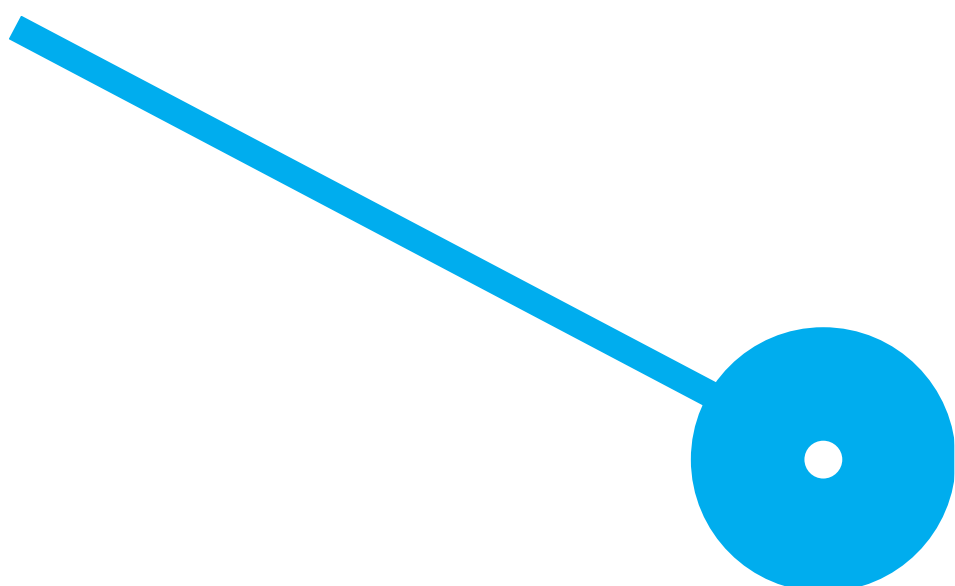
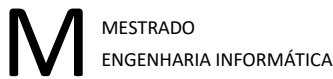


Sistema de Rastreabilidade de Madeira:  
Uma abordagem baseada em  
mecanismos semânticos para integração  
e validação dos dados das atividades de  
exploração florestal

Hugo Daniel Martins Silva

10/2024





# Sistema de Rastreabilidade de Madeira: Uma abordagem baseada em mecanismos semânticos para integração e validação dos dados das atividades de exploração florestal

Hugo Daniel Martins Silva  
8180378

## **Orientador**

Prof. Doutor Cristóvão Dinis Polido Sousa

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Informática pela Escola Superior de Tecnologia e Gestão do Instituto Politécnico do Porto.

# Declaração de integridade

Eu, Hugo Daniel Martins Silva, estudante nº 8180378, do Mestrado de Engenharia Informática da Escola Superior de Tecnologia e Gestão do Instituto Politécnico do Porto, declaro que não fiz plágio nem auto-plágio, pelo que o trabalho intitulado “Sistema de Rastreabilidade de Madeira: Uma abordagem baseada em mecanismos semânticos para integração e validação dos dados das atividades de exploração florestal” é original e da minha autoria, não tendo sido usado previamente para qualquer outro fim. Mais declaro que todas as fontes usadas estão citadas, no texto e na bibliografia final, segundo as regras de referência adotadas na instituição.

# Agradecimentos

A realização desta tese só foi possível graças ao contributo e apoio de várias pessoas a quem expresso o meu sincero agradecimento.

Em primeiro lugar, gostaria de agradecer ao meu orientador, Professor Doutor Cristóvão Dinis Polido Sousa, pela sua disponibilidade e apoio ao longo deste percurso. A sua orientação foi essencial para a concretização deste trabalho.

Deixo um especial agradecimento ao Doutor Reinaldo Silva Gomes, pela sua incansável cooperação e disponibilidade.

Agradeço também à Escola Superior de Tecnologia e Gestão e a todo o seu corpo docente e não docente. Esta instituição acolheu-me nos últimos seis anos e permitiu-me crescer tanto na minha vida pessoal quanto profissional.

Por fim, mas não menos importante, expresso o meu agradecimento aos meus amigos e familiares, especialmente aos meus pais. Tudo o que sou hoje é graças a eles.

**A todos, o meu sincero e profundo OBRIGADO!**

# Resumo

Nos últimos anos, as preocupações crescentes com a desflorestação impulsionaram a necessidade de monitorizar a origem e o histórico da madeira que chega às fábricas. Isso levou à adoção de sistemas de rastreabilidade no setor florestal. No entanto, muitos desses sistemas ainda são manuais e baseados em papel, o que os torna suscetíveis a erros e falsificações.

Com o avanço da Indústria 4.0 e a digitalização das operações florestais, surge a oportunidade para a transformação digital da rastreabilidade. Contudo, o processo de digitalização enfrenta desafios a vários níveis, sendo que um dos principais está relacionado com o facto das fontes de informação estarem dispersas entre os diversos intervenientes da cadeia de abastecimento florestal, resultando em dados imprecisos e de difícil acesso, o que dificulta uma análise completa do percurso da madeira.

Nesse contexto, o presente trabalho propõe um sistema de rastreabilidade que integra os dados das várias fases que abrangem a exploração florestal, desde a floresta até à fábrica, garantindo um fluxo contínuo de informação. O sistema é baseado numa ontologia que, além de formalizar o conhecimento necessário para a rastreabilidade, permite a identificação de erros e inconsistências através de mecanismos de *reasoning*, assegurando assim a transparência e a confiabilidade dos registos recolhidos.

Além disso, com base na ontologia instanciada, são usadas técnicas de *Graph Machine Learning* para treinar um modelo capaz de prever dados em falta e identificar relações semânticas implícitas.

A abordagem foi avaliada no contexto do projeto Floresta 4.0, tendo apresentado resultados promissores quanto à sua eficácia. Além de responder às necessidades da rastreabilidade, detetou inconsistências que até então não tinham sido identificadas pelos especialistas do domínio.

**Palavras-chave:** Ontologia, Rastreabilidade, *Graph Machine Learning*, Cadeia de Abastecimento Florestal, Transparência

# Abstract

In recent years, growing concerns about deforestation have driven the need to monitor the origin and history of the wood arriving at factories. This has led to the adoption of traceability systems in the forestry sector. However, many of these systems are still manual and paper-based, which makes them susceptible to errors and falsification.

With the advancement of Industry 4.0 and the digitalisation of forestry operations, there is an opportunity for the digital transformation of traceability. However, the digitalisation process faces challenges at various levels, one of the main issues being that the information sources are dispersed among the various stakeholders in the forest supply chain, resulting in inaccurate and hard-to-access data, which hinders a comprehensive analysis of the wood's journey.

In this context, the present work proposes a traceability system that integrates data from the various stages encompassing forestry exploitation, from the forest to the factory, ensuring a continuous flow of information. The system is based on an ontology that, in addition to formalising the knowledge necessary for traceability, allows for the identification of errors and inconsistencies through reasoning mechanisms, thereby ensuring the transparency and reliability of the collected records.

Furthermore, based on the instantiated ontology, Graph Machine Learning techniques are used to train a model capable of predicting missing data and identifying implicit semantic relations.

The approach was evaluated in the context of the Floresta 4.0 project and showed promising results in terms of its effectiveness. In addition to addressing the needs of traceability, it detected inconsistencies that had not previously been identified by domain experts.

**Keywords:** Ontology, Traceability, Graph Machine Learning, Forest Supply Chain, Transparency

# Publicações Científicas

- **Hugo Silva, Reinaldo Gomes, Cristóvão Sousa.** Classifying Forest Supply Chain Traceability Data with an Ontological Approach. *6th International Conference on Industry 4.0 and Smart Manufacturing.*

**Abstract:** In the context of Industry 4.0, effective traceability within Forest Supply Chains is essential for ensuring sustainability and transparency. This paper introduces an ontological framework for classifying Forest Supply Chain traceability data, addressing the critical need for accurate tracking from forest to final product. While existing methods often face challenges with semantic interoperability, our approach leverages ontologies to create a standardized data representation that enhances communication and decision-making among stakeholders. The proposed ontology encompasses all phases of the wood cycle and aligns with industry standards, facilitating integration into existing systems. Unlike previous studies that used simulated data, this work validates the ontology with real data, demonstrating its practical applicability. Our ontology serves as a reference for business organizations and researchers in the forestry sector, promoting efficient data management and advancing sustainable forest supply chain practices.

# Conteúdo

|  |           |
|--|-----------|
| Lista de Figuras   | viii      |
| Lista de Tabelas   | ix        |
| Exemplos de Códigos  | ix        |
| <b>1 Introdução</b>  | <b>1</b>  |
| 1.1 Objetivos . . . . .  | 3         |
| 1.2 Metodologia . . . . .  | 3         |
| 1.3 Estrutura do documento . . . . .   | 5         |
| <b>2 Rastreabilidade e Representação do Conhecimento</b>                           | <b>6</b>  |
| 2.1 Rastreabilidade . . . . .  | 6         |
| 2.1.1 Sistemas de Rastreabilidade . . . . .  | 7         |
| 2.2 Representação de Conhecimento . . . . .  | 8         |
| 2.2.1 Ontologias . . . . .   | 8         |
| 2.2.2 Grafos de Conhecimento . . . . .   | 10        |
| 2.3 Inteligência Artificial . . . . .  | 11        |
| 2.3.1 <i>Machine Learning</i> . . . . .  | 11        |
| 2.3.2 <i>Graph Machine Learning</i> . . . . .                                      | 12        |
| 2.4 Trabalhos Relacionados . . . . .   | 13        |
| 2.4.1 Análise Crítica . . . . .  | 14        |
| <b>3 Modelo de Representação de Conhecimento para a Rastreabilidade de Madeira</b> | <b>16</b> |
| 3.1 Abordagem Técnico-Científica . . . . .   | 17        |
| 3.2 Conceptualização do Domínio . . . . .  | 18        |
| 3.2.1 Caracterização da Cadeia de Abastecimento Florestal . . . . .                | 18        |
| 3.3 Representação Conceptual da Ontologia . . . . .                                | 22        |
| <b>4 Especificação dos Artefactos</b>  | <b>26</b> |
| 4.1 Arquitetura da Solução . . . . .   | 27        |
| 4.2 Formalização da Ontologia . . . . .  | 29        |
| 4.3 Especificação do Algoritmo de <i>Graph Machine Learning</i> . . . . .          | 33        |
| 4.3.1 <i>Feature Engineering</i> e Definição de <i>Metapaths</i> . . . . .         | 34        |
| 4.3.2 Treino . . . . .   | 35        |
| <b>5 Caso de Estudo</b>  | <b>36</b> |

|          |   |           |
|----------|---|-----------|
| 5.1      | Caracterização do Caso de Estudo . . . . .                            | 37        |
| 5.2      | Resultados do Sistema de Rastreabilidade . . . . .                    | 39        |
| 5.3      | Resultados da Aplicação do Algoritmo GML . . . . .                    | 42        |
| <b>6</b> | <b>Conclusões e Trabalho Futuro</b>                                   | <b>45</b> |
| 6.1      | Reflexão Crítica . . . . .  | 45        |
| 6.2      | Trabalho Futuro . . . . .   | 46        |
|          | <b>Bibliografia</b>   | <b>47</b> |
| <b>A</b> | <b>Anexos</b>   | <b>52</b> |
| A.1      | Resultados do Sistema de Rastreabilidade . . . . .                    | 52        |
| A.2      | Evolução do Treino do Algoritmo de Graph Machine Learning (GML) . . . | 55        |
| A.3      | Regras da Ontologia . . . . .   | 56        |

# Lista de Figuras

|    |   |    |
|----|---|----|
| 1  | Fases da metodologia de trabalho . . . . .  | 4  |
| 2  | Rastreabilidade bidirecional . . . . .  | 7  |
| 3  | Excerto de uma ontologia de rastreabilidade . . . . .   | 9  |
| 4  | Principais tarefas de <i>Graph Machine Learning</i> . . . . .   | 12 |
| 5  | Interação entre a abordagem semântica e a abordagem de IA . . . . .                                       | 18 |
| 6  | Fases da cadeia de abastecimento florestal . . . . .  | 19 |
| 7  | Atividades típicas na exploração de madeira . . . . .   | 20 |
| 8  | Fluxo de informação do processo de exploração florestal . . . . .   | 21 |
| 9  | Representação conceptual do conhecimento explícito . . . . .  | 23 |
| 10 | Representação conceptual do conhecimento implícito . . . . .  | 25 |
| 11 | <b>a) Arquitetura Floresta 4.0 b) Arquitetura Floresta 4.0 com o sistema de rastreabilidade</b> . . . . . | 27 |
| 12 | Arquitetura do sistema de rastreabilidade . . . . .   | 29 |
| 13 | Fluxo de criação da ontologia . . . . .   | 30 |
| 14 | Exemplo da necessidade de relações inversas . . . . .   | 35 |
| 15 | Resultados da arquitetura Floresta 4.0 . . . . .  | 37 |
| 16 | <i>Precision e Recall</i> , Adaptado de Encord (2024) . . . . .   | 39 |
| 17 | Origem da carga desconhecida . . . . .  | 41 |
| 18 | Carga de madeira adulterada . . . . .   | 41 |
| 19 | Resultado do algoritmo <i>MetaPath2Vec</i> . . . . .  | 42 |
| 20 | Classificação de madeira (Certificada/Não Certificada) . . . . .  | 52 |
| 21 | Desperdício de recursos identificados . . . . .   | 52 |
| 22 | Inconsistências identificadas nos dados . . . . .   | 53 |
| 23 | Ações necessárias para melhorar a sustentabilidade . . . . .  | 53 |
| 24 | Transparência da rastreabilidade ao longo das diversas fases . . . . .                                    | 53 |
| 25 | Inconsistências Coordenadas GPS . . . . .   | 54 |
| 26 | Transportes com excesso de peso . . . . .   | 54 |
| 27 | Evolução do treino do modelo . . . . .  | 55 |

# Lista de Tabelas

|    |  |    |
|----|--|----|
| 1  | Diferença de abordagem entre a literatura e a solução proposta . . . . . | 15 |
| 2  | Contribuições de cada perspectiva científica . . . . .                   | 17 |
| 3  | Questões de competência . . . . .  | 22 |
| 5  | Relações semânticas da ontologia . . . . .                               | 25 |
| 6  | Excerto das regras SWRL da ontologia . . . . .                           | 32 |
| 7  | Vantagens e desvantagens do <i>MetaPath2Vec</i> e HGNNs . . . . .        | 34 |
| 8  | Resultados das classificações da ontologia . . . . .                     | 40 |
| 9  | Relações semânticas obtidas pelo algoritmo <i>MetaPath2Vec</i> . . . . . | 43 |
| 10 | Resultados das previsões . . . . .                                       | 44 |
| 11 | Regras SWRL (Parte 1) . . . . .  | 56 |
| 12 | Regras SWRL (Parte 2) . . . . .  | 57 |
| 13 | Regras SWRL (Parte 3) . . . . .  | 58 |

# Exemplos de Código

|   |   |    |
|---|---|----|
| 1 | Excerto da ontologia criada em LinkML . . . . .                         | 30 |
| 2 | Excerto dos <i>Shapes</i> gerados para validação de dados RDF . . . . . | 30 |
| 3 | <i>Metapaths</i> do algoritmo . . . . .                                 | 34 |

# Lista de Abreviaturas e Siglas

**AAS** Asset Administration Shell. 46

**CSV** Comma-Separated Value. 30

**DL** Lógica Descritiva. 32

**DSR** Design Science Research. 3

**DSS** Sistema de Apoio à Decisão. 3, 17, 18, 26, 28, 37, 38, 41

**ERP** Enterprise Resource Planning. 7

**FSC** Forest Stewardship Council. 21

**GML** Graph Machine Learning. vii, 3, 12, 13, 17, 26, 33, 34, 45, 55

**GPS** Global Positioning System. 21

**HGNNs** Heterogeneous Graph Neural Networks. 33, 34

**HTTP** Hypertext Transfer Protocol. 28

**IA** Inteligência Artificial. viii, 2, 3, 8, 11, 17, 18, 45, 46

**IoT** Internet das Coisas. 2, 7

**JSON** JavaScript Object Notation. 28, 30

**KG** Grafos de conhecimento. 10, 17

**KOS** Sistema de Organização de Conhecimento. 26

**LPG** Labeled Property Graph. 10, 11

**MAE** Erro Absoluto Médio. 43

**OWL** Web Ontology Language. 10, 29, 30

**PEFC** Programme for the Endorsement of Forest Certification. 21

**RDF** Resource Description Framework. 10, 11, 28, 30

**RFID** Radiofrequência. 2, 7, 13

**SHACL** Shapes Constraint Language. 30

**SWRL** Semantic Web Rule Language. 32, 56–58

**W3C** World Wide Web Consortium. 10

**XML** Extensible Markup Language. 27

**YAML** Ain't Markup Language. 30

# Capítulo 1

## Introdução

A crescente degradação ambiental e a exploração ilegal de recursos naturais exigem uma resposta imediata das indústrias para garantir práticas sustentáveis. A consciência global sobre a necessidade de tais práticas aumentou significativamente, impulsionada por problemas ambientais como mudanças climáticas, degradação de ecossistemas e perda de biodiversidade [1]. Nesse contexto, as indústrias enfrentam um compromisso crescente com a sustentabilidade. Um exemplo disso é o recente acordo alcançado na União Europeia para a implementação de um regulamento sobre sustentabilidade. Esse regulamento estabelece critérios rigorosos sobre a forma como as indústrias devem gerir os seus produtos ao longo do seu ciclo de vida [2].

Além das políticas governamentais, também se verifica uma crescente pressão por parte dos consumidores, que estão cada vez mais conscientes e exigem maior transparência sobre os produtos e os seus impactos ambientais [3]. Este cenário não só exige uma adaptação das indústrias às regulamentações, mas também obriga a que sejam fornecidas informações sobre a origem do produto e dos materiais utilizados na sua produção.

Nesse sentido, a rastreabilidade é um fator crítico de sucesso para promover a transparência e a sustentabilidade. Definida como a capacidade de rastrear e documentar toda a trajetória de um produto, desde a sua origem até à entrega ao consumidor final, a rastreabilidade oferece uma visão clara e detalhada de cada etapa do ciclo de vida do produto [4]. Essa capacidade oferece vantagens como a redução do desperdício de recursos e a deteção de problemas ao longo do processo produtivo. Além disso, fortalece a confiança dos consumidores, que podem verificar a autenticidade e a sustentabilidade dos produtos que adquirem [5].

Na indústria florestal, a rastreabilidade é particularmente importante, pois assegura que a madeira comercializada provém de fontes sustentáveis e legais [6]. Essa garantia é fundamental para o combate aos problemas de exploração ilegal e desflorestação que têm crescido significativamente e impactam a preservação das florestas como fonte de biodiversidade e recursos naturais [6]. Apesar da sua importância para o setor, a rastreabilidade ainda enfrenta desafios, uma vez que grande parte dos seus processos ainda são manuais e baseados em papel [7]. Essa abordagem obsoleta limita o acesso e a atualização dos dados, o que origina, conseqüentemente, a possibilidade de erros e falsificações quanto à origem da madeira [8].

Diante da necessidade de garantir a rastreabilidade de forma mais digitalizada na indústria

florestal, a literatura propôs sistemas mais automatizados utilizando tecnologias como a Identificação por Radiofrequência (RFID) [9]. A abordagem desses sistemas consiste em colocar uma etiqueta na árvore que depois permite a recolha de dados em tempo real, nas diversas fases pela qual a madeira passa até chegar à fábrica. Assim, todo o percurso é monitorizado e registado com precisão e transparência [10].

Apesar das vantagens e dos bons resultados que os sistemas alcançaram, a sua adoção pelas entidades florestais nunca avançou. Para isso, contribui o custo elevado da implementação das tecnologias RFID e a logística que é necessária para marcar cada uma das árvores [11]. Além disso, o facto da cadeia de abastecimento florestal envolver vários intervenientes, cada um com diferentes interesses e práticas, torna a criação de um sistema unificado ainda mais difícil [9, 12].

Considerando o desafio de encontrar uma solução que permita a migração dos métodos manuais para sistemas automatizados e digitais, este trabalho propõe um sistema de rastreabilidade baseado no fluxo de exploração florestal. O objetivo deste sistema é integrar os dados de várias fontes de informação<sup>1</sup> de modo a garantir a rastreabilidade. Essas fontes estão, tipicamente, dispersas entre os diferentes intervenientes da cadeia de abastecimento florestal e produzem dados de forma fragmentada (por exemplo, o registo de árvores cortadas é feito num sistema e o registo do seu transporte noutra). Quando integrados corretamente, esses dados permitem construir uma visão completa do percurso da madeira. Ao contrário dos trabalhos propostos na literatura, que exigem investimentos em novas tecnologias e processos manuais para a marcação de árvores, o sistema proposto neste trabalho tira partido das tecnologias já existentes na cadeia de abastecimento florestal. Dessa forma, é evitada a necessidade de grande reengenharia de sistemas.

De forma geral, espera-se que o sistema proposto seja capaz de garantir a rastreabilidade da madeira sob duas perspetivas:

- Da floresta à fábrica: Monitoriza cada etapa do ciclo de vida da madeira, desde a colheita até à fábrica, contribuindo para a garantia da legalidade e a eficiência logística.
- Da fábrica à floresta: Permite verificar a origem da madeira e identificar potenciais erros ou inconformidades, promovendo a transparência e a conformidade com práticas de sustentabilidade.

Com base nessas perspetivas, além da monitorização, a solução será capaz de inferir se a madeira que chega à fábrica tem uma origem considerada sustentável, atribuindo um grau de confiança à sua transparência.

Para responder aos desafios da solução proposta, surge a seguinte questão de investigação: **De que forma podemos estruturar e organizar os dados das atividades de exploração florestal, integrando conhecimento específico do domínio e diferentes fontes de informação, num modelo que permita determinar a rastreabilidade da madeira, classifique-a quanto à sua sustentabilidade e identifique falhas na consistência dos dados?**

Para responder a esta questão, serão seguidas duas abordagens: uma abordagem sócio-semântica<sup>2</sup> e uma abordagem de Inteligência Artificial (IA). A abordagem sócio-semântica

---

<sup>1</sup>Sistemas de monitorização baseados em IoT e plataformas *Track and Trace*

<sup>2</sup>Abordagem que combina a representação do conhecimento com o contexto social

permitirá a representação do conhecimento, contribuindo para a estruturação e integração de dados provenientes de diversas fontes. Essa abordagem fundamenta-se na formalização de conceitos e relações relevantes no domínio florestal, facilitando a troca de informações entre os intervenientes e promovendo a interoperabilidade entre sistemas. Por sua vez, a abordagem de IA complementar a abordagem sócio-semântica, permitindo melhorar a representação do conhecimento ao identificar relações implícitas entre os conceitos envolvidos. Além disso, será também utilizada para descobrir dados em falta nos registos de rastreabilidade.

## 1.1 Objetivos

O principal objetivo deste trabalho é desenvolver um modelo de rastreabilidade e implementar uma solução baseada no fluxo de exploração florestal<sup>3</sup>. Esta solução deve integrar os dados de todas as atividades que ocorrem desde a floresta até à fábrica, fornecendo, por meio de mecanismos de rastreabilidade, o histórico da origem da madeira, permitindo assim uma gestão mais eficiente e transparente de todo o processo.

A solução desenvolvida é baseada numa ontologia, implementada em duas fases. Na primeira fase, formaliza-se o conhecimento necessário para assegurar a rastreabilidade da madeira, considerando os principais *standards* da indústria e a experiência dos especialistas do domínio. Posteriormente, a ontologia é refinada para incluir mecanismos de validação e deteção de inconsistências, permitindo identificar falhas nos dados, como inconformidades e erros operacionais. Este processo garante a integridade e a confiança na informação, assegurando que a solução de rastreabilidade monitoriza de forma eficaz todas as etapas do ciclo de vida da madeira.

Para alcançar o objetivo principal, podem ser descritos os seguintes objetivos secundários:

1. Desenvolver uma ontologia que conceptualize as classes, propriedades, relações e regras relevantes para o domínio.
2. A partir do resultado do objetivo 1, desenvolver um artefacto que permita organizar e estruturar os dados de modo a extrair novo conhecimento com base nas regras e atributos de qualidade definidos na ontologia.
3. Desenhar e implementar uma arquitetura que integre o artefacto desenvolvido no objetivo 2 com um Sistema de Apoio à Decisão (DSS).
4. Aplicar técnicas de *GML* de modo a obter um modelo capaz de identificar novas relações semânticas e efetuar a previsão de dados em falta.

## 1.2 Metodologia

A elaboração da metodologia de trabalho é essencial no desenvolvimento de trabalhos técnico-científicos. Nesse sentido, o plano de trabalho desta dissertação seguiu uma abordagem baseada nos princípios da metodologia *Design Science Research (DSR)* [13]. Durante todas as fases do projeto, existiu uma participação ativa dos especialistas do

---

<sup>3</sup>O fluxo de exploração florestal refere-se às etapas que envolvem a gestão, a colheita e o transporte de madeira, desde a floresta até à fábrica. Neste processo, não estão incluídas as etapas de transformação da madeira em produtos finais, nem a venda subsequente aos consumidores.

domínio, de modo que a solução desenvolvida atendesse às suas necessidades e expectativas.

A DSR é uma abordagem que se concentra no desenvolvimento e avaliação de artefactos inovadores que visam resolver problemas práticos e contribuir para o avanço do conhecimento académico [14]. O resultado da aplicação desta metodologia é um artefacto (modelo, método, tecnologia, etc.) que visa atingir um determinado objetivo específico. Além da sua aplicação prática, a DRS também procura oferecer novas teorias que enriquecem o conhecimento existente e orientam trabalhos futuros [15].

A Figura 1 apresenta as quatro fases do trabalho realizado, nomeadamente: i) Estudo do Domínio, ii) Análise do Problema, iii) Desenvolvimento dos Artefactos e iv) Validação dos Artefactos. Essas fases são compostas por um conjunto de tarefas, cujos os resultados contribuem para o progresso do trabalho.

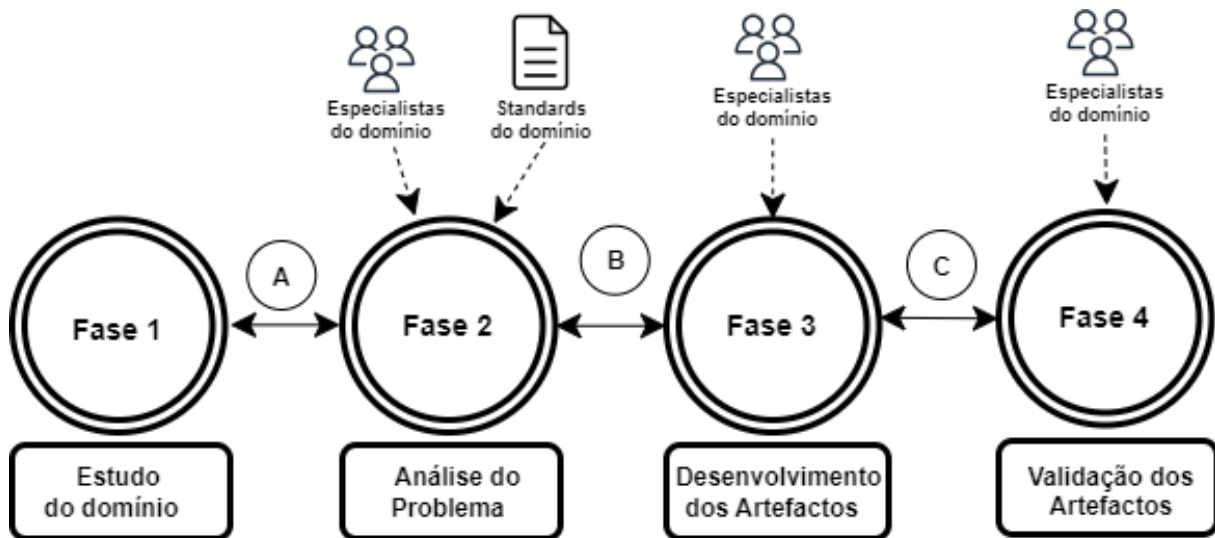


Figura 1: Fases da metodologia de trabalho

A fase i) consistiu no estudo do domínio e envolveu a compreensão do modelo de negócio, as suas necessidades e problemas. Também foi realizada uma revisão da literatura para conhecer os principais trabalhos publicados sobre o tema da dissertação. O resultado dessa fase foi um conhecimento mais abrangente do domínio e dos principais desafios.

Na fase ii), a principal tarefa foi a análise detalhada da cadeia de abastecimento florestal. Essa análise foi realizada por meio de sessões de trabalho com especialistas do domínio e consistiu na identificação dos diversos tipos e formatos de dados, assim como os principais fluxos de informação. O resultado dessa fase foi a definição dos objetivos e a especificação dos requisitos dos artefactos a serem desenvolvidos.

A fase iii) consistiu no desenvolvimento dos artefactos da solução, resultando na criação de um protótipo. Durante este processo, a fase 2 foi analisada de novo para esclarecer problemas e garantir que os artefactos desenvolvidos estavam alinhados com os requisitos.

A fase iv), por sua vez, envolveu a avaliação da solução desenvolvida através de um caso de estudo. Na avaliação, foi considerada a utilidade e a precisão dos artefactos, assim como o *feedback* dos especialistas do domínio.

Com base nas fases apresentadas, foram definidos três ciclos de interação entre elas, identificados pelas letras A, B e C. O ciclo A, entre as fases i) e ii), assegurou a relevância do problema. O ciclo B, entre as fases ii) e iii), garantiu que os artefactos desenvolvidos atendiam às necessidades dos *stakeholders*. Por fim, o ciclo C, entre as fases iii) e iv), assegurou o rigor e utilidade dos artefactos. Dessa forma, foi adotada uma abordagem incremental e interativa, permitindo que o resultado de uma fase posterior contribua para o resultado de uma fase anterior. Assim, foi assegurado que os artefactos foram avaliados e evoluíram de forma progressiva até atingir um grau alto de maturidade.

## 1.3 Estrutura do documento

Este documento está dividido em 6 capítulos, cada um abordando um tema relevante para a dissertação.

- **Capítulo 1 - *Introdução***: É descrita a motivação e o contexto do trabalho, bem como os seus objetivos, os resultados esperados e a metodologia de trabalho.
- **Capítulo 2 - *Rastreabilidade e Representação do Conhecimento***: São apresentados os principais conceitos teóricos que fundamentam e ajudam a compreender o âmbito da dissertação. Além disso, é realizada uma revisão da literatura relacionada.
- **Capítulo 3 - *Modelo de Representação de Conhecimento para a Rastreabilidade de Madeira***: Apresenta o modelo de representação de conhecimento desenvolvido para a rastreabilidade de madeira. São descritas a abordagem técnico científica e a representação conceptual da ontologia.
- **Capítulo 4 - *Especificação dos Artefactos***: São apresentados e descritos os artefactos desenvolvidos.
- **Capítulo 5 - *Caso de estudo***: É apresentado o caso de estudo em que se insere esta dissertação. São também apresentados os resultados e é discutido a utilidade dos artefactos desenvolvidos ao longo do projeto.
- **Capítulo 6 - *Conclusões e Trabalho Futuro***: São apresentados, em forma de conclusão, os resultados obtidos durante a dissertação. É também realizada uma reflexão crítica e descrito o trabalho futuro.

# Capítulo 2

## Rastreabilidade e Representação do Conhecimento

Para desenvolver um sistema eficiente de rastreabilidade, é essencial compreender as abordagens já existentes e os desafios que estas enfrentam. A rastreabilidade tem ganho destaque na literatura recente, impulsionada pela crescente necessidade de garantir a origem sustentável dos produtos [16].

Neste capítulo, apresentam-se os principais conceitos relacionados à rastreabilidade. Em seguida, discute-se como a organização e a representação do conhecimento podem aumentar a eficiência de um sistema de rastreabilidade. Abordam-se também técnicas de IA, com ênfase em *Graph Machine Learning* e no seu potencial contributo, tanto para a rastreabilidade como para a representação do conhecimento. Por fim, são apresentados os principais trabalhos relacionados com os objetivos deste estudo.

### 2.1 Rastreabilidade

Rastreabilidade pode ser definida como a capacidade de identificar e seguir o percurso de um produto ou informação ao longo de todo o seu ciclo de vida [17]. No contexto industrial, esse conceito refere-se à capacidade de monitorizar a localização atual e toda a trajetória de uma matéria-prima, desde a sua origem até à sua transformação em produto final [18].

De forma mais pragmática, rastreabilidade procura, ao longo da jornada de um determinado produto, recolher informação que permita compreender os seguintes aspetos [19]:

- **What:** Refere-se ao próprio *item* e informação subjacente. Isso inclui a identificação do produto, categoria, a composição, o número de lote e quaisquer atributos ou características específicas que sejam relevantes para a sua monitorização .
- **When:** Relaciona-se com eventos que correspondem a momentos e datas associados às diferentes etapas do ciclo de vida do produto. Inclui a data e a hora de produção, armazenamento, transporte ou qualquer outro evento.
- **Where:** Diz respeito às localizações físicas ou virtuais associadas ao produto durante o seu ciclo de vida.

- **Why:** Define a razão para a rastreabilidade. Pode incluir a necessidade de garantir a segurança e a qualidade do produto, cumprir com regulamentações e normas, ou atender a requisitos específicos de clientes e mercados.
- **Who:** Refere-se às pessoas ou ativos envolvidos em cada etapa do ciclo de vida do produto, como fabricantes, fornecedores e operadores de logística.

Existem vários tipos de rastreabilidade, contudo os dois mais comuns são a rastreabilidade direta (*forward traceability*) e a rastreabilidade inversa (*backward traceability*). A rastreabilidade direta refere-se ao fluxo de informações que acompanha o produto desde a sua origem até ao consumidor final, permitindo monitorizar todas as etapas intermédias de produção e distribuição [20]. Já a rastreabilidade inversa foca no caminho oposto, sendo utilizada principalmente para identificar a origem de problemas, uma vez que possibilita rastrear o produto de volta ao fornecedor ou à matéria-prima [20].

A Figura 2 ilustra um exemplo dos dois tipos de rastreabilidade anteriormente descritos.

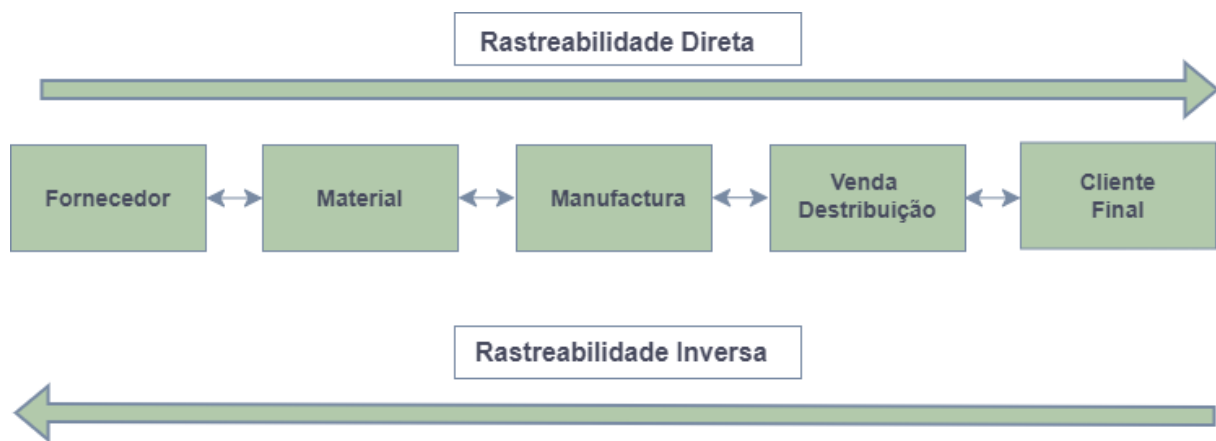


Figura 2: Rastreabilidade bidirecional

### 2.1.1 Sistemas de Rastreabilidade

Um sistema de rastreabilidade pode ser definido como o conjunto de processos, ferramentas e tecnologias que permitem acompanhar um produto, matéria-prima ou informação ao longo de todo o seu ciclo de vida, desde a origem até ao consumidor final [21].

Os primeiros sistemas de rastreabilidade que surgiram eram manuais, com os registos a serem feitos em papel. Com o avanço das tecnologias de informação e o surgimento de sistemas ERP, a rastreabilidade tornou-se mais digital e integrada em várias atividades da cadeia de abastecimento [22]. Atualmente, a evolução desses sistemas é marcada pela adoção de tecnologias como RFID e IoT, que permitem a recolha de dados em tempo real e melhoram a visibilidade e o controlo ao longo de todos os processos da cadeia de abastecimento [23].

Apesar dos avanços tecnológicos, são poucos os setores que já possuem sistemas de rastreabilidade com um grau elevado de maturidade [24]. O setor alimentar e o farmacêutico são as exceções. Isso deve-se fundamentalmente à sua natureza, na qual os regulamentos exigem um nível elevado de monitorização e controlo [24]. Nos demais setores, a diversidade de agentes envolvidos e a complexidade das cadeias de abastecimento dificultam a adoção de soluções eficazes. Além disso, a necessidade de integração de diferentes fontes

de informação, por vezes de vários países, representa também um desafio. Esse desafio deve-se à falta de padronização dos dados, que resulta em terminologia e semântica distintas que geram ambiguidades [25].

## 2.2 Representação de Conhecimento

A representação de conhecimento é uma área da IA que se dedica a estudar como organizar e estruturar o conhecimento de forma a torná-lo acessível e utilizável por sistemas computacionais. É essencial para que os computadores possam compreender e manipular informações de modo a tomar decisões e resolver problemas complexos de forma automatizada [26].

Existem diversas áreas de aplicação da representação de conhecimento, como por exemplo:

- **Web Semântica:** Permite a integração de dados de diferentes fontes na *web*, facilitando a procura e organização de informação.
- **Processamento de Linguagem Natural:** Aplicações que analisam e interpretam texto humano, permitindo a extração de informações relevantes e compreensão semântica.
- **Sistemas Especialistas:** Ferramentas que simulam o raciocínio humano em áreas específicas como diagnóstico médico e tomada de decisões. Muitas vezes, são sistemas baseados em Q&A<sup>1</sup>.

Na área da rastreabilidade, um dos problemas mais comuns é a falta de orientações adequadas sobre quais elementos devem ser partilhados e como a comunicação deve ser estruturada entre os intervenientes [27]. Nesse sentido, a representação de conhecimento pode desempenhar um papel fundamental na definição de estruturas comuns, facilitando a troca de informações bem definidas [27].

Para estruturar e organizar o conhecimento, são utilizadas diversas técnicas, cada uma adequada a diferentes tipos de problemas e necessidades [26]. Neste trabalho, serão abordadas duas das mais comuns: as ontologias e os grafos de conhecimento.

### 2.2.1 Ontologias

Por definição, uma ontologia refere-se a uma representação formal e estruturada do conhecimento de um domínio específico [28]. Este conhecimento é partilhado por uma comunidade que pode ser, por exemplo, todos os intervenientes de uma cadeia de abastecimento.

As ontologias criam um modelo semântico que permite compreender de maneira mais abrangente e precisa o contexto de um domínio. Através da definição do modelo, é possível inferir conhecimento por meio de “raciocínios” automatizados, que descobrem novas informações com base em regras e relações estabelecidas.

No contexto dos sistemas de rastreabilidade, as ontologias podem ser uma mais-valia, pois promovem a interoperabilidade, facilitam a comunicação eficiente entre *stakeholders* e permitem adaptações a mudanças no domínio [29].

---

<sup>1</sup>Q&A refere-se a sistemas que permitem a interação através de perguntas e respostas.

Embora nem todas as ontologias sejam definidas com a mesma metodologia ou linguagem, a maioria possui os seguintes elementos:

- **Classes:** Representam os conceitos principais ou categorias dentro do domínio. São utilizadas para agrupar indivíduos que partilham características semelhantes [30].
- **Propriedades:** Também designadas como *data properties*, descrevem as características das classes.
- **Relações:** Também conhecidas como *object properties*, definem como as classes se relacionam entre si.
- **Restrições:** Definem limites ou regras para as propriedades e relações.
- **Indivíduos:** Instâncias específicas de uma classe.

A Figura 3 ilustra um excerto de uma ontologia conceptual que poderia servir de base para um sistema de rastreabilidade. Esta ontologia define, de maneira estruturada, os conceitos essenciais para a rastreabilidade de um produto, facilitando a recolha e a organização das informações ao longo de todo o seu ciclo de vida. Posteriormente, essa ontologia conceptual poderia ser formalizada, permitindo o seu uso computacional.

Além da representação “estática”, que inclui classes como Produto, Origem, Fornecedor e Material a ontologia também possui classes para representar conhecimento implícito<sup>2</sup>. No exemplo, essas classes são chamadas de Rastreabilidade Consistente e Rastreabilidade Inconsistente. A classe Rastreabilidade Consistente representaria os produtos cuja rastreabilidade é completa. Já a classe Rastreabilidade Inconsistente, representaria produtos com rastreabilidade afetada devido a lacunas ou erros nos dados.

O conhecimento implícito seria obtido por meio das regras definidas na ontologia, que classificam os dados com base em factos. Por exemplo, uma regra poderia estabelecer que, se um produto tem uma origem conhecida, mas é constituído por materiais cuja origem é desconhecida então a sua rastreabilidade é inconsistente.

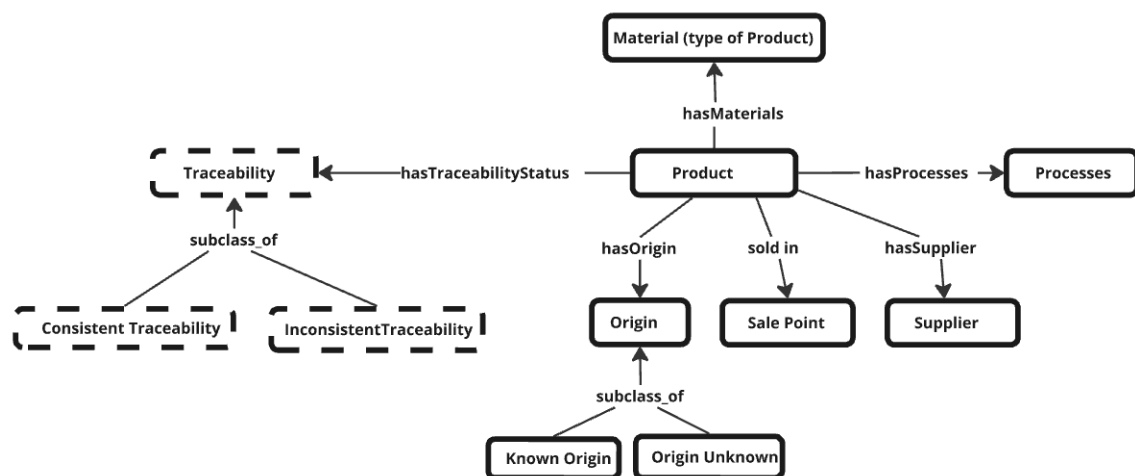


Figura 3: Excerto de uma ontologia de rastreabilidade

<sup>2</sup>Conhecimento que não é explicitamente representado na ontologia, mas que pode ser inferido através de mecanismos de *reasoning*

De forma a possibilitar a interpretação de ontologias, de modo que possam ser compreendidas e processadas por máquinas, diversas linguagens foram desenvolvidas ao longo dos anos. Entre as principais, está a *Web Ontology Language (OWL)*.

A OWL é uma linguagem declarativa para a criação e manipulação de ontologias na *semantic web*. Desenvolvida para expressar de forma detalhada e precisa as relações e propriedades de conceitos num domínio específico, a OWL permite a definição de hierarquias complexas de classes e propriedades, além de possibilitar a imposição de restrições e a realização de inferências sobre os dados. É projetada para trabalhar em conjunto com a linguagem *Resource Description Framework (RDF)* que é o padrão da W3C<sup>3</sup> para a descrição conceitual e modelação de informação [31].

Desenvolver uma ontologia em OWL, além das vantagens mencionadas, permite também que esta seja publicada na W3C, o que facilita a sua integração com outras ontologias [32].

## 2.2.2 Grafos de Conhecimento

Grafos de conhecimento (KG) são estruturas de dados que representam informações e relações em forma de grafos<sup>4</sup>, onde os nós representam entidades e as arestas representam as relações entre essas entidades [33]. Abordagens orientadas a grafos têm vindo a ganhar popularidade na área da rastreabilidade devido à sua capacidade de modelar e visualizar ligações complexas entre objetos [34].

Normalmente, os KG resultam da aplicação de ontologias como camada esquemática. As ontologias oferecem um vocabulário estruturado e regras semânticas que orientam a organização dos dados. Dessa forma, os grafos de conhecimento além de representarem entidades e relações, também incorporam o significado semântico dessas relações. No domínio da rastreabilidade, isso é essencial, pois é necessário para além da identificação das relações entre diferentes entidades, compreender o contexto e o significado dessas interações [35]. Outro aspeto importante é a escalabilidade, uma vez que sistemas de rastreabilidade lidam com grandes volumes de dados provenientes de diversos pontos da cadeia de abastecimento. Estruturas orientadas a grafos, são altamente escaláveis, permitindo que novas entidades e relações sejam incorporadas de forma eficiente, sem a necessidade de alterações complexas [36].

Existem vários modelos de grafos utilizados na construção de KGs, sendo os mais comuns o RDF e o *Labeled Property Graph (LPG)* [37]:

RDF é um modelo que representa informações como triplos no formato “sujeito-predicado-objeto”. Cada nó do grafo representa uma entidade, e as arestas descrevem as relações entre essas entidades. Esse modelo é padronizado e favorece a interoperabilidade entre diferentes sistemas e fontes de dados, facilitando a integração e a troca de informações.

Já os grafos LPG são grafos mais flexíveis que permitem que tanto os nós quanto as arestas tenham rótulos e propriedades associadas. Num grafo LPG, cada nó pode representar uma entidade, como um produto ou um processo, e pode ter várias propriedades, como data de produção ou localização. Da mesma forma, as arestas que conectam os nós também

---

<sup>3</sup><https://www.w3.org/>

<sup>4</sup>Grafos são estruturas utilizadas para modelar relações entre objetos. Um grafo é composto por um conjunto de nós e arestas que os conectam.

podem ter rótulos que descrevem o tipo de relação, como “transportado-por”, e podem incluir propriedades adicionais, como a data de transporte e distância percorrida [38].

A principal diferença entre os dois modelos está na forma como representam e manipulam os dados. Enquanto os grafos RDF utilizam uma estrutura de triplos, o que é ideal para desafios de interoperabilidade, integração de sistemas e inferência de conhecimento implícito, os grafos LPG oferecem uma modelação mais rica e detalhada que permite consultas de informações de forma simples e direta [37].

### 2.2.2.1 Validação de dados em Grafos de Conhecimento

A validação e a integridade dos dados são componentes essenciais no contexto de grafos de conhecimento. Essa validação garante que os grafos representam de forma correta as entidades e as suas relações. Nesse contexto, a linguagem SHACL é a linguagem *standard* do W3C para validar grafos RDF e garantir que os dados atendem a restrições específicas. O SHACL permite definir formas (*shapes*) que descrevem as regras de integridade para os dados, como restrições sobre tipos de dados, cardinalidade e relações entre as entidades [39].

No domínio da rastreabilidade em que a precisão dos dados é especialmente importante, com o SHACL é possível verificar se os grafos de conhecimento seguem o modelo preestabelecido, evitando inconsistências e erros. Por exemplo, pode-se garantir que uma relação entre duas entidades seja validada de acordo com as regras semânticas definidas pela ontologia, ou que a quantidade de recursos relacionados a uma atividade esteja dentro dos limites esperados.

## 2.3 Inteligência Artificial

IA refere-se à capacidade de sistemas computacionais realizar tarefas que normalmente requerem inteligência humana. Essas tarefas podem ser raciocínio, aprendizagem e tomada de decisões [40]. Dentro das diversas subáreas de IA, existe o *machine learning* que, além de dotar sistemas de comportamentos autónomos semelhantes aos humanos, permite que estes aprendam face a novos contextos e tomem decisões.

No contexto da rastreabilidade, há necessidade de dotar o sistema de capacidade de decisão, mesmo que limitada face à ausência de dados que permitam uma inferência útil. Deste modo, a área de *machine learning* pode trazer contributos adicionais para a definição de um modelo de rastreabilidade mais resiliente e flexível.

### 2.3.1 *Machine Learning*

O *machine learning* é uma subárea da IA que se concentra no desenvolvimento de algoritmos e modelos que permitem que os sistemas aprendam a partir de dados. Esses sistemas conseguem realizar tarefas como previsão e classificação sem serem especificamente programados para o efeito [41].

Na rastreabilidade, os métodos de *machine learning* podem oferecer os seguintes benefícios:

- **Detecção de Anomalias:** Identifica comportamentos atípicos, como fraudes ou desvios de qualidade. Por exemplo, pode detetar um aumento súbito nas devoluções

de produtos, sinalizando possíveis problemas de qualidade na produção.

- **Preenchimento de Dados em Falta:** Consegue prever valores ou registos em falta com base em padrões. Por exemplo, se faltarem registos sobre o transporte de um lote de madeira, um modelo de *machine learning* pode descobrir esses registos utilizando dados de lotes semelhantes.
- **Previsão da Procura:** Consegue prever a procura de produtos, permitindo um planeamento mais eficaz, de modo a reduzir o desperdício de recursos. Como exemplo, é possível prever o aumento na procura de madeira durante determinadas épocas do ano, contribuindo para a gestão de *stocks* de forma adequada.

### 2.3.2 Graph Machine Learning

Na sua essência, GML é a aplicação de técnicas *machine learning* em dados estruturados no formato de grafo [42]. Ao contrário de outros formatos de dados, como tabelas ou sequências, onde as relações entre entidades não são modeladas, no GML, as conexões e interações são tidas em conta de modo a melhorar a qualidade das previsões.

Os algoritmos de GML possuem uma vasta gama de aplicações em diversos domínios. Essas aplicações estão normalmente associadas a uma das seguintes tarefas:

- **Node Classification:** Determinar o rótulo ou categoria de um nó com base nas suas conexões no grafo e nas características dos nós vizinhos.
- **Link Prediction:** Prever se há potenciais ligações (arestas) entre nós. Por exemplo, sugerir possíveis amigos numa rede social.
- **Community Detection:** Encontrar subgrupos no grafo com padrões de ligações mais fortes entre si. Por exemplo, identificar comportamentos semelhantes em clientes e encontrar utilizadores com interesses comuns em redes sociais.

A Figura 4 ilustra as três tarefas acima descritas.

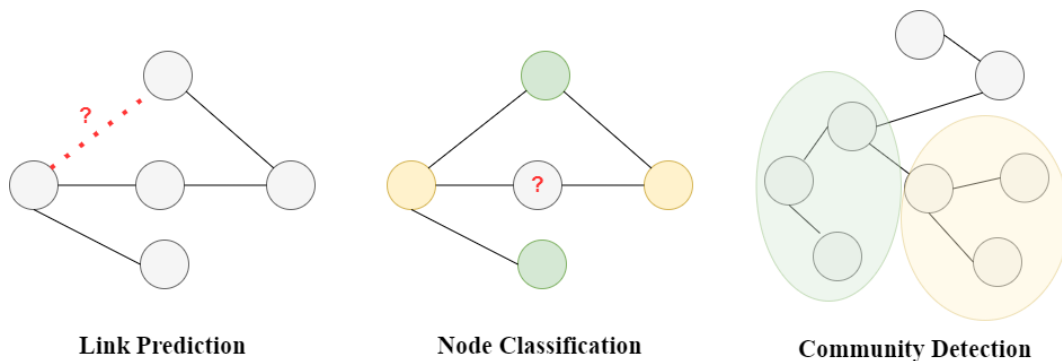


Figura 4: Principais tarefas de *Graph Machine Learning*

Para realizar essas tarefas existem as seguintes técnicas:

- **Graph Neural Networks (GNNs):** Redes neuronais projetadas especificamente para trabalhar em grafos, representando relações entre nós de maneira eficaz.
- **Graph Convolutional Networks (GCNs):** Um tipo específico de GNN que utiliza convoluções para agregar informações de nós vizinhos.

- **Graph Attention Networks (GATs):** Redes que utilizam mecanismos de atenção para ponderar a importância dos vizinhos na agregação de informações.
- **Node2Vec e DeepWalk:** Técnicas de aprendizagem de representação que transformam nós em vetores num espaço contínuo, preservando a estrutura do grafo.

Essas técnicas têm como objetivo gerar *embeddings*, que são representações vetoriais numéricas dos nós. Os *embeddings* permitem que as características e as relações entre os nós sejam representadas de maneira a serem facilmente analisadas. Um dos exemplos de análise que pode ser feita é a proximidade semântica [43].

A proximidade semântica é uma medida de similaridade entre nós que indica quão próximos estes estão em termos de significado ou função. Essa medida é útil em estruturas de dados baseadas em ontologias, pois permite a análise de interações entre diferentes entidades de modo a descobrir novas relações que não seriam evidentes à primeira vista. A eficácia dessa abordagem tem sido demonstrada em diversas aplicações nas áreas da biologia e biomedicina [43].

Em suma, comparativamente ao *machine learning* tradicional, o GML pode ser mais eficaz em sistemas de rastreabilidade baseados em ontologias, uma vez que explora de forma clara tanto a estrutura como o contexto dos dados.

## 2.4 Trabalhos Relacionados

Nos últimos 25 anos, a ciência florestal tem lidado extensivamente com a questão da rastreabilidade. Isso deve-se principalmente ao aumento da desflorestação e à exploração ilegal de madeira, que torna essencial garantir que esta seja extraída de forma sustentável [44].

Apesar do grande interesse da comunidade científica, muitas publicações limitam-se a abordar a questão de forma teórica, sem considerar a viabilidade prática das soluções propostas. Além disso, há trabalhos que se concentram apenas em fases específicas da exploração florestal, como a colheita ou o transporte. Somente algumas abordagens consideram a cadeia de abastecimento como um todo, levando em conta todas as fases que a constituem desde a colheita até à venda ao consumidor final.

Essas abordagens têm em comum o uso da tecnologia RFID, que permite a inserção de uma *tag* na árvore para a sua identificação. À medida que a árvore passa pelas diferentes fases até chegar à fábrica, a *tag* é lida por sensores instalados nas máquinas, que assim recolhem toda a informação sobre o seu percurso.

As três principais publicações no âmbito dos sistemas de rastreabilidade são resumidas a seguir:

- **Mtibaa et al.[45]:** Primeiro sistema de rastreabilidade proposto, constituído por três módulos: gestão de atividades, integração de dados de diferentes fontes e um portal *web* para consulta de informação pelos utilizadores.
- **Appelhanz et al.[46]:** Apresentam um sistema de rastreabilidade focado na transparência das informações sobre a origem da madeira. Definem um modelo de arquitetura de dados em quatro camadas e introduzem, pela primeira vez, o conceito de *backward traceability*.

- **Figorilli et al.[47]:** Estudo mais recente que, com base na arquitetura do trabalho [46], introduz a utilização da tecnologia *blockchain* para armazenar dados sobre o movimento da madeira.

Apesar dos estudos mostrarem o potencial das soluções propostas, a necessidade de aplicar manualmente a *tag* nas árvores e os elevados custos da tecnologia RFID têm dificultado a adoção desses sistemas em contextos reais [11]. Prova disso, é que as soluções apresentadas foram apenas testadas em ambientes simulados, sem considerar que os dados recolhidos podem apresentar diferentes formatos e proprietários.

Embora não proponham soluções concretas, outros trabalhos [48, 49] exploraram a rastreabilidade de madeira e destacam a necessidade de padronização de dados para a implementação de um sistema eficaz. É mencionado que apenas com a integração de várias fontes de informação, como sensores de monitorização, sistemas GIS, entre outros, é possível garantir o histórico completo da madeira.

Nesse sentido, a literatura apresenta duas publicações que formalizam, através de ontologias, o conhecimento sobre a cadeia de abastecimento florestal:

- **Blake et al.[50]:** A ontologia descreve as principais características dos produtos de madeira, os participantes envolvidos na sua exploração e os serviços que oferecem ou utilizam.
- **Efthymios Lallas et al.[51]:** Tal como o trabalho de [50], esta ontologia descreve os principais conceitos da cadeia de abastecimento florestal. A principal diferença é que os autores introduziram, pela primeira vez, o conceito de repositório de informação, que armazena os registos das várias fases que a madeira percorre, desde a floresta até à fábrica.

As ontologias propostas, apesar de oferecerem uma base sólida para a compreensão da cadeia de abastecimento florestal, não respondem de forma abrangente aos cinco termos essenciais para a rastreabilidade da madeira: porque, o que, quando, onde e quem. Essa limitação pode ser justificada pelo facto de terem sido desenvolvidas para outros objetivos, como a regulamentação e a gestão eficiente de recursos.

### 2.4.1 Análise Crítica

A revisão da literatura revela que, apesar da relevância do tema, os sistemas de rastreabilidade propostos falham em considerar de forma abrangente os diversos intervenientes e fontes de dados existentes na cadeia de abastecimento florestal. No que diz respeito às ontologias desenvolvidas, constata-se que estas não atendem plenamente às necessidades de rastreabilidade, o que pode ser atribuído a uma formalização inadequada e a uma ênfase excessiva na representação dos produtos, em detrimento da modelação dos processos e eventos associados à madeira.

Em contraste com os estudos anteriores, que se concentram principalmente nas tecnologias e em como podem ser utilizadas para assegurar o histórico da madeira, este trabalho visa desenvolver um sistema de rastreabilidade baseado num modelo semântico. O objetivo é integrar e validar dados provenientes de diferentes fontes de informação já existentes no setor florestal.

A Tabela 1 apresenta as principais diferenças entre as abordagens apresentadas na literatura e a abordagem proposta nesta dissertação.

| <b>Foco</b>                             | <b>Solução Proposta</b>  | <b>Literatura</b>  |
|---|--|--|
| <b>Objetivo</b>                         | Integrar dados de diferentes fontes de informação usando um modelo semântico como base | Propor tecnologias e arquiteturas específicas para a rastreabilidade |
| <b>Validação dos dados</b>              | Implementa um mecanismo de validação dos dados   | Não aborda a necessidade de validação dos dados                      |
| <b>Adoção prática</b>                   | Projetada para ser prática e aplicável em contextos reais                              | Soluções testadas apenas em ambientes simulados                      |
| <b>Identificação de inconsistências</b> | Identifica erros e inconsistências nos registros de rastreabilidade                    | Não existe detecção de possíveis erros                               |

Tabela 1: Diferença de abordagem entre a literatura e a solução proposta

## Capítulo 3

# Modelo de Representação de Conhecimento para a Rastreabilidade de Madeira

A rastreabilidade de madeira envolve a monitorização e o registo de todas as fases do seu ciclo de vida, desde a árvore em pé até à entrada na fábrica para processamento. Devido à complexidade e diversidade de entidades envolvidas, é necessário integrar várias fontes de dados, incluindo dados sobre origem, processamento, transporte e entrega. Apenas com essa integração é possível garantir que a rastreabilidade é realizada de forma completa e transparente.

Nesse cenário, a representação do conhecimento desempenha um papel essencial ao fornecer uma estrutura comum para descrever as entidades e as suas relações ao longo da cadeia de abastecimento. Essa abordagem não apenas facilita a interoperabilidade entre diferentes sistemas e tecnologias, como também assegura uma interpretação uniforme e precisa dos dados envolvidos.

Este capítulo descreve a abordagem técnico-científica adotada. Em seguida, é realizada a conceptualização do domínio. Por fim, é apresentada a representação conceptual do modelo de conhecimento na forma de uma ontologia.

### 3.1 Abordagem Técnico-Científica

Apesar dos avanços na digitalização da floresta, também conhecidos como a revolução digital no setor florestal, que se têm traduzido na incorporação de sensores nas máquinas que recolhem dados em tempo real do teatro de operações, a rastreabilidade de madeira ainda enfrenta vários desafios. O principal reside na existência de vários sistemas de monitorização e recolha de dados, que geram informações dispersas e dificultam a análise precisa de todas as fases da cadeia de abastecimento florestal. Nesse sentido, integrar todas as informações dispersas num único sistema de rastreabilidade é essencial.

Para isso, é proposto um modelo de representação de conhecimento para a rastreabilidade de madeira, desenvolvido com base numa abordagem sócio-semântica. Esse modelo é projetado para consolidar e organizar as informações provenientes de diferentes fontes, fornecendo uma visão coerente e integrada do ciclo de vida da madeira.

A abordagem sócio-semântica combina princípios de engenharia do conhecimento com a experiência dos especialistas do domínio e os principais *standards* da indústria, resultando num artefacto (ontologia) que além de permitir a formalização do conhecimento, também permite a realização de inferências lógicas.

Para além da abordagem sócio-semântica, serão também utilizadas técnicas de IA, nomeadamente *Graph Machine Learning*. Essas técnicas irão utilizar o modelo semântico instanciado para treinar um modelo de *machine learning* que consiga identificar relações semânticas ocultas e prever valores ausentes. O objetivo é melhorar o conhecimento disponível e permitir uma tomada de decisão mais informada.

A Tabela 2, resume as contribuições de cada perspetiva científica para o modelo de representação do conhecimento e para o problema em geral.

|                      | Abordagem Semântica   | Abordagem IA  |
|----------------------|---|---|
| <b>Contribuições</b> | <ul style="list-style-type: none"><li>• Modelo Semântico (Ontologia)</li></ul>  | <ul style="list-style-type: none"><li>• Modelo de GML projetado para previsão e classificação</li></ul>   |
| <b>Objetivos</b>     | <ul style="list-style-type: none"><li>• Integrar os dados das diferentes fontes de informação de modo a garantir a rastreabilidade</li><li>• Identificar erros e inconsistências nos registos</li><li>• Classificar a madeira que chega à fábrica como sustentável ou não sustentável</li></ul> | <ul style="list-style-type: none"><li>• Preencher lacunas no conhecimento existente através da identificação de relações ocultas</li><li>• Efetuar a previsão de valores ou registos em falta</li></ul> |

Tabela 2: Contribuições de cada perspetiva científica

A Figura 5 ilustra como os artefactos das duas abordagens descritas funcionam em conjunto. O artefacto semântico instanciado (KG) é populado com dados de um DSS através de um *pipeline* e faz a respetiva inferência do conhecimento implícito. Após esse processo, é realizado o treino e teste do algoritmo de GML. Desta etapa resulta um modelo treinado para duas tarefas: identificação de novas relações semânticas e previsão de dados em falta.

É importante realçar que, neste trabalho, a tarefa de previsão de dados em falta (número 4 da Figura 5) foi testada apenas para avaliar a viabilidade da abordagem proposta. Durante o processo de carregar dados para ontologia, não foi realizado o preenchimento de registos ausentes, utilizando-se apenas os dados presentes no DSS.

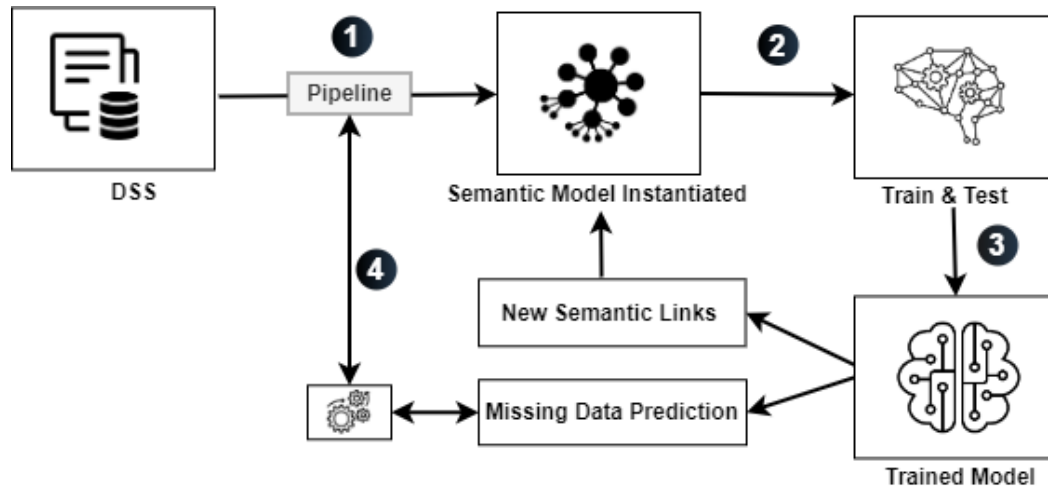


Figura 5: Interação entre a abordagem semântica e a abordagem de IA

## 3.2 Conceptualização do Domínio

A conceptualização do domínio é essencial para a criação de um modelo de representação de conhecimento, pois fornece uma compreensão detalhada e estruturada do contexto em que o modelo será aplicado. Compreender o domínio implica reconhecer as nuances e complexidades específicas do problema, incluindo as interações entre os diferentes componentes e as variáveis envolvidas. Isso permite não apenas identificar o problema com clareza, mas também desenvolver uma abordagem bem fundamentada para resolvê-lo. Ao mapear os conceitos-chave e a suas relações, é possível criar um modelo que reflete de forma precisa a realidade do domínio, facilitando a integração de dados, a análise de processos e a especificação de soluções eficazes.

### 3.2.1 Caracterização da Cadeia de Abastecimento Florestal

As florestas desempenham um papel essencial no ambiente, na economia e na sociedade [44]. Além de contribuírem significativamente para a proteção da biodiversidade, representam uma fonte valiosa de recursos naturais. Em Portugal, as florestas ocupam 3,2 milhões de hectares, o que corresponde a cerca de 36% do território nacional [52].

Embora as florestas portuguesas sejam caracterizadas por um elevado número de proprietários e uma pequena dimensão média das parcelas, constituem um setor importante para a economia do país. Anualmente, a indústria florestal contribui com cerca de 982 milhões de euros para a economia e é responsável por aproximadamente 80 000 empregos [52].

Apesar destes indicadores positivos, a indústria florestal em Portugal ainda enfrenta vários desafios. Normalmente, a exploração florestal é subcontratada a várias empresas, e por

isso, existe heterogeneidade de sistemas de monitorização. Isso leva a que, a transparência da rastreabilidade de madeira seja difícil de garantir ao longo de toda a cadeia de abastecimento.

Tipicamente, a cadeia de abastecimento florestal é composta pelas seguintes fases, ilustradas na Figura 6.

- **Gestão Florestal:** Fase que visa assegurar a saúde e sustentabilidade das florestas, garantindo a manutenção dos ecossistemas e a disponibilidade contínua de árvores para futuras colheitas.
- **Colheita:** Envolve as atividades de corte e extração. Destas atividades, normalmente, resulta biomassa que é usada para a produção de energia.
- **Armazenamento e Transporte:** Consiste nas atividades que levam a madeira e a biomassa para fábricas ou parques de armazenamento.
- **Processamento e Manufatura:** Refere-se ao processo de transformação da matéria-prima em produtos finais, como móveis, pasta e papel.
- **Venda e Distribuição:** Engloba a comercialização e entrega dos produtos ao consumidor final.

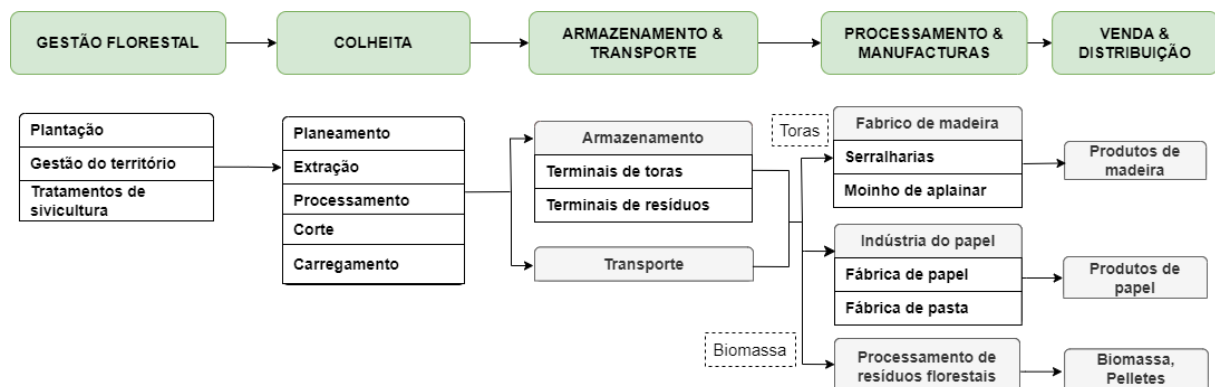


Figura 6: Fases da cadeia de abastecimento florestal

Das fases descritas, existem graus diferentes de transparência e eficiência na rastreabilidade. As fases de “Processamento e Manufatura” e “Venda e Distribuição” são aquelas em que a rastreabilidade alcança um grau de transparência maior. Isso deve-se ao facto de que a transformação da madeira em produtos finais ocorre em ambientes controlados, onde a monitorização está mais sistematizada e padronizada.

Por outro lado, as fases anteriores da cadeia, especialmente a colheita e o subsequente armazenamento e transporte, são as fases em que a rastreabilidade enfrenta maiores desafios. Num cenário simples, nestas duas fases, podem existir quatro fontes de informação fragmentadas:

- **Sistema de monitorização da máquina de corte:** Regista dados sobre a extração da madeira, como a localização e o tipo de árvore cortada.
- **Sistema de monitorização da máquina de recheça:** Regista informações sobre a operação de carga e a movimentação da madeira após o corte.

- **Sistema de monitorização do transporte (*Track and Trace*):** Acompanha a madeira durante o transporte para o local de armazenamento ou para a fábrica, fornecendo dados sobre a localização e o estado do transporte.
- **Guia de entrada na fábrica:** Regista a entrada da carga de madeira na fábrica. Normalmente, este registo é feito no sistema de informação da fábrica e armazena dados como o peso e a espécie da carga.

Concretamente, este trabalho concentra-se nas fases de colheita, armazenamento e transporte. Nestas fases, também designadas por exploração florestal, a madeira é considerada um produto e não uma matéria-prima que pode originar um produto final.

O processo típico de exploração florestal (Figura 7) abrange atividades que vão desde o planeamento até à entrega da madeira nos diversos tipos de indústria (fábricas de papel, móveis, etc.). No início do processo, é realizado o planeamento das áreas onde ocorrerão as operações florestais. Após o planeamento, as máquinas de corte e recheça são deslocadas para o terreno. A máquina de corte, conhecida como *Harvester*, é responsável por cortar as árvores e transformá-las em toras, podendo gerar biomassa como subproduto. Já a máquina de recheça, conhecida por *Forwarder*, transporta as toras da floresta para um local acessível aos camiões, onde forma uma pilha. A partir dessa pilha, a madeira é carregada e transportada para os centros de consumo ou, em alguns casos, para parques de armazenamento temporário, conforme as necessidades logísticas.

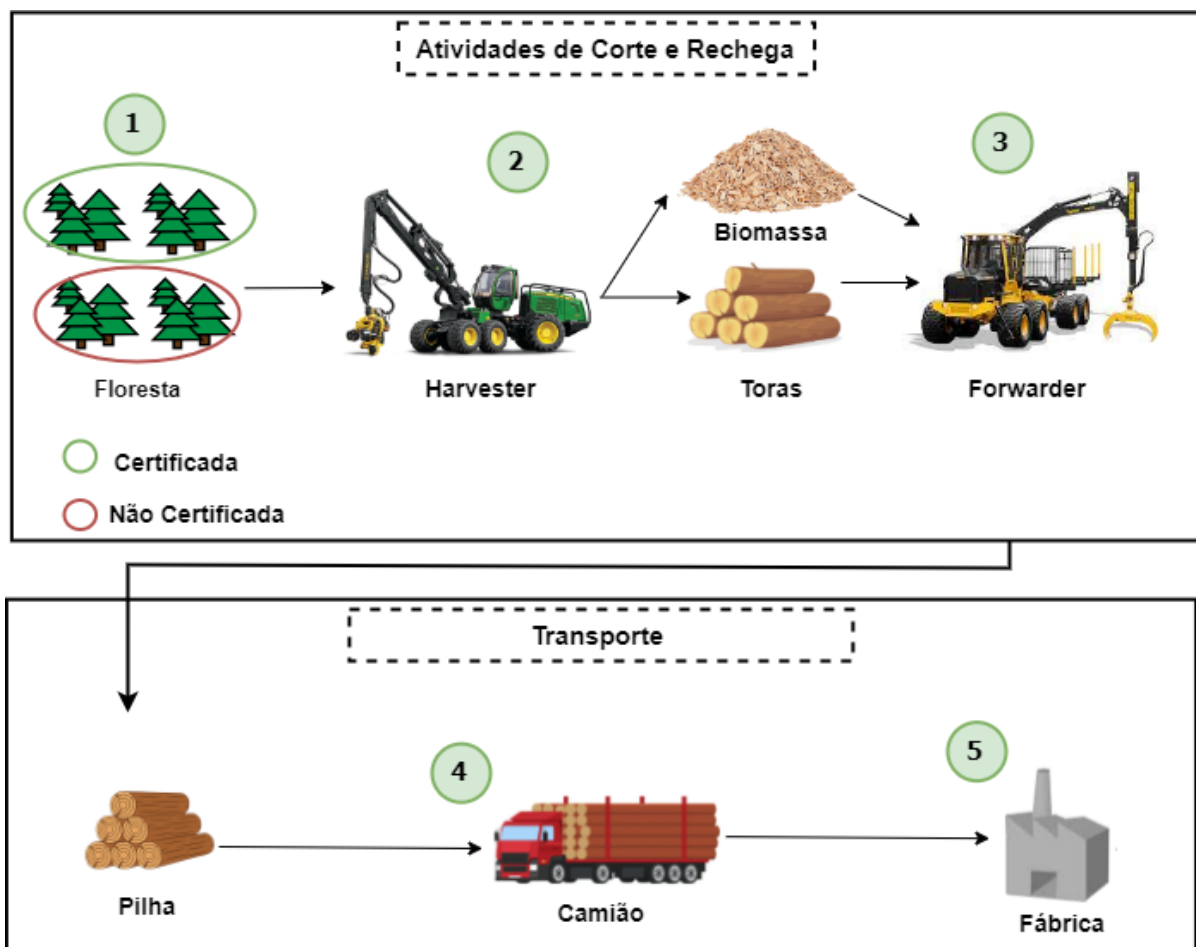


Figura 7: Atividades típicas na exploração de madeira

Existem dois tipos de madeira que são comercializadas: a madeira certificada e a não certificada. A madeira certificada provém de florestas geridas de forma sustentável e cumpre normas estabelecidas por entidades certificadoras, como o FSC <sup>1</sup> ou o PEFC <sup>2</sup> [53]. Esta certificação garante que as práticas de colheita respeitam critérios ambientais, sociais e económicos. Por outro lado, a madeira não certificada não oferece garantias sobre a origem ou os métodos utilizados na sua extração, o que pode levantar questões sobre a sua sustentabilidade e impacto ambiental [53].

De modo a garantir que esta distinção é feita de forma eficaz, é necessário um fluxo de informação bem delineado, que englobe todas as atividades envolvidas. Esse fluxo (Figura 8) integra as fontes de informação já mencionadas (sistemas de monitorização, *Track and Trace*, etc.) e inicia-se com a recolha de dados provenientes da máquina de corte, que fornece a localização específica de cada árvore cortada. Após a atividade de corte estar terminada, a madeira é transportada pelo *Forwarder* até ao camião. Durante esta operação, são recolhidos os dados GPS da máquina de modo que a movimentação da madeira seja rastreada. O fluxo de informação prossegue com o registo dos dados do camião, desde o ponto de carga até ao destino final. Essa fase inclui a monitorização da localização e do estado da carga durante o transporte. Por fim, é feita a recolha dos dados de entrada da madeira na fábrica assegurando assim que todo o percurso da madeira é devidamente registado.

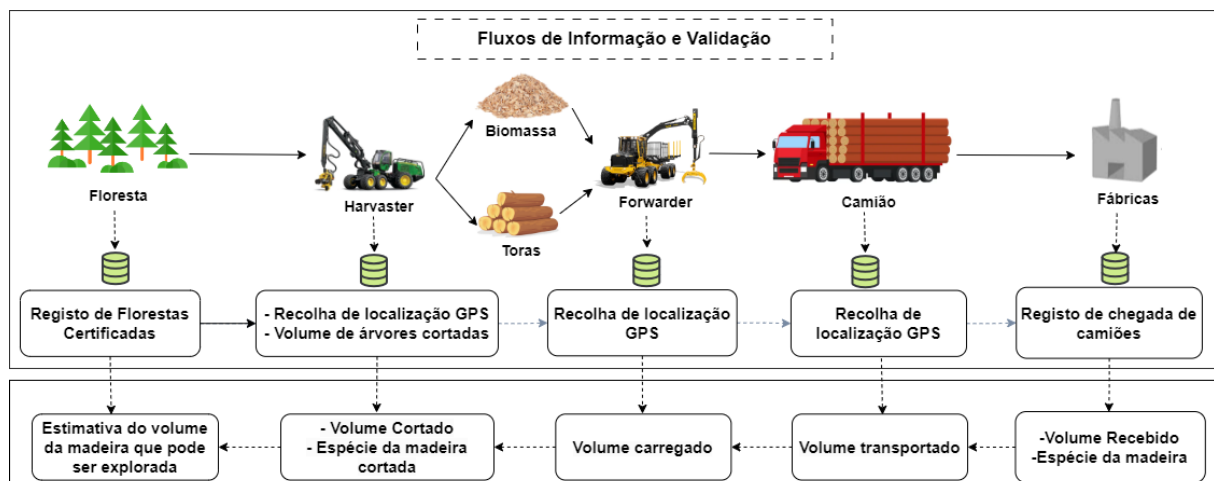


Figura 8: Fluxo de informação do processo de exploração florestal

Apesar da estrutura bem definida deste fluxo, é importante ter em conta que podem surgir lacunas. Essas lacunas podem ser falhas na transmissão de dados ou problemas técnicos nos sistemas de monitorização. Além disso, a precisão das localizações GPS pode ser afetada por vários fatores, como obstruções de sinal e condições climáticas.

Para mitigar esses problemas, é essencial a validação de dados. Essa validação deve operar numa ótica de *backward traceability*, permitindo verificar a integridade dos dados em todas as fases. Por outras palavras, deve possibilitar a rastreabilidade desde a fábrica até à árvore em pé na floresta, identificando inconsistências que possam afetar a precisão e a confiabilidade da origem da madeira.

<sup>1</sup><https://www.pt.fsc.org/pt-pt>

<sup>2</sup><https://www.pefc.pt/>

### 3.3 Representação Conceptual da Ontologia

A representação conceptual de uma ontologia é um aspeto importante na construção de modelos de representação de conhecimento porque organiza de forma explícita os conceitos, semanticamente estruturados através de relações. Para que essa representação atenda às necessidades do domínio, é essencial fazer a especificação de requisitos, que, na Engenharia Ontológica, são as questões de competência [54].

As questões de competência relacionadas à rastreabilidade da madeira que orientaram o desenvolvimento da ontologia estão apresentadas na Tabela 3.

| ID   | Questão  |
|------|--|
| CQ1  | Qual máquina está associada a uma determinada atividade?   |
| CQ2  | Quais foram as datas e localizações das atividades realizadas?   |
| CQ3  | Quais medições estão associadas a uma determinada atividade?   |
| CQ4  | Quais foram os principais pontos de falha na rastreabilidade?  |
| CQ5  | Quais inconsistências foram detetadas nos registos de rastreabilidade e em quais fases?                            |
| CQ6  | Quais cargas de madeira podem ser rastreadas até à sua origem e qual é o grau de confiança da rastreabilidade?     |
| CQ7  | Qual é a quantidade de recursos perdidos durante uma atividade de exploração de madeira?                           |
| CQ8  | Que fatores determinam a distinção entre madeira certificada e não certificada?                                    |
| CQ9  | Quais cargas de madeira podem ser classificadas como certificadas?   |
| CQ10 | Que ações podem ser implementadas para garantir a qualidade e sustentabilidade do processo de extração de madeira? |
| CQ11 | Qual o agente responsável por determinada atividade?   |
| CQ12 | Que tipo de <i>output</i> pode gerar uma atividade específica?   |
| CQ13 | Que eventos foram registados no âmbito de uma atividade?   |
| CQ14 | Qual a medição de um determinado <i>output</i> ao longo do seu ciclo de vida?                                      |
| CQ15 | Um <i>output</i> foi consumido/transformado dentro do prazo estipulado?  |
| CQ16 | Que tipo de <i>input</i> está associado a uma determinada atividade e em que pode ser transformado?                |

Tabela 3: Questões de competência

Essas questões foram elaboradas de forma a responder aos seguintes aspetos fundamentais:

- **Why:** Justifica as razões por trás da rastreabilidade da madeira, como a certificação e a transparência.
- **Where:** Refere-se à localização das atividades, identificando onde as operações ocorrem.
- **When:** Aborda a sequência de eventos e atividades relacionadas ao ciclo de vida da madeira.
- **Who:** Identifica as entidades envolvidas em cada atividade, como agentes e máquinas.
- **What:** Identifica o objeto a ser rastreado, neste caso, uma carga de madeira<sup>3</sup>.

<sup>3</sup>A rastreabilidade de madeira é realizada com uma abordagem em lote, o que significa que não se rastreia uma árvore individualmente, mas sim um lote (ou carga) de árvores.

Considerando os aspectos apresentados, a ontologia foi dividida em duas partes: uma que representa conhecimento explícito, formalizado através de classes, propriedades e relações bem definidas e outra que representa o conhecimento implícito, também estruturado em classes.

Para garantir a conformidade com normas e práticas reconhecidas, vários conceitos da ontologia foram mapeados do *Traceability Vocabulary*<sup>4</sup>. Este vocabulário, embora ainda não seja oficial, oferece uma normalização importante para os conceitos utilizados na rastreabilidade ao longo da cadeia de abastecimento. Ao alinhar a ontologia com este vocabulário, foi possível garantir que os conceitos utilizados são consistentes e compatíveis com as práticas e normas de rastreabilidade globalmente adotadas.

A Figura 9 ilustra a parte da ontologia conceptual que representa o conhecimento explícito.

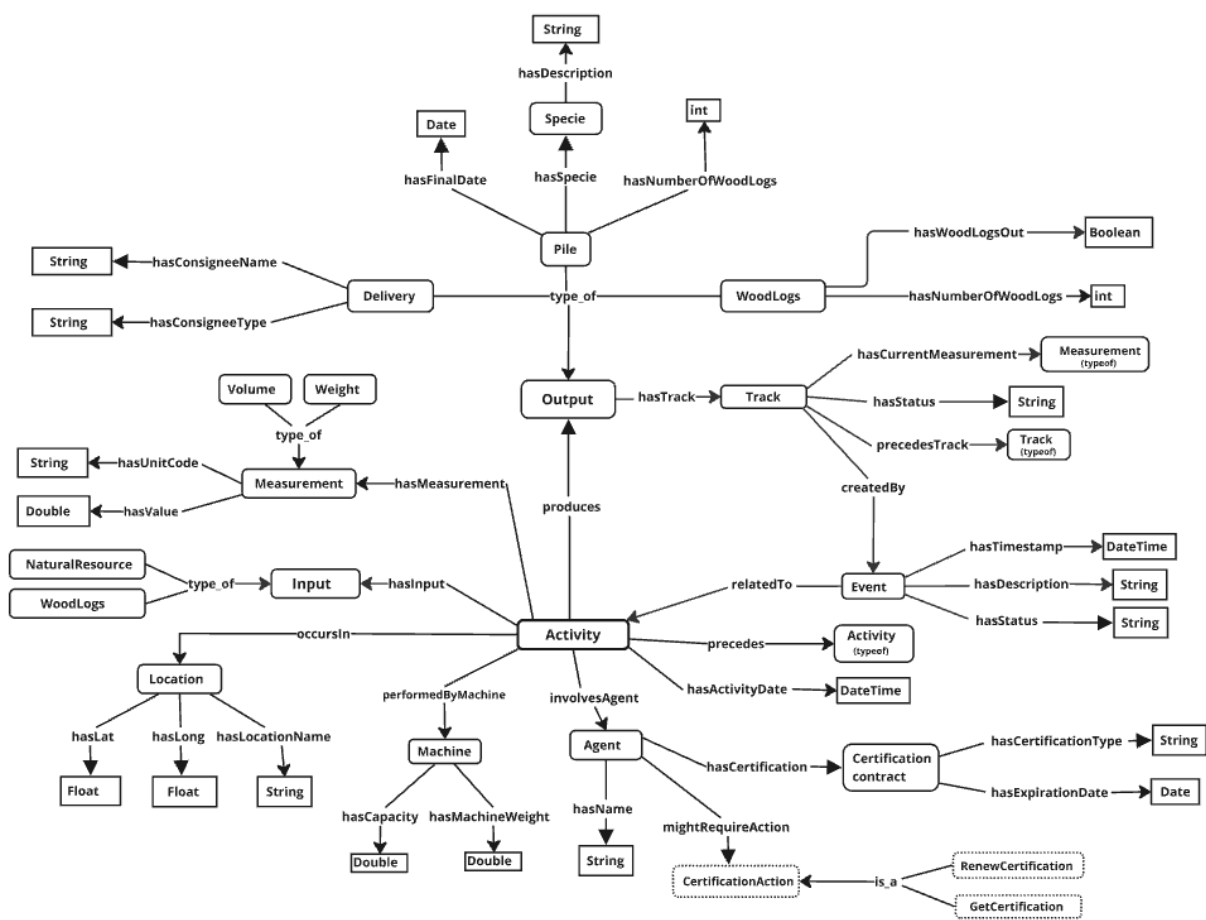


Figura 9: Representação conceptual do conhecimento explícito

Este segmento da ontologia é composto pelas seguintes classes:

- **Activity:** Refere-se às atividades que ocorrem na cadeia de abastecimento, como corte, recheia e transporte.
- **Agent:** Representa as entidades que participam nas atividades, normalmente designadas de prestadores de serviços.

<sup>4</sup><https://w3c-ccg.github.io/traceability-vocab/>

- **Machine:** Corresponde aos equipamentos utilizados nas diversas atividades, como *Forwarders*, *Harvesters* e caminhões.
- **Location:** Indica os locais onde as atividades ocorrem.
- **Event:** Representa momentos específicos relacionados a uma atividade, como, por exemplo, o início do transporte ou o fim de uma atividade de corte.
- **Output:** Refere-se aos resultados gerados pelas atividades, como toras, pilhas e cargas entregues na fábrica.
- **Track:** Representa a monitorização e a gestão das quantidades de um *output*, como por exemplo, a quantidade atual de uma pilha.
- **Measurement:** Descreve dados quantitativos das atividades, como o volume de madeira cortado ou o peso transportado.
- **Specie:** Indica a espécie de árvore que compõe a pilha, como Pinheiro, Eucalipto ou Carvalho.
- **Input:** Representa os recursos envolvidos em cada atividade. No corte, os recursos são naturais, como as árvores em pé. Na recheia, o *input* são as árvores cortadas. No transporte, o *input* é a pilha de madeira pronta para ser transportada.
- **CertificationContract:** Descreve os tipos de certificação/licença que um agente pode possuir.
- **CertificationAction:** Refere-se às ações de certificação necessárias para garantir práticas sustentáveis na exploração de madeira.

Já o segmento da ontologia que representa o conhecimento implícito (Figura 10) é composto por:

- **TraceabilityAims:** Define os objetivos e finalidades principais do sistema de rastreabilidade de madeira.
- **InconsistentData:** Refere-se às atividades que apresentam medições que não estão de acordo com as medições anteriores.
- **ConsistentData:** Refere-se às atividades que possuem medições inferidas como consistentes.
- **WoodClassification:** Representa a inferência de uma carga de madeira como certificada ou não certificada.
- **TraceabilityTrustLevel:** Indica o grau de confiança da rastreabilidade de uma carga (Alto, Médio ou Baixo).
- **OperationalInefficiency:** Representa atividades nas quais ocorreram ineficiências operacionais, como, por exemplo, árvores cortadas que ficaram na mata durante uma atividade de recheia.
- **InconsistencyTracking:** Representa atividades ou eventos onde existiram inconsistências ou falhas na rastreabilidade.

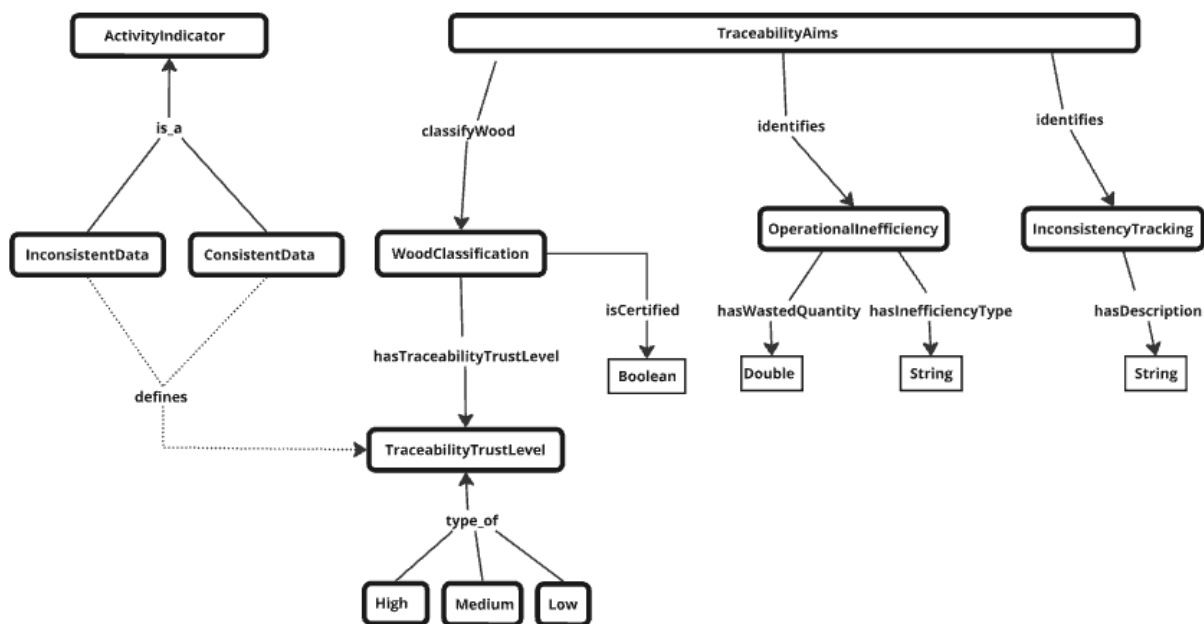


Figura 10: Representação conceitual do conhecimento implícito

As interações entre essas classes são estabelecidas pelas relações apresentadas na Tabela 5.

| Relação  | Descrição  |
|--|--|
| <i>Activity involves Agent</i>   | Atividade é realizada por um agente  |
| <i>Activity occursIn Location</i>  | Atividade ocorre numa determinada localização  |
| <i>Activity performedBy Machine</i>  | Atividade é realizada por uma máquina  |
| <i>Activity hasMeasurement Measurement</i>                                 | Atividade tem medições associadas  |
| <i>Activity precedes Activity</i>  | Indica que uma atividade pode ter uma atividade precedente   |
| <i>Activity produces Output</i>  | Atividade produz um resultado, como toras ou pilhas de madeira                                     |
| <i>Activity hasInput Input</i>   | Indica que uma atividade depende de um <i>input</i> , como árvores, toras ou pilhas                |
| <i>Output hasTrack Track</i>   | Indica que o resultado produzido por uma atividade é monitorizado através da classe <i>Track</i>   |
| <i>Agent hasCertification CertificationContract</i>                        | Agente pode possuir uma certificação/licença   |
| <i>Event relatedTo Activity</i>  | Um evento está relacionado a uma atividade   |
| <i>Track precedesTrack Track</i>   | Indica que a monitorização do estado de um <i>output</i> pode incluir uma monitorização precedente |
| <i>WoodClassification hasTraceabilityTrustLevel TraceabilityTrustLevel</i> | Indica que uma carga de madeira classificada tem um nível de confiança na rastreabilidade          |

Tabela 5: Relações semânticas da ontologia

# Capítulo 4

## Especificação dos Artefactos

A solução desenvolvida foi projetada para funcionar de forma integrada com um DSS. O objetivo é mapear e integrar os dados que estão presentes no DSS para o sistema de rastreabilidade, desenvolvido sobre a forma de Sistema de Organização de Conhecimento (KOS). Após estar carregado com os dados, o sistema de rastreabilidade irá extrair o conhecimento implícito e disponibilizá-lo no DSS. Desta forma, toda a informação é combinada num único sistema o que facilita o seu acesso e análise.

Neste capítulo são descritos os artefactos desenvolvidos. Na Secção 4.1 é apresentada a arquitetura do sistema de rastreabilidade e o contexto em que ela se insere. Na Secção 4.2 é descrito o processo de formalização da ontologia. Por fim, na Secção 4.3 é apresentada a abordagem para seleção e treino do algoritmo de GML.

## 4.1 Arquitetura da Solução

Como descrito na Secção 3.2.1 deste documento, os dados sobre as atividades da cadeia de abastecimento florestal encontram-se dispersos por vários sistemas de monitorização. No entanto, projetos como o RePlant<sup>1</sup> têm explorado formas de disponibilizar esses dados num único Sistema de Apoio à Decisão. Desses esforços resultou a arquitetura “Floresta 4.0”, ilustrada na Figura 11.a), que inclui os seguintes componentes:

- **Plataformas *Harvester e Forwarder***: Componentes responsáveis pela recolha de dados dos sensores das máquinas de corte e recheга em tempo real. Esses dados são disponibilizados em ficheiros XML, seguindo o formato StanForD 2010<sup>2</sup>.
- **Plataforma *Track and Trace***: Componente que recolhe os dados do transporte de cargas de madeira, disponibilizando-os em ficheiros XML no formato PapiNet<sup>3</sup>.
- **Plataforma *IoT***: Componente que funciona como um *gateway*, ou seja, permite que o acesso a todas as plataformas de recolha de dados seja feito através de um único serviço. É responsável por agregar e disponibilizar toda a informação ao Sistema de Apoio à Decisão.
- **Sistema de Apoio à Decisão**: Componente que recebe dados das diversas plataformas e transforma-os em informação útil para os diversos intervenientes na cadeia de abastecimento florestal.

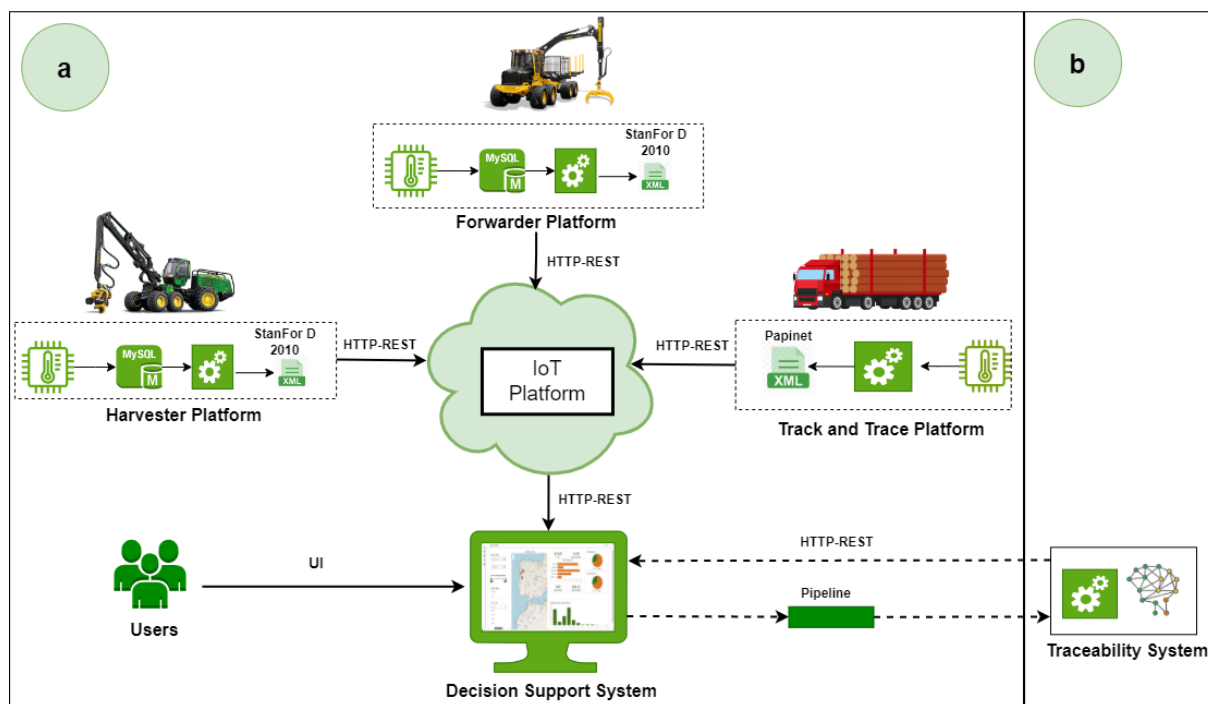


Figura 11: a) Arquitetura Floresta 4.0 b) Arquitetura Floresta 4.0 com o sistema de rastreabilidade

Com base nesta arquitetura, foi pensado desenvolver o sistema de rastreabilidade como um componente independente, representado na Figura 11.b). Este componente comu-

<sup>1</sup><https://replant.pt/en/>

<sup>2</sup><https://www.skogforsk.se/english/projects/stanford/stanford-2010/>

<sup>3</sup><https://www.papinet.org/>

nica de forma bidirecional com o Sistema de Apoio à Decisão, recolhendo os dados para a rastreabilidade e disponibilizando os resultados das inferências e da integração para análise.

A arquitetura do sistema de rastreabilidade é detalhada na Figura 12, sendo composta pelos seguintes componentes:

- **Pipeline de Extração e Integração dos Dados na Ontologia:** Componente responsável pelo preenchimento da ontologia. Utiliza conectores MySQL<sup>4</sup> e MongoDB<sup>5</sup> para estabelecer as conexões com as bases de dados do DSS e extrair os dados que serão utilizados para a rastreabilidade. Após a extração dos dados, a biblioteca Apache Jena<sup>6</sup> é utilizada para realizar a transformação e integração desses dados, que são então carregados para a ontologia. É também função do *pipeline* a criação das instâncias de configuração, quando a ontologia é populada pela primeira vez. Essas instâncias pertencem às classes *TraceabilityTrustLevel* e *CertificationAction*. Para a classe *TraceabilityTrustLevel*, são criadas três instâncias de configuração: *High*, *Medium* e *Low*. Para a classe *CertificationAction*, são criadas duas instâncias: *RenewCertification* e *GetCertification*.
- **Motor de Inferência:** Componente responsável por realizar inferências sobre os dados carregados na ontologia. Para isso, utiliza o *reasoner* Pellet<sup>7</sup>. Baseado em Java<sup>8</sup>, este motor de inferência pode ser usado em conjunto com bibliotecas Jena ou OWL API e oferece funcionalidades como a verificação da consistência de ontologias, classificação de taxonomias e verificação de implicações lógicas (como p. ex. regras e axiomas) [55]. Além disso, o Pellet pode operar de forma incremental, permitindo que novos dados sejam adicionados à ontologia sem a necessidade de reiniciar o processo de inferência. Essa capacidade é especialmente útil numa ontologia de rastreabilidade, pois permite que o sistema se adapte rapidamente a mudanças nos dados, gerando inferências atualizadas de forma eficiente e consumindo poucos recursos computacionais.
- **Pipeline para Disponibilização dos Resultados:** Este componente é responsável por disponibilizar os resultados das inferências e da integração de dados ao DSS. Utiliza o Apache Jena Fuseki<sup>9</sup>, um servidor SPARQL<sup>10</sup> que permite a consulta de dados RDF em diferentes formatos de resposta, como JSON e XML. Para obter as informações, o DSS efetua um *request* HTTP contendo a *query* SPARQL apropriada e recebe a resposta no formato JSON.

Apesar dos dados de todas as plataformas estarem no DSS, permanecem dispersos pelos vários módulos que este possui, de modo a atender às diversas necessidades de análise dos utilizadores. Essa dispersão reflete-se, por exemplo, na coexistência de registos em dois tipos de bases de dados (NoSQL e MySQL). Além disso, existem inconsistências ao nível da terminologia. Por exemplo, as pilhas são referidas como “*load\_target*” nos registos das atividades de chegada e “*pile\_reference*” nos relatórios de transporte. Essas inconsistências

<sup>4</sup><https://www.mysql.com/products/connector/>

<sup>5</sup><https://mvnrepository.com/artifact/org.mongodb/mongo-java-driver/3.12.14>

<sup>6</sup><https://jena.apache.org/documentation/ontology/>

<sup>7</sup><https://github.com/stardog-union/pellet>

<sup>8</sup><https://dev.java/>

<sup>9</sup><https://jena.apache.org/documentation/fuseki/>

<sup>10</sup>SPARQL- Linguagem de consulta para dados em formato RDF.

dificultam a interpretação dos dados e destacam a importância do mapeamento semântico, que garante a correta integração de todas as fontes de informação.

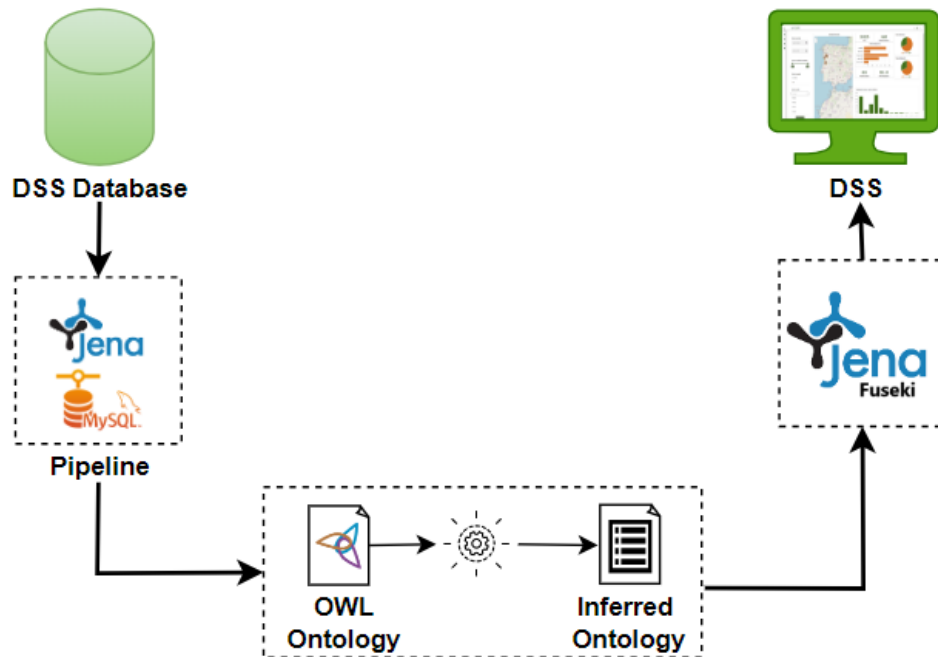


Figura 12: Arquitetura do sistema de rastreabilidade

## 4.2 Formalização da Ontologia

De modo a transformar a ontologia conceptual descrita na Secção 3.3 num artefacto formal que é utilizado na arquitetura do sistema de rastreabilidade, foi necessário realizar um processo de formalização. Esse processo consistiu em desenvolver a ontologia em OWL e foram realizadas as seguintes ações:

- Definição das classes, assegurando que estas refletiam a estrutura hierárquica desejada.
- Definição das associações entre classes através de *object properties*, identificando as *Domain Classes* e as *Range Classes* apropriadas.
- Definição das *data properties*, que representam os atributos das classes, especificando os seus *Range*, que incluem tipos literais, como *strings*, valores numéricos e datas.
- Identificação das restrições semânticas relevantes para o domínio e para os principais objetivos da ontologia.
- Definição das restrições lógicas, utilizando axiomas e regras. Estas restrições permitem que a ontologia classifique as cargas de madeira e identifique inconsistências nos dados.

Para a definição das classes, *data properties* e *object properties*, foi utilizada a linguagem LinkML<sup>11</sup>. O LinkML é uma linguagem de modelação, baseada em Python<sup>12</sup>, que permite

<sup>11</sup><https://linkml.io/linkml/generators/index.html>

<sup>12</sup><https://www.python.org/>

descrever estruturas de dados através de um ficheiro YAML. Este ficheiro YAML, pode depois, ser convertido em vários formatos, como JSON, CSV e RDF, através de um *LinkML Generator*<sup>13</sup>.

O Exemplo de Código 1 apresenta um excerto do ficheiro YAML no qual a estrutura da ontologia foi criada.

```

1  File:
2  class_uri: ex:Pile
3  is_a: Output
4  description: Represents a pile of logs
5  slots:
6    - hasNumberOfWoodLogs
7    - hasSpecie
8    - hasFinalDate
9    - hasTrack
10
11 slots:
12 hasNumberOfWoodLogs:
13   range: integer
14   minimum_value: 0
15 hasSpecie:
16   range: Specie
17   required: true
18 hasFinalDate:
19   range: date
20 hasTrack:
21   range: Track

```

Exemplo de Código 1: Excerto da ontologia criada em LinkML

Este ficheiro foi depois transformado em OWL através do *Generator*, como ilustrado na Figura 13. Durante este processo de transformação, foi igualmente gerado um ficheiro *Shapes Constraint Language (SHACL)*<sup>14</sup>, que contém as restrições, também chamadas de “*shapes*”, especificando os tipos de propriedades, valores permitidos, cardinalidade e relacionamentos entre entidades.

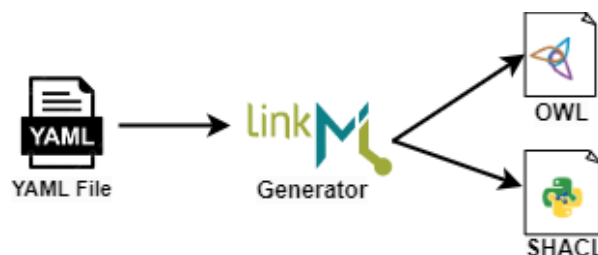


Figura 13: Fluxo de criação da ontologia

O Exemplo de Código 2 apresenta um excerto do ficheiro SHACL gerado.

```

1  @prefix ex: <http://example.org/ontology/hs_ontology/> .
2  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
3  @prefix sh: <http://www.w3.org/ns/shacl#> .

```

<sup>13</sup>Código que transforma um *schema* LinkML definido em formato YAML em outros modelos de dados, como OWL, JSON ou RDF.

<sup>14</sup>Linguagem recomendada pelo W3C para descrever e validar a estrutura de dados em grafos RDF.

```

4 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
5
6 ex:Pile a sh:NodeShape ;
7   sh:closed true ;
8   sh:description "Represents a pile of logs" ;
9   sh:ignoredProperties ( rdf:type ) ;
10  sh:property [ sh:class ex:Track ;
11              sh:nodeKind sh:BlankNodeOrIRI ;
12              sh:order 3 ;
13              sh:path ex:hasTrack ],
14            [ sh:datatype xsd:date ;
15              sh:maxCount 1 ;
16              sh:order 2 ;
17              sh:path ex:hasFinalDate ],
18            [ sh:datatype xsd:integer ;
19              sh:maxCount 1 ;
20              sh:minInclusive 0 ;
21              sh:order 0 ;
22              sh:path ex:hasNumberOfWoodLogs ],
23            [ sh:class ex:Specie ;
24              sh:maxCount 1 ;
25              sh:minCount 1 ;
26              sh:nodeKind sh:BlankNodeOrIRI ;
27              sh:order 1 ;
28              sh:path ex:hasSpecie ] ;
29  sh:targetClass ex:Pile .
30
31 ex:CertificationContract a sh:NodeShape ;
32   sh:closed true ;
33   sh:description "Represents an equipment" ;
34   sh:ignoredProperties ( rdf:type ) ;
35   sh:property [ sh:datatype xsd:date ;
36                 sh:maxCount 1 ;
37                 sh:order 1 ;
38                 sh:path ex:hasExpirationDate ],
39             [ sh:datatype xsd:string ;
40               sh:maxCount 1 ;
41               sh:order 0 ;
42               sh:path ex:hasCertificationType ] ;
43   sh:targetClass ex:CertificationContract .
44
45 ex:OperationalInefficiency a sh:NodeShape ;
46   sh:closed true ;
47   sh:ignoredProperties ( rdf:type ) ;
48   sh:property [ sh:datatype xsd:string ;
49                 sh:maxCount 1 ;
50                 sh:order 1 ;
51                 sh:path ex:hasInefficiencyType ],
52             [ sh:datatype xsd:integer ;
53               sh:maxCount 1 ;
54               sh:order 0 ;
55               sh:path ex:hasWastedQuantity ] ;
56   sh:targetClass ex:OperationalInefficiency .

```

Exemplo de Código 2: Excerto dos *Shapes* gerados para validação de dados RDF

Esse ficheiro foi útil principalmente na fase de desenvolvimento do *pipeline* que recolhe os dados do DSS e preenche a ontologia, uma vez que permitiu validar a conformidade

dos dados, assegurando que estes cumpriam as restrições definidas. Embora o *reasoner* Pellet pudesse ser utilizado para essa validação, o uso do ficheiro SHACL possibilitou essa verificação de forma direta, sem a necessidade de inicializar o *reasoner*, o que tornou o processo mais ágil e eficiente nesta etapa.

Após a criação das classes e propriedades, foram definidas as restrições lógicas. Nesta fase, foi utilizada Lógica Descritiva (DL) para os axiomas e *Semantic Web Rule Language (SWRL)* para as regras. Isso deve-se ao facto de que o LinkML permite a modelação de classes, propriedades e alguns axiomas simples, mas não possibilita a modelação de restrições lógicas mais complexas. Na Tabela 6 são apresentadas algumas<sup>15</sup> das regras definidas para cenários de inferência.

| Objetivo   | Regra SWRL  |
|--|---|
| Identificar Prestadores de Serviço com licenças expiradas  | <pre> Agent(?e) ∧ hasCertification(?e, ?c) ∧ hasExpirationDate(?c, ?expiryDate) ∧ MinimumDate(?md) ∧ hasDate(?md, ?minimumDate) ∧ swrlb:greaterThan(?expiryDate, ?minimumDate) ∧ CurrentDate(?cd) ∧ hasDate(?cd, ?currentDate) ∧ swrlb:lessThan(?expiryDate, ?currentDate) ⇒ mightRequireAction(?e, renew_certification) </pre>                                 |
| Identificar ineficiências operacionais na atividade de recheia, nomeadamente árvores que foram cortadas e permaneceram na floresta | <pre> Activity(?e1) ∧ produces(?e1, ?o1) ∧ Pile(?o1) ∧ hasNumberOfWoodLogs(?o1, ?l1) ∧ precedes(?e1, ?e2) ∧ produces(?e2, ?o2) ∧ WoodLogs(?o2) ∧ hasNumberOfWoodLogs(?o2, ?l2) ∧ swrlb:lessThan(?l1, ?l2) ∧ swrlb:subtract(?diff, ?l2, ?l1) ⇒ OperationalInefficiency(?e1) ∧ hasInefficiencyType(?e1, 'RESOURCEWASTAGE') ∧ hasWastedQuantity(?e1, ?diff) </pre> |
| Identificar inconsistências no relatório de atividades   | <pre> WoodClassification(?o1) ∧ produces(?e1, ?o1) ∧ Delivery(?o1) ∧ precedes(?e1, ?e2) ∧ precedes(?e2, ?e3) ∧ InconsistentData(?e3) ∧ InconsistentData(?e2) ⇒ hasTraceabilityTrustLevel(?o1, low) </pre>   |
| Classificar o grau de confiança com base na consistência das medições  | <pre> WoodClassification(?o1) ∧ produces(?e1, ?o1) ∧ Delivery(?o1) ∧ precedes(?e1, ?e2) ∧ precedes(?e2, ?e3) ∧ InconsistentData(?e3) ∧ InconsistentData(?e2) ⇒ hasTraceabilityTrustLevel(?o1, low) </pre>   |

Tabela 6: Excerto das regras SWRL da ontologia

<sup>15</sup>Tabela completa no Anexo A.3

## 4.3 Especificação do Algoritmo de *Graph Machine Learning*

Após a formalização da ontologia e adição de regras, foi aplicado um algoritmo de GML de modo a obter um modelo capaz de identificar novas relações e efetuar a previsão de dados em falta. Este processo envolveu as seguintes ações:

- Seleção do algoritmo.
- Preparação e pré processamento dos dados.
- Treino do algoritmo.
- Avaliação e validação dos resultados.

Para a seleção do algoritmo, limitou-se o âmbito do problema. O objetivo concentrou-se no apoio à identificação de dados incompletos e à descoberta de relações semânticas implícitas. Considerando que se está perante um modelo de informação associativo, a identificação de valores em falta poderá ser realizada através da exploração de relações existentes. Simultaneamente, essa exploração pode permitir a descoberta de novas relações que ampliem o conhecimento existente. Assim, com base nessa análise e considerando fatores como a estrutura do grafo, o volume de dados e a facilidade de entender e explicar os resultados, foram pré-selecionados 2 algoritmos: *MetaPath2Vec* e *Heterogeneous Graph Neural Networks* (HGNNs).

O *MetaPath2Vec* é um algoritmo projetado para gerar representações (*embeddings*) de nós em grafos heterogêneos. Baseia-se na ideia de que diferentes *metapaths*<sup>16</sup> entre nós podem ser explorados de modo a obter novas relações. O seu funcionamento começa com a definição dos *metapaths*. A partir desses *metapaths*, o algoritmo realiza *random walks*<sup>17</sup> no grafo, gerando sequências de nós que seguem esses caminhos. Essas sequências são então utilizadas para treinar um modelo de linguagem<sup>18</sup>, Skip-gram<sup>19</sup>, que produz representações vetoriais para os nós, obtendo a proximidade semântica entre eles [56].

Da mesma forma que o *MetaPath2Vec*, as HGNNs são projetadas para lidar com grafos heterogêneos. O principal objetivo é agregar informações dos vizinhos com base nas várias relações, o que enriquece a representação semântica dos nós. Este processo, conhecido como passagem de mensagens, consiste em obter informações dos vizinhos (atributos, categoria, etc.) por meio de diferentes relações, expressas como *metapaths*. Cada *metapath* define um caminho específico que pode ser percorrido entre diferentes tipos de nós, ponderando a importância de cada vizinho com base no tipo de relação. Desta forma, a rede neuronal aprende representações mais informativas e complexas [57].

A principal diferença entre os 2 algoritmos reside na complexidade dos modelos. O *MetaPath2Vec* adota uma perspectiva mais baseada em linguística, utilizando *random walks* e modelos de linguagem para aprender representações de nós. Por outro lado, as HGNNs baseiam-se em princípios de redes neuronais, utilizando a propagação de mensagens para

---

<sup>16</sup>Sequências de tipos de nós e arestas que refletem interações dentro do grafo.

<sup>17</sup>Movimentos aleatórios entre os nós num grafo.

<sup>18</sup>Algoritmos que analisam e geram texto em linguagem natural, aprendendo padrões e contextos a partir de grandes conjuntos de dados textuais.

<sup>19</sup><https://medium.com/@corymaklin/word2vec-skip-gram-904775613b4c>

atualizar as representações dos nós de forma iterativa, sendo mais flexíveis e capazes de capturar padrões complexos e não lineares.

A Tabela 7 apresenta algumas vantagens e desvantagens desses algoritmos. Com base nessa análise, optou-se por selecionar o algoritmo *MetaPath2Vec*, principalmente devido à sua eficácia em conjuntos de dados menores. Além disso, o foco deste algoritmo nos *metapaths* torna a interpretação dos resultados mais simples, uma vez que cada *metapath* define caminhos explícitos entre os nós. Em contrapartida, nas HGNNs, a consideração de múltiplos tipos de relações e vizinhos resulta em representações mais complexas, o que dificulta a compreensão das interações entre os nós.

| <i>MetaPath2Vec</i>   | HGNNs  |
|---|--|
| Fácil de interpretar e implementar  | Difícil de interpretar devido à complexidade da rede neuronal e à consideração de múltiplos tipos de relações e vizinhos |
| Tende a ser eficaz em conjuntos de dados relativamente pequenos   | Precisa de um grande conjunto de dados para treinar um modelo eficaz   |
| Limitado a obter padrões locais   | Capaz de obter interações complexas e padrões não lineares   |
| A flexibilidade na definição de <i>metapaths</i> permite ajustar a sua modelação para contextos específicos | A complexidade na estrutura da rede pode dificultar a adaptação a novos contextos sem <i>retraining</i> profundo         |

Tabela 7: Vantagens e desvantagens do *MetaPath2Vec* e HGNNs

### 4.3.1 *Feature Engineering* e Definição de *Metapaths*

Após a seleção do algoritmo, foi realizada a transformação do grafo de conhecimento para que pudesse ser treinado e interpretado. Nesse processo, designado por *feature engineering*, foi usada a biblioteca NetworkX<sup>20</sup> para a construção e manipulação do grafo e a biblioteca PyTorch Geometric<sup>21</sup> para a conversão da estrutura do grafo em tensores. Os tensores são uma estrutura de dados multidimensional que oferece uma forma eficiente e flexível de representar grafos em algoritmos GML.

Com o processo de *feature engineering* concluído, foram definidos os *metapaths*. Neste contexto, os *metapaths* têm um papel importante na configuração do algoritmo, pois determinam os caminhos que este irá explorar de modo a aprender as representações dos nós. No Exemplo de Código 3, são apresentados os *metapaths* criados, que são baseados nas relações semânticas do grafo de conhecimento.

```

1 metapaths = [
2     ("activity", "involves", "agent"),
3     ("agent", "has_certification", "certificationContract"),
4     ("certificationContract", "isAssignedTo", "agent"),
5     ("agent", "involved_in", "activity"),
6     ("activity", "performedBy", "machine"),
7     ("machine", "performs", "activity"),
8     ("activity", "produces", "output"),
9     ("output", "delivery_classified_as", "woodClassifier"),

```

<sup>20</sup><https://networkx.org/>

<sup>21</sup><https://pytorch-geometric.readthedocs.io/en/latest/>

```

10 | (“woodClassifier”, “classify_delivery_as”, “output”),
11 | (“output”, “produced_by”, “activity”),
12 | (“activity”, “data_classified_as”, “eventClassifier”),
13 | (“eventClassifier”, “classify_data_as”, “activity”),
14 | ]

```

Exemplo de Código 3: *Metapaths* do algoritmo

De notar que, para um *random walk* completo ser realizado, é necessário a definição de relações inversas. A Figura 14 ilustra essa necessidade: um nó do tipo *Activity* é conectado pela relação “*involves*” a um nó do tipo *Agent*; assim que o *metapath* chega ao nó *Agent*, precisa voltar ao nó *Activity* para se movimentar de novo, uma vez que o nó *Agent* não possui mais nenhuma relação.

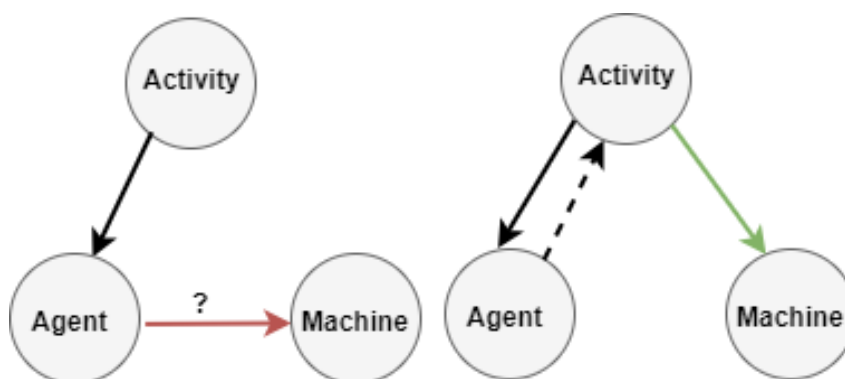


Figura 14: Exemplo da necessidade de relações inversas

### 4.3.2 Treino

Após a definição dos *metapaths*, foi realizado o treino do algoritmo. Esse treino teve como objetivo criar um modelo que aprenda e calcule a proximidade semântica entre nós. Em termos de configuração, foi definido que o algoritmo iria executar 10 iterações (*epochs*). Para cada iteração, o modelo foi configurado para ajustar as suas representações com base na perda calculada entre as previsões realizadas e os valores reais. Os dados de treino e teste foram divididos na porção 80/20 (80% dos dados para treino e 20% para teste).

Para determinar o número adequado de iterações, inicialmente, o modelo foi testado com 100 iterações. Observou-se que, a partir da décima, as melhorias eram mínimas, o que justificou a escolha final de 10 iterações como valor ideal. Em termos de métricas, o algoritmo obteve uma *Accuracy* máxima de 62%.

# Capítulo 5

## Caso de Estudo

Em vários projetos desenvolvidos recentemente, têm sido realizados esforços para aumentar a eficiência e a digitalização do setor florestal em Portugal. A introdução de sensores nas máquinas florestais e sistemas de monitorização em tempo real permitiu a recolha de dados diretamente no terreno, oferecendo uma visão em tempo real das operações. Estes avanços demonstraram, de forma clara, que a digitalização é o futuro a seguir para o setor. Como resultado, novas soluções tecnológicas, como plataformas *Track and Trace*, foram adotadas para digitalizar todo o processo de exploração florestal.

Com a digitalização das operações, surgem também novos desafios. Um deles, é a forma como esta pode ser utilizada para garantir a rastreabilidade de madeira. Com base neste contexto, neste capítulo, será descrito o processo de teste e validação dos artefactos desenvolvidos. O objetivo é validar se a solução proposta responde efetivamente ao desafio num caso de estudo real.

## 5.1 Caracterização do Caso de Estudo

Como descrito na Secção 3.2.1, a rastreabilidade da madeira só é possível se houver um fluxo de informação bem definido desde a floresta até à fábrica. Nesse sentido, o projeto Agenda Transform – Floresta 4.0<sup>1</sup>, que conta com a participação do INESC TEC<sup>2</sup>, tem como objetivo promover a investigação e o desenvolvimento de soluções que garantam a transparência e a eficiência desse fluxo de informação. Perante este objetivo, foram definidas duas etapas.

Na primeira etapa, foram definidas e implementadas as tecnologias necessárias para viabilizar o fluxo de informação. Essa fase, desenvolvida em grande parte no âmbito do projeto RePlant, resultou na criação da arquitetura “Floresta 4.0”, descrita na Secção 4.1. À data do início deste trabalho, essa arquitetura já se encontrava numa fase avançada.

Como resultado dessa arquitetura, foi desenvolvido um DSS que disponibiliza informação aos diferentes intervenientes da cadeia de abastecimento florestal (Figura 15). Essa informação, recolhida em tempo real, é especialmente útil para o dono da floresta e para os prestadores de serviços, como empresas de transporte e extração de madeira. Para o dono da floresta, a informação disponibilizada facilita a gestão eficiente dos recursos e assegura que a extração está em conformidade com as práticas de sustentabilidade. Para os prestadores de serviços, essa informação permite otimizar as operações logísticas, resultando na redução de custos e maior eficiência operacional.

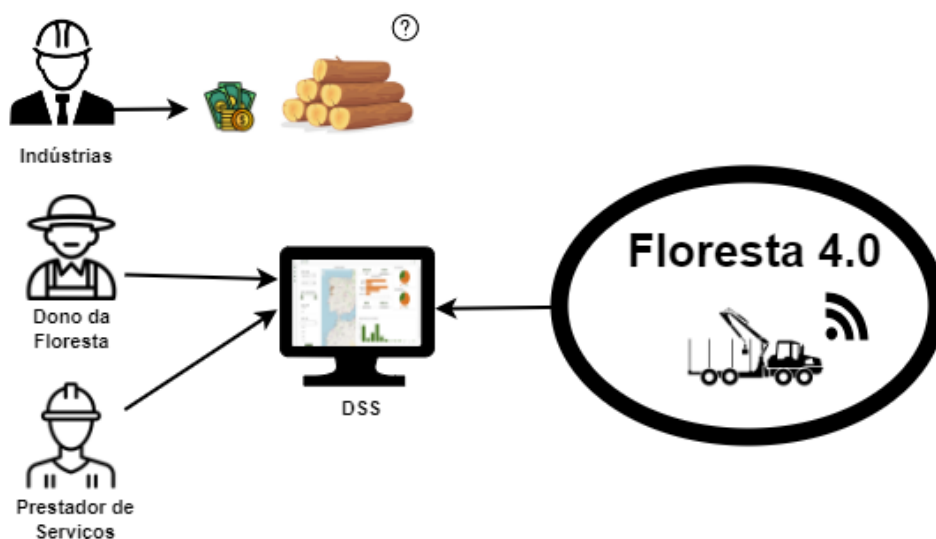


Figura 15: Resultados da arquitetura Floresta 4.0

Apesar dos resultados positivos da primeira etapa, existia um interveniente na cadeia de abastecimento florestal que não era diretamente beneficiado pela informação gerada: as indústrias florestais. Estas entidades, que compram a madeira como matéria-prima para produzir outros produtos, têm necessidades específicas. O que lhes interessa é o histórico de movimentos da madeira, ou seja, a sua rastreabilidade. Através dessa informação, as indústrias conseguem determinar se a carga de madeira que entra na fábrica deve ser classificada como certificada ou não certificada. Essa distinção influencia diretamente o

<sup>1</sup><https://transform.forestwise.pt/floresta-4-0/>

<sup>2</sup><https://www.inesctec.pt/en>

preço a ser pago pela matéria-prima e, além disso, pode determinar o uso final da madeira, já que a madeira não certificada não pode ser utilizada como matéria-prima para produtos que saem para o mercado com “rótulo” de sustentáveis.

Com base nesse contexto, surgiu a segunda etapa, que consistiu na integração dos dados de todas as fases, de modo a obter um fluxo de informação bem delineado. Para isso, foi utilizado o sistema de rastreabilidade proposto neste trabalho. Este sistema recolheu os dados armazenados na base de dados do DSS, integrou-os fazendo o respetivo mapeamento semântico e procedeu à inferência do conhecimento implícito. Os *items* seguintes fornecem uma visão geral do tipo e volume de dados recolhidos no âmbito do caso de estudo.

- **Atividades Registadas:** 707 atividades registadas, incluindo corte, recheia e transporte.
- **Cargas de Madeira Entregues na Fábrica:** 590 entregas de madeira, distribuídas por três fábricas de produção de papel (todas pertencentes à mesma entidade).
- **Floresta (Matas):** A madeira foi extraída de 3 matas diferentes.
- **Espécies:** As cargas de madeira rastreadas possuem 3 espécies.
- **Agentes:** 8 agentes envolvidos, dos quais 3 não possuem certificação.
- **Máquinas:** 17 máquinas envolvidas nas atividades, entre máquinas de corte, recheia e transporte.
- **Eventos Registados:** 1144 eventos distribuídos pelas 707 atividades.
- **Medições:** 3316 medições registadas.
- **Inconsistências e Ruído:** 15 registos inconsistentes e 4 registos duplicados.
- **Espaço Temporal:** Registos obtidos durante o período de 5 meses.

Perante estas características, é possível analisar que o conjunto de dados não tem um volume muito significativo. No entanto, possui uma variabilidade bastante boa. Essa variabilidade facilita a observação do comportamento do sistema em diferentes situações, como por exemplo, regiões com características distintas. Esse fator é importante dado que o nível de inclinação de um terreno pode influenciar a precisão das medições que são usadas para definir o nível de confiança da rastreabilidade.

É importante referir que, neste projeto, a diferença entre dados e informação reside fundamentalmente na forma como o DSS apresenta os dados ao utilizador. Por exemplo, as coordenadas (latitude e longitude) das atividades de corte são transformadas em mapas de volume (*heatmaps*) para facilitar a análise. Assim, o DSS na prática, funciona como um repositório de dados que possui ferramentas para os apresentar como informação útil para a tomada de decisão.

## 5.2 Resultados do Sistema de Rastreabilidade

De modo a avaliar a resposta do sistema de rastreabilidade, foram tidos em conta 2 fatores: a utilidade do sistema e a avaliação das classificações da ontologia. Na avaliação das classificações, os resultados das inferências foram comparados com os dados de referência fornecidos pelos especialistas. Essa comparação permitiu avaliar a capacidade da solução em classificar corretamente os dados recolhidos, incluindo as cargas de madeira certificadas ou não certificadas), a sequência de atividades (correta ou incorreta) e a consistência das medições (consistentes ou inconsistentes).

As medidas selecionadas para avaliar as classificações foram a *precision* e o *recall* (Figura 16 [58]). Estas duas medidas são bastante utilizadas no contexto das ontologias como forma de medição da qualidade dos resultados obtidos e identificação de áreas de melhoria [59].

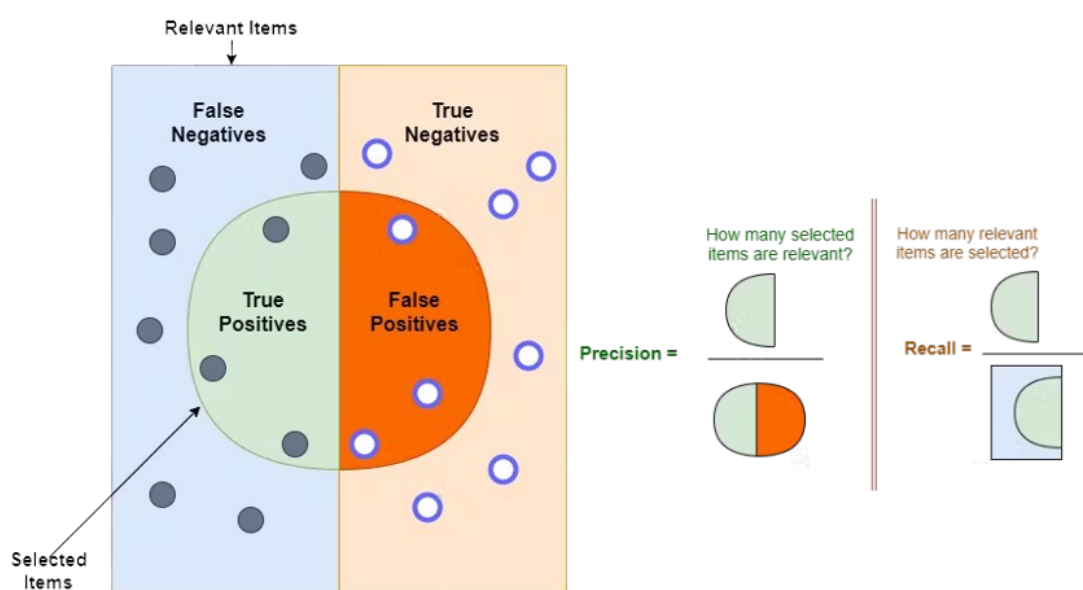


Figura 16: *Precision e Recall*, Adaptado de Encord (2024)

A *precision* é uma medida que indica a proporção de resultados inferidos que são relevantes entre o total de resultados classificados como positivos. É calculada através do rácio entre os verdadeiros positivos (VP) e a soma dos falsos positivos (FP) com os verdadeiros positivos (VP + FP) [60]:

$$Precision = \frac{VP}{VP + FP} \quad (\text{EQ. 1})$$

Já o *recall* é uma medida que indica a proporção de resultados relevantes que foram corretamente identificados. É calculado através do rácio entre os verdadeiros positivos (VP) e a soma dos verdadeiros positivos (VP) com os falsos negativos (FN) [60]:

$$Recall = \frac{VP}{VP + FN} \quad (\text{EQ. 2})$$

Para efetuar o cálculo destas medidas de modo a avaliar a precisão e o acerto com que a ontologia inferiu os resultados, foram tidos em conta os seguintes conceitos:

- **Verdadeiros Positivos (VP):** Resultados obtidos que coincidem com os resultados de referência.
- **Falsos Positivos (FP):** Resultados obtidos que não estão presentes nos resultados de referência. Estes casos representam, por exemplo, cargas de madeira que foram classificadas incorretamente como certificadas, quando não deviam ter sido.
- **Falsos Negativos (FN):** Resultados que, apesar de estarem nos resultados de referência, não foram corretamente classificados. Representam, por exemplo, inconsistências ou ineficiências operacionais nas atividades que não foram devidamente identificadas pela solução.

Na Tabela 8 é apresentada a comparação entre os resultados obtidos e os resultados de referência nas diferentes categorias de classificação.

| Inferência  | VP   | FP | FN |
|---|------|----|----|
| Cargas de madeira certificadas  | 67   | 21 | 0  |
| Cargas de madeira não certificadas  | 470  | 0  | 32 |
| Atividades com medições inconsistentes  | 23   | 0  | 0  |
| Atividades com medições consistentes  | 671  | 0  | 0  |
| Identificação das ações que os agentes devem cumprir de modo a garantir atividades sustentáveis | 3    | 0  | 0  |
| Ineficiências operacionais  | 2    | 0  | 0  |
| Classificação do grau de confiança na rastreabilidade   | 537  | 21 | 0  |
| Classificação de registos inconsistentes  | 11   | 0  | 27 |
| Classificação de atividades que não garantem práticas sustentáveis                              | 6    | 0  | 0  |
| <b>Total</b>  | 1790 | 42 | 59 |

Tabela 8: Resultados das classificações da ontologia

Com base nos valores obtidos e nas fórmulas já apresentadas, têm-se os seguintes resultados:

$$Precision = \frac{1790}{1790 + 42} = \frac{1790}{1832} \approx 0.976 \text{ ou } 97.6\%$$

$$Recall = \frac{1790}{1790 + 59} = \frac{1790}{1849} \approx 0.968 \text{ ou } 96.8\%$$

Os resultados obtidos demonstraram a eficácia do sistema de rastreabilidade desenvolvido no caso de estudo em que foi aplicado. A *precision*, situada em 97.6%, indica que a maioria das classificações realizadas está correta. Já o *recall*, que se apresenta em torno de 96.8%, destaca a capacidade de identificar todas as instâncias que deveriam ser classificadas.

Apesar dos valores obtidos, é importante destacar a presença de falsos negativos e falsos positivos. No total, foram identificados 59 falsos negativos, dos quais 32 estão relacionados

a cargas de madeira que não foram classificadas. Este valor representa aproximadamente 7% das cargas de madeira não certificadas. Já em termos de falsos positivos, foram contabilizados 42 casos. Desses, 21 estão relacionados com cargas de madeira certificada, o que representa cerca de 30% das classificações certificadas.

Esses resultados indicam que, embora o sistema apresente uma precisão elevada, existem áreas de melhoria, especialmente na redução dos falsos negativos e positivos. Avaliando os dados de *input*, ou seja, o conhecimento explícito, é possível identificar que grande parte dos resultados falsos pertencem a instâncias cujos dados não estão completos (por exemplo, atividades que deveriam ter uma atividade precedente e não têm). Isso pode indicar que o sistema apresenta algumas dificuldades na classificação de instâncias cujo ciclo de informação está incompleto.

Além das medidas de desempenho, como *precision* e *recall*, o sistema de rastreabilidade foi avaliado do ponto de vista da sua utilidade. A utilidade refere-se à capacidade do sistema de atender às necessidades dos especialistas do domínio. Nesse sentido, a avaliação foi feita tendo em consideração a contribuição do sistema para os objetivos do caso de estudo.

Neste contexto, o sistema permitiu uma visão sobre a rastreabilidade da madeira em duas perspectivas: *backward* e *forward*. A rastreabilidade *forward* permitiu a visualização de todo o percurso da madeira desde a floresta até à fábrica. Já rastreabilidade *backward* permitiu validar os dados históricos e identificar inconsistências que afetam a transparência da origem da madeira.

A Figuras 17 e 18, retiradas do DSS, apresentam 2 exemplos dessas inconsistências que foram identificadas pelo sistema de rastreabilidade. Essas inconsistências não estavam presentes na primeira versão dos resultados de referência e só foram detetadas após a integração completa dos dados com as sequências de atividades adequadas.

A Figura 17 representa uma carga que chegou à fábrica cuja origem da madeira é desconhecida.

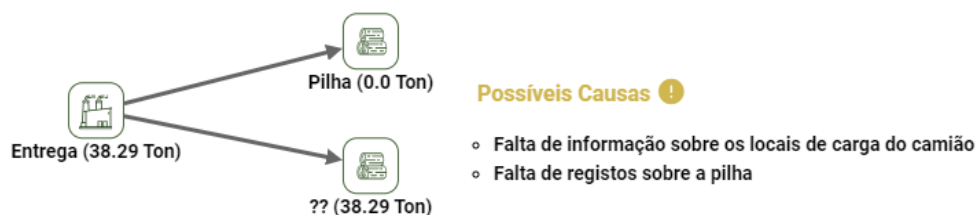


Figura 17: Origem da carga desconhecida

Já na Figura 18 é ilustrado um transporte que, durante o percurso entre o local de carga e a fábrica, carregou madeira que não foi devidamente rastreada.

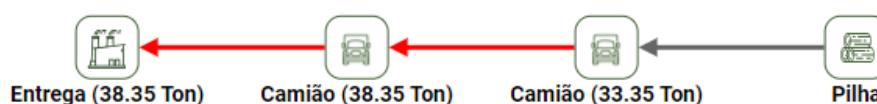


Figura 18: Carga de madeira adulterada

Em suma, o sistema de rastreabilidade desenvolvido demonstrou ser uma ferramenta útil para a rastreabilidade de madeira. Embora os resultados obtidos sejam bastante bons,

com uma alta precisão e *recall*, é importante ressaltar que a eficácia do sistema foi avaliada com um conjunto limitado de dados. Além disso, existem fatores não considerados que podem influenciar o desempenho do sistema. Um exemplo desses fatores é as condições meteorológicas, que podem afetar o peso da madeira e, conseqüentemente, as medições realizadas durante as atividades. É importante mencionar que este fator não foi considerado porque os sensores embutidos nas máquinas atualmente não recolhem dados sobre a humidade da madeira.

### 5.3 Resultados da Aplicação do Algoritmo GML

Após o teste e validação do sistema de rastreabilidade, foi aplicado sobre este o algoritmo de *Graph Machine Learning* com o objetivo de prever novas relações semânticas entre conceitos e preencher dados em falta. Como resultado, obteve-se um conjunto de *embeddings* que determinam a proximidade semântica entre os nós. Esses *embeddings* foram transformados pelo algoritmo t-SNE<sup>3</sup> em 2D (Figura 19), de modo a facilitar a sua análise.

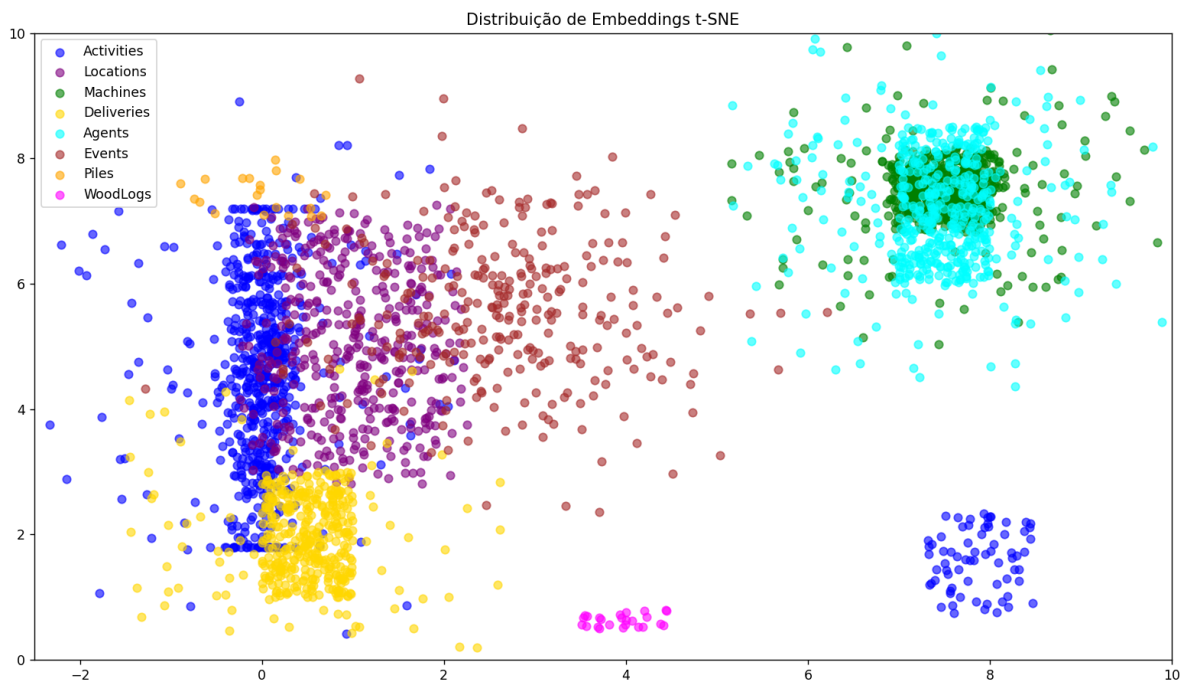


Figura 19: Resultado do algoritmo *MetaPath2Vec*

Observando o gráfico e tendo como base o facto de que os nós com proximidade semântica têm grande probabilidade de ter uma relação subjacente [61], podem-se retirar as seguintes ilações:

- **Atividades de Corte:** As atividades de corte, localizadas no canto inferior direito do gráfico, estão claramente distantes em termos de similaridade em relação ao restante das atividades. Essas atividades têm como característica diferenciadora o facto de não possuírem uma atividade precedente, no entanto, deveriam estar mais próximas, em termos de similaridade, do restante das atividades.

<sup>3</sup>O T-SNE é uma técnica estatística de redução de dimensionalidade utilizada para visualizar dados de alta dimensão num espaço de menor dimensão, geralmente 2D ou 3D.

- **Toras de Madeira:** As toras de madeira deveriam ter uma posição mais próxima das atividades de corte.
- **Atividades, Localizações e Entregas:** A posição dos nós reflete corretamente a relação semântica que existe entre estas classes.
- **Agentes e Máquinas:** O facto dos nós ocuparem posições semelhantes no gráfico sugere uma relação semântica entre esses dois conceitos.
- **Eventos e Localização:** A localização dos nós sugere uma relação semântica entre os eventos e as localizações das atividades.

Com base nessa análise, a Tabela 9 apresenta as duas relações que foi possível identificar a partir dos resultados do algoritmo *MetaPath2Vec*.

| Relação                           | Relevância  |
|-----------------------------------|-------------|
| <i>Machine belongsTo Agent</i>    | Média/Baixa |
| <i>Event hasLocation Location</i> | Baixa       |

Tabela 9: Relações semânticas obtidas pelo algoritmo *MetaPath2Vec*

Para o preenchimento de dados em falta, foi adotada uma abordagem baseada na similaridade dos nós, utilizando o método de previsão do *Random Forest*<sup>4</sup>. Primeiro, identificam-se os cinco nós com maior proximidade semântica do nó cujo valor se pretende obter. Em seguida, para prever o valor em falta, utiliza-se a média para variáveis numéricas e a moda para variáveis categóricas.

Para medir a qualidade das previsões foram utilizadas duas métricas: o Erro Absoluto Médio (MAE) e a *Accuracy*.

O MAE é uma medida que quantifica o erro médio entre os valores reais e os valores previstos [62]. É calculado através da fórmula EQ. 3:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{EQ. 3})$$

Já a *Accuracy* é uma medida que avalia a proporção de previsões corretas em relação ao total de previsões realizadas [63]. A sua fórmula está representada na EQ. 4:

$$\text{Accuracy} = \frac{\text{Número de previsões corretas}}{\text{Total de previsões}} \quad (\text{EQ. 4})$$

A Tabela 10 apresenta os resultados das previsões obtidas nos 4 testes realizados. Nas tarefas de regressão, o modelo estimou o peso com um MAE de 6.3 unidades, considerando um intervalo de valores medidos entre 30 e 42. No volume, o MAE foi de 33.5 unidades sendo que os volumes medidos variam entre 250 e 300. Já nas tarefas de classificação, a previsão da localização onde ocorreu uma atividade alcançou uma precisão de 86%, enquanto a previsão da máquina que realizou a atividade teve uma precisão de 73%. Estes resultados demonstram que o modelo apresentou um desempenho aceitável nas tarefas de regressão e um bom nível de precisão nas tarefas de classificação.

<sup>4</sup><https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>

| <b>Variável Dependente</b>                                     | <b><i>Accuracy</i></b> | <b>MAE</b> |
|--|------------------------|------------|
| Medição do peso de caminhão durante a atividade de transporte  | -                      | 6.3        |
| Medição de volume associado a atividade uma atividade de corte | -                      | 33.5       |
| Identificação da localização onde ocorreu uma atividade        | 86%                    | -          |
| Identificação da máquina que realizou uma atividade            | 73%                    | -          |

Tabela 10: Resultados das previsões

Em resumo, a aplicação do algoritmo teve resultados pouco satisfatórios na identificação de novas relações semânticas. Apesar de ser possível identificar duas relações novas, estas não têm grande relevância para a melhoria do conhecimento que já estava explícito na representação conceptual da ontologia. Além disso, o gráfico que ilustra a proximidade semântica dos nós apresenta também algumas deficiências na representação do conhecimento já existente, o que pode ser atribuído à baixa precisão (cerca de 60%) apresentada pelo modelo durante a fase de treino.

Relativamente à capacidade de prever dados em falta, a abordagem adotada demonstrou resultados satisfatórios, especialmente considerando que, embora o conjunto de dados não seja muito grande, apresenta uma boa variabilidade.

# Capítulo 6

## Conclusões e Trabalho Futuro

### 6.1 Reflexão Crítica

A rastreabilidade é cada vez mais um fator importante para a sustentabilidade do setor florestal uma vez que garante a origem da madeira e promove práticas adequadas na gestão dos recursos naturais. No contexto atual onde a procura por práticas e produtos sustentáveis é cada vez maior, a rastreabilidade permite acesso a informações claras e transparentes sobre a proveniência da madeira. Isso fortalece a confiança dos consumidores, assegura a conformidade com normas ambientais e reduz o risco de desmatamento ilegal.

No entanto, apesar da importância que a rastreabilidade tem no setor, existem desafios que dificultam a sua implementação de forma eficaz. A diversidade de abordagens e sistemas de monitorização utilizados nas diferentes fases da cadeia de abastecimento resulta num panorama fragmentado, onde os dados são dispersos e pouco padronizados. Essa falta de uniformidade faz com que a comunicação entre os diversos agentes envolvidos seja bastante difícil. Como consequência, não existe uma integração eficiente dos dados de modo a estabelecer um fluxo contínuo de informação que abranja todo o percurso da madeira, desde a floresta até à fábrica.

Nesse sentido, o trabalho descrito ao longo deste documento propôs um sistema de rastreabilidade com o objetivo de integrar os dados dispersos, permitindo uma visão completa do ciclo de vida da madeira. Este sistema foi desenvolvido para abranger tanto rastreabilidade na perspectiva *forward* como *backward*. Assim, além da transparência na informação sobre a origem da madeira, o sistema também possibilita a identificação de inconsistências nos dados e a correção de eventuais problemas.

A solução desenvolvida baseia-se numa abordagem sócio-semântica que utiliza uma ontologia para formalizar os conceitos e relações essenciais para a rastreabilidade da madeira. Esta ontologia foi criada de forma colaborativa, envolvendo especialistas do domínio, o que garantiu a incorporação dos conhecimentos práticos da área. Os principais *standards* da indústria também foram considerados na definição dos conceitos, o que permite a escalabilidade e a adaptação da solução a diferentes contextos.

Além da abordagem semântica, a solução incorporou técnicas de IA, através da aplicação de um algoritmo de GML. Este algoritmo treinou um modelo capaz de identificar novas relações e prever dados em falta com base na proximidade semântica dos nós. O objetivo

era melhorar o conhecimento existente e possibilitar a obtenção de dados que permitissem uma tomada de decisão mais informada.

Os resultados obtidos demonstraram que a solução permite a rastreabilidade da madeira de forma eficaz, tendo sido validada por especialistas do domínio. A existência de uma estrutura e de um *pipeline* que garantem a integração e organização dos dados possibilitou a criação de um fluxo de informação completo. Desse fluxo, além da rastreabilidade, foi possível identificar inconsistências nos dados recolhidos. No entanto, nem todos os resultados atingiram as expectativas inicialmente definidas. Na abordagem de IA, não foi possível obter relações que ampliassem o conhecimento do domínio já existente. Além disso, a abordagem para a previsão de dados em falta pode não ser adequada em situações onde a proximidade semântica seja insuficiente.

Em suma, de forma geral, os objetivos definidos para a dissertação foram atingidos. Foi possível implementar um sistema de rastreabilidade baseado num modelo semântico e inferir conhecimento implícito através de mecanismos de *reasoning*. Além disso, foi também possível identificar novas relações semânticas (ainda que sem grande relevância) e prever dados em falta através da utilização de IA.

## 6.2 Trabalho Futuro

Como trabalho futuro, pretende-se melhorar os resultados obtidos pela abordagem de IA. Isso inclui a exploração de algoritmos GML mais avançados, como HGNNs, e o aumento da qualidade e quantidade dos dados utilizados para o treino.

É também pretendido incorporar no modelo semântico mecanismos que permitam lidar com o impacto das condições climáticas nas medições realizadas ao longo das atividades. Variáveis como a humidade afetam o peso da madeira, o que pode comprometer o grau de confiança que é atribuído à rastreabilidade. Dessa forma, a consideração dessas variáveis irá permitir ao modelo realizar inferências de forma mais precisa e adaptativa, aumentando a fiabilidade dos resultados.

Por fim, pretende-se integrar o modelo semântico desenvolvido em *digital twins* estendendo a representação de *Asset Administration Shell (AAS)*. O AAS é a representação digital padronizada de um ativo<sup>1</sup> na Indústria 4.0 e possui uma estrutura modular que permite a inclusão de diversos submodelos para diferentes aspectos do ativo [64]. Nesse sentido, o modelo semântico desenvolvido será incorporado como um submodelo do ASS, categorizando os factos sobre a rastreabilidade.

---

<sup>1</sup>Ativo neste contexto pode ser máquinas, produtos, processos etc.

# Bibliografia

- [1] S. A. Hazaea, J. Zhu, S. F. Khatib, A. H. Bazhair, and A. A. Elamer, “Sustainability assurance practices: A systematic review and future research agenda,” *Environmental Science and Pollution Research*, vol. 29, no. 4, pp. 4843–4864, 2022.
- [2] Worldfavor, “The eu’s new digital product passport (dpp): Everything you need to know.” <https://blog.worldfavor.com/the-eus-new-digital-product-passport-dpp-everything-you-need-to-know>, 2023. Accessed: 2024-10-01.
- [3] EY, “Sustainability and traceability will be a future competitive advantage.” [https://www.ey.com/en\\_dk/consulting/sustainability-and-traceability-will-be-a-future-competitive-advantage](https://www.ey.com/en_dk/consulting/sustainability-and-traceability-will-be-a-future-competitive-advantage), 2023. Accessed: 2024-10-01.
- [4] P. Olsen and M. Borit, “How to define traceability,” *Trends in food science & technology*, vol. 29, no. 2, pp. 142–150, 2013.
- [5] G. M. Razak, L. C. Hendry, and M. Stevenson, “Supply chain traceability: A review of the benefits and its relationship with supply chain resilience,” *Production Planning & Control*, vol. 34, no. 11, pp. 1114–1134, 2023.
- [6] I. Tzoulis and Z. Andreopoulou, “Emerging traceability technologies as a tool for quality wood trade,” *Procedia Technology*, vol. 8, pp. 606–611, 2013.
- [7] S. M. Mc Carthy, M. Tambe, C. Kiekintveld, M. Gore, and A. Killion, “Preventing illegal logging: Simultaneous optimization of resource teams and tactics for security,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, 2016.
- [8] H. van der Ven and B. Cashore, “Forest certification: the challenge of measuring impacts,” *Current Opinion in Environmental Sustainability*, vol. 32, pp. 104–111, 2018.
- [9] A. Kaulen, L. Stopfer, K. Lippert, and T. Purfürst, “Systematics of forestry technology for tracing the timber supply chain,” *Forests*, vol. 14, no. 9, p. 1718, 2023.
- [10] S. Appelhanz, V.-S. Osburg, W. Toporowski, and M. Schumann, “Traceability system for capturing, processing and providing consumer-relevant information about wood products: system solution and its economic feasibility,” *Journal of Cleaner Production*, vol. 110, pp. 132–148, 2016.
- [11] A. Kaulen, L. Stopfer, K. Lippert, and T. Purfürst, “Systematics of forestry technology for tracing the timber supply chain,” *Forests*, vol. 14, no. 9, 2023.

- [12] M. Molinaro and G. Orzes, “From forest to finished products: The contribution of industry 4.0 technologies to the wood sector,” *Computers in Industry*, vol. 138, p. 103637, 2022.
- [13] J. v. Brocke, A. Hevner, and A. Maedche, *Introduction to Design Science Research*, pp. 1–13. Springer, 09 2020.
- [14] A. Hevner, S. Chatterjee, A. Hevner, and S. Chatterjee, “Design science research in information systems,” *Design research in information systems: theory and practice*, pp. 9–22, 2010.
- [15] P. Reis, “Ciência do artificial e design science research,” *Agência UFRJ de Inovação*, vol. 22, pp. 1–10, 2019.
- [16] J. Laporte and M. Vandenhoute, “Traceability: A management tool for business and governments,” *FAO FLEFGT Programme. Technical Paper (FAO) eng no. 1*, 2016.
- [17] F. Dabbene, P. Gay, and C. Tortia, “Traceability issues in food supply chain management: A review,” *Biosystems engineering*, vol. 120, pp. 65–80, 2014.
- [18] J. Sunny, N. Undralla, and V. Madhusudanan Pillai, “Supply chain transparency through blockchain-based traceability: An overview with demonstration,” *Computers Industrial Engineering*, vol. 150, p. 106895, 2020.
- [19] R. Pant, G. Prakash, and J. A. Farooque, “A framework for traceability and transparency in the dairy supply chain networks,” *Procedia-Social and Behavioral Sciences*, vol. 189, pp. 385–394, 2015.
- [20] M. H. Jansen-Vullers, C. A. van Dorp, and A. J. Beulens, “Managing traceability information in manufacture,” *International journal of information management*, vol. 23, no. 5, pp. 395–413, 2003.
- [21] P. Olsen and M. Borit, “How to define traceability,” *Trends in Food Science Technology*, vol. 29, no. 2, pp. 142–150, 2013.
- [22] A. Parreño-Marchante, A. Alvarez-Melcon, M. Trebar, and P. Filippin, “Advanced traceability system in aquaculture supply chain,” *Journal of Food Engineering*, vol. 122, pp. 99–109, 2014.
- [23] R. Mehannaoui, K. N. Mouss, and K. Aksa, “IoT-based food traceability system: Architecture, technologies, applications, and future trends,” *Food Control*, vol. 145, p. 109409, 2023.
- [24] S. P. Gayialis, E. P. Kechagias, G. A. Papadopoulos, and D. Masouras, “A review and classification framework of traceability approaches for identifying product supply chain counterfeiting,” *Sustainability*, vol. 14, no. 11, p. 6666, 2022.
- [25] F. Ameri, E. Wallace, R. Yoder, and F. Riddick, “Enabling traceability in agri-food supply chains using an ontological approach,” *Journal of Computing and Information Science in Engineering*, vol. 22, no. 5, p. 051002, 2022.
- [26] E. LBK, “O que é: Knowledge representation (representação de conhecimento),” 2024. Acessado em 06 outubro 2024.
- [27] M. Ouertani, S. Baïna, L. Gzara, and G. Morel, “Traceability and management of dispersed product knowledge during design and manufacturing,” *Computer-Aided*

- Design*, vol. 43, no. 5, pp. 546–562, 2011. Emerging Industry Needs for Frameworks and Technologies for Exchanging and Sharing Product Lifecycle Knowledge.
- [28] R. Studer, V. R. Benjamins, and D. Fensel, “Knowledge engineering: Principles and methods,” *Data & knowledge engineering*, vol. 25, no. 1-2, pp. 161–197, 2002.
- [29] T. Pizzuti, G. Mirabelli, M. A. Sanz-Bobi, and F. Gómez-González, “Food track trace ontology for helping the food traceability control,” *Journal of Food Engineering*, vol. 120, pp. 17–30, 2014.
- [30] M. A. M. Nieto, “An overview of ontologies,” *Universidad De Las Américas Puebla, Interactive and Cooperative Technologies Lab*, 2003.
- [31] OWL Working Group, “OWL 2 Web Ontology Language Document Overview: W3C Recommendation 27 October 2009.” <https://www.w3.org/TR/owl2-overview/>, 2009. [Online; accessed 9-October-2024].
- [32] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider, S. Rudolph, *et al.*, “Owl 2 web ontology language primer,” *W3C recommendation*, vol. 27, no. 1, p. 123, 2009.
- [33] D. Du, X. Ren, Y. Wu, J. Chen, W. Ye, J. Sun, X. Xi, Q. Gao, and S. Zhang, “Refining traceability links between vulnerability and software component in a vulnerability knowledge graph,” in *Web Engineering: 18th International Conference, ICWE 2018, Cáceres, Spain, June 5-8, 2018, Proceedings 18*, pp. 33–49, Springer, 2018.
- [34] M. Kuhn, E. T. Kaminski, and J. Franke, “Track and trace: Integrating static and dynamic data in a hybrid graph-based traceability model,” *Procedia CIRP*, vol. 112, pp. 250–255, 2022.
- [35] R. Jing and P. Li, “Knowledge graph for integration and quality traceability of agricultural product information,” *Frontiers in Sustainable Food Systems*, vol. 8, p. 1389945, 2024.
- [36] J. Guia, V. G. Soares, and J. Bernardino, “Graph databases: Neo4j analysis,” in *ICEIS (1)*, pp. 351–356, 2017.
- [37] R. Angles, H. Thakkar, and D. Tomaszuk, “Mapping rdf databases to property graph databases,” *IEEE Access*, vol. 8, pp. 86091–86110, 2020.
- [38] N. Baken, “Linked data for smart homes: Comparing rdf and labeled property graphs,” in *LDAC2020—8th Linked Data in Architecture and Construction Workshop*, pp. 23–36, 2020.
- [39] World Wide Web Consortium (W3C), “Shacl: Shapes constraint language,” 2017. Acessado em: 6 dez. 2024.
- [40] Parlamento Europeu, “O que é a inteligência artificial e como funciona,” 2020. Acesso em: 6 out. 2024.
- [41] Z.-H. Zhou, *Machine learning*. Springer nature, 2021.
- [42] T. Gaudet, B. Day, A. R. Jamasb, J. Soman, C. Regep, G. Liu, J. B. Hayter, R. Vickers, C. Roberts, J. Tang, *et al.*, “Utilizing graph machine learning within drug discovery and development,” *Briefings in bioinformatics*, vol. 22, no. 6, p. bbab159, 2021.

- [43] M. Kulmanov, F. Z. Smaili, X. Gao, and R. Hoehndorf, “Semantic similarity and machine learning with ontologies,” *Briefings in bioinformatics*, vol. 22, no. 4, p. bbaa199, 2021.
- [44] E. F. Strategy, “New eu forest strategy for 2030,” 2021.
- [45] F. Mtibaa and A. Chaabane, “Forestry wood supply chain information system using rfid technology,” in *IIE Annual Conference. Proceedings*, p. 1562, Institute of Industrial and Systems Engineers (IISE), 2014.
- [46] S. Appelhanz, V.-S. Osburg, W. Toporowski, and M. Schumann, “Traceability system for capturing, processing and providing consumer-relevant information about wood products: system solution and its economic feasibility,” *Journal of Cleaner Production*, vol. 110, pp. 132–148, 2016. Special Volume: Improved resource efficiency and cascading utilisation of renewable materials.
- [47] S. Figorilli, F. Antonucci, C. Costa, F. Pallottino, L. Raso, M. Castiglione, E. Pinci, D. Del Vecchio, G. Colle, A. R. Proto, G. Sperandio, and P. Menesatti, “A blockchain implementation prototype for the electronic open source traceability of wood along the whole supply chain,” *Sensors*, vol. 18, no. 9, 2018.
- [48] L. Mirowski, A. Smith, M. Hunt, and P. Turner, “A review of systems and technologies for timber traceability,” tech. rep., University of Tasmania, 2014. [Online; accessed 9-October-2024].
- [49] J. Scholz, A. De Meyer, A. S. Marques, T. M. Pinho, J. Boaventura-Cunha, J. Van Orshoven, C. Rosset, J. Künzi, J. Kaarle, and K. Nummilla, “Digital technologies for forest supply chain optimization: existing solutions and future trends,” *Environmental Management*, vol. 62, pp. 1108–1133, 2018.
- [50] J. Blake and W. Pease, “Development of an ontology to improve supply chain management (scm) in the australian timber industry,” in *Semantic Web Technologies and E-Business: Toward the Integrated Virtual Organization and Business Process Automation*, pp. 360–383, IGI Global, 2007.
- [51] E. Lallas, A. Karageorgos, and G. Ntalos, “An ontology based approach for regulatory compliance of eu reg. no 995/2010 in greece,” *Engineering Proceedings*, vol. 9, no. 1, p. 38, 2021.
- [52] PEFC, “A floresta portuguesa,” 2024. Acesso em: 17 de Setembro de 2024.
- [53] Projeto Draft, “O que é madeira certificada?,” 2024. Acesso em: 19 set. 2024.
- [54] D. L. d. Silva, R. R. Souza, and M. B. Almeida, “Princípios metodológicos para construção de ontologias: uma abordagem interdisciplinar,” 2009.
- [55] S. Union, “Pellet: A description logic reasoner,” 2024. Accessed: 2024-09-23.
- [56] Y. Dong, N. V. Chawla, and A. Swami, “metapath2vec: Scalable representation learning for heterogeneous networks,” in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 135–144, 2017.
- [57] J. Hu, B. Hooi, and B. He, “Efficient heterogeneous graph learning via random projection,” *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–14, 2024.

- [58] Encord, “Classification metrics: Accuracy, precision, recall,” 2024. Acessado em: 28 set. 2024.
- [59] P. Shvaiko and J. Euzenat, “Ontology matching: State of the art and future challenges,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 158–176, 2013.
- [60] D. M. Powers, “Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation,” *arXiv preprint arXiv:2010.16061*, 2020.
- [61] M. Jiang, Y. Chen, and L. Chen, “Link prediction in networks with nodes attributes by similarity propagation,” *arXiv preprint arXiv:1502.04380*, 2015.
- [62] ScienceDirect, “Mean absolute error,” n.d. Acessado: 29 de setembro de 2024.
- [63] E. AI, “Classification metrics: Accuracy, precision, recall,” 2024. Acessado: 2024-09-29.
- [64] M. A. P. Company, “What is the asset administration shell?.” <https://www.mhp.com/en/insights/blog/post/asset-administration-shell>, 2023. Acessado: 2024-09-30.

# Apêndice A

## Anexos

### A.1 Resultados do Sistema de Rastreabilidade

Classificação da madeira e respetivo nível de confiança na rastreabilidade (Figura 20).

| Fábrica                                | Peso        | Espécie   | Data de Receção | Hora de entrada | Prestador de Serviços | Classificação da Carga ↓ | Nível de Confiança  | Ver detalhes |
|--|-------------|-----------|-----------------|-----------------|-----------------------|--------------------------|---|--------------|
| Complexo Industrial de Setúbal         | 39.35 (Ton) | Pinheiro  | 2024-04-28      | 12:46:45        | Costa LDA             |                          | <div><div style="width: 100%;"></div></div>                           |              |
| Complexo Industrial de Setúbal         | 40.83 (Ton) | Pinheiro  | 2024-04-28      | 13:59:41        | Costa LDA             |                          | <div><div style="width: 100%;"></div></div>                           |              |
| Complexo Industrial da Figueira da Foz | 38.93 (Ton) | Eucalipto | 2024-03-04      | 08:57:54        | Costa LDA             |                          | <div><div style="width: 60%; background-color: #ffc107;"></div></div> |              |
| Complexo Industrial da Figueira da Foz | 40.18 (Ton) | Eucalipto | 2024-03-04      | 10:52:20        | Costa LDA             |                          | <div><div style="width: 100%;"></div></div>                           |              |
| Complexo Industrial da Figueira da Foz | 41.02 (Ton) | Eucalipto | 2024-03-04      | 11:16:06        | Interbrigas           |                          | <div><div style="width: 60%; background-color: #ffc107;"></div></div> |              |
| Complexo Industrial da Figueira da Foz | 40.1 (Ton)  | Eucalipto | 2024-03-04      | 16:21:40        | Interbrigas           |                          | <div><div style="width: 60%; background-color: #ffc107;"></div></div> |              |
| Complexo Industrial da Figueira da Foz | 40.37 (Ton) | Eucalipto | 2024-03-04      | 17:30:56        | Costa LDA             |                          | <div><div style="width: 100%;"></div></div>                           |              |
| Complexo Industrial da Figueira da Foz | 40.56 (Ton) | Eucalipto | 2024-03-04      | 17:55:37        | Interbrigas           |                          | <div><div style="width: 100%;"></div></div>                           |              |
| Complexo Industrial da Figueira da Foz | 38.07 (Ton) | Eucalipto | 2024-03-04      | 18:45:39        | Interbrigas           |                          | <div><div style="width: 100%;"></div></div>                           |              |
| Complexo Industrial da Figueira da Foz | 40.75 (Ton) | Eucalipto | 2024-03-04      | 19:52:21        | Interbrigas           |                          | <div><div style="width: 60%; background-color: #ffc107;"></div></div> |              |

Figura 20: Classificação de madeira (Certificada/Não Certificada)

Desperdício de recursos verificados nas atividades de exploração florestal (Figura 21).

| Atividade | Tipo                    | Máquina            | Data                | Número de toras perdidas | Peso       |
|-----------|-------------------------|--------------------|---------------------|--------------------------|------------|
| Recheça   | Desperdício de Recursos | (S) Forwarder 2024 | 2024-04-22 19:00:00 |                          | 7.20 (Ton) |
| Recheça   | Desperdício de Recursos | (S) Forwarder 2024 | 2024-06-18 12:00:00 |                          | 7.20 (Ton) |

Figura 21: Desperdício de recursos identificados

Inconsistências detetadas nos registos de rastreabilidade (Figura 22).

CLASSIFICAÇÃO DE MADEIRA    REGISTROS INCONSISTENTES    INEFICIÊNCIAS OPERACIONAIS    LICENCIAMENTO PRESTADORES DE SERVIÇOS

Carga com excesso de peso     Inconsistências coordenadas GPS

| Fábrica  | Data de Receção | Máquina     | Inconsistência    | Ver detalhes |
|----------|-----------------|-------------|-------------------|--------------|
| Santarém | 2024-04-16      | (S5) Camião | Origem da pilha   |              |
| Santarém | 2024-04-19      | (S8) Camião | Origem da pilha   |              |
| Santarém | 2024-04-19      | (S5) Camião | Origem da pilha   |              |
| Santarém | 2024-04-25      | (S) Camião  | Origem da pilha   |              |
| Santarém | 2024-06-20      | (S5) Camião | Origem da pilha   |              |
| Santarém | 2024-04-18      | (S8) Camião | Registo Duplicado |              |
| Santarém | 2024-04-18      | (S4) Camião | Registo Duplicado |              |
| Santarém | 2024-04-19      | (S4) Camião | Registo Duplicado |              |
| Santarém | 2024-04-25      | (S2) Camião | Registo Duplicado |              |
| Santarém | 2024-06-22      | (S2) Camião | Registo Duplicado |              |

Linhas por página: 10    1-10 de 20

Figura 22: Inconsistências identificadas nos dados

Ações que os prestadores de serviço devem implementar para melhorar a sustentabilidade das suas operações (Figura 23).

CLASSIFICAÇÃO DE MADEIRA    REGISTROS INCONSISTENTES    INEFICIÊNCIAS OPERACIONAIS    LICENCIAMENTO PRESTADORES DE SERVIÇOS

| Prestador de Serviços | Descrição                                   | Data de fim | Ação Necessária    |
|-----------------------|---|-------------|--------------------|
| ESilva                | Prestador de Serviço opera sem certificação | N/A         | Obter Certificação |
| JCoimbra              | Prestador de Serviço opera sem certificação | N/A         | Obter Certificação |
| Interbrigas           | Prestador de Serviço opera sem certificação | N/A         | Obter Certificação |

Linhas por página: 20    1-3 de 3

Figura 23: Ações necessárias para melhorar a sustentabilidade

Visualização do nível de confiança na rastreabilidade ao longo das diversas fases pelas quais a madeira passa até chegar à fábrica (Figura 24).

Rastreabilidade



Figura 24: Transparência da rastreabilidade ao longo das diversas fases

Atividades de corte com inconsistências detetadas nas coordenadas GPS (Figura 25).

Carga com excesso de peso  Inconsistências coordenadas GPS

Colunas Procurar

| Mata    | Data       | Hora     | Máquina            | Inconsistência                          | Ver detalhes      |
|---------|------------|----------|--------------------|---|-------------------|
| T001_EG | 2024-03-04 | 12:00:00 | (S) Harvester 2024 | Registadas árvores fora da área da mata | <a href="#">🔍</a> |

1-1 de 1

**Atividade de corte com dados inconsistentes**

Figura 25: Inconsistências Coordenadas GPS

Transportes com excesso de peso (Figura 26).

Carga com excesso de peso  Inconsistências coordenadas GPS

Colunas Procurar

| Fábrica  | Data de Receção | Máquina     | Inconsistência         | Discrepância <span style="font-size: small;">↓</span> |
|----------|-----------------|-------------|------------------------|---|
| Santarém | 2024-03-16      | (S) Camião  | Camião com peso a mais | 6.05% <span style="font-size: small;">📊</span>        |
| Santarém | 2024-03-12      | (S8) Camião | Camião com peso a mais | 2.05% <span style="font-size: small;">📊</span>        |
| Santarém | 2024-03-09      | (S) Camião  | Camião com peso a mais | 9.05% <span style="font-size: small;">📊</span>        |
| Santarém | 2024-03-06      | (S2) Camião | Camião com peso a mais | 8.05% <span style="font-size: small;">📊</span>        |
| Santarém | 2024-03-10      | (S2) Camião | Camião com peso a mais | 10.05% <span style="font-size: small;">📊</span>       |

Linhas por página: 5 1-5 de 120 < > >>

Figura 26: Transportes com excesso de peso

## A.2 Evolução do Treino do Algoritmo de GML

A evolução do treino do algoritmo de GML é apresentada na Figura 27, onde é possível observar uma melhoria contínua do modelo na capacidade de prever a proximidade semântica entre os nós.

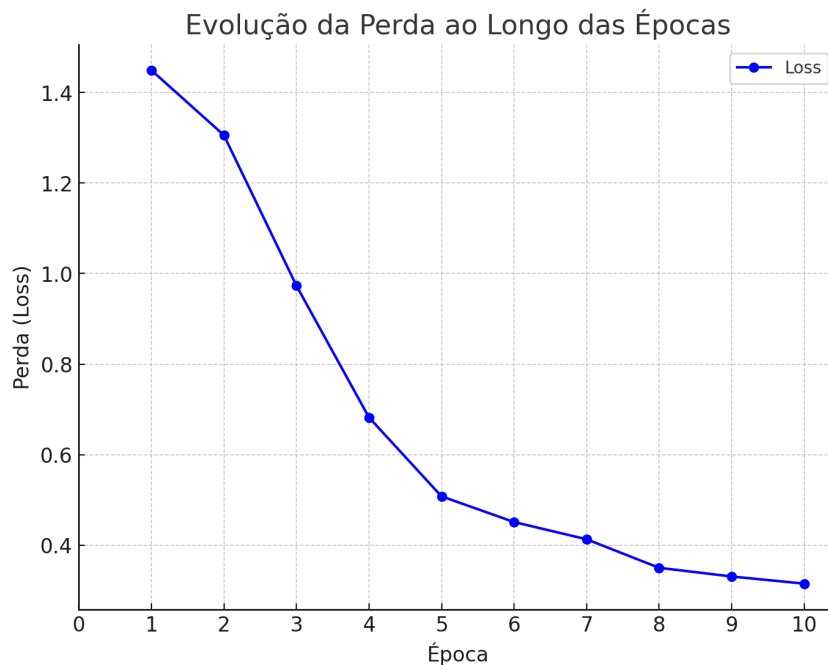


Figura 27: Evolução do treino do modelo

## A.3 Regras da Ontologia

| Objetivo   | Regra SWRL  |
|--|---|
| Identificar Prestadores de Serviço com licenças expiradas      | <pre> Agent(?e) ∧ hasCertification(?e, ?c) ∧ hasExpirationDate(?c, ?expiryDate) ∧ MinimumDate(?md) ∧ hasDate(?md, ?minimumDate) ∧ swrlb:greaterThan(?expiryDate, ?minimumDate) ∧ CurrentDate(?cd) ∧ hasDate(?cd, ?currentDate) ∧ swrlb:lessThan(?expiryDate, ?currentDate) ⇒ mightRequireAction(?e, renew_certification) </pre>                                 |
| Identificar ineficiências operacionais na atividade de recheça | <pre> Activity(?e1) ∧ produces(?e1, ?o1) ∧ Pile(?o1) ∧ hasNumberOfWoodLogs(?o1, ?l1) ∧ precedes(?e1, ?e2) ∧ produces(?e2, ?o2) ∧ WoodLogs(?o2) ∧ hasNumberOfWoodLogs(?o2, ?l2) ∧ swrlb:lessThan(?l1, ?l2) ∧ swrlb:subtract(?diff, ?l2, ?l1) ⇒ OperationalInefficiency(?e1) ∧ hasInefficiencyType(?e1, 'RESOURCEWASTAGE') ∧ hasWastedQuantity(?e1, ?diff) </pre> |
| Identificar inconsistências no relatório de atividades         | <pre> Activity(?e) ∧ produces(?e, ?o) ∧ Delivery(?o) ∧ precedes(?e, ?e2) ∧ Activity(?e2) ∧ produces(?e2, ?o2) ∧ WoodLogs(?o2) ⇒ InconsistencyTracking(?e) ∧ hasDescription(?e, 'IFORWARDING') </pre>  |
| Identificar inconsistências GPS nas atividades de corte        | <pre> Activity(?e1) ∧ produces(?e1, ?o1) ∧ WoodLogs(?o1) ∧ hasLogsOut(?o1, ?value) ∧ swrlb:equal(?value, true) ⇒ InconsistentData(?e1) ∧ InconsistencyTracking(?e1) ∧ hasDescription(?e1, 'TREE_OUT_FOREST') </pre>   |

Tabela 11: Regras SWRL (Parte 1)

| Objetivo  | Regra SWRL   |
|---|--|
| Definir o grau de confiança da rastreabilidade como baixo                       | <pre> WoodClassification(?o1) ∧ produces(?e1, ?o1) ∧ Delivery(?o1) ∧ precedes(?e1, ?e2) ∧ precedes(?e2, ?e3) ∧ InconsistentData(?e3) ∧ InconsistentData(?e2) ⇒ hasTraceabilityTrustLevel(?o1, low) </pre>  |
| Classificação de Madeira como não certificada                                   | <pre> Activity(?e1) ∧ produces(?e1, ?o1) ∧ Delivery(?o1) ∧ precedes(?e1, ?e2) ∧ Activity(?e2) ∧ produces(?e2, ?o2) ∧ Pile(?o2) ∧ precedes(?e2, ?e3) ∧ Activity(?e3) ∧ produces(?e3, ?o3) ∧ WoodLogs(?o3) ∧ involvesAgent(?e1, ?agent1) ∧ hasCertification(?agent1, ?c1) ∧ hasExpirationDate(?c1, ?ag1ExpirationDate) ∧ CurrentDate(?cd) ∧ hasDate(?cd, ?currentDate) ∧ swrlb:lessThan(?ag1ExpirationDate, ?currentDate) ⇒ WoodClassification(?o1) ∧ isCertified(?o1, false) </pre> |
| Classificação de madeira como certificada                                       | <pre> Activity(?e1) ∧ produces(?e1, ?o1) ∧ Delivery(?o1) ∧ precedes(?e1, ?e2) ∧ Activity(?e2) ∧ produces(?e2, ?o2) ∧ Pile(?o2) ∧ precedes(?e2, ?e3) ∧ Activity(?e3) ∧ produces(?e3, ?o3) ∧ WoodLogs(?o3) ∧ involvesAgent(?e1, ?agent1) ∧ hasCertification(?agent1, ?c1) ∧ hasExpirationDate(?c1, ?ag1ExpirationDate) ∧ CurrentDate(?cd) ∧ hasDate(?cd, ?currentDate) ∧ swrlb:lessThan(?ag1ExpirationDate, ?currentDate) ⇒ WoodClassification(?o1) ∧ isCertified(?o1, true) </pre>  |
| Identificar registos duplicados   | <pre> Activity(?e) ∧ produces(?e, ?o1) ∧ Delivery(?o1) ∧ hasMeasurement(?e, ?m1) ∧ Weight(?m1) ∧ hasValue(?m1, ?v1) ∧ swrlb:greaterThan(?v1, 60.00) ⇒ InconsistencyTracking(?e) ∧ hasDescription(?e, 'DUPLICATED_DATA') </pre>   |
| Identificar ações necessárias para promover práticas de exploração sustentáveis | <pre> Agent(?e) ∧ hasCertification(?e, ?c) ∧ hasExpirationDate(?c, ?expiryDate) ∧ MinimumDate(?md) ∧ hasDate(?md, ?minimumDate) ∧ swrlb:equal(?expiryDate, ?minimumDate) ⇒ mightRequireAction(?e, get_certification) </pre>  |

Tabela 12: Regras SWRL (Parte 2)

| Objetivo   | Regra SWRL  |
|--|---|
| Identificar transportes com cargas acima do peso         | <pre> Activity(?e) ^ relatedTo(?ev, ?e) ^ Event(?ev) ^ relatedTo(?ev2, ?e) ^ Event(?ev2) ^ hasMeasurement(?ev, ?m1) ^ hasMeasurement(?ev2, ?m2) ^ hasValue(?m1, ?v1) ^ hasValue(?m2, ?v2) ^ swrlb:notEqual(?v2, ?v1) ⇒ InconsistencyTracking(?e) ^ hasDescription(?e, 'TRUCK_WITH_INCREASED_LOAD') </pre>   |
| Identificar transportes com madeira não rastreada        | <pre> Activity(?e) ^ produces(?e, ?o1) ^ Delivery(?o1) ^ hasMeasurement(?e, ?m1) ^ Weight(?m1) ^ Track(?t) ^ createdBy(?t, ?e) ^ precedesTrack(?t, ?t2) ^ hasCurrentMeasurement(?t2, ?m2) ^ hasValue(?m1, ?v1) ^ hasValue(?m2, ?v2) ^ swrlb:lessThan(?v1, 60.00) ^ swrlb:greaterThan(?v1, ?v2) ⇒ InconsistencyTracking(?e) ^ hasDescription(?e, 'DELIVERY_WITH_TWO_PILES') </pre> |
| Definir o grau de confiança da rastreabilidade como alto | <pre> WoodClassification(?o1) ^ produces(?e1, ?o1) ^ Delivery(?o1) ^ precedes(?e1, ?e2) ^ precedes(?e2, ?e3) ^ ConsistentData(?e1) ^ ConsistentData(?e2) ^ ConsistentData(?e3) ⇒ hasTraceabilityTrustLevel(?o1, high) </pre>  |
| Identificar entregas de madeira sem registo de origem    | <pre> Activity(?e) ^ precedes(?e, ?e2) ^ hasActivityDate(?e2, ?eventDate) ^ MinimumDate(?minDate) ^ swrlb:equal(?eventDate, ?minDate) ^ InconsistencyTracking(?e) ^ hasDescription(?e, 'NO_HARVEST_RECORD') </pre>  |
| Identificar atividades com medições inconsistentes       | <pre> Activity(?e1) ^ precedes(?e1, ?e2) ^ Activity(?e2) ^ produces(?e2, ?o) ^ WoodLogs(?o) ^ hasMeasurement(?e1, ?m1) ^ hasMeasurement(?e2, ?m2) ^ Volume(?m1) ^ Volume(?m2) ^ hasValue(?m1, ?v1) ^ hasValue(?m2, ?v2) ^ swrlb:lessThan(?v2, ?v1) ⇒ InconsistentData(?e1) </pre>   |

Tabela 13: Regras SWRL (Parte 3)