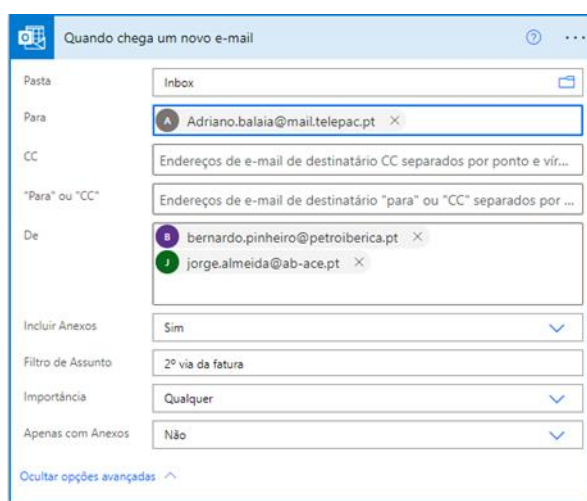


Figura 3.3 Quando chega um novo email

Uma vez que se sabe quem é o fornecedor vai-se desde já criar uma variável com o nome de cada fornecedor (Figura 3.4), pois não é possível ter apenas um único fluxo para todos os fornecedores. De facto, cada fornecedor tem vários *triggers* diferentes, e podendo alguns fornecedores ter diferentes *emails* para a mesma função, torna-se difícil criar um único fluxo capaz de responder a todos eles.

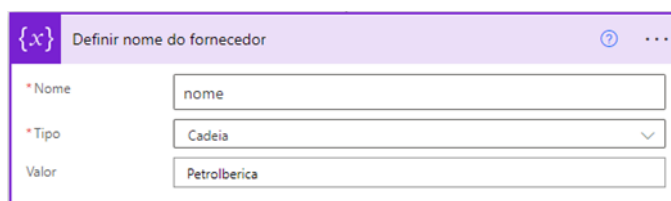


Figura 3.4 Variável para definir fornecedor

O documento que chega via *email* é guardado numa pasta no OneDrive<sup>7</sup>. Em seguida o AI Builder recolhe as informações indicadas para esta etapa do fluxo, sendo, no entanto, utilizado dois modelos diferentes do AI Builder, tal como ilustra a Figura 3.5 e a Figura 3.6.

<sup>6</sup> Ver no anexo F, a lista de *triggers* identificados para iniciar o fluxo.

<sup>7</sup> Ver no anexo G, processo para guardar ficheiro numa pasta do OneDrive.

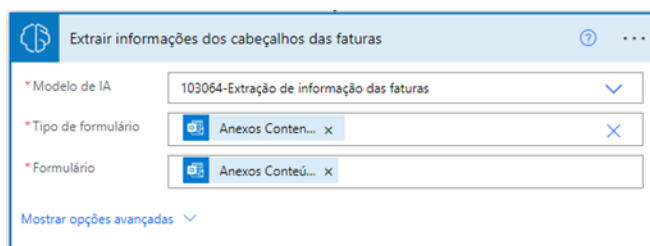


Figura 3.5 Modelo customizado para extrair cabeçalhos das faturas

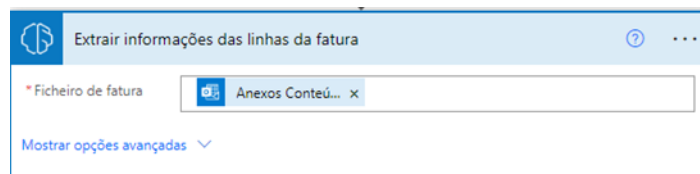


Figura 3.6 Modelo automático para extrair linhas das faturas

O primeiro modelo criado é um modelo customizado, isto é, é criado e treinado manualmente. A criação deste primeiro modelo é explicada pelo facto de o modelo automático não ter uma boa precisão para as variáveis gerais da fatura como, por exemplo, NIF, NºFatura, Valor do Iva, Valor s/Iva. Este modelo é treinado utilizando 52 ficheiros, correspondendo a 56% do total de ficheiros, já que o AI Builder necessita de pelo menos cinco ficheiros de cada *layout*.

O segundo modelo utilizado é um modelo pré-criado de IA de processamento de faturas e extrai dados-chave para ajudar a automatizar o processamento de faturas. Este modelo é utilizado para recolher os dados sobre os itens da fatura, como, por exemplo, a descrição do produto comprado, as quantidades, o valor unitário, o montante final de cada produto. Este modelo apresenta melhores resultados, quando comparado aos do modelo customizado, como é visível mais à frente.

Depois de recolher os dados indicados na Figura 3.7<sup>8</sup>, o fluxo guarda todos os dados recolhidos num ficheiro Excel<sup>9</sup>, que se encontra também no OneDrive.

<sup>8</sup> Ver no anexo H, processo para adicionar os dados das linhas das faturas ao ficheiro Excel.

<sup>9</sup> Ver nos anexos I e J, ficheiro Excel com os dados inseridos.



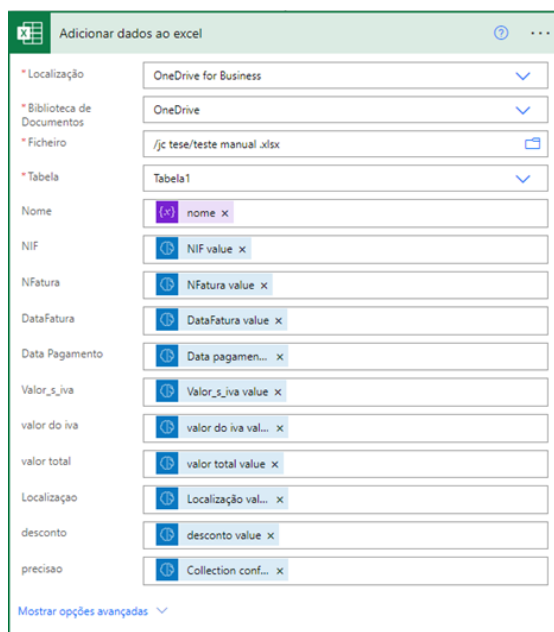


Figura 3.7 Adicionar dados dos cabeçalhos ao ficheiro Excel

Para finalizar, o fluxo tem uma condição para avisar o utilizador do sistema caso a precisão seja inferior a 80%, tal como ilustra a Figura 3.8. Caso este seja inferior, o fluxo guarda esse ficheiro numa pasta reservada para guardar esses mesmos ficheiros<sup>10</sup>, e informa o utilizador para verificar os dados extraídos<sup>11</sup>. Ainda, contempla uma condição para o caso de a data de pagamento estar ultrapassada (Figura 3.9), e, neste caso, envia um aviso para o utilizador<sup>12</sup>.

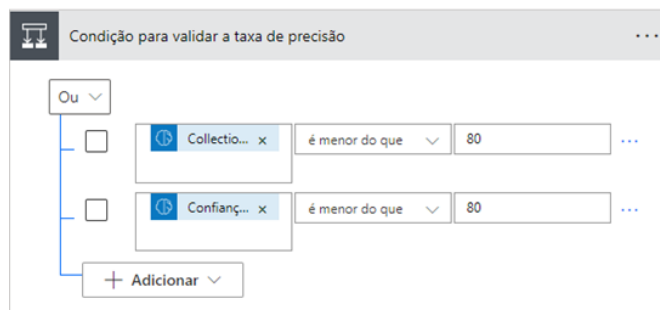


Figura 3.8 Condição para validar a taxa de precisão

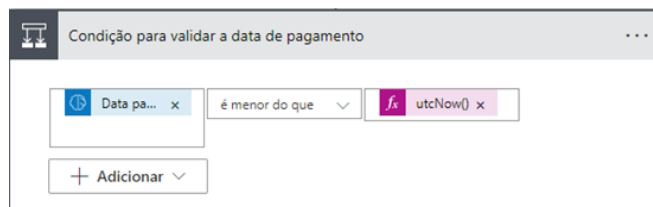


Figura 3.9 Condição para validar data de pagamento

<sup>10</sup> Ver no anexo K, processo para reservar os ficheiros com baixa precisão.

<sup>11</sup> Ver no anexo L, processo para informar o utilizador sobre a baixa precisão de extração.

<sup>12</sup> Ver no anexo M, processo para informar o utilizador sobre a data-limite de pagamento.

Ambos os modelos, o automático e o customizado, utilizam soluções de reconhecimento de texto e soluções de OCR, baseados nos Serviços Cognitivos do Microsoft Azure, para conseguir identificar textos impressos em imagem ou em documentos.

Desta forma a decisão passa por juntar ambos os modelos. Do modelo automático retiram-se os campos relativamente aos itens da fatura e do modelo customizado retiram-se os campos de identificação da fatura, obtendo-se assim o fluxo final (Figura 3.10).

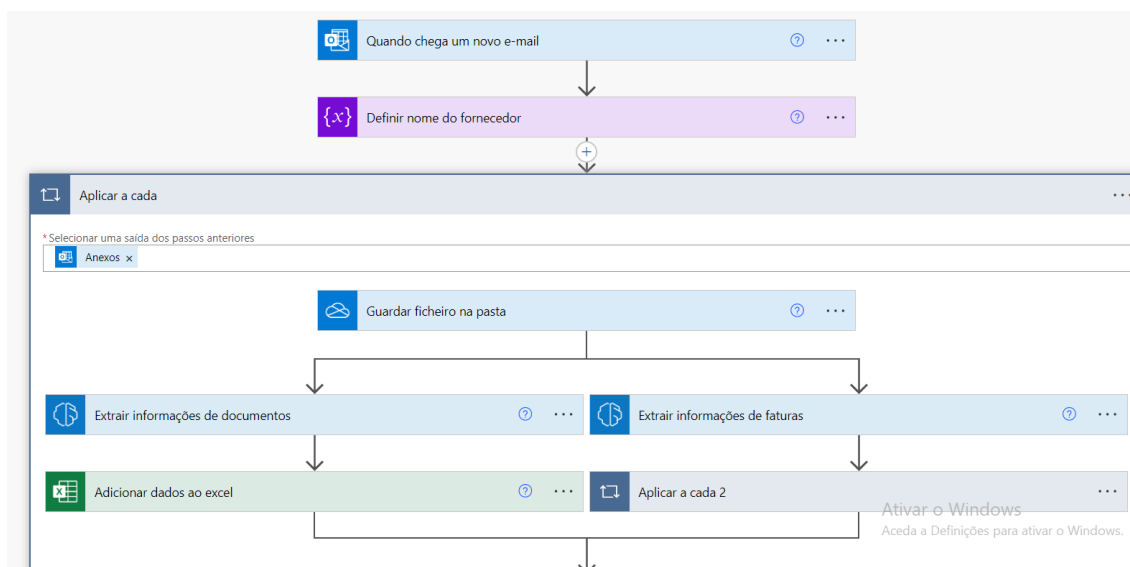


Figura 3.10 Fluxo final

### 3.3.2. Modelação do *dashboard*

Nesta fase é explicada a modelação do *dashboard*. A informação que consta no *dashboard* tem por base os dados recolhidos através do fluxo.

O *dashboard* é composto por 3 páginas<sup>13</sup>, a primeira página é referente aos fornecedores, a segunda página do *dashboard* é referente aos produtos que são comprados, por fim a terceira página diz respeito ao fluxo que foi criado. O *dashboard* segue uma série de regras de boas práticas de visualização, como, por exemplo, a escolha certa de uma *palette* de cores, manter tudo alinhado, ser consistente e manter o *dashboard* simples e de fácil leitura (Russo & Ferrari, 2017).

No que diz respeito à *palette* de cores, para o projeto é escolhida uma paleta de cores composta por cinco cores, onde a principal core da *palette* é o laranja, pois é a cor lo logotipo da empresa<sup>14</sup>.

<sup>13</sup> Ver nos anexos N, O os *mockups* das páginas do *dashboard*

<sup>14</sup> Ver anexo P, *palette* de cores usada

Quanto ao tipo de visualizações, optou-se por gráficos de fácil leitura e de fácil compreensão para os utilizadores, como é o caso dos gráficos de comparação (barras, colunas) e por gráficos de variações no tempo (gráfico de linhas), também se adicionou as hipóteses de *drill down* e *drill up* como é explicada no próximo capítulo.

As funcionalidades de cada página são explicadas no próximo capítulo.

### 3.4. Avaliação

#### 3.4.1. Avaliação do fluxo

Nesta fase, avaliam-se os modelos construídos, o modelo construído é analisado quanto à sua precisão e quanto à satisfação global dos utilizadores.

A precisão (Eq 3.1) do modelo é calculada através do rácio entre o número de registos bem classificados e o número total de registos. Assim, para calcular a precisão deste modelo, é calculada a precisão do modelo customizado e a precisão do modelo automático, usando-se a seguinte fórmula.

$$Precisão = \frac{VP}{VP + FP} \quad (Eq. 3.1)$$

Adicionalmente, a satisfação global dos potenciais utilizadores é medida através de uma serie de três questões e, como referido anteriormente, a satisfação média global tem de ser superior a 7 numa escala de 1 (completamente insatisfeito) a 10 (completamente satisfeito).

- Q1: Numa escala de 1 a 10, como avalia a sua satisfação com a facilidade de utilização deste fluxo?
- Q2: Numa escala de 1 a 10, como avalia a sua satisfação em relação às funcionalidades do fluxo?
- Q3: Numa escala de 1 a 10, como você avalia a sua satisfação geral do fluxo?

#### 3.4.2. Avaliação do *dashboard*

Nesta fase, é avaliado o *dashboard* construído, é analisado quanto à satisfação global por parte dos utilizadores. Para recolher essa satisfação é feito um questionário com quatro perguntas com uma escala de 1 (completamente insatisfeito) a 10 (completamente satisfeito).

- Q1: Numa escala de 1 a 10, como avalia a sua satisfação com a utilidade deste *dashboard*?
- Q2: Numa escala de 1 a 10, como avalia a sua satisfação em relação aos indicadores apresentados neste *dashboard*?
- Q3: Numa escala de 1 a 10, qual a sua satisfação em relação à *palette* de cores utilizada neste *dashboard*?

- Q4: Numa escala de 1 a 10, qual a sua satisfação com o *dashboard* no seu global?

### **3.5. Implementação**

A implementação deste projeto passa pela adoção do mesmo por parte da empresa Adriano Luz Duarte Balaia, ou seja, pretende-se que a empresa em causa aplique o fluxo e o *dashboard* no seu modelo de negócio, de forma a ajudar o gerente da empresa a controlar melhor os gastos da mesma e a ter toda a informação relativamente ao processo de faturação de forma automatizada.

A implementação deste projeto na empresa requer por parte da empresa a adoção de novas ferramentas como o Microsoft Power Automate ou similar, e o Microsoft Power BI ou similar.

Deste modo, a implementação deste projeto traduz-se na disponibilização de protótipos do fluxo e do *dashboard*, para que a empresa possa testar e avaliar durante algum tempo a sua eficiência e eficácia. Adicionalmente, o presente documento é também disponibilizado para que a empresa possa compreender tudo o que foi realizado e a razão para tal.

## 4. Resultados e Discussão

Este capítulo apresenta os resultados, sendo apresentados os resultados do fluxo, bem como os resultados de cada página do *dashboard*, bem como a sua explicação.

### 4.1. Resultado do fluxo

Em primeiro é analisado o modelo construído manualmente, este modelo apresentou uma precisão geral de 90%, como vemos pela Figura 4.1<sup>15</sup>. Como referido acima, este modelo apresentou uma precisão baixa nos atributos das linhas da fatura<sup>16</sup>.

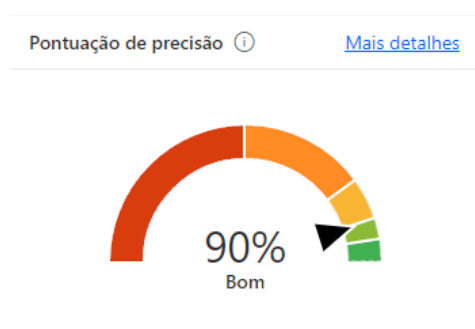


Figura 4.1 Pontuação geral do modelo customizado

Após a avaliação do modelo customizado, é avaliado o modelo automático, este modelo é um modelo pré-criado do AI Builder. Nos atributos gerais da fatura apresentou resultados inferiores ao modelo customizado, com uma precisão de 80% (Figura 4.2). Por outro lado, nos atributos dos itens da fatura este modelo apresentou resultados superiores aos do modelo customizado (Figura 4.3).

total	Precisão
geral	80%
nome	100%
nºfat	79%
data	94%
dataP	90%
nif	44%
valT	87%
val iva	93%
val s/ iva	73%
desconto	0%
localização	64%

Figura 4.2 Pontuação do modelo automático nos cabeçalhos das faturas

<sup>15</sup> Ver anexo Q, pontuação do modelo customizado para cada atributo.

<sup>16</sup> Ver anexo R, precisão do modelo customizado na extração das linhas das faturas

total	Precisão
Tabela 1	87%
cod_prod	74%
descrição	88%
quantidade	90%
unidade	90%
valor_unit	90%
montante	90%

Figura 4.3 Pontuação do modelo automático nas linhas das faturas

Uma vez avaliado o modelo quanto à sua precisão, avalia-se agora quanto à sua satisfação por parte de dois potenciais utilizadores, um gerente (R1) e o responsável pela contabilidade (R2) (Tabela 4.1), que realizaram um pequeno teste de experimentação.

Tabela 4.1 Avaliação sobre a satisfação dos utilizadores

<b>P1</b>	Numa escala de 1 a 10, como avalia a sua satisfação com a facilidade de utilização deste fluxo?
<b>R1</b>	9." Fácil de utilizar e de perceber as suas funcionalidades."
<b>R2</b>	7." Um pouco confuso no início."
<b>P2</b>	Numa escala de 1 a 10, como avalia a sua satisfação em relação às funcionalidades do fluxo?
<b>R1</b>	9. "Atende às minhas necessidades de forma excelente, consigo saber sempre onde estão as faturas de cada fornecedor caso necessite delas. "
<b>R2</b>	9. "O fluxo permitiu-me fazer outras tarefas, em vez de ficar a recolher as faturas do <i>email</i> manualmente, agora é tudo feito automaticamente e assim não perco tempo a fazer esse processo."
<b>P3</b>	Numa escala de 1 a 10, como você avalia a sua satisfação geral do fluxo?
<b>R1</b>	8." Muito satisfeito, deixa o trabalho bem mais facilitado"
<b>R2</b>	8. "Torna o processo de faturação mais rápido e não tenho de me preocupar com as tarefas demoradas e entediantes."

Assim, através das respostas dadas pelos utilizadores do fluxo (o gerente e o responsável pela contabilidade), conclui-se que o fluxo é útil e traz mais valor para a empresa. Este fluxo tem uma satisfação média global de 8 pontos, o que revela o sucesso desta parte do projeto. Esta média é calculada através da pontuação média das perguntas.

## 4.2. Resultado do *dashboard*

Em primeiro é analisada a satisfação dos potenciais utilizadores do *dashboard*, um gerente (R1) e o responsável pela contabilidade (R2) (Tabela 4.2), que realizaram um pequeno teste de experimentação.

Tabela 4.2 Avaliação da satisfação global dos utilizadores do *dashboard*

P1	Numa escala de 1 a 10, como avalia a sua satisfação com a utilidade deste <i>dashboard</i> ?
R1	9." Extremamente satisfeito, num pequeno ecrã consigo perceber onde estão os gastos da empresa."
R2	9." Muito satisfeito, é um ótimo aliado para a empresa, com ele conseguimos analisar melhor a nossa situação."
P2	Numa escala de 1 a 10, como avalia a sua satisfação em relação aos indicadores apresentados neste <i>dashboard</i> ?
R1	9." Estou extremamente satisfeito com os indicadores, são indicadores importantes para o obter respostas aos problemas."
R2	7." Satisfeito, mas podia ter mais alguns indicadores, como previsões de pagamentos, previsões de preços. No seu global corresponde ao pedido solicitado."
P3	Numa escala de 1 a 10, qual a sua satisfação em relação à paleta de cores utilizada neste <i>dashboard</i> ?
R1	6." As cores em algumas visualizações eram iguais ou muito parecidas, mas no seu global dava para perceber as diferenças entre elas."
R2	8." Muito satisfeito com as cores, dá para fazer uma boa distinção das cores."
P4	Numa escala de 1 a 10, qual a sua satisfação com o <i>dashboard</i> no seu global?
R1	8." Muito satisfeito, é fácil de trabalhar com ele, de interagir com ele e de entender os gráficos."
R2	9." Muito satisfeito, é uma ferramenta de análise de dados útil, apesar de ter algumas dificuldades com as tecnologias, é fácil perceber as diferentes análises."

Assim, através das respostas dadas pelos utilizadores do *dashboard* (o gerente e o responsável pela contabilidade), conclui-se que o *dashboard* é intuitivo e útil para os potenciais utilizadores. Este *dashboard* tem uma satisfação média de 8 pontos, o que revela o sucesso desta parte do projeto.

### 4.2.1. Resultado e explicação da página dos fornecedores

Os dados que são usados principalmente para criar a primeira página do *dashboard*, que é uma página focada nos gastos com os fornecedores, são os da tabela Faturas Cabeçalhos bem como as fórmulas DAX que se encontram na secção 3.2.2 do capítulo anterior.

Como resultado, o *dashboard* de controlo dos gastos por fornecedor, apresentado na Figura 4.4, é criado.

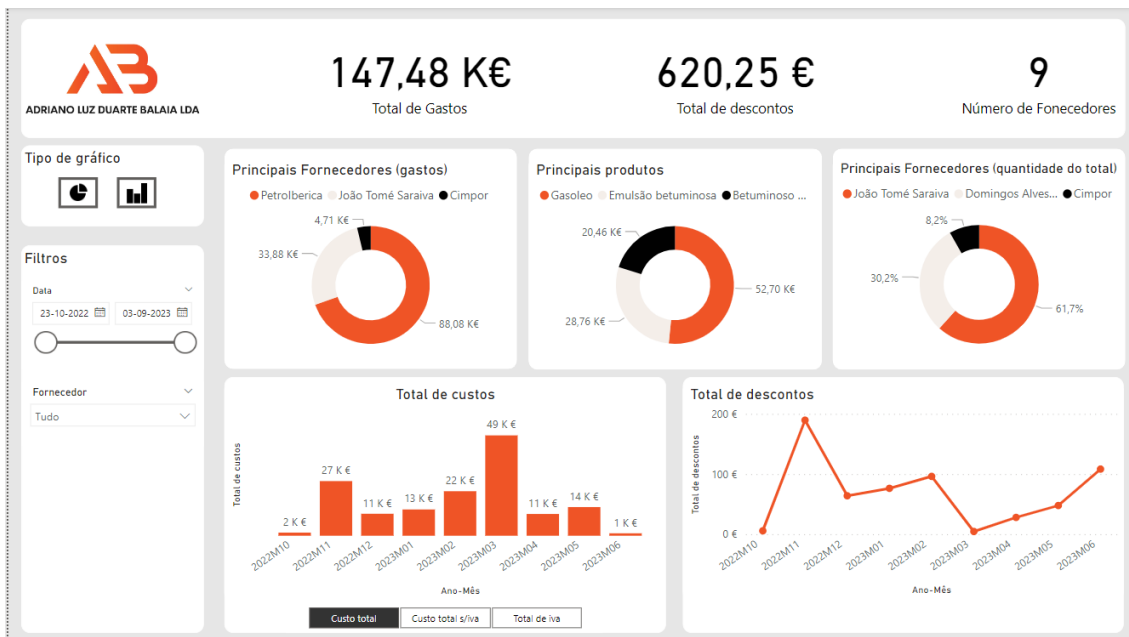


Figura 4.4 Dashboard página dos fornecedores

A proposta desta página é composta por vários elementos visuais, assim, esta página começa por mostra alguns indicadores pertinentes para a empresa (Figura 4.5), tais como o total de gastos, o total de descontos que a empresa obteve e o número de fornecedores durante o período selecionado na linha cronológica. Estes cartões mostram ao gerente da empresa uma visão mais global da situação da mesma.



Figura 4.5 Gasto total, total de descontos e nº de fornecedores

Em seguida a página apresenta os três principais fornecedores, isto é, permite ver quais são os fornecedores com que a empresa mais gasta, apresenta os três principais produtos que a empresa compra e apresenta os três principais fornecedores da empresa em termos de quantidades compradas.

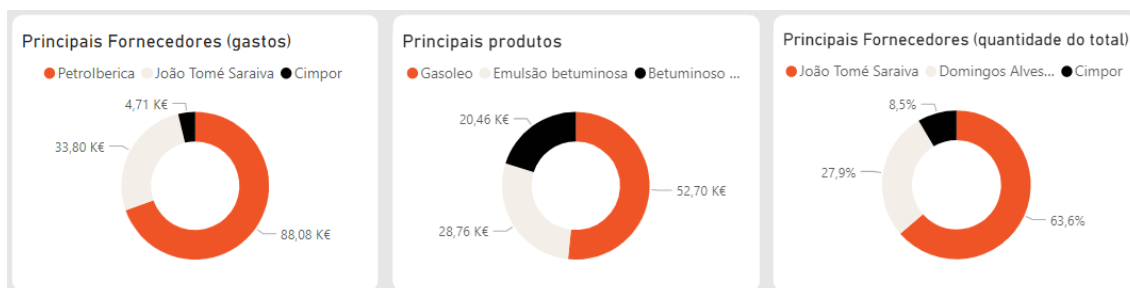


Figura 4.6 Gráficos referentes aos principais fornecedores e principais produtos



Esta análise pode ser alterada através dos botões da Figura 4.7, que permite alterar de gráficos em anel para gráficos de colunas, de forma a obter um conhecimento menos resumido em relação aos indicadores apresentados acima.



Figura 4.7 Botões para alterar forma do gráfico

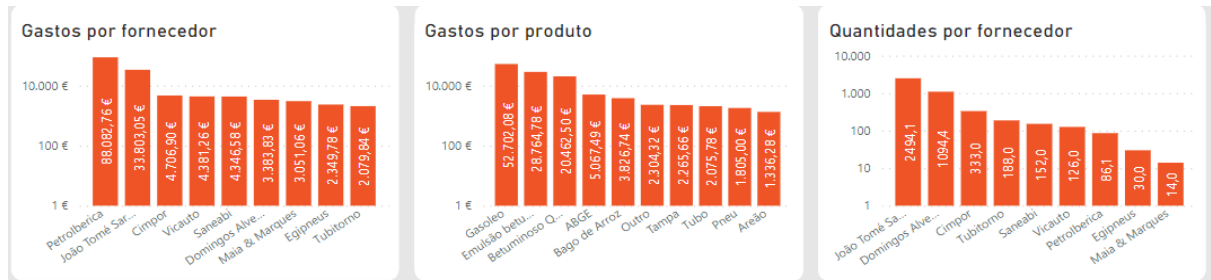


Figura 4.8 Gastos e quantidades por fornecedores e gastos por produto

Por último esta página tem dois gráficos, Figura 4.9 e Figura 4.10, de forma a detalhar em pormenor as informações da Figura 4.5, assim é possível ver o total de gastos e o total de descontos ao longo dos meses.

Através das funcionalidades do *drill down* e do *drill up* o gerente pode agrupar os dados por ano e por mês.



Figura 4.9 Total de descontos por ano e por mês

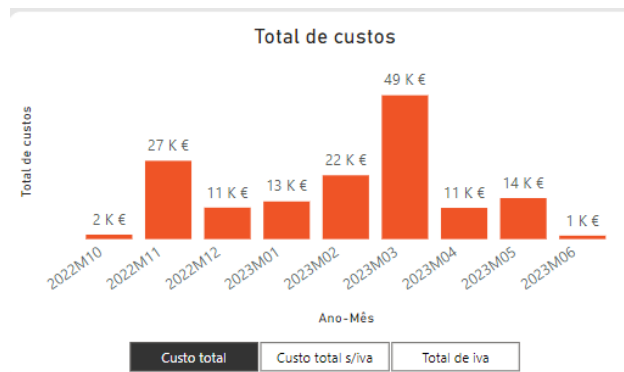


Figura 4.10 Gasto total por ano e por mês

O gráfico do total de gastos apresenta ainda três botões para escolher qual gasto o gerente quer visualizar, pode assim escolher entre o gasto total (Figura 4.10), o gasto total sem o Iva<sup>17</sup> e o gasto total do Iva<sup>18</sup>.

Esta página de controlo de fornecedores conta também com dois filtros<sup>19</sup>, o primeiro filtro é a linha cronológica que permite ao gestor ver a informação referente ao período selecionado e conta com um filtro para filtrar a informação por fornecedor.

#### 4.2.2. Resultado e explicação da página dos produtos

Para a elaboração da 2ª página do *dashboard*, que se trata de uma página focada nos produtos consumidos por parte da empresa, são considerados principalmente os dados que se encontram na tabela Faturas Linhas bem como as fórmulas DAX que se encontram na secção 3.2.2 do capítulo anterior.

Desta forma resultou o *dashboard* para controlo dos produtos, que é apresentado na Figura 4.11.

<sup>17</sup> Anexo S, gasto total sem o Iva.

<sup>18</sup> Anexo T, gasto total do Iva.

<sup>19</sup> Anexo U, filtros da página dos fornecedores.

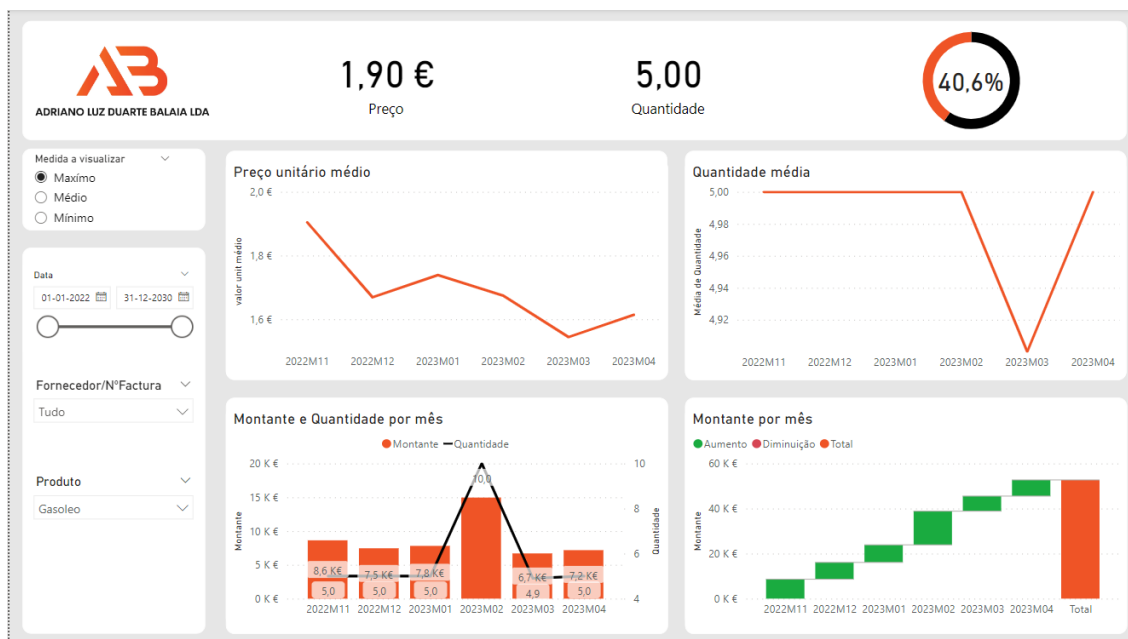


Figura 4.11 Dashboard página dos produtos consumidos

A proposta desta página é composta por vários elementos visuais, desta forma a página começa por mostrar alguns indicadores sobre os produtos consumidos (Figura 4.12), tais como o preço e a quantidade média, máxima ou mínima, durante o tempo que indicar na barra cronológica. Mostra também o percentual que o produto escolhido representa no gasto total da empresa. O gerente pode escolher qual a medida a visualizar através dos três botões criados para cada medida<sup>20</sup>.

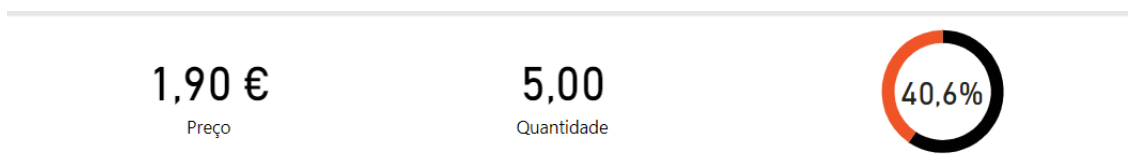


Figura 4.12 Preço, quantidade e contributo percentual no gasto total

Em seguida a página apresenta dois gráficos para o gerente, um gráfico com o preço unitário médio do produto escolhido (Figura 4.13) e um gráfico que apresenta a quantidade média encomendada do mesmo produto (Figura 4.14). Estes dois gráficos fornecem informações importantes para o gerente da empresa, pois consegue ver o preço médio do produto, ver em que alturas do ano o produto se tornou mais caro ou mais barato e consegue perceber também a quantidade média que consome do produto.

<sup>20</sup> Anexo V, botão para escolher a medida a visualizar.

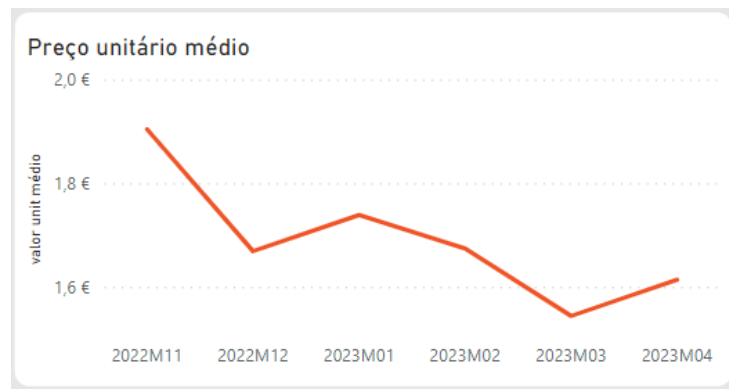


Figura 4.13 Preço unitário médio

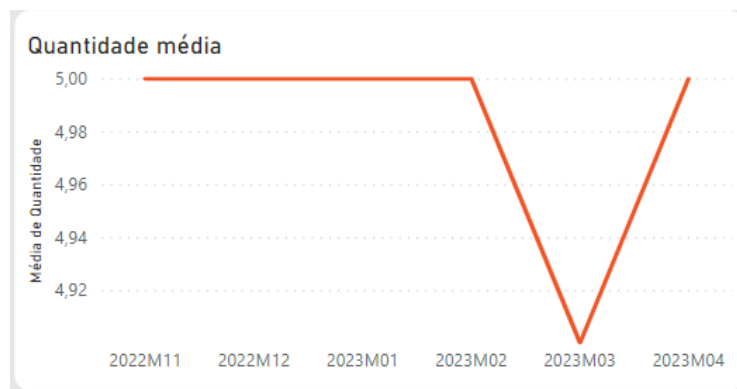


Figura 4.14 Quantidade média

Nesta página é ainda possível ver um gráfico com o preço e a quantidade por mês (Figura 4.15), desta forma e ao contrário dos gráficos anteriores é possível visualizar quanto foi gasto por mês para adquirir o produto e a quantidade total encomendada por mês do mesmo produto.

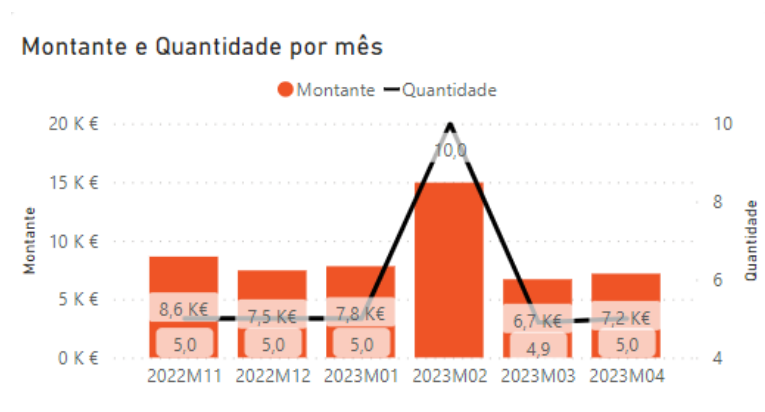


Figura 4.15 Valor total e quantidade total por mês

Por último esta página apresenta um gráfico em cascata (Figura 4.16), que permite ao gestor ver como o valor inicial gasto evolui ao longo do período do tempo até atingir o seu gasto final. Este gráfico é bastante útil para perceber as variações que existiram ao longo do tempo, além de ser um gráfico de leitura e de visualização acessível.

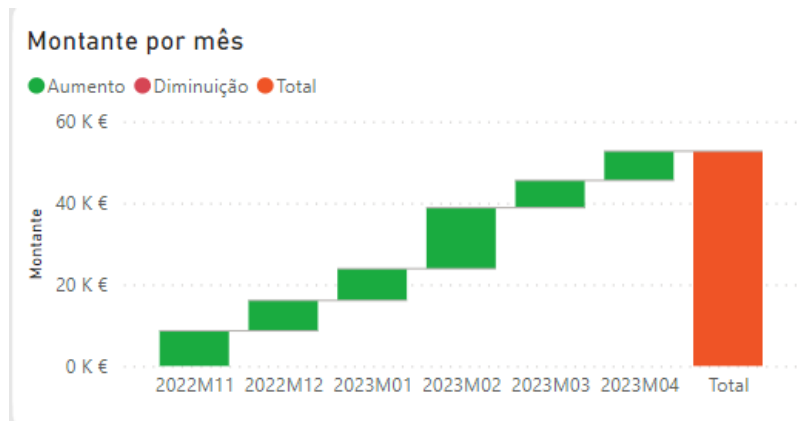


Figura 4.16 Gráfico cascata montante gasto por mês

Esta página pode ainda ser filtrada por um conjunto de três filtros dinâmicos<sup>21</sup>, o primeiro filtro é a linha cronológica, que permite ao gerente escolher qual o período a visualizar, em seguida tem um filtro para o gerente filtrar pelo nome do fornecedor ou se preferir pode pesquisar pelo número da fatura<sup>22</sup>, por último o gerente pode também filtrar pelo produto<sup>23</sup>.

#### 4.2.3. Resultado e explicação da página de controlo

Para a elaboração da última página deste *dashboard*, que se trata de uma página focada no controlo do fluxo que foi criado no capítulo anterior, são considerados os dados que se encontram na tabela Faturas Cabeçalhos e na tabela Faturas Linhas, bem como as fórmulas DAX que se encontram na secção 3.2.2 do capítulo anterior.

Desta forma resultou o *dashboard* de controlo do fluxo, que é apresentado na Figura 4.17.

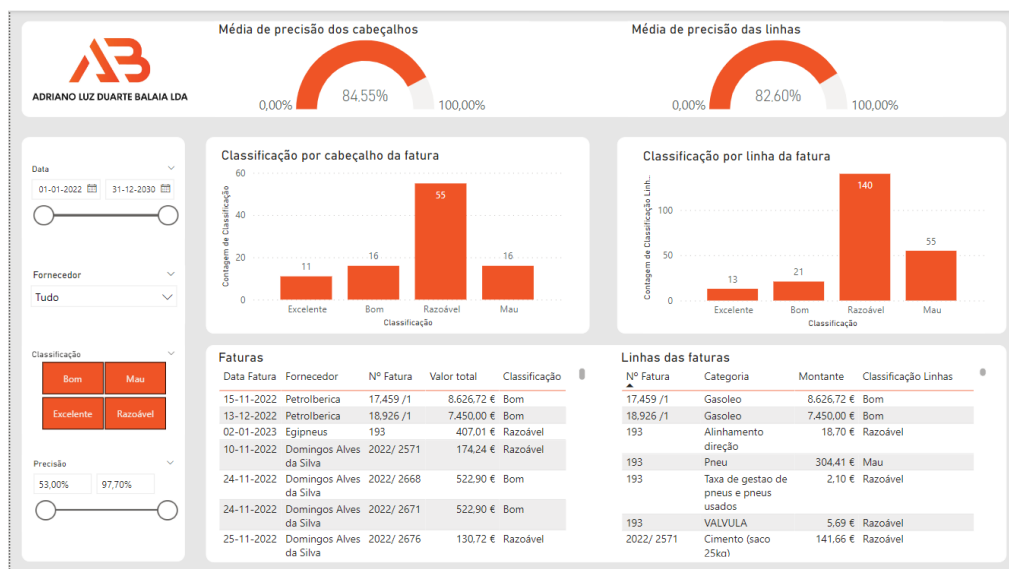


Figura 4.17 Dashboard de controlo de fluxo

<sup>21</sup> Anexo W, filtros da página do produto.

<sup>22</sup> Anexo X, este filtro deixa o utilizador filtrar por nome do fornecedor ou pelo número da fatura associada ao fornecedor.

<sup>23</sup> Anexo Y, este filtro permite filtrar por produto.

Esta página é composta por vários elementos visuais, que servem para dar ao gerente uma visão de como o fluxo está a funcionar, assim, no início da página são apresentados dois gráficos de medição (Figura 4.18), o primeiro gráfico controla a média de precisão do fluxo ao extrair os elementos dos cabeçalhos das faturas, já o segundo gráfico serve para controlar a média de precisão do fluxo ao extrair as linhas das faturas.

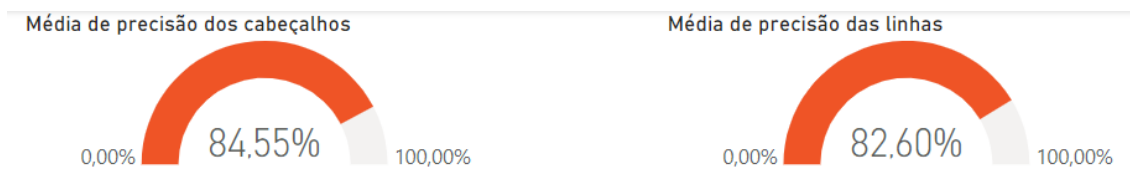


Figura 4.18 Média de precisão na extração dos cabeçalhos e das linhas

Em seguida a página mostra dois gráficos de colunas para ver o número total de extrações por categoria (Figura 4.19), isto é, cada extração está categorizada conforme a Tabela 3.3, de forma a perceber melhor como está o fluxo a funcionar.

Estes dois gráficos fornecem *insights* valiosos para o gerente, como, por exemplo, é possível ver que a maior parte dos ficheiros se encontra entre as categorias Razoável, Bom e Excelente.

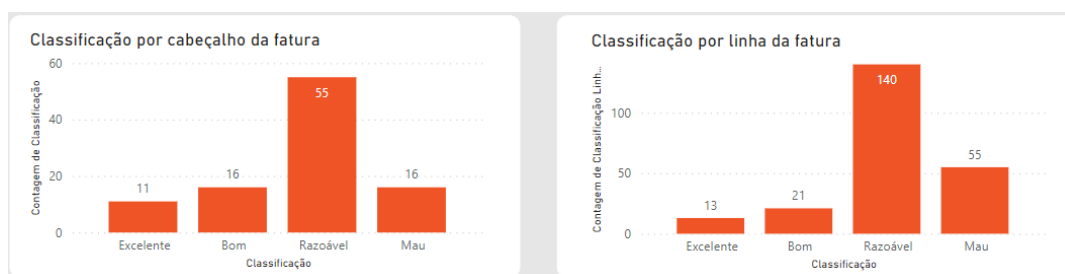


Figura 4.19 Número total de extrações por categoria

Para finalizar os elementos visuais, esta página fornece também duas tabelas com as informações recolhidas (Figura 4.20), a primeira tabela está organizada por data de emissão da fatura. Esta tabela permite ao gestor ver os principais elementos de cada fatura numa só tabela. A segunda é uma tabela de controlo das linhas das faturas.

Faturas					Linhas das faturas			
Data Fatura	Fornecedor	Nº Fatura	Valor total	Classificação	Nº Fatura	Categoria	Montante	Classificação Linhas
15-11-2022	Petrolberica	17,459 /1	8.626,72 €	Bom	17,459 /1	Gasoleo	8.626,72 €	Bom
13-12-2022	Petrolberica	18,926 /1	7.450,00 €	Bom	18,926 /1	Gasoleo	7.450,00 €	Bom
02-01-2023	Egipneus	193	407,01 €	Razoável	193	Alinhamento direção	18,70 €	Razoável
10-11-2022	Domingos Alves da Silva	2022/ 2571	174,24 €	Razoável	193	Pneu	304,41 €	Mau
24-11-2022	Domingos Alves da Silva	2022/ 2668	522,90 €	Bom	193	Taxa de gestao de pneus e pneus usados	2,10 €	Razoável
24-11-2022	Domingos Alves da Silva	2022/ 2671	522,90 €	Bom	193	VALVULA	5,69 €	Razoável
25-11-2022	Domingos Alves da Silva	2022/ 2676	130,72 €	Razoável	2022/ 2571	Cimento (saco 25ko)	141,66 €	Razoável

Figura 4.20 Tabelas de informações sobre as faturas

Nesta página também é possível encontrar um conjunto de filtros<sup>24</sup>, de forma a dinamizar a informação a visualizar. O primeiro filtro é a linha cronológica. Em seguida, há a hipótese de filtrar por fornecedor, assim é possível ver apenas os dados relativos ao fornecedor indicado.

O terceiro filtro é para filtrar por categoria de precisão, ao contrário do filtro abaixo, ele exhibe os dados pertencentes à categoria selecionada.

Por último, há um filtro que permite filtrar por valor de precisão. Com este filtro é possível visualizar apenas os dados que se encontrem dentro do intervalo de precisão especificado.

---

<sup>24</sup> Anexo Z, filtros para a página de controlo do fluxo.





## 5. Conclusão, limitações e pistas futuras

### 5.1. Sumário

O tema da automatização de processos é cada vez mais relevante, visando este projeto responder à questão como automatizar o processo de registo de faturas e controlo de gestão numa empresa de construção civil.

Para o efeito, definem-se como objetivos, primeiro, criar um fluxo para recolher, ler e extrair informações de forma automática das faturas dos fornecedores e, segundo, elaborar um *dashboard*, de forma a auxiliar o gestor da empresa a perceber melhor onde estão alocados os gastos da mesma.

Os resultados permitem responder à questão de investigação, evidenciando que as ferramentas da Microsoft utilizadas permitem automatizar os processos. Primeiro, o fluxo com uma complexidade média, permite automatizar o processo, na medida que guarda e extrai as informações das faturas que cheguem por *email* de forma automática, o fluxo tem uma precisão de 90% no processo de extração. Depois, o *dashboard* dinâmico com três ecrãs, o primeiro ecrã para controlar os gastos com os fornecedores, outro para controlar os produtos e por último um ecrã para controlar o fluxo, com a possibilidade de *drill-down* e *drill-up* nos primeiros dois ecrãs e com a possibilidade de aplicação de filtros que permite uma boa interatividade.

Por fim, ambas as ferramentas apresentaram uma satisfação por parte dos potenciais utilizadores elevada o que traduz o sucesso deste projeto.

### 5.2. Contributos

Este projeto foi planeado a pensar na situação atual da empresa e, tinha como finalidade ajudar a empresa a transformar as suas rotinas e procedimentos, de forma a melhorar a eficiência e a eficácia da mesma e de forma a colocar a empresa numa situação mais favorável em relação aos seus competidores. Este projeto permitiu avançar com a transformação digital de alguns setores da empresa.

Com a conclusão do mesmo, foi possível identificar potenciais vantagens que as tecnologias usadas neste projeto trazem para as empresas, sendo as mais significantes para a empresa a redução de gastos, isto é, a empresa ao automatizar o processo de faturação dos fornecedores, deixou de gastar dinheiro em páginas de papel para guardar toda as faturas e informações. A outra vantagem para a empresa é a libertação do colaborador para outras funções, o tempo que é gasto a realizar este processo, deixa de ser um problema para a empresa.

Além disto, a criação do *dashboard* permite ao gerente ter uma ideia mais clara sobre os gastos da mesma, de forma a perceber melhor onde a empresa gasta os seus fundos. Concluo, que a utilização desta ferramenta por parte da empresa é uma mais-valia, pois é capaz de fornecer informações valiosas de forma simples e clara. Apesar de este projeto ainda ser uma fase inicial, tem o potencial de evoluir, pois é a base de uma ferramenta para a construção de um modelo, que servirá quando a empresa tiver o processamento de faturas totalmente digitalizado.

Por fim, este trabalho é um bom contributo para a comunidade científica e para o mundo empresarial, pois é um trabalho que para a área científica apresenta uma metodologia que conjuga IA com *low code* e que é aplicada a situações do mundo real. Para o mundo empresarial, este trabalho deixa o contributo de fornecer um guia de aplicação de tecnologias de IA no mundo empresarial, neste caso no setor da construção civil.

### **5.3. Limitações**

Este projeto contou com algumas limitações quer a nível dos dados, quer a nível das ferramentas. A nível dos dados a maior limitação que existiu foi a limitação de dados, isto é, para se compreender melhor este projeto era necessário um maior conjunto de dados, ao qual não foi possível aceder. Outra limitação que existiu sobre os dados recolhidos, é o facto de este projeto trabalhar parcialmente com os dados da empresa, isto é, apenas foram considerados para este trabalho os dados que chegaram via *email*, assim foram excluídos todos os dados no formato físico. A empresa visada no projeto está a trabalhar para reduzir cada vez mais o número de documentos no formato físico, de forma a eliminar esta limitação.

A nível das ferramentas, a maior limitação foi o tempo, isto é, o Power Automate e a funcionalidade do AI Builder funcionavam através da licença experimental, na qual o período máximo de utilização era de três meses. Esta foi uma grande limitação, pois a parte de construção, avaliação do fluxo e extração dos dados teve de ser planeada e feita de forma a não ultrapassar esse período.

Outra limitação que este estudo teve foi o facto de as conclusões poderem ser influenciadas pela relação privilegiada com o órgão de gestão da empresa, podendo assim influenciar as conclusões que o mesmo apresenta em relação à satisfação global do fluxo e do *dashboard*.

#### **5.4. Trabalho futuro**

Esta investigação sem dúvida que merece um trabalho futuro, quem sabe se a curto prazo ou a longo prazo, mas fica a vontade de trabalhar alguns aspetos que não foram abordados. Assim ficam aqui alguns tópicos a considerar num eventual futuro trabalho:

- Pesquisar outros artigos e estudos sobre a conjugação de ferramentas de *low code* com algoritmos de IA
- Aumentar a base de dados e o número de fornecedores, para ver o impacto nos resultados
- Criar um fluxo que consiga responder a todos os fornecedores, isto é, não criar um fluxo específico para cada fornecedor, aumentando assim a capacidade de generalização do fluxo
- Melhorar o modelo de extração dados, aumentando o número de exemplares de cada *layout*, de forma a melhorar a precisão de extração do modelo.
- Evolução do *dashboard*, melhorar alguns indicadores e alguns gráficos do *dashboard*.

Ainda como trabalho futuro sugere-se que se crie um fluxo que esteja ligado à plataforma de faturação da empresa, de forma a recolher e extrair dados sobre os clientes, de forma a construir um *dashboard* mais global sobre a empresa.



## Referencias

- Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. (2021). Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions. *IEEE Access*, 9, 72894–72936. <https://doi.org/10.1109/ACCESS.2021.3072900>
- Bilski, A. (2011). A review of artificial intelligence algorithms in document classification. *International Journal of Electronics and Telecommunications*, 57(3), Artículo 3. Scopus. <https://doi.org/10.2478/v10177-011-0035-6>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). CRISP-DM 1.0 step-by-step data mining guide. Em *Springer*.
- Chazhour, A., & Sarobin, V. R. (2022). Intelligent automation of invoice parsing using computer vision techniques. *Multimedia Tools and Applications*, 81(20), Artículo 20. <https://doi.org/10.1007/s11042-022-12916-x>
- Chugh, R., Macht, S., & Hossain, R. (2022). Robotic Process Automation: A review of organizational grey literature. *International Journal of Information Systems and Project Management*, 10(1), Artículo 1. <https://doi.org/10.12821/ijispm100101>
- Eikebrokk, T. R., & Olsen, D. H. (2020). Robotic Process Automation and Consequences for Knowledge Workers; a Mixed-Method Study. Em M. Hattingh, M. Matthee, H. Smuts, I. Pappas, Y. K. Dwivedi, & M. Mäntymäki (Eds.), *Responsible Design, Implementation and Use of Information and Communication Technology* (pp. 114–125). Springer International Publishing. [https://doi.org/10.1007/978-3-030-44999-5\\_10](https://doi.org/10.1007/978-3-030-44999-5_10)
- Esposito, F., Ferilli, S., Basile, T. M. A., & Di Mauro, N. (2005). Intelligent document processing. *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, 1100-1104 Vol. 2. <https://doi.org/10.1109/ICDAR.2005.144>

- Esposito, F., Malerba, D., & Lisi, F. A. (2000). Machine Learning for Intelligent Processing of Printed Documents. *Journal of Intelligent Information Systems*, 14(2), 175–198. <https://doi.org/10.1023/A:1008735902918>
- Ha, H. T., & Horák, A. (2022). Information extraction from scanned invoice images using text analysis and layout features. *Signal Processing: Image Communication*, 102. Scopus. <https://doi.org/10.1016/j.image.2021.116601>
- Herm, L.-V., Janiesch, C., Helm, A., Imgrund, F., Hofmann, A., & Winkelmann, A. (2022). A framework for implementing robotic process automation projects. *Information Systems and E-Business Management*, 21(1), 1–35. <https://doi.org/10.1007/s10257-022-00553-8>
- IBM Documentation*. (2021, junho 17). <https://www.ibm.com/docs/en/cloud-paks/cp-biz-automation/20.0.x?topic=processing-what-is-automation-document>
- Jadli, A., & Hain, M. (2020). *Automatic Document Classification using Deep Feature Selection and Knowledge Transfer*. 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2020. Scopus. <https://doi.org/10.1109/IRASET48871.2020.9092256>
- Khare, A., Singh, S., Mishra, R., Prakash, S., & Dixit, P. (2022). E-Mail Assistant—Automation of E-Mail Handling and Management using Robotic Process Automation. *2022 International Conference on Decision Aid Sciences and Applications (Dasa)*, 511–516. <https://doi.org/10.1109/DASA54658.2022.9765017>
- Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. *Keele, UK, Keele Univ.*, 33.
- Leno, V., Polyvyanyy, A., Dumas, M., La Rosa, M., & Maggi, F. M. (2021). Robotic Process Mining: Vision and Challenges. *Business & Information Systems Engineering*, 63(3), Artigo 3. <https://doi.org/10.1007/s12599-020-00641-4>
- Lievano-Martínez, F. A., Fernández-Ledesma, J. D., Burgos, D., Branch-Bedoya, J. W., & Jimenez-Builes, J. A. (2022). Intelligent Process Automation: An Application

- in Manufacturing Industry. *Sustainability (Switzerland)*, 14(14). Scopus.  
<https://doi.org/10.3390/su14148804>
- Ling, X., Gao, M., & Wang, D. (2020a). Intelligent document processing based on RPA and machine learning. *2020 Chinese Automation Congress (CAC)*, 1349–1353.  
<https://doi.org/10.1109/CAC51589.2020.9326579>
- Manjunath, A. A., Nayak, M. S., Nishith, S., Pandit, S. N., Sunkad, S., Deenadhayalan, P., & Gangadhara, S. (2022). Automated invoice data extraction using image processing. *IAES International Journal of Artificial Intelligence*, 12(2), Artigo 2. Scopus. <https://doi.org/10.11591/ijai.v12.i2.pp514-521>
- Melchert, F., Klesse, M., & Winter, R. (2004). *Aligning Process Automation and Business Intelligence to Support Corporate Performance Management*. 4053–4063. Scopus.
- Nicolaieff, L., Kandi, M. M., Zegaoui, Y., & Bortolaso, C. (2022). Intelligent Document Processing with Small and Relevant Training Dataset. *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 1–7.  
<https://doi.org/10.1109/ISCV54655.2022.9806100>
- Extract information from any business document automatically to enable end-to-end business process automation| Automation Anywhere.* (nd).  
<https://www.automationanywhere.com/rpa/intelligent-document-processing>
- Power Automate.* (nd).<https://powerautomate.microsoft.com/pt-pt/guidedtour/power-platform/power-automate/5/1>
- Russo, M., & Ferrari, A. (2017). *Power BI visualization best practices*.  
<https://www.sqlbi.com/p/power-bi-dashboard-design-course/>
- Satav, M. S., Varade, T., Kothavale, D., Thombare, S., & Lokhande, P. (2020). Data Extraction From Invoices Using Computer Vision. *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, 316–320.  
<https://doi.org/10.1109/ICIIS51140.2020.9342722>

- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Silva, A. J., Cortez, P., Pereira, C., & Pilastrri, A. (2021). *Business analytics in industry 4.0: A systematic review*. <https://doi.org/10.1111/exsy.12741>
- Tian, J., Li, L., Tian, J., & Li, L. (2022). Research on artificial intelligence of accounting information processing based on image processing. *Mathematical Biosciences and Engineering*, 19(8), Artigo 8. <https://doi.org/10.3934/mbe.2022391>
- V-aangie, Alexbuckgit, Vivek, K., & Cellier, A. (2023). *AI Builder Overview*. <https://learn.microsoft.com/pt-pt/ai-builder/overview>
- Yindumathi, K. M., Chaudhari, S. S., & Aparna, R. (2020). Analysis of Image Classification for Text Extraction from Bills and Invoices. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–6. <https://doi.org/10.1109/ICCCNT49239.2020.9225564>
- Zhang, N., & Liu, B. (2019). Alignment of business in robotic process automation. *International Journal of Crowd Science*, 3(1), Artigo 1. <https://doi.org/10.1108/IJCS-09-2018-0018>



# Anexo

```
= Table.AddColumn("#Colunas com Nome Mudado", "Classificação", each if [Precisão] >= 0.95 then "Excelente" else if [Precisão] >= 0.9 then "Bom" else if [Precisão] >= 0.8 then "Razoável" else "Mau")
```

## Anexo A Criação da coluna classificação

1 DIM\_CALENDÁRIO =

Data	Ano	N_Mês	N_Semana	Dia	N_Dia_Semana	Ano-Mês	Mês	Dia_Semana
01/01/2024	2024	1	1	1	2	2024M01	JAN	SEG
02/01/2024	2024	1	1	2	3	2024M01	JAN	TER
03/01/2024	2024	1	1	3	4	2024M01	JAN	QUA
04/01/2024	2024	1	1	4	5	2024M01	JAN	QUI
05/01/2024	2024	1	1	5	6	2024M01	JAN	SEX
06/01/2024	2024	1	1	6	7	2024M01	JAN	SÁB
07/01/2024	2024	1	2	7	1	2024M01	JAN	DOM
08/01/2024	2024	1	2	8	2	2024M01	JAN	SEG
09/01/2024	2024	1	2	9	3	2024M01	JAN	TER
10/01/2024	2024	1	2	10	4	2024M01	JAN	QUA
11/01/2024	2024	1	2	11	5	2024M01	JAN	QUI
12/01/2024	2024	1	2	12	6	2024M01	JAN	SEX
13/01/2024	2024	1	2	13	7	2024M01	JAN	SÁB
14/01/2024	2024	1	3	14	1	2024M01	JAN	DOM
15/01/2024	2024	1	3	15	2	2024M01	JAN	SEG

## Anexo B Tabela Calendário

índice	bt
1	Custo total s/iva
2	Total de iva
3	Custo total

## Anexo C Tabela para auxiliar visualização (tabela AUX\_BT\_Custo)

Índice	BT
1	Mínimo
2	Médio
3	Maxímo

## Anexo D Tabela para auxiliar visualização (tabela AUX\_BT)

```

1 DIM_CALENDÁRIO =
2
3 VAR MINDATE = "01/01/2022"
4 VAR MAXDATE = "31/12/2030"
5 RETURN
6
7 ADDCOLUMNS(
8
9     CALENDAR(MINDATE,MAXDATE),
10    "Ano", YEAR([Date]),
11    "N_Mês", MONTH([Date]),
12    "N_Semana", WEEKNUM([Date]),
13    "Dia", DAY([Date]),
14    "N_Dia_Semana", WEEKDAY([Date]),
15    "Ano-Mês", YEAR([Date])&"M"&RIGHT("0"&MONTH([Date]),2),
16    "Mês", UPPER(FORMAT([Date],"MMM")),
17    "Dia_Semana", UPPER(FORMAT([Date],"DDD"))
18 )

```

#### Anexo E Código de criação da tabela Calendário

Nome	Email	Assunto
<b>CIMPOR</b>	AMFLopes@cimpor.com	cópia da fatura VD-IBAN-NIB-AGREPOR Agregados SA
<b>Tubitorno</b>	tubitorno@gmail.com service@sagecloud.email	fatura_linde; ou fact tube/...
<b>Saneabi</b>	saneabi@hotmail.com	fatura a crédito
<b>Domingos Alves da Silva</b>	geral@doalsil.com	2º via ou cópia da fatura
<b>EGIPNEUS</b>	geral@egipneus.com	faturas
<b>João Tomé Saraiva</b>	geral@joaotomesaraiva.pt	cópia da fatura nº da fatura
<b>Maia e Marques</b>	maiamarques@maiamarques.p ou tiago@maiamarques.net	envio da fatura nº da fatura
<b>Petroiberica</b>	jorge.almeida@ab-ace.pt ou <b>bernardo.pinheiro@petroiberica.pt</b>	2º via da fatura já emitida
<b>Vicauto</b>	rui.almeida@vicauto.pt geral@vicauto.pt carlos.sousa@vicauto.pt	Fatura

#### Anexo F Tabela com os triggers

\* Seleccionar uma saída dos passos anteriores

Anexos x

Guardar ficheiro na pasta

\* Caminho da Pasta

\* Nome do Ficheiro

\* Conteúdo do Ficheiro

#### Anexo G Guardar ficheiro numa pasta DO OneDrive

\* Seleccionar uma saída dos passos anteriores

items x

Adicionar dados dos itens ao excel
?
...

\* Localização v

\* Biblioteca de Documentos v

\* Ficheiro

\* Tabela v

nome {x} nome x

Nfatura  ID da Fatura x

cod\_prod  Código de pro... x

descrição  Descrição do it... x

quantidade  Quantidade do... x

unidade  Unidade do item x

valor\_unit  Preço unitário ... x

montante  Montante do it... x

precisao T  Confiança de t... x

[Mostrar opções avançadas](#) v

### Anexo H Adicionar dados das linhas no ficheiro Excel

Nome	NIF	NFatura	DataFatura	Data Pagament	Valor_s_iva	valor do iva	valor total	Localização	desconto	precisao	classificaç
Epígneus	503413909	2085	26/10/2022	25/11/2022	185,37 €	42,64 €	228,01 €	Guarda	0,00 €	83,50%	Razoável
Vicauco	502 852 321	22001/014199	07/11/2022	07/12/2022	968,00 €	222,64 €	1 190,64 €	Viseu	0,00 €	88,90%	Razoável
Epígneus	503413909	2191	08/11/2022	08/12/2022	219,51 €	50,49 €	270,00 €	Guarda	0,00 €	71,60%	Mau
Domingos Alves da Silva	500087113	2022/ 2571	10/11/2022	10/12/2022	141,66 €	32,58 €	174,24 €	Guarda	1,43 €	80,10%	Razoável
Cimpor	501 755 098	3301007260	14/11/2022	14/11/2022	372,37 €	85,65 €	458,02 €	Fornos de Algodres	0,00 €	94,40%	Bom
Petrolberica	502840153	98 /6	14/11/2022	13/01/2023	9 164,30 €	2 107,79 €	11 272,09 €	Casal de Comba	0,00 €	85,40%	Razoável
Tubitorno	PT504225596	2022E/1200	14/11/2022	14/12/2022	122,07 €	28,08 €	150,15 €	Guarda	25,52 €	75,50%	Mau
Petrolberica	502840153	17,459 /1	15/11/2022	15/12/2022	7 013,59 €	1 613,13 €	8 626,72 €	Casal de Comba	0,00 €	94,80%	Bom
Tubitorno	PT504225596	2022E/1222	18/11/2022	18/12/2022	104,73 €	24,09 €	128,82 €	Guarda	26,19 €	83,10%	Razoável
Domingos Alves da Silva	500087113	2022/ 2671	24/11/2022	24/12/2022	184,59 €	42,46 €	227,05 €	Guarda	46,15 €	92,30%	Bom
Domingos Alves da Silva	500087113	2022/ 2668	24/11/2022	24/12/2022	240,53 €	55,32 €	295,85 €	Guarda	60,13 €	92,20%	Bom
Tubitorno	PT504225596	2022E/1253	24/11/2022	24/12/2022	14,24 €	3,28 €	17,52 €	Guarda	3,56 €	81,60%	Razoável
Cimpor	501 755 098	3301007544	25/11/2022	25/11/2022	438,38 €	100,83 €	539,21 €	Fornos de Algodres	0,00 €	97,70%	Excelente
Domingos Alves da Silva	500087113	2022/ 2676	25/11/2022	25/12/2022	106,28 €	24,44 €	130,72 €	Guarda	26,57 €	85,10%	Razoável
João Tomé Saraiva	PT506887260	JTS/298	30/11/2022	30/12/2022	2 358,54 €	542,46 €	2 901,00 €	Guarda	0,00 €	65,40%	Mau
Saneabi	502804858	FY/32096	30/11/2022	29/01/2023	120,00 €	27,60 €	147,60 €	Castelo Branco	0,00 €	81,40%	Razoável
Maia & Marques	501499806	600222/3676	06/12/2022	02/03/2023	180,00 €	41,40 €	221,40 €	Fundão	0,00 €	93,80%	Bom
Vicauco	502 852 321	22001/015748	09/12/2022	08/01/2023	397,00 €	91,31 €	488,31 €	Viseu	0,00 €	90,30%	Bom
Vicauco	502 852 321	22001/015837	12/12/2022	11/01/2023	502,00 €	115,46 €	617,46 €	Viseu	0,00 €	86,80%	Razoável
Petrolberica	502840153	18,926 /1	13/12/2022	12/01/2023	6 056,91 €	1 393,09 €	7 450,00 €	Casal de Comba	0,00 €	94,70%	Bom
Vicauco	502 852 321	22001/015973	14/12/2022	13/01/2023	78,00 €	17,94 €	95,94 €	Viseu	0,00 €	85,00%	Razoável
Saneabi	502804858	FY/32409	15/12/2022	13/02/2023	313,80 €	72,17 €	385,97 €	Castelo Branco	0,00 €	82,50%	Razoável
Saneabi	502804858	FY/32408	15/12/2022	13/02/2023	21,00 €	4,83 €	25,83 €	Castelo Branco	0,00 €	81,50%	Razoável
Maia & Marques	501499806	600222/3830	19/12/2022	17/02/2023	65,00 €	14,95 €	79,95 €	Fundão	0,00 €	89,60%	Razoável

### Anexo I Excel com dados extraídos dos cabeçalhos das faturas

nome	Nfatura	cod_pr	descrição	categoria	quantidade	valor_uni	montante	preciso T	classificação
Cimpor	3501001107		Bago de Arroz	Bago de Arroz	36,06	11,50 €	418,14 €	87,30%	Razoável
Cimpor	3501000724		Bago de Arroz	Bago de Arroz	39,32	11,50 €	452,18 €	41,60%	Muito mau
Cimpor	3501000305		Bago de Arroz	Bago de Arroz	40,26	462,99 €	462,99 €	80,90%	Razoável
Cimpor	3301007260		Bago de Arroz	Bago de Arroz	32,38	11,50 €	372,37 €	86,30%	Razoável
Cimpor	3301007544		Bago de Arroz	Bago de Arroz	38,12	11,50 €	438,38 €	84,10%	Razoável
Cimpor	3501001963		Bago de Arroz	Bago de Arroz	36,36	11,50 €	414,69 €	46,70%	Muito mau
Cimpor	3501001647		Bago de Arroz	Bago de Arroz	35,18	11,50 €	404,57 €	32,20%	Muito mau
Cimpor	3501001883		Bago de Arroz	Bago de Arroz	37,18	427,57 €	427,57 €	84,00%	Razoável
Cimpor	3501002879		Bago de Arroz	Bago de Arroz	38,11	11,50 €	435,85 €	80,60%	Razoável
Domingos Alves da Silva	2023A11/246	5,6E+12	Cimento CEM	Cimento (saco)	48	3,98 €	189,31 €	88,30%	Razoável
Domingos Alves da Silva	2023A11/106	02688.D1	Tampa B-125 -	Tampa	1	51,14 €	48,58 €	90,80%	Bom
Domingos Alves da Silva	2023A11/106	02688.D1	Tampa B-125 -	Tampa	1	51,14 €	48,58 €	90,80%	Bom
Domingos Alves da Silva	2023A11/29	114.0194.	Caixa quadrada	Caixa	3	7,72 €	22,01 €	82,80%	Razoável
Domingos Alves da Silva	2023A11/28	1,27E+12	Tubo de	Tubo	50	5,77 €	259,76 €	92,90%	Bom
Domingos Alves da Silva	2023A11/28	1,27E+12	Tubo corrugado	Tubo	4	33,09 €	119,12 €	91,70%	Bom
Domingos Alves da Silva	2022/ 2859	1,27E+12	Tubo de	Tubo	50	5,97 €	268,76 €	84,80%	Razoável
Domingos Alves da Silva	2023A11/685	5,6E+12	Cimento CEM	Cimento (saco)	40	3,98 €	157,75 €	86,70%	Razoável
Domingos Alves da Silva	2022/ 2676	1339.1.18	Manta geotextil	Manta geotextil	86	0,70 €	106,28 €	85,70%	Razoável
Domingos Alves da Silva	2022/ 2671	1339.1.18	Manta geotextil	Manta geotextil	150	0,70 €	184,59 €	87,90%	Razoável
Domingos Alves da Silva	2022/ 2571	5,6E+12	Cimento CEM	Cimento (saco)	40	3,58 €	141,66 €	88,40%	Razoável
Domingos Alves da Silva	2023A11/392	5,6E+12	Cimento CEM	Cimento (saco)	36	3,98 €	141,98 €	82,90%	Razoável
Domingos Alves da Silva	2023A11/352	1,27E+12	Tubo PE	Tubo	50	0,79 €	35,49 €	89,00%	Razoável
Domingos Alves da Silva	2023A11/352	1,27E+12	Tubo PE	Tubo	50	0,57 €	25,61 €	89,10%	Razoável
Domingos Alves da Silva	2023A11/544	5,6E+12	Cimento CEM	Cimento (saco)	35	3,98 €	138,04 €	86,50%	Razoável
Domingos Alves da Silva	2023A11/436	5,6E+12	Cimento CEM	Cimento (saco)	40	3,98 €	157,75 €	86,90%	Razoável
Domingos Alves da Silva	2022/ 2668		Manta geotextil	Manta geotextil	150	0,70 €	184,59 €	90,90%	Bom
Domingos Alves da Silva	2022/ 2668		Manta geotextil	Manta geotextil	100	0,70 €	55,94 €	91,00%	Bom
Egipneus	193	BG5484	245/45R17 99Y	Pneu	2	152,21 €	304,41 €	38,80%	Muito mau
Egipneus	193	11	SGPU/	Taxa de gestao	2	1,05 €	2,10 €	89,00%	Razoável
Egipneus	193	AD4	Alinhamento	Alinhamento	1	18,70 €	18,70 €	89,00%	Razoável
Egipneus	193				2	2,85 €	5,69 €	89,10%	Razoável

Anexo J Excel com dados extraídos das linhas das faturas

Guardar ficheiro na pasta para verificação
?
...

\* Caminho da Pasta

/jc tese/petro/verificação

\* Nome do Ficheiro

Anexos Nome

\* Conteúdo do Ficheiro

Anexos Conteú...

Anexo K Guardar ficheiro numa pasta para verificar dados

**Tarefa para verificar dados extraídos**

\* Lista de Tarefas: Tasks

\* Título: Precisão de extração baixa

Data para Conclusão: AAAA-MM-DDThh:mm:ss

Data/Hora do Lembrete: AAAA-MM-DDThh:mm:ss

Importância: Baixa, normal ou alta.

Estado: Indica o estado ou o progresso da tarefa - não iniciada, em curso, conclu

Corpo Conteúdo

Font 12 B I U [ícones]

A precisão de extração dos dados da fatura nº [Nfatura value] é baixa ( [Collection confidence score] ). Por favor verifique os dados inseridos.

[Caminho]

Mostrar opções avançadas

Anexo L Mensagem enviada para verificar os dados extraídos

**Tarefa de alerta para data de pagamento da fatura**

\* Lista de Tarefas: Tasks

\* Título: Pagamento em atraso

Data para Conclusão: AAAA-MM-DDThh:mm:ss

Data/Hora do Lembrete: AAAA-MM-DDThh:mm:ss

Importância: Baixa, normal ou alta.

Estado: Indica o estado ou o progresso da tarefa - não iniciada, em curso, conclu

Corpo Conteúdo

Font 12 B I U [ícones]

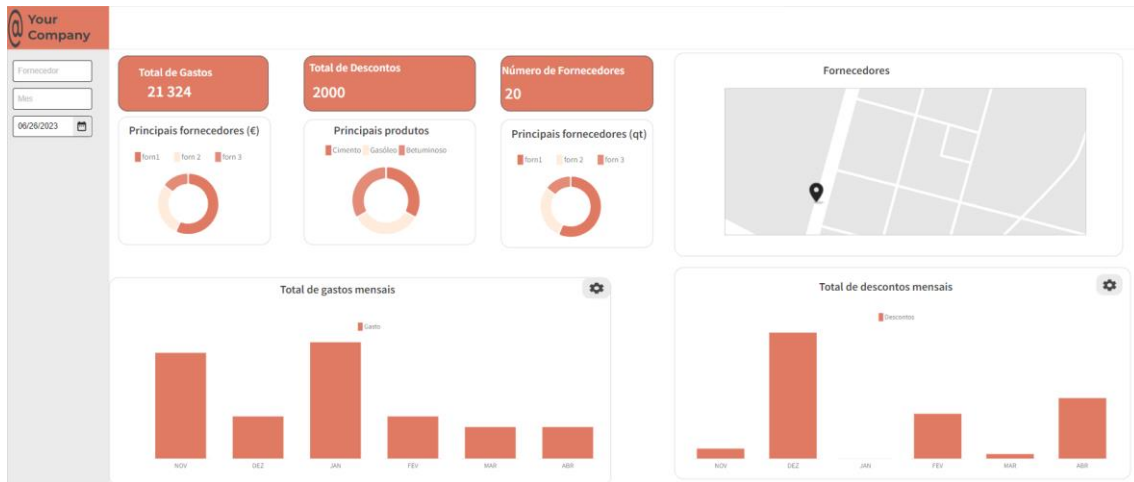
A data de pagamento da fatura [Nfatura value] expirou. Por favor, verifique se a situação já se encontra regularizada.

[Caminho]

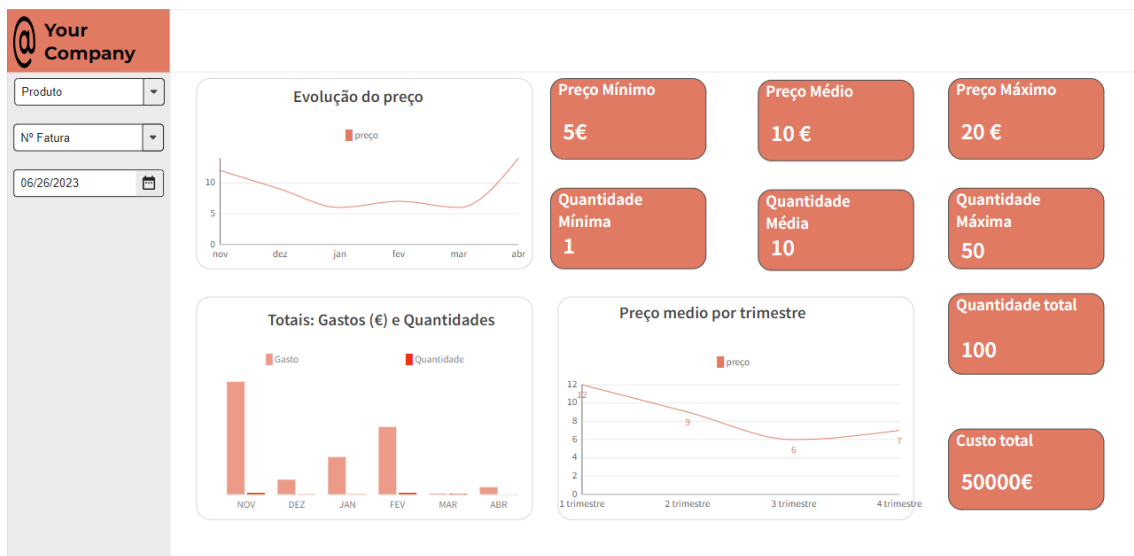
Adicionar conteúdo dinâmico

Mostrar opções avançadas

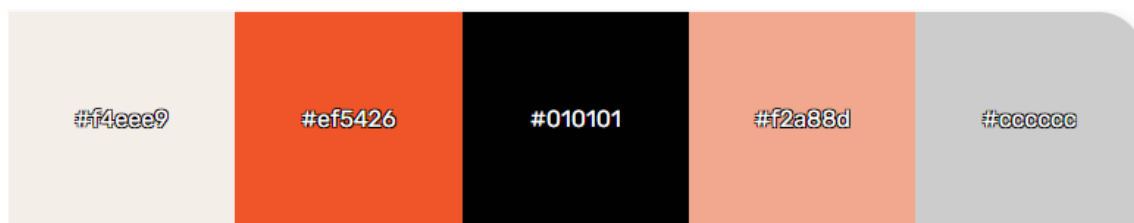
Anexo M Mensagem enviada caso data de pagamento seja ultrapassada






















Anexo N Mockup página dos fornecedores



Anexo O Mockup página do produto



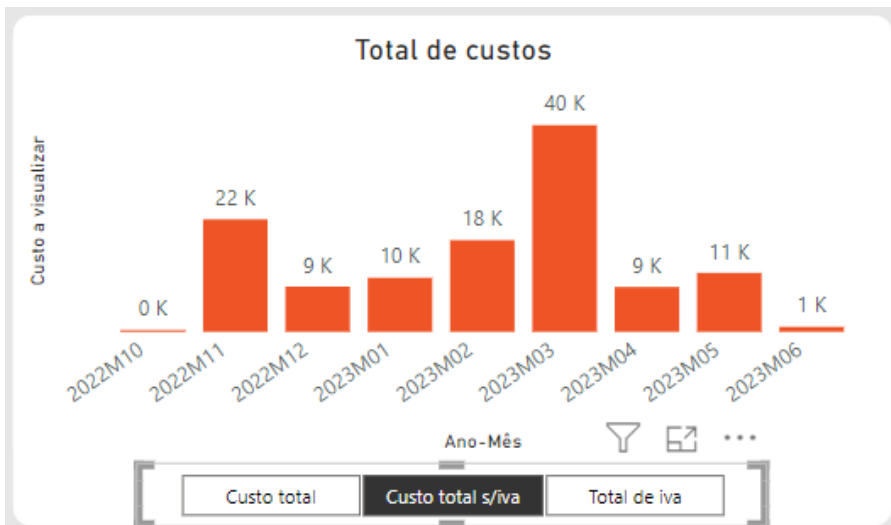
Anexo P Palette de cores utilizada

	Nome	Geral
		
	Nome	94 
	valor total	91 
	valor do iva	93 
	NIF	99 
	NFatura	92 
	Valor_s_iva	91 
	Data pagame...	99 
	DataFatura	89 
	Localização	97 
	desconto	99 

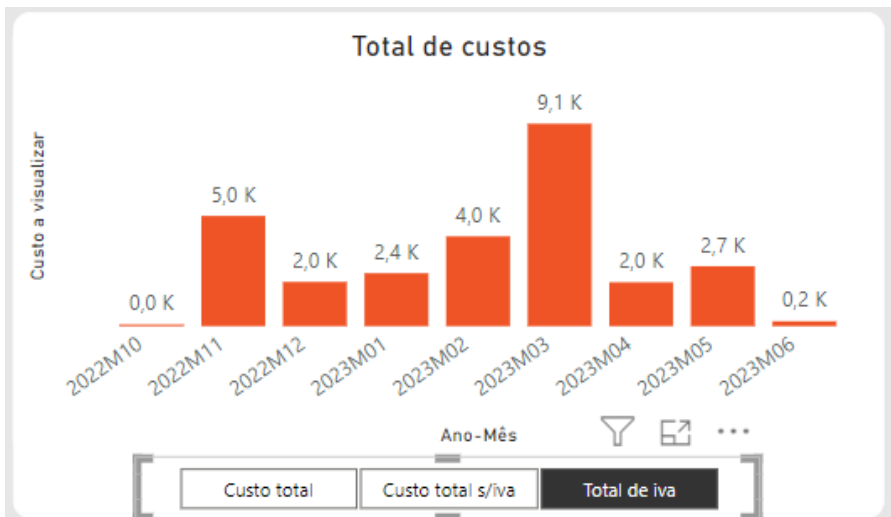
Anexo Q Precisão de cada atributo, modelo customizado

	Nome	Geral
	<b>Tabela 1</b>	82 
	cod_prod	81 
	descrição	79 
	quantidade	89 
	valor_unit	78 
	montante	82 

Anexo R Precisão das linhas das faturas, modelo customizado



Anexo S Gasto total sem iva por ano e por mês





Anexo T Gasto total de iva por ano e por mês



### Filtros

Data ▼

23-10-2022  03-09-2023 

Fornecedor ▼

Tudo ▼

Anexo U Filtros da página fornecedores

Medida a visualizar ▼

Máximo

Médio

Mínimo

Anexo V Medida a visualizar na página do produto

### Filtros

Data

01-01-2022  31-12-2030 




Fornecedor/NºFactura


Tudo 




Produto

Gasoleo 

Anexo W Filtros página produtos

Fornecedor/NºFactura 

Tudo 

-   Domingos Alves d...
-   Egipneus
-   João Tomé Saraiva
  - JTSJ/1095
  - JTSJ/1226
  - JTSJ/298
  - JTSJ/430

Anexo X Filtro por fornecedor ou por nº fatura

**Produto**



Gasoleo ^

- ABGE
- Abraçadeiras
- Alinhamento direção
- Anilha
- Areão
- Areia
- Bago de Arroz

Anexo Y Filtrar por produto

## Filtros

Data

01-01-2022  31-12-2030 



Fornecedor

Tudo 

Classificação

Bom	Mau
Excelente	Razoável

Precisão

53,00% 97,70%



Anexo Z Filtro página de controlo