



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Modelo de classificação documental automatizada, criação e encaminhamento de processos, com recurso a Aprendizagem Automática e Automação Robótica de Processos (RPA) aplicada a processos documentais na Administração Pública

Joaquim Fernando Ribeiro Muxagata

Mestrado em Digitalização na Administração Pública,

Orientadora:

Doutora Isabel Flores, Professora Auxiliar Convidada
ISCTE - Instituto Universitário de Lisboa

Setembro, 2025

Departamento de Ciência Política e Políticas Públicas

Modelo de classificação documental automatizada, criação e encaminhamento de processos, com recurso a Aprendizagem Automática e Automação Robótica de Processos (RPA) aplicada a processos documentais na Administração Pública

Joaquim Fernando Ribeiro Muxagata

Mestrado em Digitalização na Administração Pública,

Orientadora:

Doutora Isabel Flores, Professora Auxiliar Convidada,
ISCTE - Instituto Universitário de Lisboa

Setembro, 2025

Dedico este trabalho àqueles que, ao longo do meu percurso académico e profissional, contribuíram para o desenvolvimento do pensamento crítico, da perseverança e do compromisso com o serviço público. À minha família, pelo apoio constante e pela compreensão nos momentos de maior exigência. Aos docentes e colegas que, com o seu saber e partilha, enriqueceram esta jornada. Esta dedicatória estende-se também a todos os Servidores Públicos que, enfrentam os desafios da modernização com resiliência, compromisso, dedicação e sentido de missão.

Agradecimento

Agradeço, em primeiro lugar, à Professora Doutora Isabel Flores, pela orientação rigorosa, disponibilidade e estímulo intelectual ao longo deste trabalho. A sua orientação foi determinante para a consolidação da proposta e para o aprofundamento do trabalho realizado.

Agradeço também aos amigos, colegas e profissionais da Administração Pública, que, direta ou indiretamente, contribuíram com reflexões, dados e experiências que enriqueceram este estudo. A todos os docentes do Mestrado em Digitalização na Administração Pública do ISCTE, pela qualidade do ensino e pela partilha de conhecimento que marcaram este percurso académico.

À minha família e amigos, expresso profunda gratidão pelo apoio incondicional e pela paciência durante todo o período em que com eles estive menos presente e dedicado à realização deste trabalho.

Resumo

Este trabalho propõe o desenvolvimento e validação de um modelo de classificação documental automatizada e encaminhamento de processos documentais, recorrendo a técnicas de aprendizagem automática e Automação Robótica de Processos (RPA). Num contexto caracterizado por desafios como o envelhecimento dos trabalhadores, a escassez de recursos e a crescente exigência de eficiência e qualidade nos serviços públicos, este trabalho visa contribuir para a modernização da gestão documental, reduzir erros, aumentar a celeridade dos processos e libertar recursos humanos para tarefas de maior valor acrescentado.

O modelo proposto considera cinco etapas principais, definição de taxonomia e classes documentais, preparação dos documentos, classificação e segmentação, extração de atributos e encaminhamento automatizado assegurada por robôs de RPA com sistemas de gestão documental existentes.

A metodologia adotada combina investigação aplicada de natureza quantitativa e qualitativa, incluindo revisão bibliográfica, preparação de dados, desenvolvimento e validação empírica do modelo.

A avaliação do modelo conjuga métricas como precisão, *recall* e *F1-score*, com supervisão técnica e ciclos de melhoria contínua. Os resultados obtidos indicam uma redução superior a 70% no tempo de processamento documental e uma precisão global de 83% na classificação documental, validando a hipótese de que a combinação de inteligência artificial (IA) e RPA pode contribuir para melhorar a performance dos processos de gestão documental, contribuindo para a transformação digital da Administração Pública e para a melhoria contínua dos seus serviços.

Palavras-Chave: Classificação Documental Automatizada, Aprendizagem Automática, RPA, Transformação Digital; Administração Pública

Abstract

This study addresses the growing complexity of administrative documentation in the public sector by proposing a model for automated document classification and workflow routing using machine learning and RPA. In response to challenges such as workforce aging, limited resources, and rising efficiency demands, the model aims to modernize document management, streamline operations, and reallocate human effort to higher-value tasks.

The proposed approach enables the dynamic assignment of documents to appropriate departments or workflows based on extracted attributes and predefined business rules. Integration with existing systems is achieved through RPA, which simulates human interactions with legacy interfaces, thereby avoiding disruptive changes to IT infrastructure.

The combination of attribute extraction, rule-based routing, and system integration contributes to gains in operational efficiency, error reduction, and service quality.

The model comprises taxonomy and document class definition, document preparation, classification and segmentation, attribute extraction, and automated routing via RPA integrated with existing document management systems. The methodology blends quantitative and qualitative applied research, including literature review, data preparation, and empirical validation.

Model evaluation combines metrics such as precision, recall, and F1-score, with technical supervision and continuous improvement cycles. The results indicate a reduction of 70% in document processing time and an overall classification accuracy of 83%, validating the hypothesis that the combination of AI and RPA can enhance the performance of document management processes, contributing to the digital transformation of Public Administration and the continuous improvement of its services and demonstrates the potential of intelligent automation to transform document-centric processes in the public sector.

Keywords: Automated Document Classification, Machine Learning, RPA, Digital Transformation, Public Administration

Índice

Agradecimento	iii
Resumo	v
Abstract	vii
Glossário Técnico	xiii
Glossário de Siglas	xv
CAPÍTULO 1	1
Introdução	1
1.1. Enquadramento	1
1.2. Identificação do Problema de Ineficiência na Classificação Documental	2
1.3. Objetivos, Questões e Hipóteses de Pesquisa	5
1.4. Hipótese	5
1.5. Metodologia	6
CAPÍTULO 2	7
Revisão da Literatura	7
2.1. Gestão Documental na Administração Pública	7
2.2. Enquadramento Histórico e Normativo da Gestão Documental	8
2.3. O Papel da Inteligência Artificial na Gestão Documental	10
2.4. Casos de Uso e Boas Práticas Internacionais	13
2.5. A IA e o Repositionamento das Pessoas no Setor Público	14
2.6. Aprendizagem Automática e Classificação Documental	15
2.7. Automação Robótica de Processos	15
CAPÍTULO 3	17
Desenvolvimento do Modelo	17
3.1. Etapas do Modelo	17
3.1.1. Definição de Classes Documentais e Taxonomia	18
3.1.1.1. Criação da Taxonomia Documental	19
3.1.1.2. Definição de Classes Documentais do Nível 1 da MEF	21
3.1.1.3. Definição de Tipos de Documentos	22
3.1.1.4. Identificação e Definição de Atributos	22
3.1.1.5. Definição da Lista de Atributos	23
3.1.1.6. Definição de Padrões e Expressões Regulares	23
CAPÍTULO 4	29
Implementação do Modelo	29
4.1. Arquitetura Técnica da Solução	29
4.2. Preparação dos Dados	30
4.3. Treino e Teste do Modelo	32

4.4. Extração de Atributos	34
4.5. Encaminhamento Automatizado	36
CAPÍTULO 5	37
Avaliação do Modelo	37
5. Objetivos da Avaliação do Modelo	37
5.1. Metodologia de Avaliação	37
5.2. Validação da Extração de Atributos	39
5.3. Exceções, Erros e Auditoria	39
5.4. Resultados Obtidos	40
5.4.1. Desempenho do Modelo de Classificação Documental	40
5.4.2. Eficiência na Extração de Atributos	42
5.4.3. Impacto Operacional e Redução de Tempo	42
5.5. Eficiência Operacional	43
5.6. Supervisão e Melhoria Contínua	44
CAPÍTULO 6	45
Conclusões	45
Referências Bibliográficas	48
Anexos	53
Anexo A – Análise de frequência de classificação documental para o Nível 2 do classificador do MEF (116 516 amostras) provenientes do SGD de um serviço do Ministério das Finanças	53
Anexo B – Análise de frequência de classificação documental para o Nível 3 do classificador do MEF (116 516 amostras), provenientes do SGD de um serviço do Ministério das Finanças	54
Anexo C - Tipos de documentos frequentes em serviços públicos	60
Anexo D – Exemplo da Proposta de Taxonomia Documental e Atributos por Tipo de Documento	63
Anexo E – Exemplo de regras de encaminhamento	64
Anexo F – Exemplo de Expressões Regulares	66
Anexo G – Classificador Inteligente	68
Anexo H – Fluxo robotizado	69
Anexo I – Automatizar o fluxo documental	73
Anexo J – Configuração dos extratores de atributos automatizados	74
Anexo K – <i>Script</i> para treino do modelo de classificação do MEF	76
Anexo L – Nuvem de palavras do documento	77

Índice de tabelas

Tabela 1. Frequência de classificações documentais de Nível 1 na amostra inicial.....	21
Tabela 2. Distribuição por classe dos documentos para treino e teste do modelo de classificação.	31
Tabela 3. Resultados do teste de classificação aplicado ao conjunto de teste.....	40
Tabela 4. Distribuição da amostra inicial pelas classes de Nível 2 do classificador do MEF.....	53
Tabela 5. Distribuição da amostra inicial pelas classes de Nível 3 do classificador do MEF.....	55
Tabela 6. Tipos de documentos administrativos.	60
Tabela 7. Tipos de documentos jurídicos e normativos.	60
Tabela 8. Tipos de documentos financeiros.	61
Tabela 9. Tipos de documentos de recursos humanos.....	61
Tabela 10. Tipos de documentos técnico-operacionais.....	62
Tabela 11. Tipos e documentos de comunicação e informação.	62
Tabela 12. Exemplo de regras de encaminhamento por tipo de documento.	64
Tabela 13. Exemplo de regras de encaminhamento por entidade remetente.....	64
Tabela 14. Exemplo de regras de encaminhamento por classificação documental.	65
Tabela 15. Exemplo de regras de encaminhamento por atributos específicos.	65
Tabela 16. Lista de expressões regulares para extração de atributos.	66

Índice de figuras

Figura 1. Árvore da problemática abordada.....	3
Figura 2. Principais marcos da classificação documental na AP portuguesa.	8
Figura 3. Proposta de modelo.....	18
Figura 4. Exemplo da taxonomia e do conjunto mínimo de atributos a extraír.....	22
Figura 5. Exemplo de expressões regulares para complementar a extração de atributos.....	23
Figura 6. Visão geral da arquitetura técnica proposta.	30
Figura 7. Exemplo de configuração da extração de atributos por tipo de documento previsto na taxonomia.....	34
Figura 8. Exemplo de configuração de expressões regulares para extração de atributos.	35
Figura 9. Resultado de teste de classificação.	41
Figura 10. Exemplo da taxonomia definida.	63
Figura 11. Exemplo da interface de gestão das expressões regulares para extração de atributos.	66
Figura 12. Exemplo de configuração da interface de treino do classificador inteligente para identificar o tipo de documento.	68
Figura 13. Exemplo de ativação dos classificadores por tipo de documento.	68
Figura 14. Exemplo do fluxo robotizado de preparação e otimização dos documentos.	69
Figura 15. Exemplo do fluxo de classificação automatizada do tipo de documento.....	70
Figura 16. Exemplo do fluxo de classificação do Nivel 1 da MEF.....	71
Figura 17. Exemplo do fluxo de extração de atributos previstos na taxonomia.....	72
Figura 18. Exemplo da automatização do interface gráfico.	73

Figura 19 Gestão dos tipos de documentos na taxonomia.	74
Figura 20 Exemplo de marcação visual dos atributos para reconhecimento e extração automatizada.	75
Figura 21. Código em python para treinar o modelo de classificação (Nível 1 da MEF).	76
Figura 22. Gráfico de nuvem de palavras do documento.	77

Glossário Técnico

Para facilitar a compreensão dos termos técnicos e administrativos utilizados ao longo deste trabalho, apresenta-se abaixo um glossário com as principais termos e respetivos significados. Este glossário pretende apoiar a leitura e maior clareza na interpretação dos conceitos abordados, especialmente no contexto da Administração Pública e das tecnologias aplicadas à gestão documental.

Termo	Definição
Administração Pública	Conjunto de entidades que prestam serviços públicos e gerem recursos do Estado.
Aprendizagem Automática (Machine Learning)	Técnica de IA que permite aos sistemas aprenderem com dados e melhorarem sem programação explícita.
Automação Robótica de Processos (RPA)	Tecnologia que usa robôs de software para automatizar tarefas em sistemas digitais.
Classificação Documental	Processo de classificação de documentos segundo critérios funcionais ou temáticos.
Classificador MEF	Modelo hierárquico da DGLAB com três níveis (funções, subfunções e atividades).
Entidade Nomeada	Elemento de dados com significado semântico (ex.: Assunto, data), extraído automaticamente de documentos.
Expressões Regulares	Padrões de texto utilizados para identificar e extrair dados específicos em documentos.
Extração de Atributos	Identificação automatizada de informações relevantes (ex.: datas, nomes, números) em documentos.
Processamento de Linguagem Natural (PLN)	Área da IA que permite aos algoritmos compreender e gerar linguagem humana.
Segmentação Documental	Divisão de documentos em agrupamentos para facilitar a análise e extração de dados.
Sistema de Gestão Documental (SGD)	Plataforma usada para armazenar, organizar e encaminhar documentos numa organização pública.
Taxonomia Documental	Estrutura hierárquica que organiza tipos de documentos e seus atributos para facilitar a extração automatizada.
Treino e Teste de Modelo	Etapas de desenvolvimento de modelos de IA, treino com dados conhecidos e teste com dados não considerados no treino.
Workflow Documental	Sequência de etapas que um documento percorre desde a receção até ao arquivamento.

Glossário de Siglas

Para facilitar a compreensão das siglas utilizadas ao longo deste trabalho, apresenta-se um glossário com as principais siglas e respetivos significados.

Sigla	Significado
ADSE	Assistência na Doença aos Servidores do Estado
BEP	Bolsa de Emprego Público
BERT	Bidirectional Encoder Representations from Transformers
Bi-LSTM	Bidirectional Long Short-Term Memory
BPM	Business Process Management (Gestão de Processos de Negócio)
CGA	Caixa Geral de Aposentações
COFOG	Classification of the Functions of Government
CRISP-DM	Cross Industry Standard Process for Data Mining
CRF	Campos Aleatórios Condicionais
DGAEP	Direção-Geral da Administração e do Emprego Público
DGLAB	Direção-Geral do Livro, dos Arquivos e das Bibliotecas
DR	Diário da República
EIPA	European Institute of Public Administration
EO	Entidade Orçamental
EU-OSHW	Agência Europeia para a Segurança e Saúde no Trabalho
F1-score	Métrica de avaliação de modelos (harmonia entre precisão e recall)
GPT	Generative Pre-trained Transformer
HMM	Modelos Ocultos de Markov
IA	Inteligência Artificial
IBAN	Número Internacional de Conta Bancária
MEF	Macroestrutura Funcional
NIF	Número de Identificação Fiscal

Sigla	Significado
OCR	Reconhecimento Ótico de Caracteres
PCM	Presidência do Conselho de Ministros
PLN	Processamento de Linguagem Natural
RPA	Automação Robótica de Processos (<i>Robotic Process Automation</i>)
SGD	Sistema de Gestão Documental
SOFE	Sistema de Otimização da Formação e Ensino
URL	<i>Uniform Resource Locator</i> (endereço de página na internet)

CAPÍTULO 1

Introdução

A modernização da Administração Pública (AP) é um desafio comum em vários países da Organização para a Cooperação e Desenvolvimento Económico (OCDE), em regra focados na melhoria da eficiência e eficácia dos serviços públicos, com o propósito de responder às expectativas dos cidadãos e das empresas e também para responder aos desafios demográficos por conta do envelhecimento e da dispersão territorial. (Comissão Europeia, 2016)

Ainda que as reformas administrativas variem de país para país, a digitalização dos serviços públicos, bem como a transferência de responsabilidades e recursos para níveis locais de Governo (descentralização e subsidiariedade), são abordagens frequentes. (Comissão Europeia, 2016)

Neste contexto, a transformação digital tem sido um tema central nas estratégias de modernização e eficiência dos serviços públicos, não se resumindo a uma questão tecnológica, antes envolvendo um processo de mudança da cultura organizacional, exigindo a revisão de processos, a redefinição de papéis e a criação de novas competências nos servidores públicos. Esta transformação pode agilizar e simplificar o trabalho e promover ganhos de eficiência. (Berryhill, J. et al., 2019)

Espera-se que este trabalho contribua para a implementação nos serviços públicos de sistemas automáticos, realçando os benefícios e as oportunidades proporcionados pela transformação digital, designadamente com recurso à Inteligência Artificial (IA) e à Automação Robótica de Processos (RPA). A tendência de redução do número de efetivos nos mapas de pessoal e a inversão da pirâmide etária na estrutura de trabalhadores (DGAEP, 2024) pode ter os seus efeitos parcialmente mitigados pela automação de processos como a classificação documental.

1.1. Enquadramento

A gestão documental na AP é uma realidade caracterizada por um volume considerável de documentos, respetivos processos e inerentes atividades de classificação de documentos e extração de atributos. A classificação e encaminhamento manual dos documentos é uma tarefa morosa, propensa a erros e cuja eficácia e eficiência impactam diretamente na agilidade e qualidade dos serviços públicos prestados.

A modernização da AP contribui para que os serviços prestados aos cidadãos e empresas sejam eficientes, acessíveis e de elevada qualidade. (Hammad et al., 2025) Num contexto de oferta e procura crescentes por serviços públicos digitais, impulsionada por fatores como a dispersão territorial e o envelhecimento da população, a modernização assume um papel diferenciador na prestação do serviço público. (Ferreira, N., 2025)

Contudo, neste processo destacam-se alguns desafios, como:

- a resistência à mudança, uma vez que a transformação digital exige uma mudança operacional e cultural profunda nos serviços públicos; (Idzi & Gomes, 2022)
- a escassez de recursos financeiros, de recursos humanos qualificados e a falta da experiência tecnológica, que impactam na implementação de projetos de modernização; (Zhang & Chen, 2024) e
- o envelhecimento das equipas, que representa também um desafio em termos de transmissão de conhecimento e de adaptação às novas tecnologias e metodologias de trabalho.

Pese embora as reformas administrativas variarem de país para país, há uma abordagem comum nos países da OCDE. (Idzi & Gomes, 2022) Por exemplo, Portugal implementou várias reformas focadas na digitalização dos serviços públicos e na descentralização administrativa, França adotou medidas para melhorar a eficiência dos serviços públicos através da modernização tecnológica e da gestão por resultados, a Alemanha focou-se na reforma dos processos administrativos e na introdução de novas práticas de gestão pública, o Reino Unido implementou a Nova Gestão Pública (*New Public Management*), que inclui a privatização de alguns serviços públicos e a introdução de práticas de gestão do setor privado, e os países nórdicos (Suécia, Dinamarca e Noruega), conhecidos pelas suas abordagens inovadoras na Administração Pública, incluíram a digitalização e a participação dos cidadãos nos processos de tomada de decisão. (Comissão Europeia, 2016)

A digitalização dos serviços públicos para melhorar a acessibilidade e a eficiência é mais do que uma necessidade operacional, contribui para dar resposta às expectativas e necessidades da sociedade contemporânea.

1.2. Identificação do Problema de Ineficiência na Classificação Documental

A burocracia, enquanto sistema organizacional baseado em regras, procedimentos e hierarquias formais, está intimamente ligada ao volume de trabalho gerado com o tratamento documental nos serviços públicos. Este desafio evidencia o papel da gestão documental na prestação de serviços à sociedade e está relacionado com a necessidade de promover transparência e rastreabilidade nas decisões administrativas. De acordo com (Bouckaert & Halligan, 2007), a gestão do desempenho no setor público europeu tem vindo a enfatizar a importância da documentação como instrumento de responsabilização e melhoria contínua, que contribui para a estratégia da governação pública.

Neste contexto é tratado um elevado volume e variedade de documentos, fator que, por si só, já constitui um desafio, ao qual acresce a necessidade de garantir a consistência na classificação e no encaminhamento dos documentos que dão entrada nos serviços, o que implica afetação permanente de recursos ao processo.

Atualmente, verifica-se erros e inconsistências na categorização de documentos e no encaminhamento das entradas documentais, impactando diretamente na celeridade e na precisão das respostas institucionais. De acordo com (Idzi & Gomes, 2022), a fragmentação de processos e a ausência de reintegração digital nos serviços públicos resultam em duplicação de tarefas, atrasos e erros. Também (Duarte et al., 2018) indicam que a falta de *accountability* documental e a ineficiência na categorização dos documentos são dificuldades acrescidas para a boa governança. Esta ineficiência operacional e a sobrecarga de trabalho comprometem a produtividade e a qualidade dos serviços prestados. (Idzi & Gomes, 2022; Duarte et al., 2018)

A árvore do problema representada na Figura 1 descreve a estrutura lógica do problema associada à classificação documental e ao encaminhamento de processos na Administração Pública, evidenciando as causas estruturais, o problema central e as consequências operacionais.

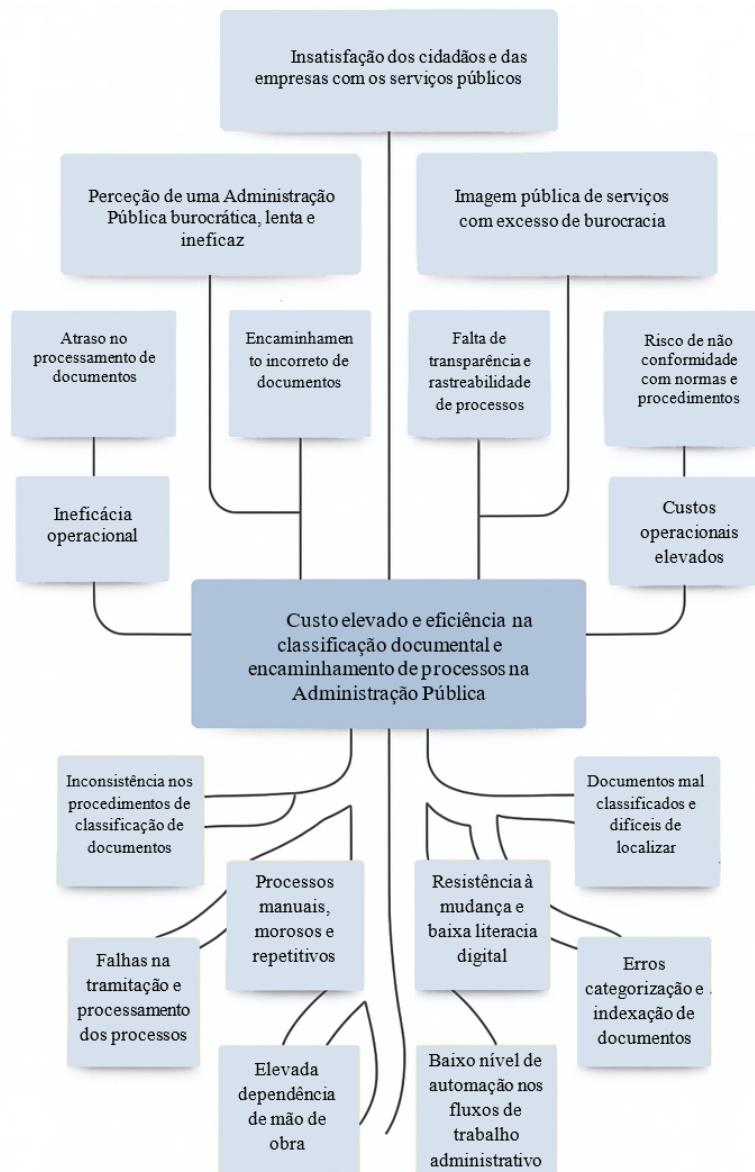


Figura 1. Árvore da problemática abordada.

Esta representação, como proposta de abordagem neste trabalho, realça a necessidade de tornar os processos de classificação e tratamento dos documentos mais ágeis e eficientes, nomeadamente através da aplicação de modelos de classificação automatizada com recurso a aprendizagem automática e robotização de processos.

O recurso a uma árvore de problemas permite representar de forma estruturada e visual a lógica causal subjacente ao problema da classificação documental e do encaminhamento de processos documentais. Este modelo conceitual é utilizado em metodologias de diagnóstico e planeamento estratégico, como o Quadro Lógico (*Logical Framework Approach*), por permitir identificar causas, consequências e o problema central de forma hierárquica e interligada. (EuropeAid, 2004)

No presente trabalho, a árvore de problemas foi desenvolvida com base na análise empírica dos processos documentais em serviços públicos e na revisão da literatura sobre gestão documental, automação e transformação digital. A estrutura da árvore contempla três níveis:

- As Raízes caraterizam as causas estruturais, onde se incluem fatores como a falta de padronização na classificação documental, a morosidade dos processos manuais, a dificuldade de escalabilidade, os erros frequentes na categorização e a ausência de integração entre sistemas. Estas causas estão alinhadas com os desafios identificados por (Penteado, P., 2015) e (Correia, M., 2023) que destacam a sobrecarga das equipas e a complexidade dos processos arquivísticos; (Penteado, 2015), (Correia, M., 2023)
- O Tronco representa o problema central do trabalho, que são os custos associados à ineficiência na classificação documental e encaminhamento de processos, que se traduzem num consumo de recursos, baixa produtividade e limitações na capacidade de resposta institucional (Hammad et al., 2025);
- Os Ramos representam as consequências e caraterizam os impactos operacionais advindos de atrasos, erros, desmotivação dos trabalhadores e risco de não conformidade, mas também consequências maiores, como a insatisfação dos cidadãos e empresas e a percepção de uma AP burocrática e lenta. Estas consequências são corroboradas por estudos sobre a relação entre eficiência administrativa e confiança institucional (Schneider, L., 2021; Idzi & Gomes, 2022)

A utilização desta representação ajuda a justificar a necessidade de intervenção tecnológica, através da aplicação de soluções baseadas em aprendizagem automática e robotização de processos, para mitigar as causas identificadas e transformar as consequências negativas em ganhos de eficiência.

1.3. Objetivos, Questões e Hipóteses de Pesquisa

O propósito deste trabalho é estudar a viabilidade da aplicação de um modelo que combine a tecnologia de RPA com a automatização dos processos de classificação e encaminhamento dos *workflows* documentais, considerando que, atualmente, muitos destes processos são executados de forma predominantemente manual, o que acarreta ineficiências e custos operacionais. Para o qual se propõe os seguintes objetivos:

1. criar um modelo de aprendizagem automática capaz de classificar documentos de acordo com o classificador documental da Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB). Para tal, será utilizada uma base de documentos previamente classificados, com origem em sistemas de gestão documental de um serviço público, para treino e teste do modelo.
2. implementar um modelo que combine a RPA com a aprendizagem automática. Esta sinergia permitirá automatizar o encaminhamento de documentos para os destinatários corretos, com o intuito de reduzir a intervenção manual nos processos documentais.
3. avaliar a eficácia do modelo proposto em termos de precisão de classificação e eficiência no encaminhamento documental.

Espera-se que o modelo desenvolvido apresente precisão na classificação documental e que o sistema de RPA melhore a eficiência do encaminhamento de documentos.

De forma a atingir os objetivos definidos, segue-se a questão que guiará este trabalho:

- Poderá um modelo automatizado que combine aprendizagem automática e RPA realizar a classificação documental de forma eficiente e com uma reduzida margem de erro?

1.4. Hipótese

As hipóteses que se testam são a possibilidade de:

- H1: Desenvolver um modelo que combine aprendizagem automática e RPA para automatizar o encaminhamento de processos documentais;
- H2: O modelo desenvolvido apresentar uma taxa de erro global inferior a 20% na classificação documental automatizada e encaminhamento de processos, sendo por isso considerado uma forma eficiente de gerir processos documentais;
- H3: A implementação do modelo reduzirá o tempo médio de processamento documental em pelo menos 50%.

1.5. Metodologia

A metodologia a adotar neste trabalho focaliza-se numa abordagem de investigação aplicada, de natureza quantitativa e qualitativa. A investigação seguirá as seguintes fases:

1. Revisão bibliográfica, com foco nas seguintes áreas:
 - Gestão Documental na Administração Pública, tanto a nível nacional quanto internacional;
 - Aprendizagem Automática e Classificação Documental;
 - Automação Robótica de Processos;
2. Preparação e Pré-processamento dos Dados:
 - Recolha de Dados, seleção de conjuntos de documentos previamente classificados, provenientes de sistemas de gestão documental de serviços públicos, como a Entidade do Tesouro e Finanças ou a Entidade Orçamental, para o treino e teste do modelo;
 - Limpeza e Normalização, para remoção de inconsistências, duplicações e ruído. Será também efetuada a normalização dos dados, incluindo a conversão de documentos não digitais (imagens) para texto através de reconhecimento ótico de caracteres (OCR);
 - Definição de Atributos e Taxonomia, tomando em consideração o classificador documental da DGLAB e as necessidades dos serviços públicos para definir os tipos de documentos, os atributos relevantes a extrair e a taxonomia documental a seguir;
3. Desenvolvimento do Modelo:
 - O *workflow* de RPA para automatizar a classificação documental, ao interagir com sistemas existentes para receber e classificar documentos, preencher metadados e encaminhar documentos de forma automatizada;
4. Avaliação e Validação do Modelo;
5. Teste de desempenho para avaliar os seguintes aspectos:
 - Precisão da classificação, através da medição da taxa de acerto do modelo na categorização correta dos documentos;
 - Eficiência no encaminhamento dos processos documentais, medição do tempo médio necessário para a leitura, classificação, extração de atributos, registo no sistema de gestão documental (SGD) e encaminhamento do processo, comparando o processo automatizado com o procedimento manual. Considerando para este efeito uma estimativa conservadora de cinco minutos por documento no processo manual, propõe-se como objetivo uma redução de pelo menos 50% do tempo despendido com o processo automatizado;
 - Impacto nos custos operacionais e estimativa da redução de custos associados à automação, nomeadamente a libertação de recursos humanos de tarefas repetitivas.

CAPÍTULO 2

Revisão da Literatura

Nas últimas décadas, assistimos a uma crescente adoção de tecnologias digitais no setor público, com destaque para IA e RPA. Estas tecnologias emergem como ferramentas que têm potencial para dar resposta aos desafios contemporâneos da gestão pública, em particular no que respeita à necessidade de maior eficiência, transparência e qualidade na prestação de serviços aos cidadãos. (Schneider, L., 2021)

A presente revisão bibliográfica tem como objetivo analisar os principais avanços desenvolvidos na área da IA, com particular enfoque na RPA e na sua aplicação em serviços públicos. Pretende-se assim estabelecer um quadro teórico e conceptual que sustente o desenvolvimento de um modelo de classificação documental automatizada para a AP portuguesa, conforme proposto no projeto de investigação.

2.1. Gestão Documental na Administração Pública

O esforço despendido pelas equipas nos processos de classificação documental é significativo. Autores como (Penteado, P., 2015) salientam desafios consideráveis, tendo em conta a variedade e diversidade de tipologias de documentos e a necessidade de conformidade com normas, referindo que a implementação de sistemas digitais pode mitigar esses desafios, automatizando processos e aumentando a eficiência. (Penteado, P., 2015)

É comummente aceite que as vantagens, da classificação documental organizada está relacionada com o acesso fácil e rápido à documentação, à sua indexação e controlo no acesso e na preservação dos documentos. Todavia, o conjunto de atividades de classificação documental tende a consumir tempo e recursos cada vez mais escassos nos serviços públicos, sendo estas tarefas realizadas de forma rotineira, mas com benefícios empiricamente reconhecidos, como a otimização do espaço físico e digital, facilidade de acesso e consulta, recuperação de informações e, consequentemente, impacto na produtividade das equipas. Autores como (Correia, M., 2023) indicam que o potencial de impacto da IA nos serviços, nas operações técnicas na área da gestão de arquivos torna-se evidente.

A integração de sistemas inteligentes gera contributos positivos e altera a forma como os documentos são classificados e acedidos para automatizar partes do fluxo de trabalho de classificação documental, reduzindo o esforço despendido pelas equipas nestas tarefas. (Correia, M., 2023)

A classificação documental é reconhecida como uma etapa relevante na gestão de documentos dos serviços públicos em Portugal, contribuindo para a organização, preservação e acesso aos documentos.

Envolve a categorização sistemática de documentos e contribui para a gestão de documentos ao longo de seu ciclo de vida. De acordo com a DGLAB, a classificação documental é fundamental para a elaboração da tabela de seleção, que define os prazos de guarda e o destino final dos documentos. Este processo ajuda a identificar os documentos que devem ser preservados permanentemente e aqueles que podem ser eliminados após um período específico. (DGLAB, 2025)

2.2. Enquadramento Histórico e Normativo da Gestão Documental

A importância da classificação documental nos serviços públicos tem sido reconhecida ao longo dos anos.



Figura 2. Principais marcos da classificação documental na AP portuguesa.

O Decreto-Lei nº 29/72, de 24 de janeiro, foi criado para resolver a acumulação de documentos físicos, dando valor probatório aos microfilmes e inserindo as entidades produtoras da documentação no processo de avaliação. Cada organismo público deveria publicar uma portaria identificando a documentação a eliminar e a conservar definitivamente. No entanto, o decreto-lei revelou as suas limitações, verificadas por reduzida adesão por parte dos serviços e por dificuldade na correta utilização das portarias. A isto acrescia o custo e a dificuldade de acesso aos microfilmes, que não contribuiu para a sua utilização, verificando-se um avolumar e acumulação de documentação.

Mais tarde, o Decreto-Lei nº 447/88, de 10 de dezembro, e a Portaria nº 192/2012, de 19 de junho, vieram estabelecer diretrizes para a gestão de documentos públicos em Portugal, e esses são apontados pelos arquivistas como marcos importantes na proteção e disponibilização da documentação dos serviços públicos, reforçando a importância da gestão e classificação dos documentos como suporte à atividade dos serviços.

Após a criação dos instrumentos de gestão documental em 1988, só em 2003 a entidade superintendente da política nacional obteve os primeiros dados quantitativos sobre a gestão de documentos na Administração Pública. De 645 organismos, 362 participaram num inquérito que revelou baixa introdução dos arquivistas nos arquivos correntes e falta de instrumentos de gestão documental. A documentação em arquivo intermédio era frequentemente armazenada em condições inadequadas, e continuava a haver eliminação ilegal e acumulação excessiva de documentos.

Em 2008, um novo inquérito foi realizado aos serviços com Plano de Gestão Documental publicado. O resultado do inquérito indicava que o plano era utilizado principalmente para efeito da eliminação de documentos em depósitos, devido à dificuldade na criação de planos de classificação. Em 2010, outro inquérito mostrou um aumento na classificação de documentos e implementação de Sistemas Eletrónicos de Gestão de Arquivos, embora ainda houvesse grandes quantidades de documentação por avaliar devido à baixa aplicação dos instrumentos de gestão documental. (Lourenço, A., 2012)

Em 2011, foi criada a Macroestrutura Funcional (MEF) para representar conceptualmente as funções dos Serviços da AP Central. A MEF é baseada numa estrutura hierárquica de dois níveis, composta por classes que representam funções no primeiro nível e subfunções no segundo, permitindo uma maior especificidade. Em 2013, uma nova versão da MEF foi publicada, incluindo também as funções da AP Local. (Lourenço, A., 2012)

Ao que se seguiu, em 2013, à "Harmonização de classes de 3º nível em planos de classificação conformes à Macroestrutura Funcional", com a inclusão de um terceiro nível de classificação, proporcionando maior especificidade, para refletir os processos de negócio dos serviços, o que veio permitir a associação a classificações de prazos de conservação e destino final para a documentação.

Este modelo vem acrescentar tarefas de classificação documental aos processos, que exigem o conhecimento das normas de classificação e que são demorados. Na prática, verificam-se duas dificuldades na sua adoção: a natural resistência à mudança; e o acréscimo de trabalho nas equipas.

A Estratégia para a Transformação Digital da Administração Pública 2021-2026, aprovada pela Resolução do Conselho de Ministros nº 131/2021, de 10 de setembro, tem como desígnio tornar a AP mais responsável às expectativas dos cidadãos e empresas, prestando serviços mais simples, integrados e inclusivos, funcionando de forma mais eficiente, inteligente e transparente, através da exploração do potencial de transformação das tecnologias digitais e da utilização inteligente dos dados.(Presidência do Conselho de Ministros, 2021)

No segundo trimestre de 2024, os dados publicados pela Direção-Geral da Administração e do Emprego Público (DGAEP), relativos à síntese estatística do emprego público, indicam a existência de 749 678 postos de trabalho no setor das Administrações Públicas, evidenciando que a média das idades destes trabalhadores é de 48,3 anos.

Ainda que a média seja um indicador relevante, pode ofuscar a variabilidade dos dados e induzir a interpretações erróneas, na medida em que, por exemplo, as áreas governativas das Finanças e da Agricultura e Alimentação apresentavam o emprego com a média etária mais alta, verificando-se que mais de metade dos trabalhadores tinham mais de 55 anos (56,3% na Agricultura e Alimentação e 52,7% nas Finanças). (DGAEP, 2024)

O envelhecimento dos trabalhadores dos serviços públicos apresenta desafios, como a necessidade de novas competências tecnológicas e a digitalização dos locais de trabalho, mas apresenta também oportunidades, como seja beneficiar da experiência e do conhecimento dos trabalhadores com maior antiguidade, potencializando um clima organizacional favorável à liderança adaptativa e à promoção de novas formas de trabalhar, com revisão de processos e procedimentos e com a automatização de tarefas. (Fernandes, C. & Ganhão, T., 2023)

Por conseguinte, é importante definir modelos que possam ser incluídos na estratégia de modernização dos serviços públicos de forma moderna e eficaz, implicando encontrar soluções que permitam agilizar os procedimentos e as atividades. As expectativas sobre os contributos da IA nas atividades que impactam com a vida do cidadão e das empresas é elevada e espera-se que a automatização e a robotização de tarefas se traduzam num contributo para as estratégias de modernização de serviços públicos.

2.3. O Papel da Inteligência Artificial na Gestão Documental

A IA tem o potencial de transformar as políticas e os serviços públicos, esperando-se, como benefícios resultantes da sua aplicação, uma melhoria do tempo de resposta e da qualidade dos serviços prestados ao cidadão, não obstante serem necessárias flexibilidade e experimentação, bem como perspetivas multidisciplinares. (Berryhill, J. et al., 2019)

O relatório “Artificial Intelligence for interoperability in the European public sector - An exploratory study” (European Commission, 2023) apresenta o resultado de um estudo realizado no contexto do Public Sector Tech Watch, que destaca a importância da IA para promover a interoperabilidade no setor público, facilitando a estruturação, normalização e ligação de um grande volume de dados na AP, tendo por base a análise do impacto dos sistemas de IA na melhoria da interoperabilidade no setor público europeu.

No relatório, foram abordadas iniciativas da Comissão Europeia, como a Lei da Europa Interoperável e a proposta de regulamento sobre IA, que visam garantir uma abordagem consistente e ética na utilização da IA. As principais conclusões destacam a importância da IA como ferramenta essencial para promover a interoperabilidade, ressaltando a necessidade de consciencialização e colaboração entre as organizações públicas para aproveitar os benefícios desta tecnologia.

O estudo inclui uma revisão da literatura e das políticas relevantes, bem como a análise de estudos de caso. Os principais resultados da análise quantitativa demonstraram que cerca de 26% dos casos no setor público utilizam IA para apoiar a interoperabilidade. A camada semântica é relevante em 91% dos casos estudados, e a utilização de ontologias e taxonomias aliadas à IA pode facilitar a interoperabilidade entre sistemas. A maioria dos casos de IA utilizados para interoperabilidade estão relacionados com processos de gestão interna, elegendo como objetivo primordial impulsionar a simplificação administrativa. Os sistemas baseados em IA realizam ações como detetar, estruturar e classificar dados para aumentar a interoperabilidade, resultando em benefícios como a eliminação de tarefas repetitivas, redução de custos e melhoria na qualidade dos serviços públicos.

O mesmo estudo identifica também desafios, como a falta de competências, a interoperabilidade e a qualidade dos dados, apresentando recomendações para apoiar os gestores públicos na implementação de soluções de IA.

A utilização da IA nos serviços públicos pode ter inúmeras aplicações. Por exemplo, pode ser utilizada para classificar documentos, organizando-os de acordo com critérios previamente definidos, modelos baseados em regras ou através de algoritmos de aprendizagem automática supervisionados, como árvores de decisão ou redes neurais, que aprendem, a partir de conjuntos de dados ou documentos já classificados, a classificar novos documentos. Os modelos podem ser complementados com algoritmos que permitem detetar objetos ou padrões nos dados, com recurso a técnicas como visão computacional e IA, relevante no âmbito deste trabalho, é a capacidade de identificar e extrair atributos e entidades com recurso a técnicas de processamento de linguagem natural (PLN), para identificar e extrair atributos necessários para a correta classificação dos documentos e para o preenchimento de metadados dos processos de negócio.

A digitalização e processamento de documentos sobre a forma de imagens “textos não digitais” é também um instrumento a considerar para, com recurso a OCR, a obtenção do texto digital dos documentos inicialmente disponibilizados em formato de imagem. O PLN pode também ser aqui utilizado para, após a identificação dos atributos, limpar e corrigir dados, proceder à respetiva normalização e à remoção de duplicações.

Estas funções demonstram como a IA pode ser uma ferramenta poderosa e versátil na melhoria dos serviços públicos, contribuindo para uma gestão mais eficiente e segura dos dados.

Cabendo destacar a importância da IA na estruturação e padronização de dados para melhorar a interoperabilidade, há, todavia, desafios a considerar, designadamente:

- Falta de precisão dos dados de entrada;
- Complexidade da estrutura orgânica e funcional dos serviços;
- Restrições orçamentais; e
- Dificuldade em encontrar profissionais com as competências necessárias.

Também identifica desafios como a falta de competências, a interoperabilidade e a qualidade dos dados. Por fim, apresenta recomendações detalhadas para auxiliar decisores políticos e gestores públicos na implementação de soluções de IA.

Assim, a aplicação de IA em contextos governamentais e administrativos servirá para:

- Classificar e categorizar dados com base em critérios específicos;
- Identificar objetos ou padrões em dados;
- Localizar entidades, recursos ou palavras-chave nos dados;
- Converter dados não digitais em digitais;
- Remover dados redundantes ou irrelevantes;
- Realizar operações matemáticas em dados;
- Fornecer suporte para conversas em linguagem natural; e
- Anonimizar e remover informações de identificação dos dados.

De entre as inúmeras aplicações práticas da IA nos serviços públicos, neste estudo destaca-se a relacionada com RPA, sistemas robotizados de classificação e ferramentas para extração de dados e informações em documentos. A capacidade de ler e extrair dados de documentos ou sistemas, e preencher formulários, fazer o *login* e interagir em aplicações de tecnologias diferentes, mover ficheiros ou pastas, extraír dados de uma página *web* e extraír informações de documentos ou de imagens, ainda que, *per si*, correspondam a atividades simples, representam muitas das atividades realizadas pelos técnicos no seu dia a dia, o que confere um enorme potencial aos modelos que tirem partido destas automatizações.

O desenvolvimento de modelos de automatização de tarefas em sistemas de informação com base em RPA pode ser um contributo para serviços públicos mais eficientes e capazes de criar valor, reduzindo o esforço humano na operacionalização dos processos.

A resistência à mudança é frequente na implementação dos sistemas de gestão documental e na adoção das metodologias de classificação documental, muitas vezes por desconhecimento ou percepção de que as tarefas de classificação documental são complexas e demoradas. Para minimizar esta resistência natural é fundamental um plano de comunicação e formação. (Vicente, M., 2013)

Autores como Rios e Cordeiro (2010) enfatizam o aumento de trabalho para as equipas com a implementação da classificação documental, bem como a exigência e o nível de conhecimento profundo das normas de classificação, que se traduzem numa maior demora do processo de classificação e no aumento natural da carga de trabalho das equipas. (Rios & Cordeiro, 2010)

Importa, portanto, encontrar formas de assegurar a correta classificação documental e, simultaneamente, mitigar estas dificuldades na sua adoção. (Penteado, P., 2015) destaca a importância e a necessidade de promover o desenvolvimento de soluções eficientes e de assegurar que os sistemas de arquivo são fidedignos no que diz respeito à produção de meta informação, fundamental para a gestão do arquivo e dos processos de negócio dos serviços públicos.

2.4. Casos de Uso e Boas Práticas Internacionais

O relatório dos *European Public Sector Awards* 2023-2024 realça a influência da IA na modernização da AP europeia, destacando projetos que promovem a interoperabilidade entre organismos públicos através da automação documental e da integração com sistemas legados. Estas iniciativas demonstram como a IA pode ser aplicada de forma ética e eficaz para melhorar a eficiência, reduzir erros operacionais e aumentar a qualidade dos serviços prestados aos cidadãos. Em particular, os projetos premiados ilustram abordagens inovadoras na classificação automatizada de documentos, extração de atributos e encaminhamento inteligente de processos, alinhando-se com os objetivos da transformação digital e da simplificação administrativa. (Paul, M. & Longinidou, E., 2024)

A digitalização do cancelamento de antecedentes criminais com o uso de algoritmos demonstra interoperabilidade técnica por meio da integração de diferentes sistemas na infraestrutura digital do Ministério da justiça espanhol. O projeto incorporou o uso de tecnologias como IA e RPA na gestão e tramitação de procedimentos. Por exemplo, na automação de procedimentos de ordens de pagamento, a IA e, mais especificamente, a aprendizagem automática foram utilizados para otimizar o processo.

Este reconhecimento europeu reforça a pertinência do modelo proposto neste trabalho, ao evidenciar que soluções baseadas em IA e RPA são já uma realidade em diversos contextos governamentais europeus, como é o caso da iniciativa do Ministério da Justiça espanhol, com recurso a RPA para garantir que vários sistemas da administração judicial possam trocar dados e informação com eficiência, automatizando processos entre plataformas com tecnologias distintas. (Paul, M. & Longinidou, E., 2024) Um outro estudo europeu identifica 143 casos de uso de IA no setor público, recolhidos através de inquéritos, entrevistas e análise documental, organizados segundo as funções de governo (COFOG). (European Commission, 2023) Entre os exemplos mais relevantes para este trabalho destacam-se os sistemas de classificação documental automatizada, extração de atributos e encaminhamento inteligente de processos, implementados em países como Estónia, Finlândia, Alemanha e Países Baixos. Estes casos comprovam o potencial da IA para melhorar a interoperabilidade entre sistemas, reduzir a carga administrativa e aumentar a eficiência dos serviços públicos, reforçando a pertinência da proposta de modelo apresentada neste trabalho. A sistematização destes casos permite não só identificar boas práticas, mas também orientar a adoção de soluções tecnológicas alinhadas com os princípios da governação digital europeia.

2.5. A IA e o Repositionamento das Pessoas no Setor Público

A adoção de iniciativas de transformação digital com sucesso nos serviços públicos está muitas vezes relacionada com a forma como as pessoas da organização e a cultura organizacional aceitam e suportam a mudança. Em Portugal, e ainda que o nível de literacia digital nos serviços públicos seja razoável, será, todavia, a transformação digital dos serviços condição necessária para consolidar as mudanças nas organizações públicas?

O artigo “Tomada de decisão nas organizações: o que muda com a Inteligência Artificial?” analisa como a IA contribui para a transformação da tomada de decisão em vários setores, mas, fundamentalmente, reflete sobre o impacto no mercado de trabalho da integração da IA nas decisões das organizações e dos serviços e como estas e estes podem alterar significativamente o papel das pessoas em determinadas atividades, promovendo uma colaboração mais eficiente entre pessoas e algoritmos. (Yu et al., 2024)

A automação e a IA podem desempenhar um papel importante no desenvolvimento de sistemas mais autónomos e automatizados na formação profissional. Ferramentas de aprendizagem adaptativa e plataformas de *e-learning* personalizadas permitem que os trabalhadores adquiram novas competências de forma contínua e eficiente. Estas tecnologias facilitam a criação de programas de formação que possam responder às necessidades específicas de cada trabalhador, promovendo o desenvolvimento profissional contínuo.

A intervenção humana continua a ser fundamental nas decisões automatizadas para assegurar a justiça e a ética. Além disso, a requalificação profissional é necessária para preparar os trabalhadores para as mudanças tecnológicas. Programas de requalificação ajudam a garantir que os trabalhadores se possam adaptar às novas exigências do mercado de trabalho, mantendo a relevância das suas competências. (Sanches, A. P., 2024) São várias as aplicações práticas da IA na AP propostas neste trabalho, e destacam-se estas duas relacionadas com o problema de RPA: ferramentas de transcrição e reconhecimento de texto e ferramentas para extração de dados em plataformas que não comuniquem com as soluções do serviço público ou que envolvam trabalho manual.

Entre as várias valências de uma solução RPA, destaca-se a capacidade de inserir dados, fazer cálculos simples, ler e extrair dados de sistemas, responder a *e-mails*, preencher formulários, abrir anexos, efetuar o *login* em algumas aplicações, mover ficheiros ou pastas, extraír dados de uma página *web* e extraír informações de ficheiros em formato PDF ou de imagem. Deste modo, evidencia-se que o recurso ao RPA é uma mais-valia para as organizações, fazendo com que estas sejam mais eficientes e, como tal, mais competitivas, capazes de criar valor e de reduzir o custo dos processos, com ganhos de eficiência e consistência.

2.6. Aprendizagem Automática e Classificação Documental

A IA representa um campo da ciência da computação dedicado ao desenvolvimento de sistemas capazes de realizar tarefas que requeriam tradicionalmente inteligência humana. Toledo (2023) define a IA como "o desenvolvimento de algoritmos e ferramentas que imitam a capacidade de aprendizagem humana, tomada de decisão e reconhecimento de padrões". Esta definição evidencia a amplitude do conceito e a sua aplicabilidade em diversos contextos, incluindo a Administração Pública. (Teixeira De Toledo & Mendonça, 2023)

Diferentemente dos *softwares* tradicionais, que operam segundo regras predefinidas, os sistemas de IA têm a capacidade de aprender com dados, de se adaptar a novas informações e melhorar o seu desempenho ao longo do tempo, sem necessidade de intervenção humana direta. (Santos & Lima, 2025) Esta característica distintiva, conhecida como aprendizagem automática, confere à IA um potencial transformador significativo.

A evolução da IA tem sido marcada por avanços significativos nas últimas décadas. Desde os primeiros sistemas periciais baseados em regras até às atuais redes neurais profundas, a IA tem expandido as suas capacidades e aplicações. De acordo com a Associação Nacional dos Especialistas em Políticas Públicas e Gestão Governamental (Anesp, 2023), esta evolução tem sido impulsionada por três fatores: o aumento da capacidade de computação, muito por conta da computação na *cloud*; o acesso a grandes volumes de dados não tratados (*big data*); e a utilização de bons algoritmos.

Neste contexto, verifica-se também uma evolução da automação em ambiente digital, muito associada à execução de tarefas e atividades em sistemas de informação, sem necessidade de intervenção humana. Conforme explica Mattos (2024), "Os resultados mostraram que as tecnologias, como a digitalização de serviços e automação de processos, têm contribuído para a melhoria da eficiência, redução de custos e maior transparência na gestão pública.". Esta definição evidencia a relação estreita entre a transformação digital onde se inclui a automação e a IA, sendo esta última frequentemente utilizada como ferramenta para implementar soluções de automação. (Mattos, T. et al., 2024)

2.7. Automação Robótica de Processos

O Institute of Electrical and Electronics Engineers (2017) Define RPA como “*software* pré-configurado que usa regras do negócio e da atividade predefinida para realizar uma execução autónoma da combinação de processos, atividades, transações e tarefas de um ou mais sistemas de *software* não relacionados, de forma a entregar um resultado ou serviço sem o uso de mão humana”. (IEEE, 2017) Trata-se de uma forma automação, que utiliza *software* para imitar as ações humanas na interação com sistemas digitais.

A RPA destaca-se como uma tecnologia que permite configurar *software* para capturar e interpretar aplicações existentes, processar transações, manipular dados, desencadear respostas e comunicar com outros sistemas digitais. (Syed et al., 2020)

É importante distinguir automação de automatização, conceitos frequentemente confundidos. Enquanto a automação diz respeito à aplicação de tecnologia sem intervenção humana, a automatização diz respeito à aplicação da tecnologia em tarefas que ainda necessitam de pessoas para serem iniciadas.

A RPA distingue-se de outras formas de automação pela sua capacidade de interagir com interfaces de utilizador da mesma forma que um humano o faria, sem necessidade de modificar os sistemas existentes. Esta característica torna-a particularmente valiosa para organizações com sistemas legados, como é frequentemente o caso na Administração Pública.

São vários os exemplos de serviços públicos que recorrem ao uso de robôs de conversação para atendimento ao cidadão, bem como à digitalização de documentos administrativos e extração de atributos de forma automatizada. A Agência Europeia para a Segurança e Saúde no Trabalho aponta benefícios, desafios e boas práticas relacionados com a automatização de tarefas com recurso à IA e à robótica para reduzir a carga de trabalho decorrente das atividades habitualmente realizadas de forma artesanal, tendo por base uma avaliação de riscos, formação e sensibilização dos trabalhadores para o uso de novas tecnologias e os benefícios que se espera obter com a sua adoção, bem como assegurar que é respeitada a regulamentação dos vários países em matéria de privacidade e proteção de dados. (European Agency for Safety and Health at Work, 2024)

A RPA utiliza robôs de *software* para automatizar tarefas repetitivas e baseadas em regras ou em processos de aprendizagem automática. Nos serviços públicos, a RPA começa a ser utilizada para automatizar o registo de entrada de dados e documentos nos processos e fluxos de trabalho, o que implica a recolha, extração e análise de dados em documentos. (Campos, D., 2023)

A RPA está a transformar o ambiente de trabalho nas organizações, e os serviços públicos estão também a seguir este rumo, dado que, ao automatizar tarefas repetitivas, é possível libertar os trabalhadores para atividades que exijam análise complexa. É, contudo, necessária uma compreensão clara dos processos a ser automatizados. A maioria das implementações de RPA inicia-se com uma prova de conceito, que demonstra os recursos e o potencial da tecnologia. (Campos, D., 2023)

Na dissertação de mestrado “O Estado da Arte na exploração do RPA segundo os estudos publicados” (Campos, D., 2023) é fornecida uma visão global do estado da arte do recurso a RPA, com base na revisão de literatura que adotou a abordagem *scoping review*, que resultou na seleção de 14 artigos para resposta à questão-chave: “Quais são as abordagens para a adoção de RPA nas organizações, que a literatura identifica, num contexto de estudo dos efeitos da sua implementação?”. Conclui-se que o RPA se tem disseminado à medida que as organizações recorrem à automatização de tarefas repetitivas, com o objetivo de reduzir custos, melhorar a eficiência e libertar funcionários para se concentrarem em trabalhos de maior valor acrescentado e pode beneficiar das capacidades avançadas da aprendizagem automática, por exemplo a identificação e classificação documentais.

CAPÍTULO 3

Desenvolvimento do Modelo

A presente proposta visa estabelecer um modelo abrangente para a classificação documental automatizada, utilizando técnicas de aprendizagem automática e RPA. O modelo proposto tem como objetivo principal otimizar a gestão documental, reduzir erros de classificação, aumentar a eficiência no encaminhamento de documentos e libertar recursos humanos para tarefas de maior valor acrescentado.

A proposta baseia-se numa abordagem estruturada em cinco etapas principais, complementadas por mecanismos de supervisão e melhoria contínua, para promover a precisão, fiabilidade e adaptabilidade do modelo ao longo do tempo, é uma adaptação do modelo CRISP-DM (Cross Industry Standard Process for Data Mining). Esta adaptação permite adequar as fases originais do CRISP-DM, como a compreensão do negócio, compreensão dos dados, preparação dos dados, modelação, avaliação e implementação ao contexto específico dos Serviços Públicos e às exigências da classificação documental automatizada com integração de RPA. A literatura indica que o CRISP-DM é suficientemente flexível para ser ajustado a diferentes domínios e necessidades, permitindo incorporar etapas adicionais ou redefinir atividades para dar resposta a objetivos específicos. (Wirth, R. & Hipp, J., 2000), (Schröer et al., 2021) Assim, preserva-se a lógica sequencial e iterativa do CRISP-DM, mas introduz ajustes que contemplam a definição de taxonomia documental, a extração de atributos e a integração com sistemas legados através de RPA.

Esta estrutura permite uma abordagem sistemática e escalável, adaptável, respeitando simultaneamente o classificador documental da DGLAB e as práticas de encaminhamento estabelecidas na organização.

3.1. Etapas do Modelo

O modelo proposto considera cinco etapas. A primeira etapa consiste na definição dos tipos ou classes de documentos a considerar pelo modelo para classificação automatizada. A esta etapa, segue-se a definição do conjunto de atributos que devem ser identificados em cada uma das classes de documentos. A terceira etapa consiste na aplicação de algoritmos de aprendizagem automática para classificação e segmentação dos documentos, ao que se segue a etapa de extração dos atributos e finalmente a integração com o SGD com recurso a agentes robóticos de RPA.

Após a aplicação do modelo é necessário assegurar, através de supervisão por técnicos especializados, o resultado do processo automatizado, com recurso a amostragem para que sejam validados os resultados. Caso se identifiquem más classificações, os conjuntos de treino e de teste devem ser enriquecidos com mais amostras para minimizar o viés de classificação.

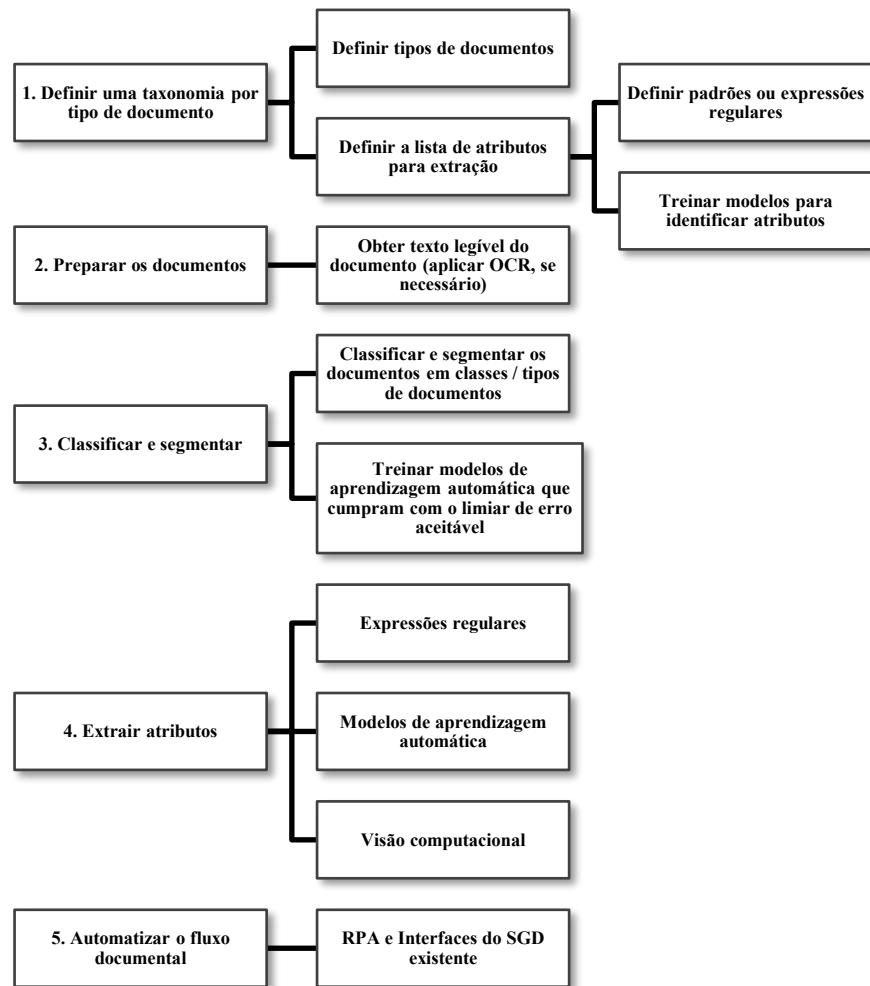


Figura 3. Proposta de modelo.

3.1.1. Definição de Classes Documentais e Taxonomia

A primeira etapa do modelo consiste na criação de uma taxonomia e na definição dos tipos de documentos a serem processados pelo sistema. Esta fase permite estabelecer as bases de todo o processo de classificação automatizada.

Na senda de realizar o processamento inteligente de documentos de forma automatizada com recurso a PLN, de acordo com Vashisht (2025), é possível extrair, classificar e analisar dados constantes nos documentos com precisão e eficiência. (Vashisht et al., 2025)

Neste contexto, a taxonomia refere-se à classificação sistemática de entidades com base em relações hierárquicas ou associativas com o objetivo de obter os valores correspondentes aos atributos identificados como necessários para cada tipo de documento, valores esses que muitas vezes estão contidos no texto dos documentos.

Para obter os valores dos atributos de forma automatizada, recorre-se ao conceito de entidade, que se refere a qualquer elemento de dados com significado autónomo e relevância semântica, que possa ser identificado, extraído e classificado automaticamente: São exemplos de entidades:

- Entidade pessoas (ex.: nome do requerente, técnico responsável pelo processo, dirigente máximo do serviço);
- Entidade serviços (ex.: empresas públicas, hospitais, escolas tribunais);
- Entidade geográfica (ex.: cidades, países, moradas, distrito, concelho, freguesia, código postal);
- Entidade tempo (ex: data de entrada do documento, prazo de resposta, período de vigência do contrato, data de publicação em Diário da República, data de publicação);
- Entidades Temáticas (ex.: rubrica orçamental, classificação económica);
- Entidade imposto (ex.: IVA, IRS, IUC, IRC).

A definição de entidades numa taxonomia permite reconhecer padrões e extraer dados estruturados a partir de documentos não estruturados (como PDF, imagens digitalizadas ou *e-mails*) e associar automaticamente esses dados a atributos que serão depois utilizados no processo de classificação documental automatizado.

3.1.1.1. Criação da Taxonomia Documental

A taxonomia documental permite definir uma estrutura conceptual que orienta a categorização sistemática dos dados a extraír de cada documento. Trata-se de um instrumento que, para além de contribuir para organizar a informação, permite também definir os atributos frequentes de cada tipo documental, facilitando a sua identificação e tratamento automatizado.

É através dela que se estabelece, de forma estruturada, o conjunto de elementos que devem ser reconhecidos e extraídos de cada tipo de documento. Esta definição orienta diretamente a construção das expressões regulares e dos modelos de aprendizagem automática, permitindo identificar padrões e atributos com maior precisão.

Importa referir que um mesmo tipo de documento pode corresponder a múltiplas classificações no âmbito do classificador documental MEF. Por exemplo, o tipo documental “Ofício” pode ser enquadrado em diversas classes de nível 1 do MEF (*cf.* Tabela 1), dependendo do seu conteúdo e contexto administrativo. Apesar dessa diversidade classificatória, é possível identificar um conjunto mínimo de atributos recorrentes que caracterizam este tipo de documento, como: [Assunto], [Número do ofício], [Data de emissão] e [Entidade remetente].

A definição e sistematização destes atributos é essencial para a automação do processo documental, pois permite que os algoritmos de aprendizagem automática reconheçam padrões e extraiam dados relevantes. Assim, a taxonomia documental proposta é responsável por mapear, para cada tipo de documento, o conjunto de atributos a extrair de forma automatizada, constituindo-se como um elemento estruturante do modelo para a obtenção dos atributos necessários para introdução na interface do SGD.

A taxonomia documental definida é uma estrutura hierárquica e sistemática de classificação que organiza os tipos de documentos e os seus atributos de forma lógica, com o objetivo de facilitar a extração automatizada de informação dos documentos para que seja possível, com recurso a RPA, automatizar a criação e encaminhamento das entradas documentais. Numa fase posterior, será realizada a classificação documental por algoritmos de aprendizagem automática, onde o tipo de documento e a taxonomia definida não têm impacto na classificação documental.

A taxonomia proposta é descrita brevemente no Anexo D e a título ilustrativo, na Figura 4 , onde se exemplificam as classes correspondentes aos tipos de documento considerados e os atributos mínimos da respetiva classe, tendo o propósito de guiar a extração sistemática dos atributos necessários para a integração com sistemas de gestão documental através de RPA. São considerados os princípios de organização funcional, categorização temática e normalização de metadados, para permitir escalabilidade e a adaptação, bem como considerados três níveis hierárquicos:

- Grupo funcional, para representar a macroestrutura organizacional, neste caso designada por “MF” (Ministério das Finanças), que agrupa os documentos por afinidade funcional;
- Categoria documental, para agrupar os documentos por natureza e finalidade, como por exemplo:
 - Documentos Administrativos;
 - Documentos Jurídicos e Normativos;
 - Documentos Financeiros...;
- Tipo Documental, para identificar o tipo específico de documento, como por exemplo, Despacho, Ofício, Fatura, Contrato, Portaria, entre outros.

Cada tipo de documento considera o conjunto mínimo de atributos a extraer que são considerados nas etapas de extração automatizada com recurso a mecanismos de extração de dados baseados em expressões regulares e algoritmos de aprendizagem automática.

A estrutura modular da taxonomia permite evolução manutenção e correção, com a adição de novos tipos de documentos, categorias ou atributos, sem comprometer a integridade do modelo.

3.1.1.2. Definição de Classes Documentais do Nível 1 da MEF

A definição das classes documentais a considerar no modelo é realizada em conformidade com a MEF, que organiza os documentos em três níveis hierárquicos: funções (nível 1), subfunções (nível 2) e atividades (nível 3).

A abordagem adotada neste trabalho privilegia, por razões de simplificação, a implementação inicial do nível 1 da MEF, que contém 14 classes documentais. Esta escolha permite uma primeira validação do modelo, assegurando a escalabilidade futura para os dois níveis de classificação subsequentes.

Esta escolha seguiu duas etapas metodológicas:

- Análise do classificador MEF — foi realizada uma leitura crítica e sistemática da estrutura do classificador documental, com o objetivo de identificar as categorias funcionais mais representativas da realidade administrativa dos serviços públicos;
- Mapeamento das práticas existentes — foram analisadas as práticas de classificação documental já implementadas no SGD para identificar padrões recorrentes e lacunas na categorização atual.

A tabela 1 apresenta a distribuição das classes documentais de nível 1, com base numa amostra de 116 516 documentos classificados, evidenciando a predominância das funções “Organização e Funcionamento” (000), “Planeamento” (010) e “Consultadoria, Assessoria Técnica e Contencioso” (020).

Tabela 1. Frequência de classificações documentais de Nível 1 na amostra inicial.

Código	Classificador de nível 1	Contagem
000	Organização e Funcionamento	58 211
010	Planeamento	31 581
020	Consultadoria, Assessoria Técnica e Contencioso	9 433
030	Atividades de Controlo	5 423
040	Relações Institucionais	4 070
050	Comunicação, Marketing e Relações-Públicas	2 511
060	Informação e Documentação	1 815
070	Sistemas e Tecnologias de Informação e Comunicação	806
080	Recursos Humanos	692
090	Património Instalações e Recursos Materiais e Animais	637
100	Gestão Orçamental e Contabilística	619
110	Processo Legislativo e Regulamentação	464
200	Tesouro e Finanças	245
225	Administração e Emprego Público	9
Total		116 516

Considerando a mesma amostra, na classificação de nível 2 (subfunções) são identificadas 79 classificações distintas e, no nível 3 de classificação, são identificadas 190 classificações (Anexo A e Anexo B).

3.1.1.3. Definição de Tipos de Documentos

Com base na taxonomia proposta, detalham-se os atributos mais relevantes para os tipos de documentos mais frequentes e que se propõe serem tratados de forma automatizada pelo modelo.

São exemplo de tipos de documentos frequentes os ofícios, requerimentos, pareceres técnicos, despachos, informações internas, contratos, faturas, relatórios técnicos, atas de reuniões, memorandos, requerimentos.

Para cada tipo de documento foi identificado o conjunto mínimo de atributos a recolher para efeitos de processo documental, como se detalha no Anexo D e se exemplifica na Figura 4.

3.1.1.4. Identificação e Definição de Atributos

A segunda etapa do modelo foca-se na identificação e definição dos atributos relevantes para cada tipo de documento. Estes atributos são posteriormente extraídos para facilitar a classificação, indexação e encaminhamento dos documentos.

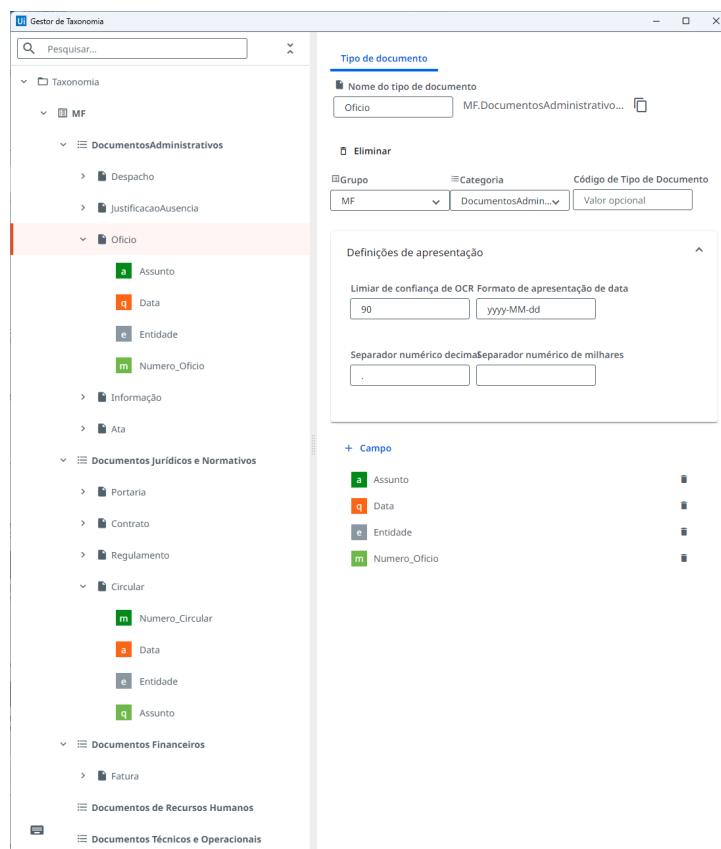


Figura 4. Exemplo da taxonomia e do conjunto mínimo de atributos a extraer.

3.1.1.5. Definição da Lista de Atributos

Para cada tipo de documento identificado na etapa anterior é estabelecida uma lista mínima de atributos a considerar no processo de extração automatizada. Pese embora se trate de uma lista reduzida, deve, todavia, ser suficientemente abrangente para permitir obter o conjunto mínimo de dados necessários para a criação e encaminhamento do processo documental.

Esta abordagem considera a análise funcional dos documentos, a sua natureza, finalidade e contexto administrativo. Os atributos foram organizados por classe de tipo de documento, de acordo com a estrutura da taxonomia. Cada atributo é descrito de acordo com um conjunto de características que permitem a sua correta identificação e extração, nomeadamente, o tipo de dado específico, como texto, número, data ou nome, o que permite determinar o formato e a natureza da informação a extrair. É também definido um formato de apresentação dos dados, particularmente útil para atributos dos quais se espera extrair datas ou números, cujo formato é normalizado no ato da extração.

Para dar resposta à extração de valores múltiplos para um atributo, a taxonomia prevê a possibilidade de um atributo conter múltiplos valores, o que pode ser útil em documentos que, por exemplo, incluem listas de participantes ou rubricas orçamentais.

Por fim, cada atributo é caracterizado com um conjunto de regras de derivação e validação de apoio à extração automatizada, regras essas que combinam expressões regulares e limiares de confiança. A lista de atributos definida na taxonomia proposta foi concebida para ser modular e extensível, bem como para permitir adaptação a novas classes de tipos de documentos ou alterações nos atributos, com vista a assegurar conformidade com ajustes nos requisitos do SGD.

3.1.1.6. Definição de Padrões e Expressões Regulares

Para complementar a extração automática de atributos com recurso a modelos de aprendizagem automática, o modelo proposto contempla uma etapa adicional para obter os atributos também através de expressões regulares, de modo a identificar informações específicas nos documentos. As expressões regulares utilizadas estão descritas de forma exaustiva no (cf. Anexo F), e, como descrição sumária, a Figura 5 exemplifica as expressões regulares implementadas.

Figura 5. Exemplo de expressões regulares para complementar a extração de atributos.

Propõe-se que as expressões regulares utilizadas para a extração de atributos sejam sistematicamente testadas, validadas e refinadas com base em amostras representativas dos documentos processados. Este processo deve seguir uma abordagem iterativa e estruturada, integrando ciclos de avaliação e melhoria contínua, de modo a garantir a precisão e a fiabilidade da extração automática. Sempre que se verifique que um atributo não pode ser corretamente extraído, devem ser introduzidas novas amostras nos conjuntos de treino e ajustadas as expressões regulares ou os modelos de aprendizagem automática, assegurando a evolução adaptativa do modelo. Esta prática contribui para a robustez do modelo e para a capacidade de responder às variações que ocorrem naturalmente no conteúdo dos documentos tratados.

3.1.2. Processamento e Preparação de Documentos

A terceira etapa do modelo envolve o processamento e preparação dos documentos para a classificação e extração de atributos. Esta fase é crucial para garantir a qualidade e consistência dos dados de entrada para os algoritmos de aprendizagem automática.

3.1.3. Classificação e Segmentação com Algoritmos de Aprendizagem Automática

A classificação e segmentação de documentos são etapas importantes para a organização e gestão eficiente da informação dos documentos na Administração Pública. A aplicação de algoritmos de aprendizagem automática permite automatizar estes processos e reduzir algumas das limitações dos métodos manuais ou baseados em regras fixas. Nesta secção, serão explorados os principais algoritmos e abordagens para a classificação e segmentação documentais, considerando as especificidades do contexto da Administração Pública.

A classificação documental pode considerar categorias ou classes predefinidas para os documentos, com base no seu conteúdo. No contexto deste trabalho, a classificação será realizada de acordo com a taxonomia documental definida na secção 3.1.1. Considerando para o efeito a aplicação de algoritmos de aprendizagem automática supervisionada, que aprendem a partir do conjunto de documentos previamente classificados que serão considerados como dados de treino. Assim, propõe-se que sejam explorados para realizar esta tarefa, os seguintes algoritmos:

- As Redes Neuronais e *Deep Learning* que de acordo com autores como (Silva, C. B., 2023) podem ser úteis neste desafio de tratar grandes volumes de dados não estruturados e extrair características do texto. Os autores salientam o potencial das Redes Neuronais Recorrentes (RNN) ou *Transformers* (e.g., BERT, GPT) como capazes de realçar dependências contextuais e semânticas nos textos dos documentos, elementos que podem contribuir para processos de classificação mais precisos e adequados para taxonomias complexas. (Silva, C. et al., 2023)

- Classificação simbólica baseada em regras, para os casos de documentos mais simples como faturas, contratos, informações, formulários, estes classificadores tiram partido de listas de palavras-chave frequentes por tipo de documento e no caso deste trabalho podem tirar partido da taxonomia definida para a sua aplicação. Esta técnica será combinada com o classificador por aprendizagem automática com recurso a redes neurais, combinando uma abordagem baseadas em regras ou palavras-chave com modelos de aprendizagem automática com o intuito de conseguir maior interpretabilidade, na classificação. Esta abordagem híbrida é considerada eficaz por autores como Kaibassova et al. (2025), em sistemas de gestão documental que contenham vocabulário técnico, estrutura semântica complexa. (Kaibassova et al., 2025)

A escolha do algoritmo ou dos algoritmos a utilizar, dependerá dos resultados obtidos na fase de teste, tendo como critérios principais o desempenho nas tarefas de classificação documental, medido através de métricas como precisão, *recall* e *F1-score*, bem como pela capacidade de interpretação dos resultados. No contexto da Administração Pública, a interpretabilidade dos modelos é particularmente relevante, uma vez que é essencial compreender os fatores que sustentam as decisões automatizadas, garantindo transparência, rastreabilidade e confiança nos processos.

A segmentação de documentos, embora frequentemente confundida com a classificação, constitui um processo distinto e essencial no tratamento documental, sobretudo no contexto da Administração Pública. Trata-se da divisão de um documento em partes logicamente estruturadas ou semanticamente relevantes, para permitir a leitura e extração de dados. Esta segmentação pode incidir sobre diferentes elementos, como seções textuais, parágrafos, cabeçalhos, tabelas, imagens ou outros componentes visuais, nestes casos será realizada uma correspondência por mancha visual e o RPA irá tirar partido da visão computacional para procurar corresponder os dados aos atributos definidos na taxonomia.

No âmbito deste trabalho, a segmentação documental aplica-se a documentos típicos dos serviços públicos (como despachos, informações, ofícios, requerimentos, faturas, entre outros) e assume três funções principais. Em primeiro lugar, permite a identificação de áreas de interesse, ou seja, a delimitação de zonas do documento onde se concentram dados relevantes para a extração de atributos específicos, como o nome do requerente, o número do processo ou a data do despacho.

Em segundo lugar, contribui para a estruturação de documentos, ao decompor conteúdos complexos em blocos de tratamento mais simples, para simplificar a análise, o tratamento automatizado e a interoperabilidade entre sistemas. Por fim, a segmentação possibilita a extração de conteúdo específico, isolando blocos de texto ou elementos visuais que contêm a informação necessária para os fins administrativos ou analíticos em causa.

Na segmentação de documentos poderão ser aplicadas técnicas de PLN e de Visão Computacional para identificar, com maior precisão, as estruturas internas dos documentos e os atributos relevantes para a extração de informação.

O processo de segmentação será integrado no fluxo de trabalho realizado pelo RPA do modelo proposto. Após a fase de pré-processamento e preparação dos documentos, estes serão submetidos a modelos de classificação documental, para procurar atribuir a cada documento a tipologia apropriada. Nas etapas subsequentes, serão utilizados modelos de segmentação sobre os mesmos documentos, para identificar entidades e extrair os valores dos atributos relevantes.

A combinação de classificação e segmentação pode permitir uma melhor compreensão da estrutura e do conteúdo dos documentos administrativos, assegurando uma preparação mais eficaz para as fases seguintes do modelo, nomeadamente a extração e validação de dados.

3.1.4. Extração de Atributos e Encaminhamento

A extração de atributos permite identificar e recuperar informações específicas dos documentos, por exemplo (entidades nomeadas, datas, números de processo, nome dos requerentes, número do ofício, número da informação, datas, ...). Esta etapa permite a extração automatizada de metadados, que serão depois utilizados na integração com o SGD.

O reconhecimento de entidades nomeadas é realizado com recurso a algoritmos baseados em modelos estatísticos como os Modelos Ocultos de Markov (HMM) e os Campos Aleatórios Condicionais (CRF) e redes neurais, incluindo arquiteturas Bi-LSTM-CRF e modelos baseados em *Transformers*. Estes algoritmos serão treinados para identificar e classificar entidades relevantes nos documentos que irão permitir extraer os valores dos atributos relevantes.

A adaptação do modelo ao vocabulário técnico e à estrutura documental do Serviço Público é essencial uma vez que se tem como pressuposto que os sistemas de gestão documental, as normas e as regras de cada serviço público podem ter diferenças entre si. (Heck, A., 2022)

No caso de documentos em formato de imagem, será aplicado OCR para converter o conteúdo visual em texto pesquisável, este trabalho é realizado no fluxo do RPA numa atividade prévia ao reconhecimento de entidades e extração de atributos.

Após a extração dos atributos, é considerada uma etapa de normalização para uniformizar os formatos dos atributos extraídos (por exemplo das datas, códigos postais ou dos valores numéricos e em atributos que careçam de um cuidado especial com a sua validação como é o caso dos Número de Identificação Fiscal (NIF), do Número Internacional de Conta Bancária (IBAN) ou da Classificação Económica da Despesa. Esta etapa permite melhorar a correspondência e integridade dos dados extraídos e responder à necessidade de conformidade com padrões normativos ou até a comparação com bases de dados, para assegurar conformidade e fiabilidade da informação extraída.

O encaminhamento automatizado de documentos no SGD existente no Serviço Público é assegurado pelo RPA que permite a integração dos dados extraídos com a interface gráfica SGD existente no serviço, desta forma é possível realizar o direcionamento ou encaminhamento de um documento para um departamento, ou trabalhador de acordo com fluxo de trabalho previsto, esta é uma etapa que permite a redução do tempo de processamento manual, e contribui para o aumento da eficiência operacional.

As regras de encaminhamento são definidas de acordo com as práticas em vigor para o encaminhamento dos processos documentais do Serviço Público e podem variar de acordo com regras, (por exemplo, "se o tipo de documento for 'requerimento' ou fatura, encaminhar para o departamento de gestão de recursos" se o tipo de documento é ofício reencaminhar para a direção) estas regras podem ser mais elaboradas (por exemplo, se o serviço de origem é um Hospital, encaminhar para o departamento de acompanhamento setorial da saúde).

O modelo pressupõe a integração com os sistemas de gestão documental e de processos que estiver a ser utilizado pelo Serviço Público com recurso ao RPA para operacionalizar esta tarefa, funcionando como uma camada de automação que interage diretamente com as interfaces dos sistemas existentes, dispensando o desenvolvimento de *webservices*, trabalhos de integração de sistemas ou alterações estruturais na infraestrutura tecnológica existente. Os robôs de RPA são capazes de replicar ações humanas nas interfaces gráficas do sistema documental, como o preenchimento de formulários, a navegação entre menus, o carregamento de ficheiros ou movimentação de documentos entre diretórios. Essa automação permite o registo e encaminhamento dos documentos independentemente do SGD existente.

3.1.5. Supervisão, Validação e Melhoria Contínua

A eficácia do modelo depende de vários fatores, designadamente da capacidade de adaptação e melhoria contínua. Neste sentido, a supervisão técnica por parte das equipas, e a validação sistemática dos resultados são fundamentais para controlar a qualidade e fiabilidade do modelo.

A supervisão será realizada pelos técnicos da equipa que gere o expediente, cuja experiência e conhecimento nesta temática é elevada, para isso recorre-se a amostragem aleatória de documentos classificados automaticamente, de forma a verificar a conformidade com o esperado e identificar eventuais erros de classificação. Esta abordagem, conforme defendido por (Heck, A., 2022) e (Rodríguez & Bezerra, 2020), não só valida a precisão do modelo, como também assegura que é robusto em diferentes contextos operacionais.

Para a validação do modelo consideram-se as métricas de precisão, o *recall* e o *F1-score*, que permitem avaliar o desempenho dos algoritmos de aprendizagem automática nas tarefas de classificação e extração de atributos. (Silva, C. et al., 2023)

A taxa de erro aceitável para o processo de classificação fixa-se num máximo de 20%, sempre que este limiar for ultrapassado, devem ser realizadas ações de melhoria dos conjuntos de treino e teste com novas amostras representativas e retrainando o modelo de aprendizagem automática. (Campos, D., 2023)

A melhoria contínua do modelo proposto inclui um ciclo iterativo de avaliação, ajustamento e revalidação, para permitir que o modelo aprenda com os resultados obtidos e se adapte. Este ciclo pressupõe uma abordagem sistemática, sustentada em evidência empírica e em práticas de supervisão que garantam a fiabilidade e a conformidade dos resultados com os objetivos definidos. (Mattos, T. et al., 2024)

A atualização periódica da taxonomia documental, a revisão das regras de encaminhamento e a reconfiguração dos robôs de RPA são atividades que contribuem para que o modelo se mantenha funcional e conforme com os sistemas de gestão documental em uso. Propõe-se que a supervisão possa ser realizada pelos técnicos da organização, com recurso a amostragem aleatória e solicitação de validação manual dos documentos classificados, para que se possam identificar erros e proceder ao enriquecimento dos conjuntos de treino e teste, sempre que se verifiquem erros superiores ao limiar definido. (Campos, D., 2023), (Rodríguez & Bezerra, 2020)

A adoção de mecanismos de supervisão e melhoria contínua é um contributo para a sustentabilidade, manutenção e a escalabilidade do modelo.

CAPÍTULO 4

Implementação do Modelo

4. Modelo

A implementação do modelo de classificação documental automatizada e encaminhamento de processos é uma etapa importante para a validação da proposta teórica definida nos capítulos anteriores. Esta fase pretende operacionalizar os conceitos definidos, recorrendo a tecnologias de aprendizagem automática e RPA, para demonstrar que a solução proposta tem viabilidade técnica e funcional, tendo por base os fundamentos teóricos e as etapas de desenvolvimento delineadas nos capítulos anteriores. O foco aqui reside na materialização da solução, com a descrição da arquitetura técnica proposta para a preparação e processamento dos documentos, processo de treino e teste dos modelos de aprendizagem automática, técnicas de extração de atributos e, por fim, a integração com o SGD existente através de RPA.

O objetivo é fornecer uma visão abrangente e técnica de como o modelo transita do conceito para a realidade operacional, abordando as escolhas tecnológicas, os desafios práticos e as soluções adotadas para garantir a sua eficácia e robustez.

4.1. Arquitetura Técnica da Solução

A arquitetura técnica da solução foi definida para ser modular, escalável e interoperável e para permitir a integração dos componentes de IA e RPA com o SGD no serviço. O modelo considera componentes de PLN, algoritmos de classificação supervisionada e agentes robóticos de RPA para integrar o resultado da classificação dos documentos e extração de atributos, com SGD em uso, recorrendo às interfaces gráficas existentes.

A classificação documental automatizada tira partido de soluções e *frameworks* existentes na ferramenta de RPA como o “Document Understanding Framework” da plataforma UiPath, que permite modelar o fluxo documental, desde a área de receção dos documentos no expediente, até à extração de atributos. Nos casos em que não é possível ler os caracteres que compõem o documento, é utilizada a tecnologia “Intelligent OCR”, que realiza a conversão de imagens de digitalizações em texto legível, trabalho que é realizado numa fase inicial do processamento do documento para garantir que todos os documentos têm, conteúdo legível e pesquisável.

Para classificar os documentos recorre-se a um classificador baseado em aprendizagem automática combinado com um classificador baseado em palavras-chave, implementado através de uma funcionalidade de aprendizagem automática da plataforma Uipath.

A extração de atributos estruturados é realizada com recurso a expressões regulares, que permite identificar padrões específicos nos documentos, como assuntos, datas, códigos.

Para as atividades de encaminhamento dos documentos, são utilizados módulos lógicos de decisão, que permitem definir regras de encaminhamento com base nos atributos extraídos.

Por fim, a integração com os sistemas de gestão documental existentes é assegurada através da interação direta com as interfaces gráficas existentes e sem necessidade de desenvolvimentos específicos, o que simplifica em muito a interação com o SGD existente. A figura 6 ilustra a visão geral da arquitetura proposta

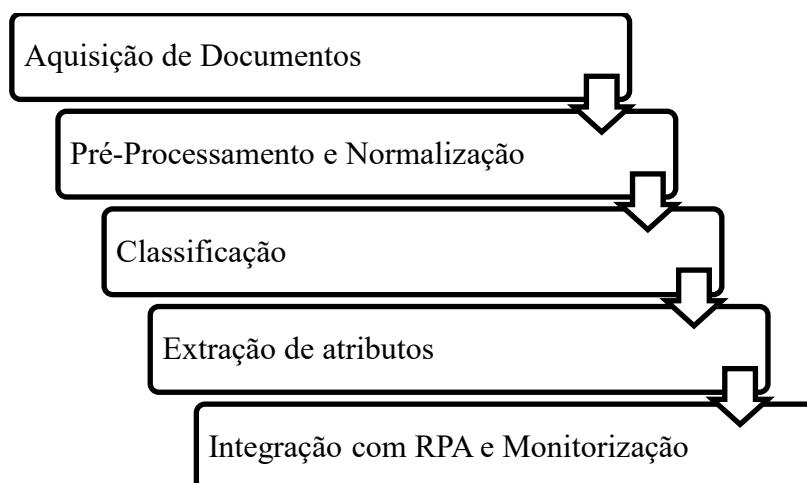


Figura 6. Visão geral da arquitetura técnica proposta.

4.2. Preparação dos Dados

A qualidade e a adequação dos dados de entrada impactam com o desempenho de qualquer modelo de aprendizagem automática. Os documentos utilizados para o treino e teste do modelo foram obtidos a partir de um repositório histórico de documentos de um serviço do Ministério das Finanças, e são representativos de uma variedade de tipos de documentos (ofícios, requerimentos, pareceres, despachos, faturas, etc.) com formatos de ficheiros múltiplos onde se incluem PDF nativos e de conteúdo legível ou resultantes de digitalização que correspondem na prática a imagens resultantes da digitalização de documentos. Sempre que se verifique que o conteúdo de um documento não corresponde a texto legível é aplicado OCR para assegurar que o conteúdo é legível e apto para as etapas que realizam PLN.

A amostra inicial continha um total de 116.516 documentos, que serviu como base para o estudo da distribuição das classificações documentais de acordo com o classificador da MEF. Esta análise permitiu identificar padrões de frequência por classe documental.

Todavia, por razões de simplificação e com o objetivo de acelerar o desenvolvimento inicial do modelo, o conjunto de documentos utilizado para treino e teste foi restringido aos documentos em formato PDF, num total de 57 223 documentos. Esta seleção incluiu também documentos digitalizados que foram previamente tratados com OCR, assegurando que todos os documentos considerados para o treino e teste do modelo estavam aptos para as etapas de vectorização, segmentação e classificação.

A distribuição dos documentos por classe documental do nível 1 da MEF é apresentada e descrita na Tabela 2, verifica-se uma predominância das classes “Recursos Humanos” (080) e “Gestão Orçamental e Contabilística” (100), que refletem áreas com elevado volume de produção documental.

Tabela 2. Distribuição por classe dos documentos para treino e teste do modelo de classificação.

Classe	# Documentos
000 - Organização e Funcionamento	5125
010 – Planeamento	290
020 - Consultadoria, Assessoria Técnica e Contencioso	653
030 - Atividades de Controlo	363
040 - Relações Institucionais	935
050 - Comunicação, Marketing e Relações-Públicas	667
060 - Informação e Documentação	545
070 - Sistemas e Tecnologias de Informação e Comunicação	394
080 - Recursos Humanos	19349
090 - Património Instalações e Recursos Materiais e Animais	3915
100 - Gestão Orçamental e Contabilística	23962
110 - Processo Legislativo e Regulamentação	252
200 - Tesouro e Finanças	660
225 - Administração e Emprego Público	113
Total	57223

Esta decisão metodológica pressupõe que se mantém a qualidade dos dados de entrada e a viabilidade técnica do modelo, sem comprometer a representatividade da amostra global. A amostra completa continua a ser relevante para a análise estatística, para a definição da taxonomia documental e para futuras iterações do modelo, nomeadamente na sua escalabilidade para níveis mais detalhados de classificação documental (nível 2 e 3 da MEF).

4.3. Treino e Teste do Modelo

O processo de treino e teste dos modelos de aprendizagem automática é estruturado em duas etapas, com objetivos distintos.

A primeira etapa classifica de forma automatizada o tipo de documento, para esta tarefa, foi utilizado um conjunto de documentos previamente classificados, que considera tipologias como ofícios, requerimentos, pareceres, faturas, contratos, entre outros (cf. Anexo C), onde se recorre a dois classificadores automatizados, um classificador generativo “*classifyDocument*” da plataforma UiPath, integrado no módulo “*Document Understanding*” que procura reconhecer padrões semânticos no texto dos documentos e avança com uma classificação do tipo de documento.

Para a classificação do tipo de documento combina-se um classificador generativo com um classificador inteligente treinado com um conjunto de documentos pré-classificados do SGD.

O processo de treino é feito com recurso a um assistente com a interface gráfica. O classificador atribui pesos às palavras-chave com base na sua frequência e relevância dentro do conjunto de treino, construindo um modelo que avalia a proximidade semântica entre modelo treinado e o conteúdo dos documentos submetidos para classificação.

Se ambas as classificações automatizadas do tipo de documento forem congruentes o fluxo avança para a fase seguinte, caso contrário o documento será marcado para tratamento manual.

A escolha desta abordagem para a classificação do tipo de documento está relacionada com uma relação tripartida entre os tipos de documentos, a taxonomia e os atributos considerados para extração automatizada. Esta etapa do modelo será supervisionada por técnicos da equipa que gere as entradas e saídas de documentos da organização, aqui, é fundamental garantir a simplicidade de atualização destas componentes do modelo são relevantes para a autonomia desta equipa na manutenção e evolução do modelo.

A segunda etapa corresponde à atribuição da classificação documental de acordo com o classificador da MEF, ao nível 1. Foi considerada uma amostra de documentos (cf. Tabela 2) para treinar o modelo, estes documentos foram extraídos do SGD e agrupados pela respetiva classe documental do classificador da MEF (cf. Tabela 1).

Optou-se pela utilização do algoritmo *Naive Bayes*, escolhido para o treino do modelo de classificação pela sua simplicidade e eficiência, como demonstrado por (Ting, S et al., 2011) num estudo comparativo de classificadores. Nesse trabalho, os autores evidenciam que o *Naive Bayes* apresenta um desempenho competitivo, face a métodos mais complexos, e demonstraram que é eficaz em tarefas de classificação documental. O treino no modelo foi realizado em *Python* (cf. Anexo K) por uma questão de performance. O modelo pré-treinado é depois utilizado no fluxo de trabalho do robô na plataforma UiPath.

4.4. Extração de Atributos

A extração de atributos permite identificar e recuperar de forma automatizada, os valores correspondentes aos atributos de cada um dos tipos de documento previstos, de acordo com o definido na taxonomia definida para cada um dos documentos a processar. Esta atividade é complementar à classificação documental do MEF, na medida em que para processar um documento num SGD é necessário alem da classificação documental, extrair um conjunto de metadados necessários ao encaminhamento do processo.

Conforme detalhado no Anexo H, a operacionalização desta etapa é realizada com recurso a técnicas de PLN, expressões regulares e modelos de aprendizagem automática supervisionada, que permitem reconhecer padrões e entidades específicas nos documentos. Os atributos a extraer foram definidos na taxonomia documental descrita no ponto 3.1.1.

Nos casos em que os documentos não permitem uma leitura do seu conteúdo por se encontram em formato imagem resultando de digitalização, é aplicada uma etapa prévia de reconhecimento ótico de caracteres, caso seja necessário, para converter o conteúdo do documento em texto pesquisável.

A extração dos atributos é realizada por dois mecanismos complementares:

- Modelos de reconhecimento de entidades nomeadas, treinados com dados representativos dos tipos de documentos frequentes no SGD, que permitem identificar entidades como assuntos, nomes de serviços, locais, datas e valores.

No exemplo descrito na figura 7 exemplifica-se a marcação visual dos atributos e respetiva correspondência com o previsto na taxonomia que um modelo de aprendizagem automática utiliza para aprender a identificar através das manchas visuais os valores dos atributos mapeados;

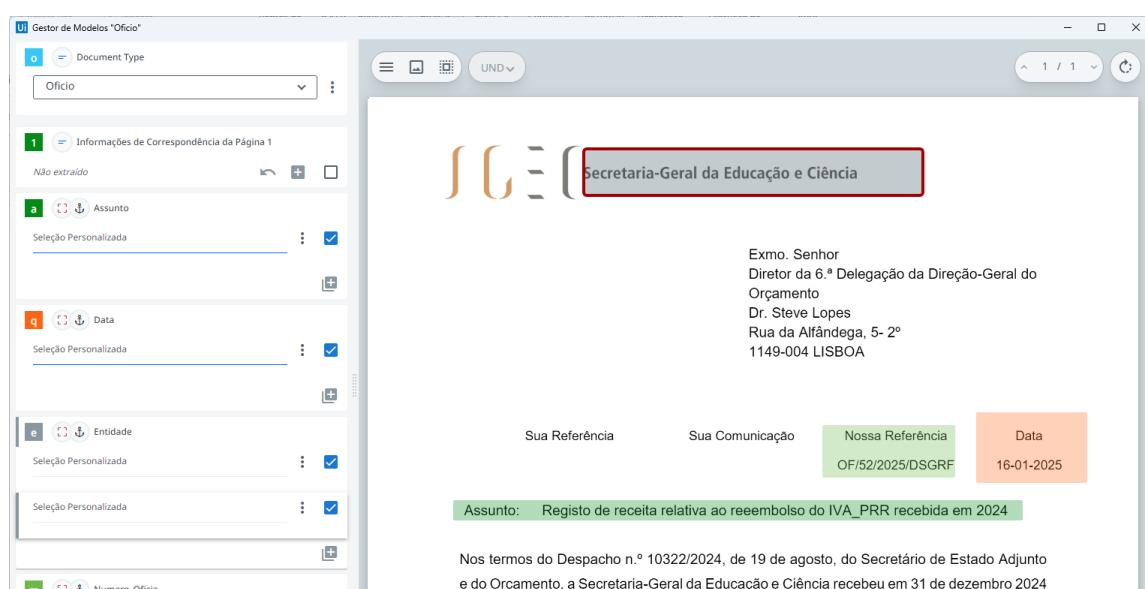


Figura 7. Exemplo de configuração da extração de atributos por tipo de documento previsto na taxonomia.

- Expressões regulares, para identificar atributos com formatos bem definidos e previsíveis, como datas, números de identificação fiscal, códigos orçamentais ou referências documentais. Foram definidas expressões regulares conforme mencionado no Capítulo 3.

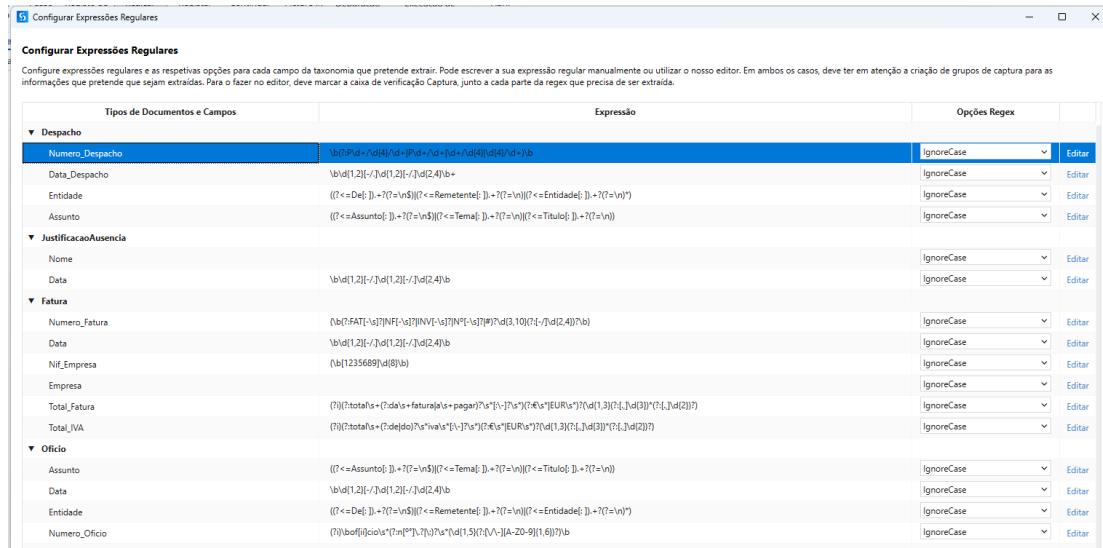


Figura 8. Exemplo de configuração de expressões regulares para extração de atributos.

Após a extração, é realizada uma etapa de normalização dos dados, para promover uniformidade dos formatos e validação de integridade dos atributos extraídos. Esta normalização ajuda na conformidade com padrões esperados para as datas, códigos postais, NIF ou classificadores económicos e Orçamentais.

Os atributos extraídos são depois utilizados para preenchimento dos campos apresentados no formulário do SGD, através da interação dos robôs RPA com as interfaces gráficas existentes. Esta etapa permite obter os valores dos atributos necessário para automatizar o registo e encaminhamento dos documentos, sem necessidade de desenvolvimentos adicionais no SGD em uso.

A eficácia da extração será avaliada com base em métricas de precisão e cobertura, comparando os resultados obtidos com conjuntos de dados anotados manualmente. O modelo considera como critério de sucesso uma taxa de extração correta superior a 80%, sendo prevista a supervisão técnica e o enriquecimento dos conjuntos de treino sempre que este limiar for inferior ao definido.

4.5. Encaminhamento Automatizado

O encaminhamento automatizado permite que após as etapas de classificação do tipo de documento e extração dos atributos relevantes de acordo com a taxonomia definida, e após a etapa de classificação documental de acordo com o Nível 1 do classificador do MEF, os documentos sejam direcionados para os destinatários de acordo com os fluxos de trabalho existentes no SGD, e com base em regras de negócio previamente definidas.

Conforme descrito no Anexo I, a operacionalização do encaminhamento é assegurada com recurso a RPA, que interage diretamente com as interfaces gráficas dos sistemas de gestão documental. Esta abordagem evita alterações estruturais no sistema existente e assegura uma integração não intrusiva e adaptável.

As regras de encaminhamento consideram como critérios o tipo de documento, a entidade remetente, o serviço de origem ou atributos específicos extraídos do conteúdo dos documentos. Por exemplo, documentos classificados como “Requerimento” ou “Faturas” são encaminhados para o Departamento de Recursos. Estas regras podem considerar uma construção de expressões lógicas condicionais compostas, como:

Se o tipo de documento for “Ofício” e a entidade remetente contiver na designação o termo “hospital”, encaminhar para o departamento de acompanhamento setorial da saúde.

A flexibilidade do modelo permite que as regras de encaminhamento sejam ajustadas, todavia, a decisão tomada na Entidade Orçamental (EO), passa por um encaminhamento específico para um grupo de gestão de entradas, o que significa que, apesar do modelo implementar o encaminhamento de forma automatizada, está apenas a ser considerado um conjunto mínimo de regras de encaminhamento. Todavia o trabalho desenvolvido permite a aplicação de diferentes critérios de encaminhamento como descrito no (Anexo F). A especialização das regras de encaminhamento deve respeitar a cultura e as práticas da organização, que consideram regras de encaminhamento por (tipo de documento, entidade remetente, classificação documental MEF do nível 1, regras compostas, atributos específicos) como caracterizado no (Anexo E).

CAPÍTULO 5

Avaliação do Modelo

A avaliação tem como objetivo fornecer um método estruturado para avaliar a eficácia e a precisão do modelo de classificação automatizada de documentos. O modelo proposto é composto por quatro etapas que consideram a definição de tipos de documentos, preparação dos documentos, classificação e segmentação, e extração de atributos.

5. Objetivos da Avaliação do Modelo

Os principais objetivos desta avaliação são verificar a precisão da classificação dos documentos, avaliar a eficácia dos algoritmos de aprendizagem automática utilizados, garantir a qualidade da extração de atributos e identificar e corrigir possíveis vieses no modelo.

Para avaliar a eficácia do modelo de classificação documental automatizado, importa definir um conjunto de métricas e indicadores para medir seu desempenho de forma precisa e objetiva, designadamente no que diz respeito ao desempenho e ao resultado obtido.

Quanto ao desempenho, importa considerar uma medida de precisão que permita evidenciar a proporção de documentos corretamente classificados em relação ao total de documentos classificados pelo modelo e uma medida de performance que permita avaliar o tempo médio necessário para classificar um documento.

Relativamente ao resultado interessa comparar a precisão do modelo com a precisão da classificação manual e o tempo médio do processo robotizado quando comparado com o processamento de um documento de forma manual, para que se possa estimar a redução nos custos operacionais associados à utilização do modelo.

5.1. Metodologia de Avaliação

Para garantir a qualidade e melhoria contínua do modelo de classificação documental automatizada, criação e encaminhamento de processos, com recurso a aprendizagem automática e RPA aplicada a processos documentais na Administração Pública, preconiza-se a realização de uma supervisão técnica com revisão manual de uma amostra aleatória de documentos classificados. A taxa de erro de classificação deve ser inferior a 20%. Se ocorrerem erros em proporção superior a este limiar, devendo os dados de treino e teste ser reforçados com mais amostras para aumentar o volume e a variabilidade e consequentemente melhorar o resultado da classificação automática, considerando a avaliação como um

processo cíclico de quatro etapas a fim de validar a precisão de classificação documental e a eficiência no encaminhamento dos processos documentais do modelo proposto.

A primeira etapa consiste em avaliar globalmente a precisão da taxonomia definida. Para isso, a equipa com a responsabilidade da gestão do registo de entradas de documentos, realiza por amostragem uma revisão, para verificar se todos os tipos de documentos relevantes foram considerados. A taxonomia documental definida deve garantir uma cobertura abrangente dos tipos de documentos previstos, assegurando a sua representatividade e utilidade prática no contexto operacional.

Na segunda etapa, o objetivo é garantir que os documentos estejam em um formato legível e pronto para processamento. Será aplicada a técnica de OCR (Reconhecimento Ótico de Caracteres) em uma amostra de documentos, seguida de uma verificação manual da precisão.

A terceira etapa envolve a avaliação da precisão dos algoritmos de classificação e segmentação. Serão utilizadas métricas como precisão, *recall* e *F1-score* em um conjunto de dados de teste. Os algoritmos devem atingir uma precisão mínima de 80% e um *F1-score* de pelo menos 80% para serem considerados eficazes como defendido por autores como Jiang et al., (2021), especialmente quando complementados por supervisão humana ou ciclos de melhoria contínua. A Precisão pretende medir a exatidão das boas classificações ou previsões do modelo e o *Recall* a sensibilidade ou taxa de boas classificações do modelo. (Mariano, 2021) , (Jiang et al., 2021)

Para o cálculo destas métricas considera-se a seguinte metodologia de apuramento:

- Precisão, pretende medir a proporção de documentos bem classificados numa classe.

$$\text{Precisão} = \frac{\text{documentos bem classificados por classe}}{\text{todos os documentos classificados como da classe}} \quad (5.1)$$

- *Recall*, pretende medir a proporção de documentos classificados corretamente.

$$\text{Recall} = \frac{\text{documentos bem classificados por classe}}{\text{todos os documentos que realmente pertencem à classe}} \quad (5.2)$$

- *F1-score*, pretende medir a média harmônica entre precisão e o recall, é como que o equilíbrio entre a capacidade do modelo classificar corretamente os documentos numa classe e a precisão dessa classificação.

$$F1\text{-score} = 2 * \frac{\text{Precisão} * \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (5.3)$$

Na quarta etapa, o objetivo é verificar a precisão da extração de atributos com recurso a expressões regulares e modelos de aprendizagem de automática. A comparação dos atributos extraídos será feita com um conjunto de dados anotado manualmente. No que diz respeito à extração de atributos, pretende-se o modelo seja capaz de identificar e extraer corretamente os atributos identificados na taxonomia para cada tipo de documento.

Para assegurar a qualidade dos resultados, será realizada uma supervisão técnica através de amostragem aleatória de documentos classificados e revisão manual por especialistas. Em caso de erros, os conjuntos de treino e teste serão enriquecidos com novas amostras para minimizar o viés.

5.2. Validação da Extração de Atributos

A validação da extração de atributos é efetuada através da comparação entre os dados extraídos automaticamente e os valores anotados manualmente numa amostra representativa de documentos, selecionada especificamente para este efeito.

A supervisão técnica será realizada pela equipa que atualmente gere o registo e controlo das entradas e saídas dos documentos do serviço, que procede à revisão manual dos resultados obtidos com recurso a amostragem aleatória. Esta abordagem permite identificar erros, inconsistências e falhas no processo de extração, para permitir ajustar o modelo e introduzir melhorias incrementais.

Este processo de validação é iterativo e contínuo, e deve ser integrado num ciclo de melhoria para assegurar que os atributos extraídos são conformes com os requisitos do SGD em uso.

5.3. Exceções, Erros e Auditoria

O tratamento de exceções e erros prevê que caso ocorram situações inesperadas, como impossibilidade de extrair atributos necessários à operacionalização da automação do SGD, erros de sistema, níveis de confiança nas etapas de classificação automatizada ou documentos que não se enquadram em nenhuma regra predefinida. Nestes casos, o documento é sinalizado e movido para uma fila de exceções para revisão manual, e neste caso é gerado um alerta por email para a divisão responsável pelo registo das entradas e saídas documentais.

As principais ações realizadas pelo robô são registadas num diário de registo detalhado, com informação relativa ao documento processado, data e hora de execução, tipo de documento e classificação do MEF atribuída, bem como dos atributos extraídos e o resultado do encaminhamento. Qualquer erro ou exceção, são também alvo de registo, indicando o número da atividade, a operação realizada e a data, desta forma é criado um diário de operações que permite auditoria, rastreabilidade e monitorização da execução do modelo.

5.4. Resultados Obtidos

A avaliação dos resultados obtidos com a implementação do modelo de classificação documental automatizada e extração de atributos foi realizada com base em métricas de desempenho e eficácia, considerando os objetivos definidos no capítulo 1 e os critérios de sucesso estabelecidos na metodologia.

A análise dos resultados está organizada em três dimensões:

- Desempenho do Modelo de Classificação Documental;
- Eficiência na Extração de Atributos;
- Impacto Operacional e Redução de Tempo.

5.4.1. Desempenho do Modelo de Classificação Documental

O modelo foi testado com uma amostra de documentos previamente classificados segundo o nível 1 do classificador da MEF. A distribuição por classe documental permitiu aferir o desempenho do modelo em diferentes contextos funcionais. As métricas utilizadas foram:

- Precisão, que mede a proporção de documentos corretamente classificados;
- Recall, que mede a capacidade de o modelo identificar os documentos pertencentes a uma determinada classe;
- F1-score, que fornece uma média harmónica entre precisão e recall, refletindo o equilíbrio entre ambos.

Os resultados (*cf.* Tabela 3) demonstraram uma precisão global de 83%, com variações por classe documental. As classes com maior volume de documentos, como “Gestão Orçamental e Contabilística” e “Recursos Humanos”, apresentaram os melhores desempenhos, enquanto classes menos representadas evidenciaram maior margem de erro, o que reforça a necessidade de enriquecimento dos conjuntos de treino.

Tabela 3. Resultados do teste de classificação aplicado ao conjunto de teste.

Classe (Classificado MEF nível 1)	Precisão	Recall	F1-Score	Documentos
000 – Organização e Funcionamento	0,54	0,85	0,66	1026
010 – Planeamento	0,47	0,29	0,35	49
020 – Consultadoria, Assessoria Técnica e Contencioso	0,28	0,1	0,14	134
030 – Atividades de Controlo	0,62	0,32	0,42	74
040 – Relações Institucionais	0,58	0,5	0,53	204
050 – Comunicação, Marketing e Relações-Públicas	0,91	0,74	0,82	131
060 – Informação e Documentação	0,48	0,54	0,5	99
070 – Sistemas e Tecnologias de Informação e Comunicação	0,81	0,61	0,7	79
080 – Recursos Humanos	0,9	0,9	0,9	3755

Classe (Classificado MEF nível 1)	Precisão	Recall	F1-Score	Documentos
090 – Património, Instalações e Recursos Materiais e Animais	0,83	0,84	0,83	814
100 – Gestão Orçamental e Contabilística	0,9	0,84	0,87	4813
110 – Processo Legislativo e Regulamentação	1	0,41	0,58	54
200 – Tesouro e Finanças	0,72	0,38	0,5	128
225 – Administração e Emprego Público	1	0,48	0,65	21
		Total		11381

O teste de classificação para estudar a amostra de documentos com o modelo de aprendizagem automática, considerou-se uma segmentação da amostra em 80% (45 842) dos documentos para treinar o modelo e 20% (11 381) para testar. Esta distribuição segue a recomendação de (Bishop, 2006) para evitar *overfitting*.

	precision	recall	f1-score	support
000 - Organização e Funcionamento	0.54	0.85	0.66	1026
010 - Planeamento	0.47	0.29	0.35	49
020 - Consultadoria, Assessoria Técnica e Contencioso	0.28	0.10	0.14	134
030 - Actividades de Controlo	0.62	0.32	0.42	74
040 - Relações Institucionais	0.58	0.50	0.53	204
050 - Comunicação, Marketing e Relações Públicas	0.91	0.74	0.82	131
060 - Informação e Documentação	0.48	0.54	0.50	99
070 - Sistemas e Tecnologias de Informação e Comunicação	0.81	0.61	0.70	79
080 - Recursos Humanos	0.90	0.90	0.90	3755
090 - Património Instalações e Recursos Materiais e Animais	0.83	0.84	0.83	814
100 - Gestão Orçamental e Contabilistica	0.90	0.84	0.87	4813
110 - Processo Legislativo e Regulamentação	1.00	0.41	0.58	54
200 - Tesouro e Finanças	0.72	0.38	0.50	128
225 - Administração e Emprego Público	1.00	0.48	0.65	21
accuracy			0.83	11381
macro avg	0.72	0.56	0.60	11381
weighted avg	0.84	0.83	0.83	11381

Figura 9. Resultado de teste de classificação.

A avaliação do desempenho do modelo de classificação foi realizada com base numa amostra de 11 381 documentos, distribuídos pelas 13 classes funcionais de nível 1 da MEF. Os resultados obtidos são apresentados através das métricas de precisão, *recall*, *F1-score* e suporte, permitindo uma análise da eficácia do modelo em cada classe documental.

A precisão global do modelo foi de 83%, o que indica que a maioria dos documentos foi corretamente classificada. A análise por classe revela variações significativas no desempenho do modelo. A classe 000 – Organização e Funcionamento apresenta um *recall* elevado (0.90), o que significa que o modelo consegue identificar corretamente a maioria dos documentos desta categoria. No entanto, a precisão mais baixa (0.54) sugere que muitos documentos de outras classes estão a ser mal classificados e como pertencentes a esta classe, o que pode indicar uma tendência de sobreclassificação ou uma definição excessivamente abrangente desta categoria. Este fenómeno pode também ser explicado por más classificações manuais realizadas no sistema de onde foram extraídos os documentos e os dados.

A classe 010 – Planeamento apresenta uma precisão perfeita (1.00), mas um *recall* moderado (0.64), o que pode indicar que, embora o modelo seja rigoroso ao atribuir documentos a esta classe, ainda falha em identificar todos os documentos que lhe pertencem. Este padrão repete-se em outras classes com menor representatividade na amostra, como 020 – Consultadoria, Assessoria Técnica e Contencioso, onde o equilíbrio entre precisão e revocação é mais instável.

As médias não ponderadas das métricas indicam um desempenho médio da precisão por classe, o que sugere um desempenho aceitável, mas com margem para melhoria, como sugerido no ponto 5.6, onde se propõe que as más classificações ou as classificações dos documentos com uma precisão reduzida, sejam propostos para classificação manual e reutilizados para alimentar os conjuntos de documentos utilizados para treino e teste do modelo, para reforçar a capacidade de identificar corretamente documentos de classes menos representadas.

Se considerarmos as médias ponderadas e o número de documentos por classe, os resultados são melhores (precisão de 0,84, *recall* de 0,83 e *F1-score* de 0,3), evidenciando um desempenho robusto nas classes com maior volume de amostras.

Estes resultados sugerem que o modelo possa ser eficaz na classificação de documentos mais frequentes, mas que a sua performance necessita de ser melhorada nas classes com menor suporte. Isto corrobora a estratégia de enriquecimento dos conjuntos de treino quando a precisão das classificações individuais for inferior a 0,80, incluindo-se no modelo, uma atividade para este efeito.

5.4.2. Eficiência na Extração de Atributos

A extração de atributos foi realizada com recurso a expressões regulares e modelos de reconhecimento de entidades nomeadas. A taxa de sucesso na extração correta dos atributos relevantes por tipo documental foi superior a 95%, cumprindo o critério de qualidade definido. A normalização dos dados extraídos contribuiu para a consistência dos metadados e para a conformidade com os requisitos dos sistemas de gestão documental.

5.4.3. Impacto Operacional e Redução de Tempo

A implementação do modelo de classificação documental automatizada, com recurso a aprendizagem automática e RPA, revelou um elevado impacto operacional, sobretudo na redução do tempo médio de processamento por documento, quando comparando com o processo manual, estimado em cinco minutos por documento para (identificação do tipo de documento, classificação documental, extração de atributos e encaminhamento), o modelo automatizado demonstrou uma redução do tempo em cerca de 70% para realizar as mesmas tarefas, o que representa ganhos relevantes em termos de produtividade e libertação de recursos humanos para tarefas de maior valor acrescentado.

5.5. Eficiência Operacional

A eficiência operacional do modelo proposto foi avaliada com base em dois critérios, o tempo médio de processamento por documento e a precisão da classificação documental. Estes indicadores ajudam a aferir o impacto da automatização na gestão documental, comparando o desempenho do modelo com os procedimentos manuais atualmente em vigor.

Atualmente, o processo manual de registo, extração de atributos, classificação documental e encaminhamento de documentos apresenta um tempo médio estimado de cinco minutos por documento, desde a receção até ao registo e encaminhamento no SGD. Este tempo inclui as etapas de identificação do tipo de documento, extração de atributos, classificação documental, preenchimento dos metadados e encaminhamento para o departamento responsável pelo tratamento do documento.

Com a implementação do modelo automatizado, que combina aprendizagem automática e RPA, estimou-se uma redução de pelo menos 50% no tempo de processamento, mas foi possível alcançar uma redução superior, o que representa ganhos em termos de produtividade e libertação de recursos humanos para tarefas de maior valor acrescentado.

Em termos de precisão da classificação documental, os testes realizados com uma amostra representativa de documentos apresentam um resultado de taxa de acerto de 83%, o que indica que a maioria dos documentos é corretamente classificada pelo modelo. Este valor supera a média de precisão observada nos processos manuais, estimada em 80%, e aproxima-se do limiar definido para o modelo (erro inferior a 20%).

Estes resultados sugerem que o modelo proposto contribui para a melhoria da eficiência operacional, e que permite reduzir o tempo de processamento dos documentos, com um nível de precisão da classificação equivalente ao atual e assegurar consistência na extração de atributos e no encaminhamento dos documentos. A integração com o SGD através de RPA permite ainda uma implementação não intrusiva, sem necessidade de alterações nos sistemas existentes. O processo robotizado inclui mecanismos de melhoria continua, e marca para processamento manual os documentos em que a confiança na classificação é inferior a 80% ou nos casos em que não é possível extrair os atributos identificados como necessários na taxonomia para o tipo de documento que está a ser tratado, esta abordagem de melhoria continua permite melhorar o resultado do modelo, reutilizando as classificações ou extrações de atributos supervisionadas como novos dados para alimentar o conjunto de treino e de teste, fatores que reforçam a viabilidade técnica e funcional da solução.

5.6. Supervisão e Melhoria Contínua

A sustentabilidade do modelo é apoiada por um ciclo sistemático de melhoria contínua para manter o modelo funcional, e para reagir a eventuais necessidades de ajuste.

A supervisão será realizada pela equipa que atualmente gere o registo e controlo das entradas e saídas de documentos que deverá aferir a conformidade do resultado das classificações automatizadas com o esperado, numa base mensal durante o primeiro ano de implementação do modelo, através de amostra aleatória 20 documentos, passando a validar a conformidade dos resultados com os critérios definidos em cada trimestre nos anos seguintes.

Esta abordagem permite identificar erros de classificação, inconsistências na extração de atributos e desvios nas regras de encaminhamento, e funciona como mecanismo de controlo de qualidade, esta é uma prática recomendada como referem alguns autores, como Heck (2022) e Rodríguez & Bezerra (2020) que sugerem também um acompanhamento das métricas de precisão, *recall* e *F1-score* dos algoritmos de classificação, para acompanhar e avaliar o desempenho dos algoritmos de aprendizagem automática uma vez que o modelo prevê a adição de amostras aos conjuntos de treino e teste e o “re-treino” dos modelos. Considera-se como aceitável uma taxa de erro global inferior ao limiar definido, e sempre que este limiar for ultrapassado, deverão ser iniciados trabalhos de melhoria, designadamente o enriquecimento dos conjuntos de treino e teste de forma complementar com novas amostras que incorporem volume e variabilidade.

Para acompanhamento da performance do modelo considera-se de forma iterativa ciclos compostos por quatro etapas:

- Avaliação dos resultados obtidos;
- Identificação de erros nos processos de classificação, extração ou encaminhamento;
- Ajuste no conjunto de treino e teste e caso necessário, nas regras de encaminhamento;
- Revalidação dos resultados de classificação e encaminhamento.

A atualização da taxonomia documental, a revisão das regras de encaminhamento bem como a reconfiguração das atividades do modelo de RPA são atividades que compõem este processo de melhoria.

CAPÍTULO 6

Conclusões

Este trabalho teve como objetivo desenvolver e validar um modelo de classificação documental automatizada, com capacidade para identificar o tipo de documento, realizar a classificação documental, considerando o classificador de nível 1 da MEF, extrair atributos de forma automatizada dos documentos e encaminhar de forma automatizada os processos documentais com recurso às interfaces gráficas dos sistemas de gestão documental existentes. Para isso foram utilizadas técnicas de aprendizagem automática e de RPA.

A investigação partiu de um problema concreto da Administração Pública, relacionado com a elevada carga de trabalho associada à gestão documental, a heterogeneidade dos documentos e a necessidade de maior eficiência e rastreabilidade. A metodologia adotada combinou revisão bibliográfica, desenvolvimento técnico e validação empírica, sustentando a proposta num modelo híbrido que alia aprendizagem automática e automação robótica.

Na revisão bibliográfica com recolha e análise de contributos teóricos e práticos sobre classificação documental, aprendizagem automática e RPA, procede-se à caracterização do estado da arte e das lacunas existentes, o que suportou a escolha metodológica e a abordagem tecnológica do modelo proposto, reforçando a pertinência da abordagem híbrida adotada.

Ao que se seguiu a caracterização da metodologia e detalhe do processo de desenvolvimento do modelo, incluindo a seleção de algoritmos, a preparação dos dados, bem como a definição de atributos e a arquitetura de integração com RPA. A clareza metodológica permitiu garantir a reproduzibilidade do estudo e a robustez dos resultados.

A precisão obtida pelo modelo, que atingiu 83%, indica que a maioria dos documentos foi corretamente classificada, refletindo a capacidade do algoritmo em identificar padrões relevantes nos dados. Esta métrica está diretamente relacionada com a taxa de erro, uma vez que esta corresponde à proporção de classificações incorretas em relação ao total de documentos processados. Assim, uma precisão de 83% corresponde a uma taxa de erro aproximada de 17%, valor que se encontra abaixo do limiar máximo definido de 20% para considerar o modelo eficiente.

Os resultados obtidos demonstram que o modelo proposto é tecnicamente viável e funcional. A precisão global de 83% na classificação documental e a redução de mais de 70% no tempo médio de processamento por documento evidenciam ganhos em eficiência, consistência e produtividade. Estes indicadores validam a hipótese de que a aplicação de tecnologias de IA e RPA pode transformar positivamente os processos de gestão documental nos serviços públicos em particular no registo e encaminhamento de novas entradas de documentos.

Contudo, a análise crítica dos resultados também evidencia algumas limitações, como o desequilíbrio das amostras por classe para o treino e teste, semelhança semântica entre classes documentais.

As classes com maior volume de documentos, como “Gestão Orçamental e Contabilística” e “Recursos Humanos”, apresentaram resultados robustos, com valores de precisão e F1-score superiores a 0,85. Por outro lado, classes menos representadas na amostra, como “Consultadoria, Assessoria Técnica e Contencioso” ou “Planeamento”, evidenciaram desempenhos mais frágeis.

Estas variações podem ser explicadas por três fatores principais. As amostras de treino e teste não estão balanceadas; a predominância de documentos pertencentes a determinadas classes pode enviesar o modelo, que aprendeu com maior eficácia os padrões dessas classes, em detrimento daquelas menos representadas e pode também existir ambiguidade semântica entre algumas classes, o que também pode explicar os resultados; o terceiro fator a considerar está relacionado com certas funções do classificador da MEF apresentarem sobreposição temática, o que dificulta a distinção clara entre elas, especialmente quando os documentos analisados não utilizam vocabulário técnico específico ou apresentam estrutura pouco definida com reflexo na qualidade das classificações manuais ou até mesmo erros ou inconsistências nas classificações originais que contribuem para introduzir ruido nos dados.

Para melhorar o desempenho do modelo, propõem-se ações como o equilíbrio dos conjuntos de treino e teste, a revisão das classificações manuais com supervisão técnica.

Este trabalho representa um passo importante na transformação digital e modernização da gestão documental nos serviços públicos, ao apontar um modelo automatizado de classificação documental integrado com RPA. Espera-se que a sua implementação melhore a eficiência, reduza erros e aumente a transparência nos processos administrativos. Contudo, para maximizar os resultados, será necessário enriquecer os conjuntos de treino, para garantir maior representatividade das amostras e mitigando vieses que possam comprometer a precisão do modelo. Além disso, a incorporação de abordagens semânticas avançadas, como modelos baseados em *Transformers* (por exemplo, BERT ou GPT), poderá elevar a compreensão contextual dos documentos, aumentando a qualidade da classificação. Outro vetor essencial será a criação de interfaces de supervisão e feedback, permitindo que técnicos validem, corrijam e alimentem o sistema com novos dados, assegurando a melhoria contínua. Em termos de trabalhos futuros, prevê-se a implementação inicial na Entidade Orçamental no primeiro semestre de 2026. Estas fases permitirão aferir a adaptabilidade do modelo, identificar requisitos específicos e promover a sua generalização como solução estratégica para a modernização da gestão documental em diferentes contextos organizacionais.

Os resultados obtidos ao longo do desenvolvimento e validação do modelo validam as hipóteses formuladas que se resumem de seguida:

- H1: Desenvolver um modelo que combine aprendizagem automática e RPA para automatizar o encaminhamento de processos documentais;

Resultado: o modelo desenvolvido demonstra viabilidade técnica e operacional ao combinar aprendizagem automática e RPA

- H2: O modelo desenvolvido apresentar uma taxa de erro global inferior a 20% na classificação documental automatizada e encaminhamento de processos, sendo por isso considerado uma forma eficiente de gerir processos documentais;

Resultado: O modelo alcançou 83% de precisão, que corresponde a uma taxa de erro de 17% que é inferior ao limiar definido na hipótese.

- H3: A implementação do modelo reduzirá o tempo médio de processamento documental em pelo menos 50%.

Resultado: Verificou-se uma redução de 70% no tempo médio de processamento.

O sistema automatizado demonstrou ser capaz de executar, com fiabilidade, as tarefas anteriormente realizadas de forma manual, e reduzir o tempo médio de processamento por documento de cinco minutos para cerca de um minuto e trinta segundos. Esta redução de tempo superior a 70% representa um ganho operacional, com impacto direto na produtividade das equipas e na capacidade de resposta institucional.

O modelo proposto é escalável e adaptável, podendo ser aplicado a diferentes contextos organizacionais mediante ajustes específicos. Constitui uma base para futuras investigações, nomeadamente na integração com sistemas legados, na evolução para classificações de nível 2 e 3 da MEF, e na incorporação de abordagens semânticas mais avançadas.

Em suma, os objetivos definidos foram alcançados e a hipótese de trabalho validada com base em evidência empírica. O modelo proposto caracteriza uma abordagem viável, escalável e alinhada com os princípios da transformação digital, e pode ser um contributo para a modernização dos serviços públicos e para a melhoria da gestão documental. A sua implementação pode representar um avanço na forma como os serviços públicos tratam, classificam e encaminham documentos, libertando as pessoas para tarefas de maior valor.

Referências Bibliográficas

- Anesp, Associação Nacional dos Especialistas em Políticas Públicas e Gestão Governamental (Ed.). (2023). *Inteligência artificial e gestão pública: Mundo em transformação*. Zabelê Comunicação.
- Berryhill, J., Heang, K., Clogher, R., & McBride, K. (2019). *Hello, World: Artificial intelligence and its use in the public sector* (OECD Working Papers on Public Governance No. 36; OECD Working Papers on Public Governance, Vol. 36). OECD. <https://doi.org/10.1787/726fd39d-en>
- Bishop, C. (2006). *Pattern recognition and machine learning*. Springer.
- Bouckaert, G., & Halligan, J. (2007). *Managing Performance* (0 ed.). Routledge. <https://doi.org/10.4324/9780203935958>
- Campos, D. (2023). *O estado da arte na exploração do RPA segundo os estudos publicados: Scoping review* [Dissertação de Mestrado, Universidade Católica Portuguesa]. <http://hdl.handle.net/10400.14/43977>
- Correia, M. (2023). *A inteligência artificial em serviços de informação—Estudo de caso em arquivos Europeus* [Dissertação de Mestrado, Faculdade de Letras da Universidade de Coimbra]. https://estudogeral.uc.pt/retrieve/267034/MagdaCorreia_versaofinal.pdf
- DGAEP. (2024). *Boletim Estatístico do Emprego Público (BOEP)*. Direção-Geral da Administração e do Emprego Público. <https://www.dgaep.gov.pt>
- DGLAB. (2025, março 11). *Classificação e avaliação de documentos*. <https://arquivos.dglab.gov.pt/servicos/classificacao-e-avaliacao/>
- Duarte, A., Boente, R., Marinho, R., & Zouain, D. (2018). Produção Acadêmica sobre Accountability: Categorizações na área de Administração Pública. *Contabilidade, Gestão e Governança*, 21(3), 459–475. https://doi.org/10.21714/1984-3925_2018v21n3a9
- EuropeAid. (2004). *Project Cycle Management Guidelines*. European Commission, EuropeAid Cooperation Office. https://www.cookislands.gov.ck/images/documents/DCD_Docs/Development-

Resources/Project-Logicial_Framework_Approaches/2004-

Europeaid_project_cycle_management_guidelines.pdf

European Agency for Safety and Health at Work. (2024, maio 22). *Strategies for safety and health in an automated world*. https://healthy-workplaces.osha.eu/sites/hwc/files/hwc/publication/Automation%20of%20tasks_infosheet_PT.pdf

European Commission. (2023). *143 AI Cases in the Public Sector are Available as Open Data*. ELISE - European Location Interoperability Solutions for e-Government. <https://interoperable-europe.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/news/143-ai-cases-public-sector-are-available-open-data>

Fernandes, C. & Ganhão, T. (2023, outubro). *O Envelhecimento e o Fator Idade na Gestão de Recursos Humanos*. Direção-Geral da Administração e do Emprego Público.
https://www.dgaep.gov.pt/upload/Estudos/2023/Estudo_Envelhecimento_TG.pdf

Ferreira, N. (2025, junho). *Modernização da Administração Pública: O Exemplo do IMPIC*
Modernization of Public Administration: The Case of IMPIC.
https://www.researchgate.net/publication/392914510_Modernizacao_da_Administracao_Publica_O_Exemplo_do_IMPIC_Modernization_of_Public_Administration_The_Case_of_IMPIC

Hammad, M., Omar, K., Abdul Halim, M., Abid, M., Javed, A., & Zafar, M. (2025). Digital Transformation in Public Sector: Enhancing Employee Motivation and Performance. *Public Organization Review*, 25(3), 1459–1481. <https://doi.org/10.1007/s11115-025-00870-9>

Heck, A. (2022). *Processamento de Linguagem Natural Aplicado a Reconhecimento de Entidades Nomeadas em Textos Legais em Português Brasileiro* [Dissertação de Mestrado, Universidade Federal de Santa Catarina]. <https://repositorio.ufsc.br/handle/123456789/233250>

Idzi, F., & Gomes, R. (2022). Digital governance: Government strategies that impact public services. *Global Public Policy and Governance*, 2(4), 427–452. <https://doi.org/10.1007/s43508-022-00055-w>

IEEE. (2017). *IEEE Guide for Terms and Concepts in Intelligent Process Automation*. IEEE.
<https://doi.org/10.1109/IEEEESTD.2017.8070671>

- Jiang, S., Hu, J., Magee, C., & Luo, J. (2021). *Deep Learning for Technical Document Classification* (Versão 5). arXiv. <https://doi.org/10.48550/ARXIV.2106.14269>
- Kaibassova, D., Mukhametzhanova, B., Tokseit, D., Kubegenova, A., & Kozhanov, M. (2025). Document analysis via combined vectorization and machine learning approaches. *International Journal of Innovative Research and Scientific Studies*, 8(4), 2195–2204. <https://doi.org/10.53894/ijirss.v8i4.8356>
- Lourenço, A. (2012). *É necessário alterar o processo de avaliação arquivística?* Faculdade de Ciências Sociais e Humanas Universidade Nova de Lisboa. https://www.academia.edu/68783267/%C3%89_Necess%C3%A1rio_Alterar_O_Processo_De_Avalia%C3%A7%C3%A3o_Arquiv%C3%ADstica
- Mariano, D. (2021). Métricas de avaliação em machine learning: Acurácia, sensibilidade, precisão, especificidade e F-score. Em D. Mariano, F. Dezordi, P. Martins, J. Xavier, T. Sousa, L. Lima, & L. Santos (Eds.), *BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional* (1.ª ed.). Alfahelix. <https://doi.org/10.51780/978-6-599-275326-15>
- Mattos, T., Neves, A., Silva, F., Medeiros, F., Ferreira, F., Filho, M., & Silva, D. (2024). Automação no setor público: Como implementar de forma eficaz. *Aplicabilidade de Tecnologias na Administração Pública*, 13–17.
- Paul, M. & Longinidou, E. (2024). *AI-Powered Interoperability: Lessons Learned from the European Public Sector Awards*. European Institute of Public Administration (EIPA). https://www.eipa.eu/wp-content/uploads/2025/01/EPSC-Briefing_AI.pdf
- Penteado, P. (2015). Gestão de documentos de arquivo na Administração Pública em Portugal – Experiências e desafios. *Acervo*, 28(2), 121–133.
- Presidência do Conselho de Ministros. (2021, setembro 10). Resolução do Conselho de Ministros n.º 131/2021: Estratégia para a Transformação Digital da Administração Pública 2021-2026. *Diário da República*, 17–34.
- Rios, E. R., & Cordeiro, R. I. D. N. (2010). Plano de classificação de documentos arquivísticos e a teoria da classificação: Uma interlocução entre domínios do conhecimento. *Perspectivas em Ciência da Informação*, 15(2), 123–139. <https://doi.org/10.1590/S1413-99362010000200009>

- Rodríguez, M., & Bezerra, B. (2020). Processamento de Linguagem Natural para Reconhecimento de Entidades Nomeadas em Textos Jurídicos de Atos Administrativos (Portarias). *Revista de Engenharia e Pesquisa Aplicada*, 5(1), 67–77. <https://doi.org/10.25286/repa.v5i1.1204>
- Sanches, A. P. (2024). Os impactos da IA e da automatização na gestão de recursos humanos. *Revista Minerva Universitária*. <https://revistaminervauniversitaria.com.br>
- Santos, L. S., & Lima, W. S. D. (2025). TRANSFORMAÇÃO DIGITAL NA ADMINISTRAÇÃO PÚBLICA BRASILEIRA: INTELIGÊNCIA ARTIFICIAL E BUSINESS INTELLIGENCE EM FOCO. *Revista ft*, 29(151), 49–50. <https://doi.org/10.69849/revistaff/ar10202510261749>
- Schneider, L. (2021). *Oportunidades e desafios da inteligência artificial no setor público: O caso do tribunal de contas de Portugal* [Dissertação de Mestrado, ISCTE – Instituto Universitário de Lisboa]. <http://hdl.handle.net/10071/24312>
- Schröer, C., Kruse, F., & Gómez, J. (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, C. B. (2023). Uma Análise Comparativa das Técnicas de Machine Learning. *Revista de Inteligência Artificial Aplicada*, 5(2), 45–60.
- Silva, C., Gomes, F., Santos, J., & Santos, C. (2023). Uma Análise Comparativa das Técnicas de Machine Learning. *Apoena*, 7, 501–511.
- Syed, R., Suriadi, S., Adams, M., Bandara, W., Leemans, S., Ouyang, C., Ter Hofstede, A., Van De Weerd, I., Wynn, M., & Reijers, H.-. (2020). Robotic Process Automation: Contemporary themes and challenges. *Computers in Industry*, 115, 103162.
<https://doi.org/10.1016/j.compind.2019.103162>
- Teixeira De Toledo, A., & Mendonça, M. (2023). A aplicação da inteligência artificial na busca de eficiência pela administração pública. *Revista do Serviço Público*, 74(2), 410–438.
<https://doi.org/10.21874/rsp.v74i2.6829>
- Ting, S, Hung, H., & Tsang, A. (2011). Is Naïve Bayes a Good Classifier for Document Classification? *International Journal of Software Engineering and Its Applications*.

https://www.researchgate.net/publication/266463703_Is_Naive_Bayes_a_Good_Classifier_for_Document_Classification

Vashisht, J., Agarwal, A., & Vashisht, M. (2025). IntelliDoc: An AI-Driven Document Intelligence Framework for Real-Time Document Processing. In M. D. Sharma, R. Sharma, & S. Peng (Eds.), *Proceedings of International Conference on Next-Generation Communication and Computing* (Vol. 1305, pp. 233–243). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-3725-6_20

Vicente, M. (2013). *O modelo de gestão documental no ensino particular e cooperativo: Análise da Escola Profissional de Educação para o Desenvolvimento – EPED* [Dissertação de Mestrado, Universidade de Lisboa]. <http://hdl.handle.net/10451/12271>.

<http://hdl.handle.net/10451/12271>

Wirth, R. & Hipp, J. (2000). *CRISP-DM: Towards a Standard Process Model for Data Mining*. 29–39.

Yu, A., Nardy, A., Hirano, H., Oliveira, J., Ribeiro, N., & Grando, N. (2024). Tomada de decisão nas organizações: O que muda com a Inteligência Artificial? *Estudos Avançados*, 38(111), 327–348. <https://doi.org/10.1590/s0103-4014.202438111.017>

Zhang, J., & Chen, Z. (2024). Exploring Human Resource Management Digital Transformation in the Digital Age. *Journal of the Knowledge Economy*, 15(1), 1482–1498.

<https://doi.org/10.1007/s13132-023-01214-y>

Anexos

Anexo A – Análise de frequência de classificação documental para o Nível 2 do classificador do MEF (116 516 amostras) provenientes do SGD de um serviço do Ministério das Finanças

Tabela 4. Distribuição da amostra inicial pelas classes de Nível 2 do classificador do MEF.

Código	Classificador de nível 2	Contagem
100.05	Alterações Orçamentais e Outros Atos de Gestão Orçamental	33 057
080.15	Prestação, Suspensão e Cessação do Trabalho	15 238
100.01	Orçamento de Estado	11 340
100.15	Contas Públicas	8 827
080.01	Recrutamento Seleção Progressão e Mobilidade	7 045
000.01	Estrutura Orgânica e Funcional	6 301
100.10	Acompanhamento Controlo e Execução Orçamental	4 989
090.01	Aprovisionamento	4 277
080.40	Remunerações, Abonos e Descontos	3 625
000.05	Desenvolvimento Organizacional	3 102
040.10	Assuntos da União Europeia	2 756
020.01	Assessoria Jurídica	2 250
080.55	Proteção na Saúde	1 864
080.10	Processos, Registros e Meio de Identificação	1 796
200.15	Tesouraria do Estado	1 781
040.05	Relação com outros Organismos e Pessoas Coletivas	1 269
080.30	Formação	891
080.05	Avaliação de Desempenho	631
030.01	Auditórias	594
060.01	Gestão de Conteúdos	494
050.01	Eventos e Cerimónias Oficiais	443
070.10	Gestão de Infraestruturas	340
090.10	Gestão de Bens Móveis	335
080.20	Acidentes em Serviço	290
090.25	Manutenção e Conservação de Imóveis e Instalações	276
010.10	Controlo Interno de atividades	237
010.15	Estatísticas	202
090.05	Compras Públicas Transversais	202
110.15	Normas e Procedimentos	200
070.05	Desenvolvimento de Sistemas	195
090.15	Gestão do Parque de Veículos	183
020.10	Contencioso Administrativo	179
030.25	Avaliações	170
090.20	Conceção, Construção e Gestão de Imóveis	134
070.20	Disponibilidade do Serviço	125
060.05	Gestão de Documentos de Arquivo	118
010.01	Planos, Programas e Projetos	76
080.60	Ação Social	72
010.05	Planeamento de atividades	65

Código	Classificador de nível 2	Contagem
010.20	Análise Económica e Financeira	57
080.35	Relações Laborais	50
020.05	Assessoria Técnica	42
080.50	Segurança Social	41
110.01	Projeto / Iniciativa	37
040.15	Relações Internacionais	36
030.20	Acompanhamento e Monitorização	29
080.25	Desenvolvimento de Carreiras	27
020.15	Contencioso Fiscal e Aduaneiro	20
000.15	Articulação Interorganizações	18
050.05	Comunicação e Imagem Institucional	17
200.10	Regularizações e Recuperações Financeiras	17
090.30	Gestão de Stocks	16
070.15	Suporte ao Serviço	14
030.10	Inquéritos e Sindicâncias e Peritagens	12
080.45	Segurança, Higiene e Saúde no Trabalho	11
020.25	Contencioso Judicial	10
200.20	Intervenção Financeira do Estado	10
000.10	Atos de Direção	9
020.30	Contencioso Orçamental e Financeiro	9
070.01	Coordenação Estratégica de STIC	9
070.25	Gestão de Segurança	9
040.01	Relação Estado / Cidadão	8
225.25	Organização, Modernização e Gestão Pública	7
060.10	Gestão de Documentação Bibliográfica	6
200.01	Dívida Pública	6
110.10	Publicitação	5
050.10	Gestão de Lojas	4
000.15	Articulação Interorganizações	3
110.05	Aprovação	2
020.20	Contencioso Comunitário	1
030.05	Inspeções	1
060.15	Gestão de Coleções Museológicas	1
200.05	Fundo de Regularização da Dívida Pública	1
225.15	Negociação Coletiva de Trabalho na Administração Pública	1
225.20	Qualificação e Desenvolvimento de Competências para a AP	1
Total		116 516

**Anexo B – Análise de frequência de classificação documental para o Nível 3
do classificador do MEF (116 516 amostras), provenientes do SGD de um
serviço do Ministério das Finanças**

Tabela 5. Distribuição da amostra inicial pelas classes de Nível 3 do classificador do MEF.

Código	Classificador de nível 3	Contagem
080.15.05	Mapas de assiduidade / Relação de faltas	12365
100.05.13	Pedidos de libertação de crédito (PLC)	12185
100.01.01	Preparação e elaboração do OE - Funcionamento e PIDDAC	11340
100.05.14	Guias de reposição / Guias de receita do Estado	8752
000.01.01	Projetos de diplomas	6220
100.15.04	Contas das administrações públicas	4368
090.01.02	Aquisição de bens não duradouros	4123
100.15.02	Contas de gerência	3704
100.05.10	Necessidades de financiamento	3282
000.05.02	Despachos internos / Circulares internas	3094
080.40.01	Vencimentos, suplementos e outros abonos	2890
080.01.05	Mobilidade de funcionários	2563
100.05.06	Outras alterações orçamentais	2463
100.05.07	Cativação / Descativação de verbas	2400
100.10.04	Relatórios de execução / Mapas de execução	2204
100.05.02	Créditos especiais	2193
100.10.03	Acompanhamento de Execução Orçamental	2016
080.01.03	Concursos de ingresso / Acesso	1705
080.10.01	Processos Individuais / Atualização de ficheiro de funcionário	1692
080.55.01	ADSE / Correspondência expedida	1638
020.01.01	Informações / Pareceres / Notas jurídicas	1427
040.10.02	Geral	1382
040.10.01	Projetos de diplomas de transposição de diretivas	1361
040.05.01	Cooperação com entidades externas	1142
080.15.04	Atestados médicos / Recuperação de vencimento de exercício	1134
080.15.07	Tolerância de ponto / Greves	975
080.01.01	Quadro de pessoal / Gestão de vagas	922
100.05.01	Gestão flexível	904
020.01.02	Pedidos de esclarecimentos / Dúvidas	813
200.15.02	Contas abertas na DGT	792
100.15.01	Conta Geral do Estado	733
080.01.06	Progressões / Mudanças de nível / Escalões de funcionários	645
080.05.03	S.I.A.D.A.P	591
030.01.02	Auditórias / Inspeções	589
200.15.01	Unidades de tesouraria do Estado	557
	Estornos de pagamentos / Receita indevidamente entregue / Devolução /	
200.15.03	Entrega de valores a outras entidades	400
080.01.04	Nomeação / Seleção / Renovação de comissão de serviço dirigentes	395
100.10.01	Controlo e publicitação de alterações orçamentais	391
080.01.02	Contratos de trabalho	374
100.05.12	Fundo de Maneio	339
080.30.03	Ações de formação externa	330
100.05.03	Dotação provisional	330
080.15.01	Horário de trabalho / Regulamentação	326
080.30.07	Certificados dos formadores / Formandos	323
060.01.01	Divulgação de informação	298

Código	Classificador de nível 3	Contagem
080.15.06	Concessão de licenças / Dispensas	291
080.20.01	Acidentes em serviço (ofícios / despesas / reembolsos)	290
100.10.05	Pareceres de natureza orçamental / financeira	284
050.01.01	Reuniões / Atas de reuniões	277
080.01.07	Comissão de serviço / Renovação / Cancelamento de licenças fora do quadro	259
080.40.02	Descontos obrigatórios e facultativos	205
090.05.01	Registo de fornecedores	202
070.10.01	Hardware	181
020.10.01	Recursos / Reclamações / Contestações	179
070.05.01	Sistemas de Informação Relatórios e documentos relacionados com as avaliações de Programas e	177
030.25.01	Medidas	170
010.15.01	Documentos relativos a indicadores de gestão	168
060.01.03	Divulgação de informação na Intranet	167
090.10.07	Telemóveis	162
090.25.01	Reparação / Obras de manutenção e conservação	159
050.01.02	Seminários / Conferências / Colóquios	154
090.01.01	Aquisição de bens duradouros	154
080.40.03	Deslocações e estadas	144
010.10.01	Relatório de atividades / Contributo para o relatório de atividades	142
080.55.02	SOFE / Correspondência recebida / expedida	141
080.01.09	Candidaturas espontâneas	129
080.40.05	Inscrição / Reinscrição (CGA, ADSE, SOFE, etc.)	129
080.40.04	Subsídio de alojamento	128
090.20.03	Contratos de arrendamento	125
070.20.01	Continuidade de negócio	120
090.25.02	Contratos de prestação de serviços (limpeza, água, etc.)	117
100.05.11	Cabimentação de despesas	111
080.10.02	Cartões / Placas de identificação / Código de Identificação (SRH)	104
040.05.03	Pedidos de divulgação de informação geral	103
070.10.02	<i>Software</i>	90
080.55.03	Seguros de saúde / acidentes de trabalho de funcionários, agentes ou outros	85
100.10.02	Publicação dos Mapas da Lei	85
110.15.02	Procedimentos contabilísticos e classificações orçamentais	82
060.05.02	Oferta / Receção de publicações Licença s/vencimento / Maternidade / Paternidade / Licença de equiparação a	80
080.15.02	bolseiro	80
080.30.02	Ações de formação interna	77
010.10.02	Outros relatórios	74
090.15.03	Seguro automóvel	73
000.01.03	Despachos de delegações e subdelegações de competências	69
070.10.03	Comunicações Plano anual de formação / Diagnóstico das necessidades de formação /	69
080.30.01	Divulgação	69
110.15.03	Normas e instruções para a orçamentação e execução orçamental	68
080.15.08	Mapas de férias	66
010.01.02	Bolsa de emprego público (BEP)	65

Código	Classificador de nível 3	Contagem
010.05.01	Plano de atividades / Contributo para o plano de atividades	65
	Inscrições para as ações de formação / Convocatórias para as ações de	
080.30.04	formação	65
090.15.02	Processo individual de viaturas	62
010.20.01	Estudos de apoio e análise efetuados pelo GEFP	57
080.01.08	Pedidos de estágios / Emprego	53
080.40.06	Descontos / Penhoras judiciais	53
090.10.02	Contratos de assistência técnica a equipamentos	51
080.40.07	Declarações diversas	49
090.15.01	Aquisição de veículos / Aluguer de viaturas	48
100.05.05	Modificação das Leis orgânicas	48
090.10.03	Autos de abate de bens (por depreciação, furto)	44
090.10.05	Reparação e manutenção de bens e equipamentos	44
020.05.01	Participação em comissões, comités e grupos de trabalho	42
080.60.02	Cofre Previdência MF / Correspondência recebida / expedida	41
110.01.01	Preparação de um Diploma	37
040.15.00	Relações Internacionais	36
200.15.04	Fundos saídos	32
080.05.01	Legislação e procedimentos (objetivos)	31
080.50.01	Regime Geral	31
080.60.01	Apoio social / Comparticipação em lares / Subsídios por morte e funeral, etc.	31
030.20.01	Correspondência com os coordenadores dos programas	29
060.01.02	Divulgação de informação no portal Web	29
080.40.08	Diplomas de regulamentação de assuntos de pessoal	27
040.05.02	Anuário do MF	23
080.35.02	Normas e orientações	23
080.25.03	Recenseamento de efetivos	22
100.15.03	Contas provisórias	22
010.10.03	Manuais de procedimentos / Auditorias	21
100.05.04	Orçamento Retificativo / Suplementar	21
020.15.00	Sem Nome Definido	20
080.30.06	Colaboração com outras entidades no âmbito da formação	20
110.15.08	Normas e regras para a realização de despesas	20
000.01.02	Leis orgânicas e diplomas complementares	19
070.05.02	Estudos e serviços	18
	Relações com organizações representativas dos trabalhadores / Reunião de	
080.35.01	trabalhadores	17
	Regularização de responsabilidades financeiras do Estado, de Entidades	
200.10.01	extintas e de situações do passado	17
090.10.01	Inventário de bens	15
090.30.01	Requisição de material	15
100.05.09	Antecipação de duodécimos	15
100.05.08	Despesas de anos anteriores	14
010.15.03	Balanço social	13
040.10.00	Assuntos da União Europeia	13
050.05.01	Folhetos / Cartazes	13
060.05.03	Normas de divulgação / Gestão Documental	13

Código	Classificador de nível 3	Contagem
090.10.06	Seguro de equipamentos diversos	13
030.10.00	Inquéritos e Sindicâncias e Peritagens	12
050.01.03	Missões estrangeiras em visita ao país / Serviço	12
110.15.01	Circulares da DGO	12
110.15.04	Informações e Pareceres da DGO	12
000.15.00	Articulação Interorganizações	11
010.01.01	Programa Simplex	11
010.15.04	Estatísticas de pessoal (greves, etc.) Documentos referentes a vistorias e inspeções médicas de avaliação das condições de trabalho	11
010.15.02	Mapas estatísticos	10
020.01.03	Intervenções em processos de inquérito ou disciplinares	10
020.25.01	Recursos contenciosos	10
200.20.00	Intervenção Financeira do Estado	10
000.10.01	Convocatórias / Atas / Reuniões de direção	9
020.30.00	Contencioso Orçamental e Financeiro	9
070.15.01	Serviço de <i>Help-Desk</i>	9
080.05.02	Concelho de Coordenação de Avaliação	9
090.20.01	Afeitação / Mudança de instalações	9
100.10.06	Contributos para a Reforma da Gestão Financeira Pública	9
000.05.01	Circulares conjuntas	8
060.05.01	Aquisição / Renovação de assinatura	8
060.05.06	Copiadores órgãos e serviços da DGO	8
070.01.02	Projetos STIC	8
070.25.01	Políticas de segurança	8
080.35.04	Reclamações sobre condições de trabalho / Outras	8
080.50.02	Regime Especial	8
060.10.00	Divulgação de informação	6
090.10.04	Cedência de mobiliário / Equipamento informático	6
200.01.00	Dívida Pública Estudos relativos ao regime jurídico da criação, fusão, reestruturação e extinção de serviços públicos	6
030.01.01	Plano anual de auditorias / Inspeções	5
060.05.08	Correspondência Geral	5
070.15.02	Gestão de incidentes de suporte	5
070.20.02	Níveis de Serviço	5
040.01.01	Indemnizações por responsabilidade do Estado	4
040.01.02	Correspondência com o Provedor de Justiça	4
050.05.02	Campanhas publicitárias	4
050.10.00	Gestão de Lojas	4
080.25.01	Caraterização de perfis profissionais	4
110.10.01	Publicação em DR	4
110.15.09	Reforma da Gestão Financeira Pública - Documentação	4
000.15.00	Seminários / Conferências / Colóquios	3
060.05.04	Planos de classificação	3
080.30.05	Dossiers de candidatura ao FSE	3
080.30.09	Manuais de formação	2

Código	Classificador de nível 3	Contagem
080.35.03	Processos disciplinares	2
080.50.03	Pensões de Sobrevivência	2
020.20.00	Contencioso Comunitário	1
030.05.00	Inspeções	1
040.05.04	Diplomas de fixação de remunerações de órgãos de gestão	1
060.05.07	Correspondência com a Secretaria-Geral e fornecedores	1
060.15.00	Gestão de Coleções Museológicas	1
070.01.01	Planeamento STIC	1
070.25.02	Incidentes de segurança	1
080.15.03	Lista de antiguidade	1
080.25.02	Descongelamento de admissões	1
080.30.08	Processo de acreditação como entidade formadora Dossiers técnicos / contabilísticos no âmbito das candidaturas a financiamentos	1
080.30.10	comunitários	1
090.30.02	Guias de entrega de material	1
110.05.01	Aprovação de um Diploma	1
110.05.02	Retificação de um Diploma	1
110.10.02	Ofícios de comunicação á INCM	1
110.15.05	Rafe - Ficheiros SIC (envio)	1
110.15.07	Normas relativas a concursos de pessoal	1
200.05.00	Fundo de Regularização da Dívida Pública	1
225.15.01	Estudos e projeções	1
225.20.01	Apreciação de pedidos de autorização do exercício de funções Pareceres e propostas relativos ao regime jurídico da criação, fusão,	1
225.25.02	reestruturação e extinção de serviços públicos	1
Total		116 516

Anexo C - Tipos de documentos frequentes em serviços públicos

Documentos Administrativos

Relacionados com decisões, comunicações e registos internos ou externos no âmbito da gestão pública.

Tabela 6. Tipos de documentos administrativos.

Tipo de Documento	Descrição
Despachos	Decisões formais emitidas por dirigentes.
Ofícios	Comunicações oficiais entre entidades públicas ou com cidadãos.
Memorandos	Comunicações internas com carácter informativo ou instrutivo.
Circulares	Instruções ou informações dirigidas a múltiplos destinatários.
Relatórios	Exposição sistemática de resultados, análises ou avaliações.
Pareceres	Pronunciamentos técnicos ou jurídicos sobre matérias específicas.
Requerimentos	Pedidos formais apresentados por cidadãos ou funcionários.
Atas	Registros oficiais de reuniões, assembleias ou sessões.
Notas informativas	Documentos breves com esclarecimentos ou orientações.
Ordens de serviço	Instruções operacionais emitidas por superiores hierárquicos.
Comunicados internos	Mensagens institucionais dirigidas aos trabalhadores.

Documentos Jurídicos e Normativos

Relacionados com normas, acordos e procedimentos com força jurídica ou regulamentar.

Tabela 7. Tipos de documentos jurídicos e normativos.

Tipo de Documento	Descrição
Leis e Decretos-Leis	Normas jurídicas aprovadas por órgãos legislativos.
Portarias	Regulamentos emitidos por membros do Governo.
Resoluções	Decisões de órgãos colegiais com efeitos normativos.
Contratos e Protocolos	Instrumentos jurídicos que formalizam acordos entre entidades.
Regulamentos Internos	Normas que regem o funcionamento interno de serviços ou organismos.
Deliberações	Decisões formais tomadas por órgãos colegiais.
Instruções normativas	Diretrizes técnicas ou administrativas com força vinculativa.
Termos de responsabilidade	Declarações formais de compromisso ou aceitação de condições.

Documentos Financeiros

Relacionados com o controlo e operações financeiras.

Tabela 8. Tipos de documentos financeiros.

Tipo de Documento	Descrição
Orçamentos	Planeamento financeiro que prevê receitas e despesas.
Balancetes	Demonstrações periódicas da situação financeira.
Faturas e Recibos	Comprovativos de operações financeiras realizadas.
Guias de pagamento	Instruções para pagamento de taxas ou serviços.
Folhas de vencimento	Registos dos salários e descontos aplicáveis aos trabalhadores.
Relatórios financeiros	Análises detalhadas da execução orçamental e financeira.
Ordens de pagamento	Autorizações formais para a realização de despesas.
Justificativos de despesa	Documentos que comprovam gastos realizados.

Documentos de Recursos Humanos

Relacionados com a gestão de pessoal, desde a contratação até à avaliação.

Tabela 9. Tipos de documentos de recursos humanos.

Tipo de Documento	Descrição
Contratos de trabalho	Acordos laborais entre o serviço público e os trabalhadores.
Processos individuais	Dossiês com informação funcional e administrativa de cada trabalhador.
Avaliações de desempenho	Registos da performance dos trabalhadores.
Pedidos de férias, licenças e ausências	Solicitações formais relacionadas com a gestão do tempo laboral.
Justificação ou declaração	Justificação de ausência ou declaração de presença.
Requerimento	Documento formal para solicitar algo
Declarações de serviço	Comprovativos de vínculo ou funções exercidas.
Registos de formação	Documentos que comprovam ações de capacitação ou qualificação.
Termos de posse	Documentos que formalizam a entrada em funções.
Mapas de assiduidade	Registos de presença e pontualidade dos trabalhadores.

Documentos Técnicos e Operacionais

Relacionados com a execução técnica, planeamento e gestão de bens e serviços.

Tabela 10. Tipos de documentos técnico-operacionais.

Tipo de Documento	Descrição
Projetos e estudos técnicos	Documentos que detalham intervenções, obras ou análises.
Mapas e plantas	Representações gráficas utilizadas em áreas como urbanismo ou ambiente.
Planos de atividade	Planeamento das ações a desenvolver num período definido.
Inventários	Registos sistemáticos de bens patrimoniais ou materiais.
Cadernos de encargos	Documentos que definem requisitos técnicos e administrativos para contratação pública.
Relatórios de execução	Documentos que acompanham o progresso de projetos ou ações.
Fichas técnicas	Descrições detalhadas de equipamentos, materiais ou procedimentos.
Registros de manutenção	Histórico de intervenções técnicas em equipamentos ou infraestruturas.

Documentos de Comunicação e Informação

Relacionados com informação e comunicação interna e externa.

Tabela 11. Tipos e documentos de comunicação e informação.

Tipo de Documento	Descrição
Comunicados de imprensa	Informações oficiais destinadas à comunicação social.
Informações internas	Notas, boletins ou comunicados dirigidos aos trabalhadores.
Comunicação	Comunicações diversas.
Campanhas informativas	Materiais destinados à sensibilização ou divulgação pública.
Boletins informativos	Publicações periódicas com atualizações institucionais.
Cartazes e folhetos	Materiais gráficos de divulgação.
Publicações institucionais	Documentos editados para fins de comunicação externa.
Convites e anúncios	Informações sobre eventos ou iniciativas públicas.

Anexo D – Exemplo da Proposta de Taxonomia Documental e Atributos por Tipo de Documento

The screenshot shows the 'Gestor de Taxonomia' application interface. On the left, a tree view displays the taxonomy structure under 'Taxonomia'. The 'Oficio' node under 'MF/DocumentosAdministrativos' is selected and highlighted in pink. The right side of the screen shows the configuration details for this selected node.

Tipo de documento

Nome do tipo de documento: Oficio
MF.DocumentosAdministrativo...

Eliminar

Grupo	Categoria	Código de Tipo de Documento
MF	DocumentosAdmin...	Valor opcional

Definições de apresentação

Limiar de confiança de OCR Formato de apresentação de data
90 yyyy-MM-dd

Separador numérico decimal Separador numérico de milhares
.

Campo

Assunto	Data	Entidade	Numero_Oficio
a	q	e	m

Figura 10. Exemplo da taxonomia definida.

Anexo E – Exemplo de regras de encaminhamento

As regras de encaminhamento são aplicadas de forma simples, ou de forma combinada para permitir dar resposta à necessidade e lógica de encaminhamento definidas pelo serviço, considerando a tipologia, conteúdo, natureza, e finalidade dos documentos.

Regras por tipo de documento

Tabela 12. Exemplo de regras de encaminhamento por tipo de documento.

Tipo Documental	Regra de Encaminhamento
Requerimento	Departamento de Gestão de Recursos
Fatura	Departamento de Gestão de Recursos
Parecer técnico	Departamento Jurídico
Despacho	Secretariado da Direção
Relatório de Execução	Departamento de Acompanhamento Setorial
Contrato	Departamento Jurídico

Regras por entidade remetente

Tabela 13. Exemplo de regras de encaminhamento por entidade remetente.

Entidade Remetente	Regra de Encaminhamento
Hospital	Departamento de Acompanhamento Setorial da Saúde
Escola	Departamento de Acompanhamento Setorial da Educação
Tribunal	Departamento de Acompanhamento Setorial do Ministério da Justiça

Regras por Classificação Documental (MEF Nível 1)

Tabela 14. Exemplo de regras de encaminhamento por classificação documental.

Classificador MEF	Regra de Encaminhamento
000 – Organização e Funcionamento	Encaminhar para o Departamento de Gestão de Recursos
010 – Planeamento	Encaminhar para o Departamento de Gestão de Recursos
020 – Consultadoria e Contencioso	Departamento Jurídico
100 – Gestão Orçamental e Contabilística	Departamento de Acompanhamento Setorial

Regras por Atributos Específicos

Tabela 15. Exemplo de regras de encaminhamento por atributos específicos.

Atributo Extraído	Regra
Rubrica Orçamental = 01.02.03.04.05.06	Departamento de Planeamento, Gestão Orçamental e Assuntos Europeus
Assunto contém “pedido de esclarecimento”	Departamento Jurídico
Data de emissão > 30 dias	Secretariado da Direção

Anexo F – Exemplo de Expressões Regulares

Exemplo da configuração dos extractores dos atributos definidos na taxonomia com recurso a expressões regulares para identificação de padrões de atributos previsíveis nos documentos:

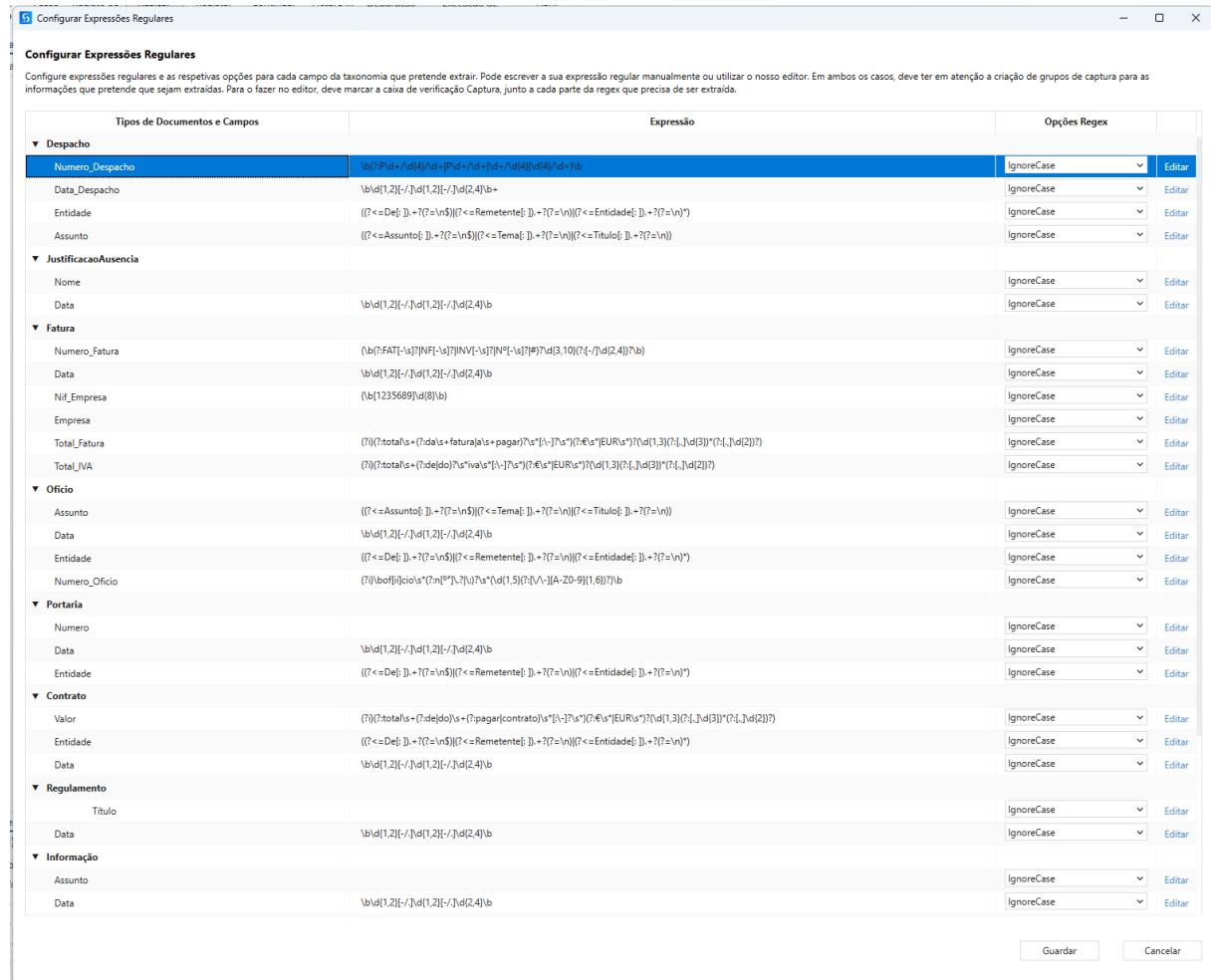


Figura 11. Exemplo da interface de gestão das expressões regulares para extração de atributos.

Exemplo de expressões regulares consideradas para identificação de padrões de atributos previsíveis nos documentos de acordo com o considerado na taxonomia:

Tabela 16. Lista de expressões regulares para extração de atributos.

Atributo a extrair	Expressão regular	Exemplo de valor a extrair
Datas	\b\d{1,2}\[-.\]\d{1,2}\[-.\]\d{2,4}\b	12/08/2025, 01-01-2024
Identificação Fiscal	\b[1235689]\d{8}\b	123456789

Atributo	a	Expressão regular	Exemplo de valor a extraír
Chave Orçamental		^\d{2}(\.\d{2}){5}\$	01.02.03.04.05.06
Código do Ministério		^\d{2}(?=.)	01 extraído de 01.02.03.04.05.06
Código da Secretaria de Estado		(?<=^\d{2}\.)\d{2}(?=.)	02 extraído de 01.02.03.04.05.06
Código do Capítulo		(?<=^\d{2}\.\d{2}\.\d{2}\.)\d{2}(?=.)	04 extraído de 01.02.03.04.05.06
Número de Documento		\bN[.]? [º]? \d{1,6}(\d{4})?\b	Nº 1234/2025, Nº 56789
Código Postal		\b\d{4}-\d{3}\b	1000-001
IBAN		\b[A-Z]{2}\d{2}[]?(d{4}[]?)\{4\}\d{1,4}\b	PT50 0002 0123 1234 5678 9015 4
Assunto		`(?<=Assunto[:]).+?(?=\\n\$)`	
Entidade Remetente		`(?<=De[:].+?(?=\\n\$)ou(?<=Remetente[:].+?(?=\\n`	
Montante Financeiro		`\b\d{1,3}(?:\d{3})*(?:,\d{2})?\s?(€ EUR)\b`	
Referência Documental		\b[A-Z]{2,5}[-_]\d{3,6}\b	DOC-12345, REF_2023
Telefone / Fax		\b\d{2,3}[-.]\d{3}[-.]\d{3,4}\b	21 345 6789, 213-456-789
URL / Link		https://[^\\s]+	https://www.eo.gov.pt
Classificação MEF		\b\d{3}(\.\d{2}){0,2}\b	100, 100.05, 100.05.13
Número de Processo		\b(?:P\d+ \d{4})\d+ P\d+ \d+ \d{4}\d{4}\d+\b	P123/2025, 1234/2025, 2025/123

Anexo G – Classificador Inteligente

Exemplo do treino do modelo de classificação inteligente para classificação do tipo de documento.

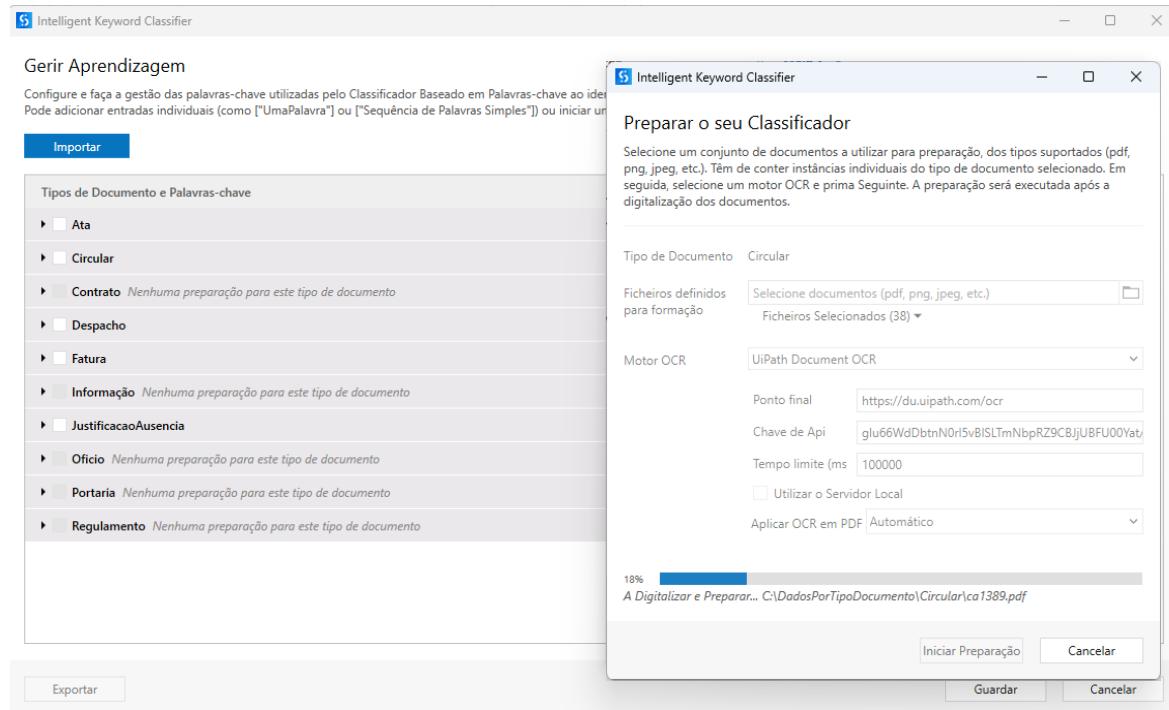


Figura 12. Exemplo de configuração da interface de treino do classificador inteligente para identificar o tipo de documento.

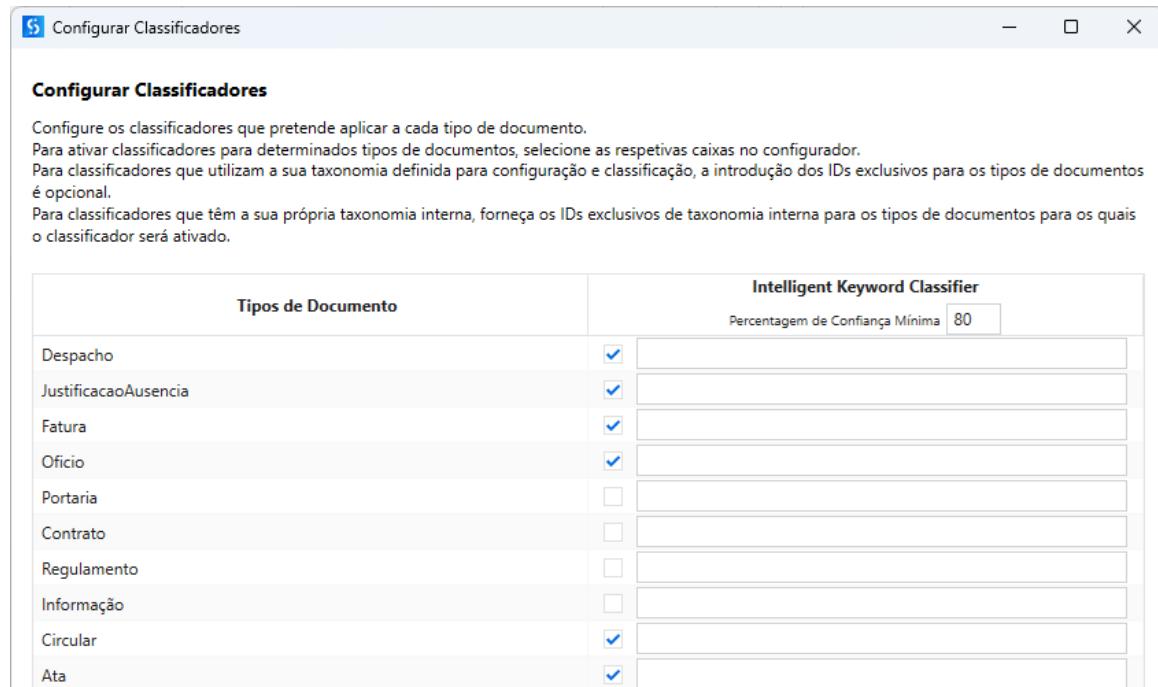


Figura 13 Exemplo de ativação dos classificadores por tipo de documento.

Anexo H – Fluxo robotizado

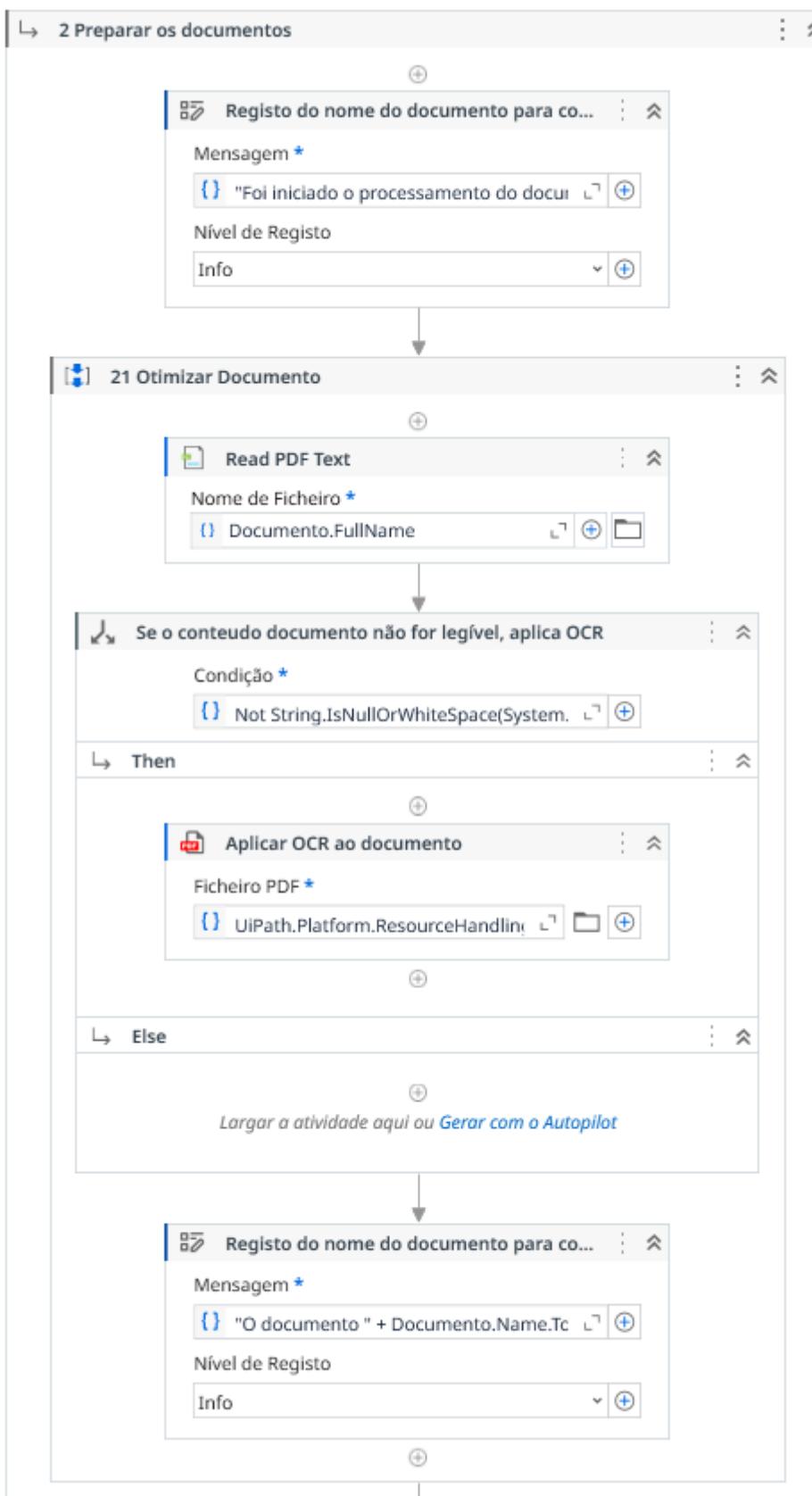


Figura 14 Exemplo do fluxo robotizado de preparação e otimização dos documentos.

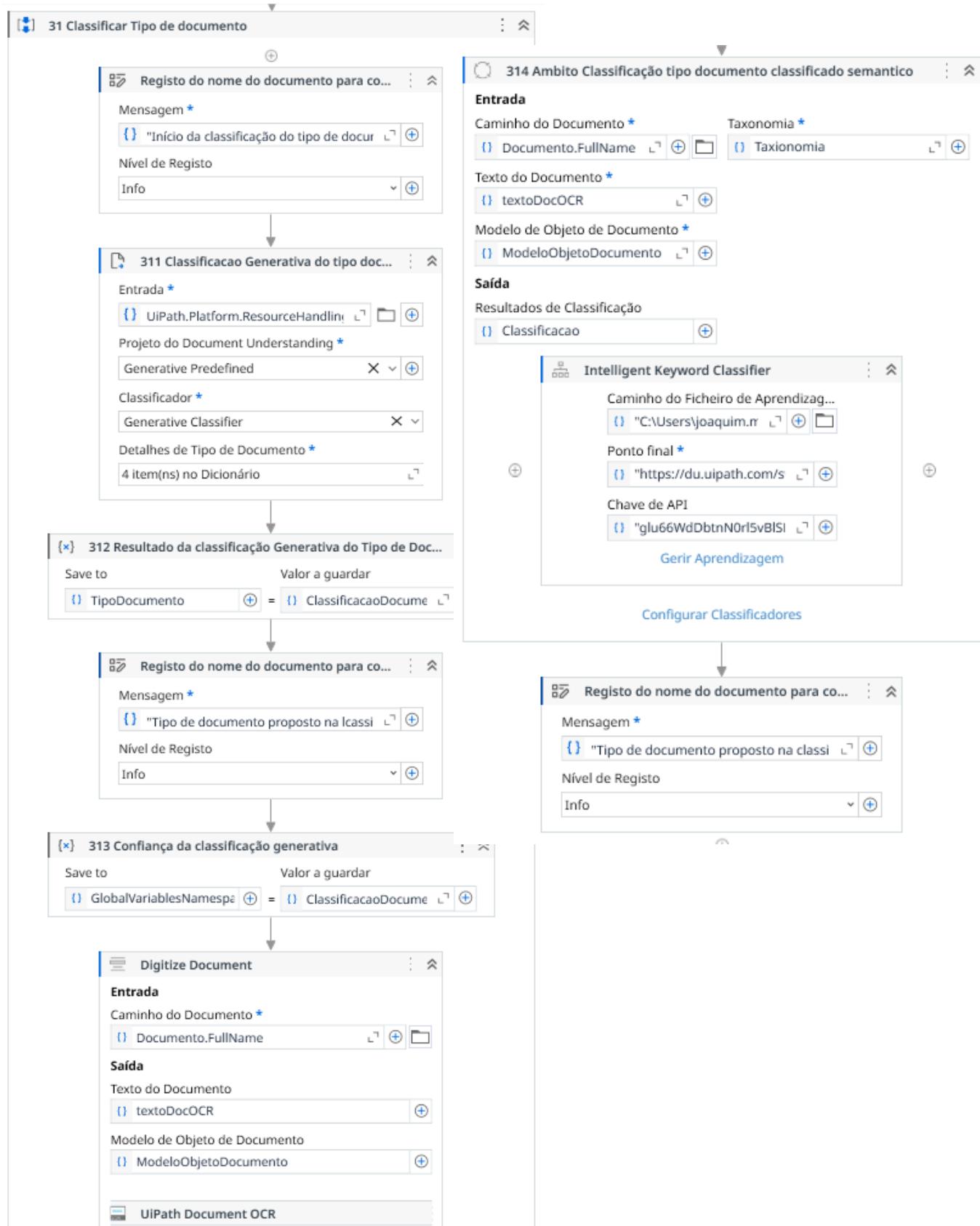


Figura 15 Exemplo do fluxo de classificação automatizada do tipo de documento.

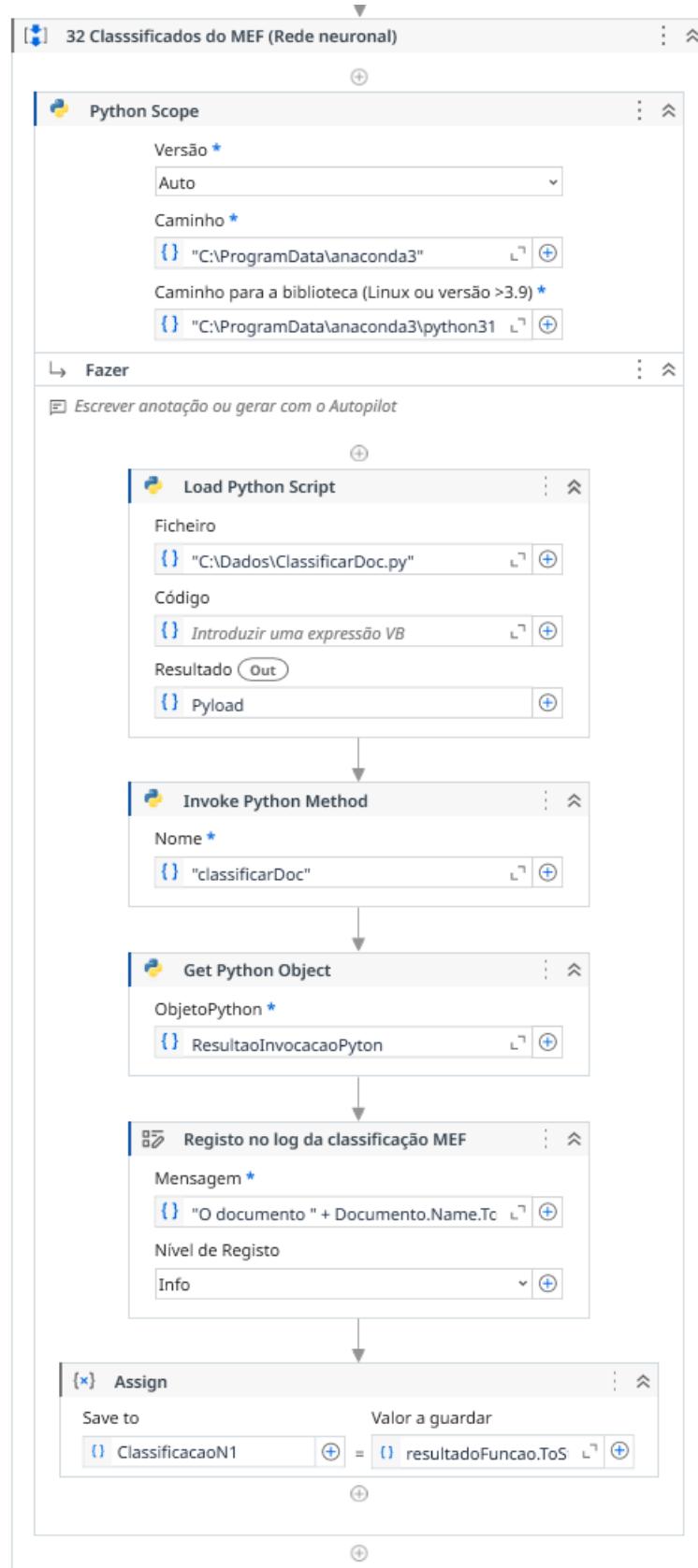


Figura 16 Exemplo do fluxo de classificação do Nível 1 da MEF.

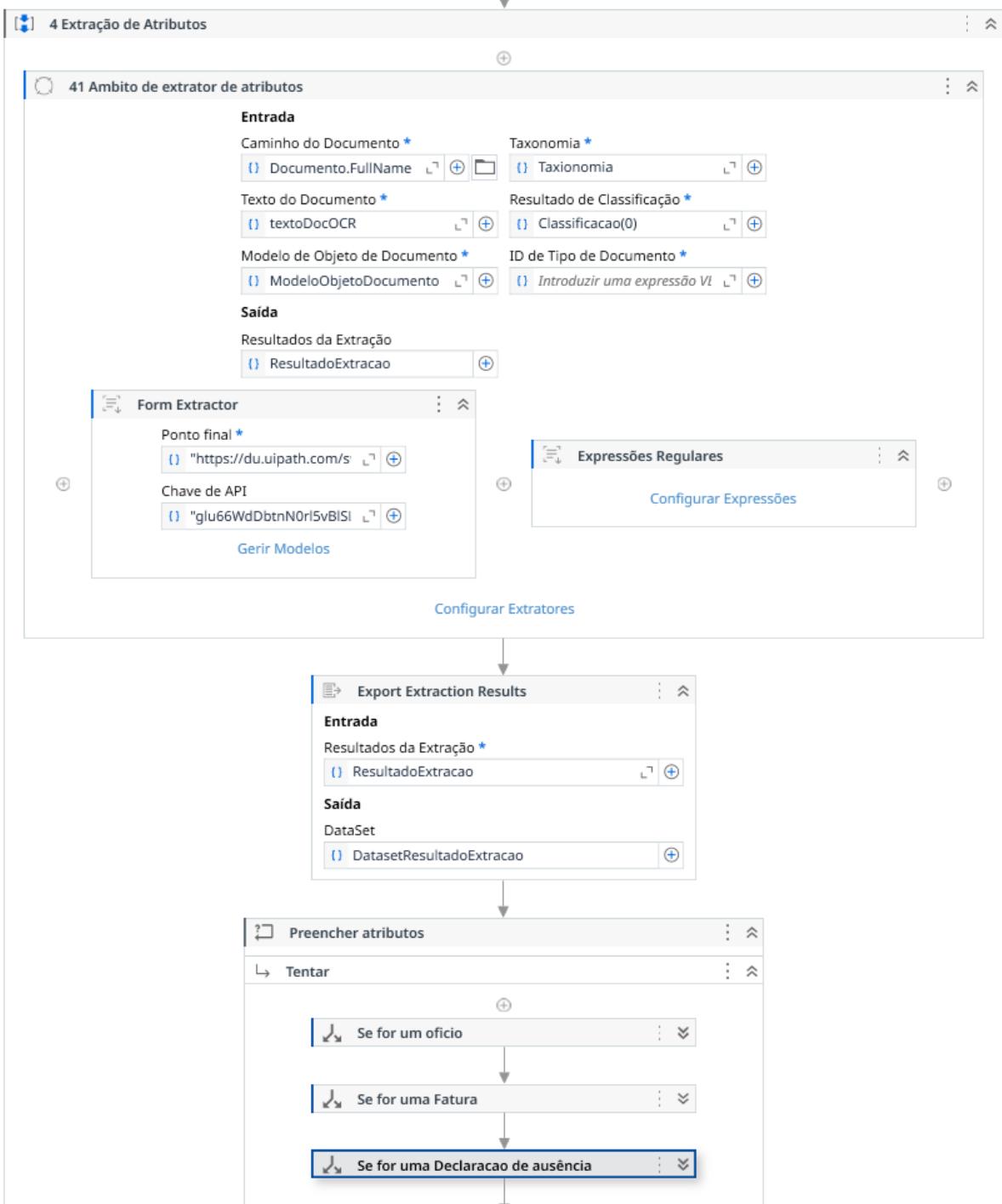


Figura 17 Exemplo do fluxo de extração de atributos previstos na taxonomia.

Anexo I – Automatizar o fluxo documental

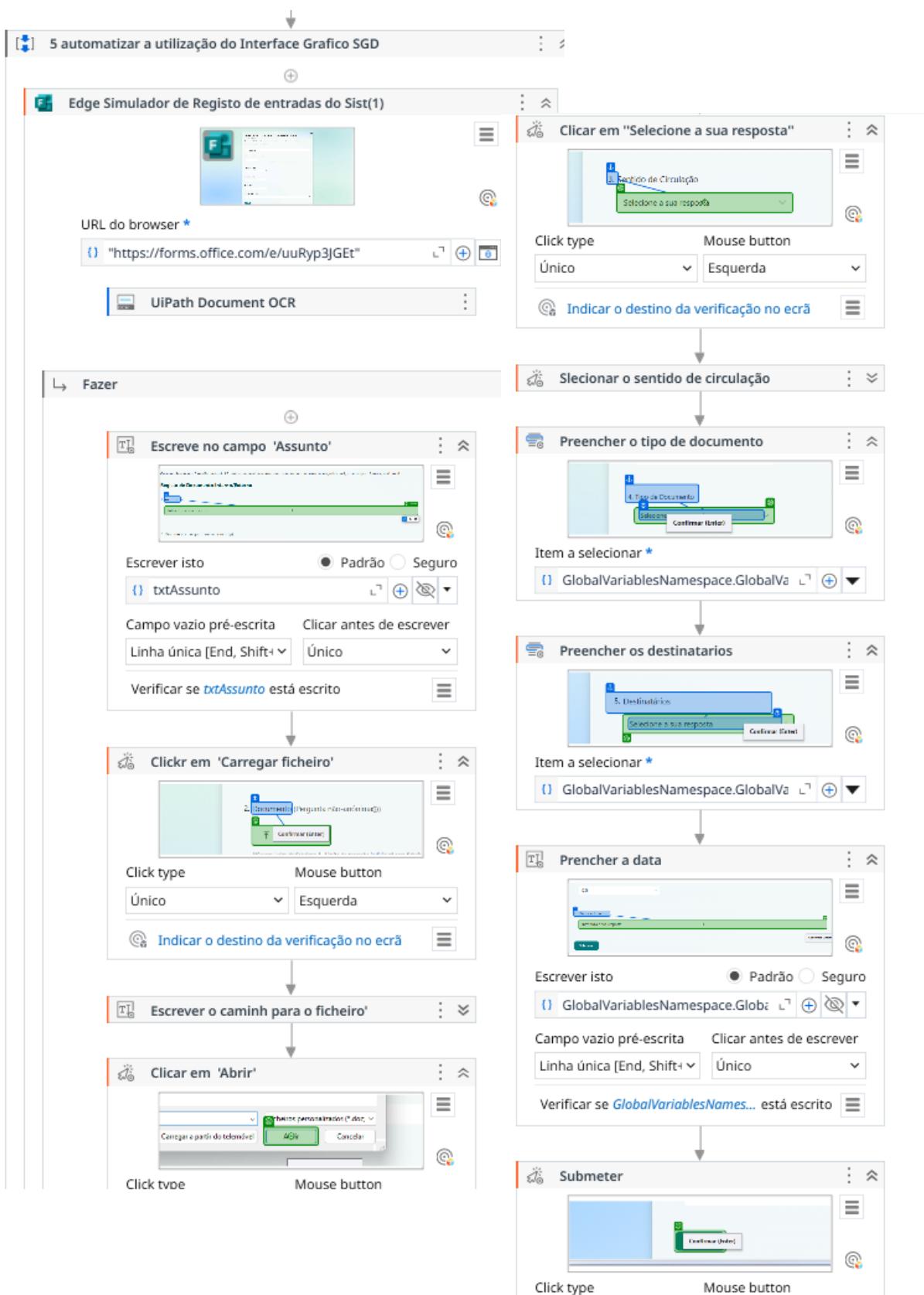


Figura 18 Exemplo da automatização do interface gráfico.

Anexo J – Configuração dos extractores de atributos automatizados

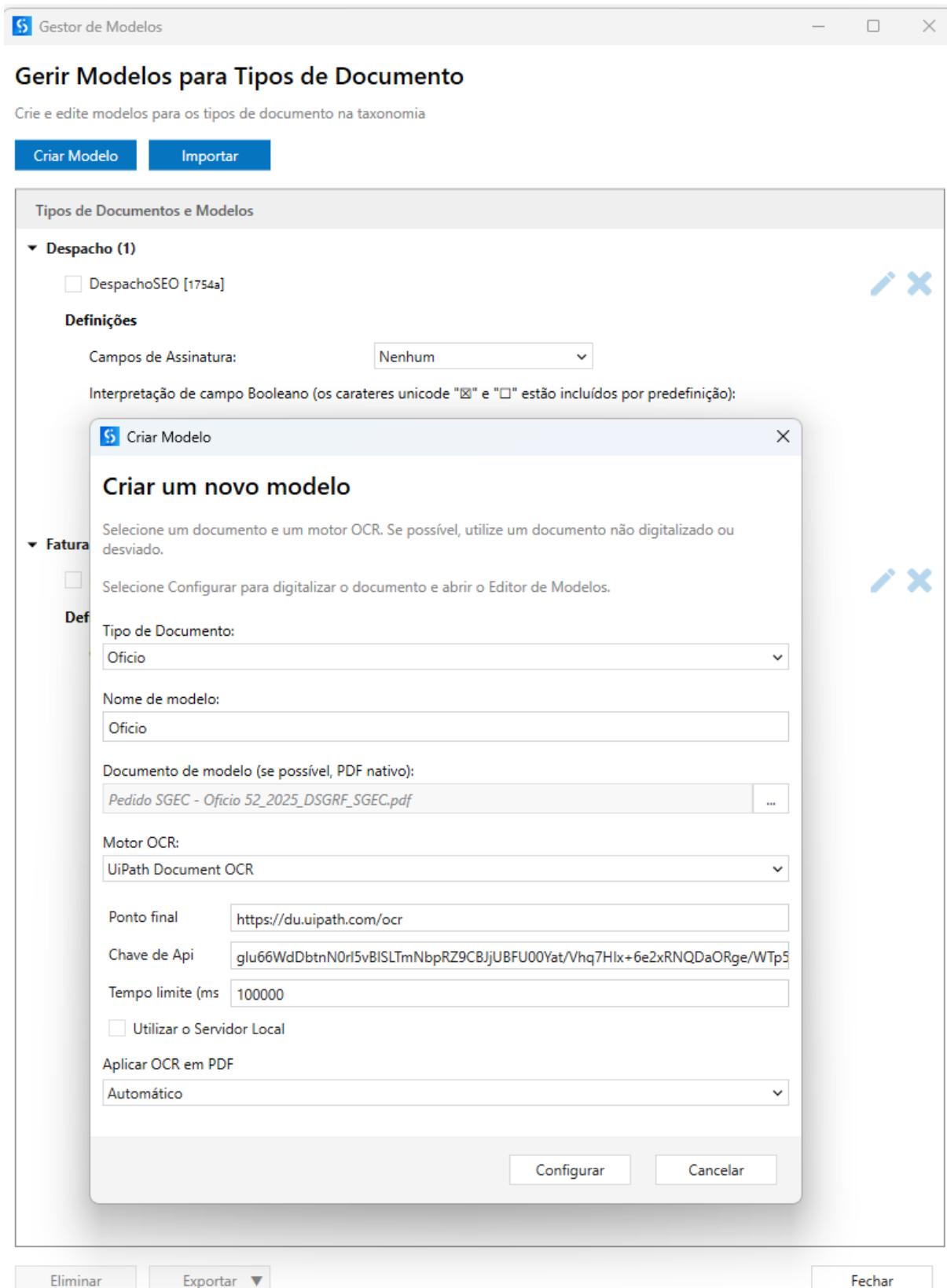


Figura 19 Gestão dos tipos de documentos na taxonomia.

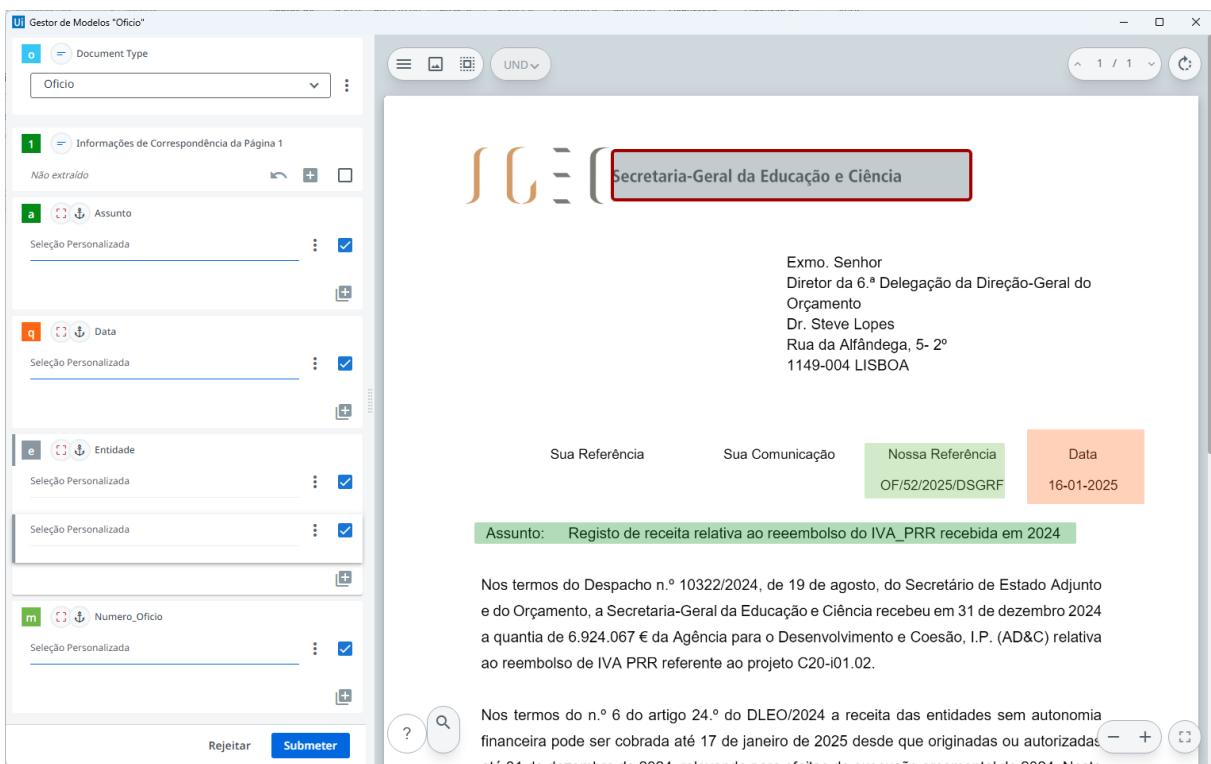


Figura 20 Exemplo de marcação visual dos atributos para reconhecimento e extração automatizada.

Anexo K – Script para treino do modelo de classificação do MEF

A abordagem utilizada para treinar o modelo de classificação do nível 1 da MEF considera o treino do modelo para um classificador *Naive Bayes*, procedendo-se às seguintes etapas:

- Extração de texto com *PyMuPDF* (*fitz*);
- Vetorização com *TfidfVectorizer* do *scikit-learn*;
- Remoção de stopwords com *NLTK*;
- Treino de modelo *Naive Bayes* com dados classificados;
- Avaliação do modelo com métricas de classificação;

```
1  # -*- coding: utf-8 -*-
2  import os
3  import fitz # PyMuPDF
4  from sklearn.feature_extraction.text import TfidfVectorizer
5  from sklearn.model_selection import train_test_split
6  from sklearn.naive_bayes import MultinomialNB
7  from sklearn.metrics import classification_report
8  import joblib # Para guardar o modelo
9
10 def extract_text_from_pdf(pdf_path):
11     try:
12         with fitz.open(pdf_path) as doc:
13             return "\n".join(page.get_text() for page in doc)
14     except Exception as e:
15         print(f"Erro ao extrair texto de {pdf_path}: {e}")
16         return ""
17
18 def load_dataset(base_dir):
19     texts, labels = [], []
20     for class_name in os.listdir(base_dir):
21         class_path = os.path.join(base_dir, class_name)
22         if os.path.isdir(class_path):
23             for file_name in os.listdir(class_path):
24                 if file_name.lower().endswith('.pdf'):
25                     file_path = os.path.join(class_path, file_name)
26                     text = extract_text_from_pdf(file_path)
27                     if text.strip():
28                         texts.append(text)
29                         labels.append(class_name)
30     return texts, labels
31
32 # Caminho para o diretório base
33 base_directory = 'C:\Dados'
34
35 # Carregar dados
36 texts, labels = load_dataset(base_directory)
37 import nltk
38 nltk.download('stopwords')
39 from nltk.corpus import stopwords
40 print("Início do treino do modelo")
41 stopwords_pt = stopwords.words('portuguese')
42 vectorizer = TfidfVectorizer(stop_words=stopwords_pt, max_features=5000)
43 # Vetorização
44 X = vectorizer.fit_transform(texts)
45
46 # Divisão treino/teste
47 X_train, X_test, y_train, y_test = train_test_split(X, labels, test_size=0.2,
48                                                 random_state=42)
49
50 # Treinar modelo
51 model = MultinomialNB()
52 model.fit(X_train, y_train)
53
54 # Avaliação
55 y_pred = model.predict(X_test)
56 print(classification_report(y_test, y_pred))
57
58 # Guardar modelo e vetor
59 joblib.dump(model, 'modelo_classificacao_MEF_N1.pkl')
60 joblib.dump(vectorizer, 'vetor_tfidf.pkl')
61
62 print("Fim de treino")
```

Figura 21. Código em python para treinar o modelo de classificação (Nível 1 da MEF).

Anexo L – Nuvem de palavras do documento

Gráfico de nuvem de palavras do documento, gerado com a ferramenta “wordclouds”.



Figura 22. Gráfico de nuvem de palavras do documento.