

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

Ontologias aplicadas à Ciência de Dados

Luiz Augusto Masset Lacombe Papa

PROJETO FINAL DE GRADUAÇÃO

CENTRO TÉCNICO CIENTÍFICO - CTC

DEPARTAMENTO DE INFORMÁTICA

Curso de Graduação em Ciência da Computação

Rio de Janeiro, Novembro de 2025



Luiz Augusto Masset Lacombe Papa

Ontologias aplicadas a Ciência de Dados

Relatório de Projeto Final, apresentado ao programa Ciência da Computação da PUC-Rio como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Marcos Vianna Villas

Departamento de Informática

Rio de Janeiro
Novembro 2025

Resumo

PAPA, Luiz Augusto Masset Lacombe. VILLAS, Marcos Vianna. Ontologias aplicadas à Ciência de Dados. Rio de Janeiro, 2025, 74p. Relatório de Projeto Final II – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho investiga a aplicação de ontologias em cada etapa do ciclo de vida de projetos de Ciência de Dados, com foco em ontologias de domínio. A partir de revisão bibliográfica e de um modelo de referência construído a partir de metodologias consolidadas (CRISP-DM, TDSP, SEMMA, KDD, Domino), foram propostas vinte assertivas que descrevem benefícios das ontologias para cada fase do ciclo de vida. A proposta foi validada por meio de questionários e entrevistas com especialistas da área.

Palavras-chave: ciência de dados, ontologia, ciclo de vida de projetos

Abstract

PAPA, Luiz Augusto Masset Lacombe. VILLAS, Marcos Vianna. Ontologies applied to Data Science. Rio de Janeiro, 2025, 74p. Final Project Report II – Department of Informatics, Pontifical Catholic University of Rio de Janeiro.

This work investigates the application of ontologies in each stage of the Data Science project lifecycle, with a focus on domain ontologies. Based on a literature review and a reference model built from established methodologies (CRISP-DM, TDSP, SEMMA, KDD and Domino), twenty assertions were proposed to describe the benefits of ontologies for each phase of the lifecycle. The proposal was validated through questionnaires and interviews with domain specialists.

Keywords: data science, ontology, project lifecycle

Sumário

1. Introdução.....	1
2. Situação Atual.....	1
2.1. Etapas da Ciência de Dados.....	1
2.2. Uso de Ontologias para Bancos de Dados.....	3
2.3. A Convergência da Ciência de Dados e das Ontologias.....	4
3. Objetivos.....	6
4. Metodologia.....	6
4.1. Etapas.....	6
5. Ciência de Dados.....	8
5.1. Por que escolher metodologias.....	8
5.2. Explicação das escolhas das metodologias.....	9
5.3. Metodologias escolhidas.....	9
5.4. Metodologias descartadas.....	17
5.5. Modelo de referência para ciclo de vida de Ciência de Dados.....	17
6. Ontologias.....	19
6.1. A Estrutura de uma Ontologia.....	19
6.2. Tecnologias para Ontologias.....	20
6.3. Web Semântica.....	20
6.4. Exemplo de Ontologia de Domínio: OWL-Time.....	21
7. Aplicação de Ontologias no Ciclo de Vida de Ciência de Dados.....	22
8. Entrevistas.....	29
8.1. Participantes.....	29
8.2. Formulário.....	30
8.3. Email.....	30
9. Resultados.....	31
9.1. Síntese das respostas objetivas.....	31
9.2. Pontos mais fortes.....	32
9.3. Pontos mais fracos.....	33
9.4. Facilitadores.....	33
9.5. Adicionais.....	34
10. Ajustes.....	35
11. Conclusão.....	38
11.1. Objetivos x Realidade.....	38
11.2. Trabalhos futuros.....	39
12. Referências.....	40
13. Apêndices.....	44
13.1. Apêndice 1: Formulário.....	44
13.2. Apêndice 2: Email.....	64
13.3. Apêndice 3: Comentários formulário.....	65

1. Introdução

A crescente importância da Ciência de Dados¹ tem transformado a maneira como as organizações abordam a tomada de decisões e a inovação. A capacidade de extrair conhecimento valioso de grandes volumes de dados heterogêneos tornou-se um diferencial competitivo crucial. No entanto, essa ascensão também trouxe desafios: a heterogeneidade dos dados, provenientes de diversas fontes e armazenados em formatos variados, dificulta a integração e a análise coesa. Além disso, a complexidade dos algoritmos de aprendizado de máquina levanta preocupações sobre a transparência e a compreensibilidade de seus resultados.

Nesse contexto, as ontologias² emergem como uma solução promissora para enfrentar esses desafios. Ao fornecerem um vocabulário compartilhado e uma estrutura conceitual clara, elas podem atuar como uma ponte entre diferentes fontes de dados, facilitando a integração semântica e a interoperabilidade. Adicionalmente, a natureza explícita das ontologias pode contribuir para a explicabilidade dos modelos de machine learning, tornando o raciocínio por trás das decisões mais transparentes.

2. Situação Atual

Nesta seção, descreve-se o cenário atual da convergência entre Ciência de Dados e ontologias. São apresentados os principais processos do ciclo de vida da área e, na sequência, demonstra-se aplicações de ontologias em bancos de dados, destacando seu papel no suporte à qualidade e integração de dados.

2.1. Etapas da Ciência de Dados

O ciclo de vida da Ciência de Dados é iterativo e dinâmico, permitindo o retorno a etapas anteriores para refinamento sempre que surgem novas percepções ou mudanças nos requisitos. Diversos modelos enfatizam essa característica cíclica, como CRISP-DM (Cross-Industry Standard Process for

¹ Campo interdisciplinar que combina estatística, computação e conhecimento de domínio para coletar, processar e analisar dados com o objetivo de extrair conhecimento e apoiar a tomada de decisões.

² Uma representação formal de conhecimento, que define um conjunto de conceitos, propriedades e as relações entre eles.

Data Mining), proposto por um consórcio de empresas [1], e TDSP (Team Data Science Process), desenvolvido pela Microsoft [2], que são modelos amplamente usados e adaptados a diferentes contextos e domínios [3][4].

As etapas mais recorrentes entre os diferentes modelos existentes são descritas a seguir.

- **Definição do problema e entendimento do negócio**

A primeira etapa crucial em qualquer projeto de Ciência de Dados é definir claramente o problema e os objetivos de negócio [5]. Isso envolve entender o contexto e transformar as necessidades da empresa em problemas que a Ciência de Dados pode ajudar a resolver.

- **Coleta de dados**

Este processo envolve a identificação de fontes de dados adequadas e a coleta de dados em diversos formatos [4].

- **Preparação e limpeza de dados**

A etapa de preparação e limpeza de dados envolve a transformação de dados brutos em um formato utilizável para análise. Esta fase envolve corrigir erros, tratar valores ausentes e garantir que os dados estejam em um formato consistente. Esta etapa frequentemente é a mais demorada de um projeto de Ciência de Dados, consumindo uma parcela significativa do tempo total [5].

- **Análise Exploratória de Dados (EDA)**

A Análise Exploratória de Dados tem como objetivo compreender as características dos dados, como tendências centrais, dispersão e distribuição, além de identificar padrões e anomalias, testar hipóteses, e verificar suposições. Ela auxilia na obtenção de uma compreensão mais profunda do conjunto de dados e das relações entre as variáveis, que correspondem aos atributos ou colunas que descrevem as observações do conjunto de dados.

- **Construção e Avaliação de Modelos**

Esta etapa envolve a seleção de algoritmos de aprendizado de máquina apropriados e a construção de modelos. Isso requer a escolha de algoritmos com base no tipo de problema e nas características dos dados [3].

- **Implantação**

Nesta etapa, o modelo treinado é integrado ao ambiente de produção, por meio de APIs ou dashboards, e passa a ser monitorado continuamente, garantindo o ciclo completo de operação em larga escala [5].

Nem todo projeto de Ciência de Dados necessariamente percorre as etapas de modelagem e implantação, como aqueles focados apenas em análise exploratória ou descritiva, que podem não demandar a construção de modelos. A modelagem é crucial para tarefas de predição e classificação, enquanto a implantação é necessária quando os resultados devem ser integrados a sistemas de produção. É importante destacar que essas metodologias são flexíveis, permitindo que etapas sejam adaptadas, repetidas ou até mesmo omitidas, sempre de acordo com os objetivos e as necessidades específicas de cada projeto [3][5].

2.2. Uso de Ontologias para Bancos de Dados

Ontologias são representações formais de conhecimento que definem categorias, propriedades e relações entre conceitos de um domínio, permitindo comunicação e entendimento compartilhado em diferentes contextos [6].

No contexto de banco de dados, as ontologias funcionam como estruturas formais que organizam e definem conceitos e relações dentro de um domínio, promovendo integração semântica, consistência de dados e interoperabilidade entre sistemas [7].

As principais aplicações em bancos de dados são apresentadas a seguir.

- **Integração e interoperabilidade**

As ontologias fornecem uma estrutura comum para a compreensão e a vinculação de dados de fontes distintas. Ao oferecerem um vocabulário compartilhado e uma representação estruturada do conhecimento, as ontologias superam os desafios da heterogeneidade de dados e permitem melhor integração e interoperabilidade semântica³ [8][9].

³ Capacidade de diferentes sistemas trocarem e interpretarem dados de forma significativa, garantindo que o conteúdo transmitido seja compreendido com o mesmo sentido por todos os envolvidos, independentemente de suas plataformas ou estruturas internas [9].

- **Consultas semânticas (OBDA - Ontology-Based Data Access)**

O acesso a dados baseado em ontologias (OBDA) permite consultar bancos de dados relacionais a partir de uma visão conceitual do domínio, representada por uma ontologia. Essa abordagem torna a interação com os dados mais intuitiva, pois os usuários podem utilizar a linguagem do próprio domínio de aplicação, resultando em buscas mais relevantes e alinhadas ao significado dos dados. O OBDA foi desenvolvido para o modelo relacional, no qual consultas em SPARQL são traduzidas em SQL por meio de mapeamentos entre os conceitos da ontologia e o esquema relacional. Embora existam pesquisas voltadas à extensão desse paradigma para outros tipos de bancos de dados, como grafos e documentos, as implementações consolidadas, como o Ontop [8][10], permanecem centradas em sistemas relacionais.

- **Melhoria na qualidade dos dados**

As ontologias podem impor consistência e integridade aos dados, fornecendo uma estrutura definida e um conjunto de regras, usando linguagens como SHACL (Shapes Constraint Language), um padrão W3C para validação de grafos RDF que define “shapes” para verificar cardinalidade, tipos e padrões de valor em seus dados, detectando inconsistências sintáticas e semânticas [11][12].

- **Governança de dados**

Ao fornecer estruturas explícitas, ontologias apoiam políticas de qualidade e conformidade, essenciais para a gestão de grandes volumes de dados [12].

Apesar dos benefícios proporcionados pelas ontologias, ainda persistem desafios técnicos relevantes, em especial o mapeamento de esquemas relacionais para ontologias [8] e a escalabilidade no processamento de grafos RDF de grande porte, na ordem de bilhões de triplas [13].

2.3. A Convergência da Ciência de Dados e das Ontologias

A convergência entre ontologias e Ciência de Dados vem sendo destacada como estratégia-chave para melhorar explicabilidade, integração e automação dos fluxos de trabalho [7][14].

As aplicações atuais dessa convergência são descritas a seguir.

- **Definição e modelagem de problemas**

As ontologias ajudam na representação conceitual do domínio, promovendo uma compreensão mais clara e abrangente logo nas etapas iniciais dos projetos [29].

- **Aquisição e integração de dados**

As ontologias oferecem uma base semântica que orienta a identificação e a coleta de fontes de dados relevantes, promovendo uma abordagem mais direcionada e coerente. O acesso a dados baseado em ontologias (OBDA) permite consultar dados diversos por meio de um vocabulário compartilhado, integrando ontologias de domínio a fontes de dados por meio de mapeamentos declarativos [8][10].

- **Engenharia de features**

As ontologias podem guiar a criação de variáveis semanticamente significativas, beneficiando algoritmos de ML, especialmente em contextos como classificação de texto ou construção de grafos de conhecimento [15].

- **Explicabilidade e avaliação de modelos**

As ontologias oferecem uma estrutura semântica que auxilia na interpretação dos resultados de modelos de machine learning, contribuindo para a explicabilidade em sistemas de IA, especialmente na abordagem neuro-simbólica⁴ [14].

- **Preparação e limpeza de dados**

As ontologias fornecem definições claras e padronizadas dos conceitos e relacionamentos presentes nos conjuntos de dados. Isso facilita a detecção de inconsistências, ambiguidade e redundâncias [11].

⁴ Combina redes neurais com representação simbólica de conhecimento (como ontologias), buscando unir a capacidade de reconhecimento de padrões e aprendizado automático dos modelos neurais com a interpretabilidade, transparência e raciocínio estruturado das abordagens simbólicas [14].

3. Objetivos

O principal objetivo deste trabalho foi de investigar como aplicar ontologias em cada etapa do ciclo de vida de Ciência de Dados, de modo a, por exemplo, aprimorar a consistência na terminologia e estrutura dos dados, a interoperabilidade entre diferentes sistemas e a explicabilidade de modelos de machine learning.

Com seus resultados buscamos contribuir para que os analistas de dados possam incorporar práticas baseadas em ontologias de forma prática e acessível, promovendo maior alinhamento semântico entre os dados, melhorando a integridade da informação, a reutilização de conhecimento e a transparência dos processos analíticos.

4. Metodologia

4.1. Etapas

O diagrama (Figura 1) apresenta um planejamento com as etapas de estudo, desenvolvimento, validação e refinamento da proposta.

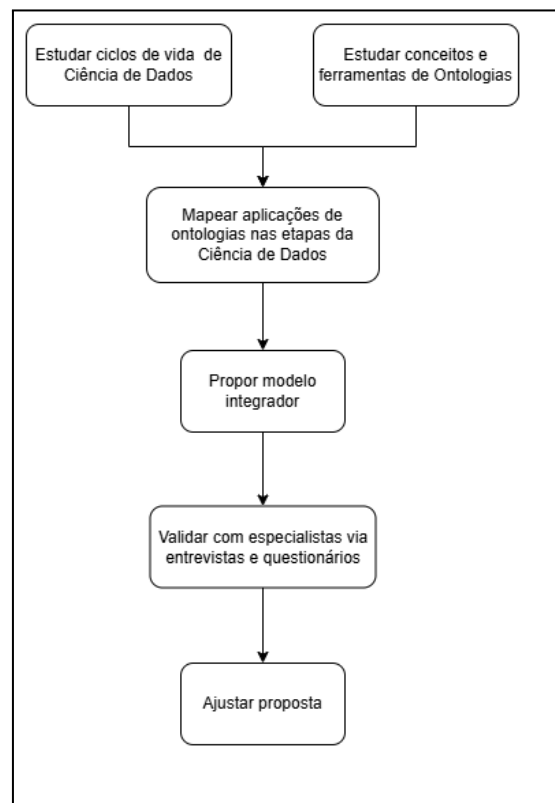


Figura 1: Etapas da Metodologia

4.1.1. Estudar ciclos de vida de Ciência de Dados

- Levantamento e análise de modelos de ciclo de vida da Ciência de Dados (CRISP-DM, TDSP, etc.)
- Identificação das principais etapas e desafios em projetos típicos de Ciência de Dados.

4.1.2. Estudar conceitos e ferramentas de Ontologias

- Revisão dos conceitos fundamentais de ontologias, incluindo OWL, RDF
- Análise de ferramentas como Protégé, Ontop, TopBraid Composer, entre outras.
- Estudo de casos de uso em integração semântica, explicabilidade e qualidade de dados.

4.1.3. Mapear aplicações de ontologias nas etapas da Ciência de Dados

- Análise da literatura sobre o uso de ontologias em cada fase do ciclo de vida da Ciência de Dados.
- Identificação de lacunas e oportunidades de melhoria nos fluxos atuais com auxílio de ontologias.

4.1.4. Propor modelo integrador

- Proposição de uma estrutura integrativa para aplicação de ontologias nas etapas de projetos de Ciência de Dados. Definição de exemplos práticos e cenários de aplicação.

4.1.5. Validar com especialistas via entrevistas e questionários

- Condução de entrevistas e aplicação de questionários com profissionais da área.
- Análise do feedback obtido.

4.1.6. Ajustar proposta

- Revisão do modelo com base nas sugestões dos especialistas.
- Consolidação de um conjunto de boas práticas e recomendações.

5. Ciência de Dados

Esta seção apresenta uma análise objetiva do ciclo de vida de projetos de Ciência de Dados, justificando a seleção de metodologias para estudo, descrevendo-as sucintamente e propondo um modelo de referência unificado.

5.1. Por que escolher metodologias

A escolha de metodologias estruturadas para guiar projetos de Ciência de Dados é essencial para aumentar as chances de sucesso. Muitos projetos falham por falta de clareza de objetivos, papéis definidos e governança. Metodologias bem definidas ajudam a suprir lacunas organizacionais, enfrentando desafios técnicos e sociais, além de orientar o gerenciamento de dados, equipe e projeto. [3]

Neste trabalho, foram escolhidas metodologias para embasar a definição de um modelo de referência, que foi utilizado posteriormente para avaliar o uso de ontologias em cada uma das etapas. O objetivo foi criar uma estrutura padronizada sobre a qual seja possível analisar como e onde as ontologias podem agregar valor ao ciclo de vida da Ciência de Dados.

5.2. Explicação das escolhas das metodologias

As metodologias selecionadas neste trabalho foram fundamentadas principalmente nos estudos [3], [16] e [17], que fornecem bases empíricas e analíticas para a escolha de abordagens consolidadas no contexto da Ciência de Dados. O estudo apresentado em [3] oferece uma análise crítica das principais metodologias existentes, identificando lacunas e oportunidades, e justificando a adoção de modelos com ampla cobertura de processos e etapas bem definidas. A pesquisa em [16], realizada em 2014, investigou as metodologias de Ciência de Dados mais utilizadas. Por fim, a pesquisa descrita em [17], conduzida com 237 profissionais da área, em 2021, evidencia as principais metodologias usadas no mercado.

Com base nesses critérios, foram selecionadas cinco metodologias para embasar a construção de um modelo de referência: CRISP-DM, TDSP, SEMMA, KDD Process e o Domino Data Science Lifecycle. As escolhas consideraram a existência de etapas bem definidas, documentação acessível e adoção comprovada no meio acadêmico e industrial [3][16][17]. Em contrapartida, a metodologia Agile Data Science Lifecycle foi descartada, embora útil como abordagem gerencial complementar, carece de definição clara das fases envolvidas em um projeto típico de Ciência de Dados, o que dificulta sua utilização como referência analítica neste estudo.

As metodologias selecionadas são brevemente descritas na Seção 5.3, enquanto a Seção 5.4 apresenta as justificativas para as descartadas. A Seção 5.5 consolida o modelo de referência, que será utilizado posteriormente na avaliação do papel das ontologias em projetos de Ciência de Dados.

5.3. Metodologias escolhidas

5.3.1. CRISP-DM

O CRISP-DM (Cross-Industry Standard Process for Data Mining) é considerado o padrão de fato para o desenvolvimento de projetos de mineração de dados e descoberta de conhecimento. As razões para sua ampla utilização incluem o fato de ser um modelo estruturado, confiável e de fácil compreensão. Um de seus principais pontos fortes é ser independente tanto da indústria, consolidando-se como um modelo de processo organizacional que consiste em seis fases iterativas [18].

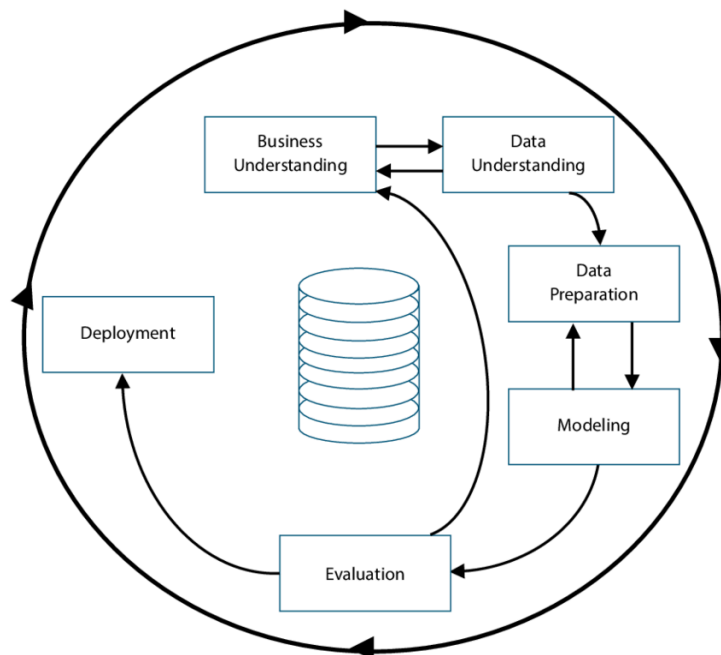


Figura 2: Diagrama CRISP-DM

Fonte:

https://www.researchgate.net/figure/Phases-of-the-CRISP-DM-reference-model-Retrieved-from-CRISP-DM-10-Step-by-Step-Data_fig5_340620896, Acesso em: 02 jun. 2025.

Adaptada de [1]

As fases do CRISP-DM são [1]:

- Entendimento do Negócio (Business Understanding): Definição dos objetivos do projeto sob uma perspectiva de negócio e conversão desse conhecimento em uma definição de problema de mineração de dados.
- Entendimento dos Dados (Data Understanding): Coleta inicial dos dados e análise exploratória para entender a estrutura, verificar problemas de qualidade e identificar hipóteses iniciais.
- Preparação dos Dados (Data Preparation): Transformação dos dados brutos no conjunto final que será usado na modelagem. Envolve seleção, limpeza e transformação dos dados. Essas tarefas podem ser repetidas múltiplas vezes e não seguem uma ordem rígida.
- Modelagem (Modeling): Seleção e aplicação de diferentes técnicas de modelagem, ajustando seus parâmetros. Algumas técnicas exigem formatos específicos dos dados, o que pode exigir retornar à etapa de preparação.

- Avaliação (Evaluation): Verificação se os modelos atendem aos objetivos de negócio e se alguma questão importante foi esquecida.
- Implantação (Deployment): Torna os resultados utilizáveis, seja com relatórios simples ou aplicação em sistemas. O cliente geralmente executa essa fase, mas é preciso direcioná-lo a como usar os modelos criados.

5.3.2. TDSP

O TDSP (Team Data Science Process) é uma metodologia ágil e iterativa de ciência de dados desenvolvida pela Microsoft para orientar o desenvolvimento eficiente de soluções de análise preditiva e aplicações inteligentes. Sendo uma consolidação de melhores práticas da Microsoft e de outras empresas da indústria, o TDSP visa aprimorar a colaboração em equipe e o aprendizado contínuo. Ele é composto por um ciclo de vida definido, uma estrutura de projeto padronizada, infraestrutura e ferramentas, e pode ser utilizado de forma flexível em conjunto com outros processos como o CRISP-DM [2].

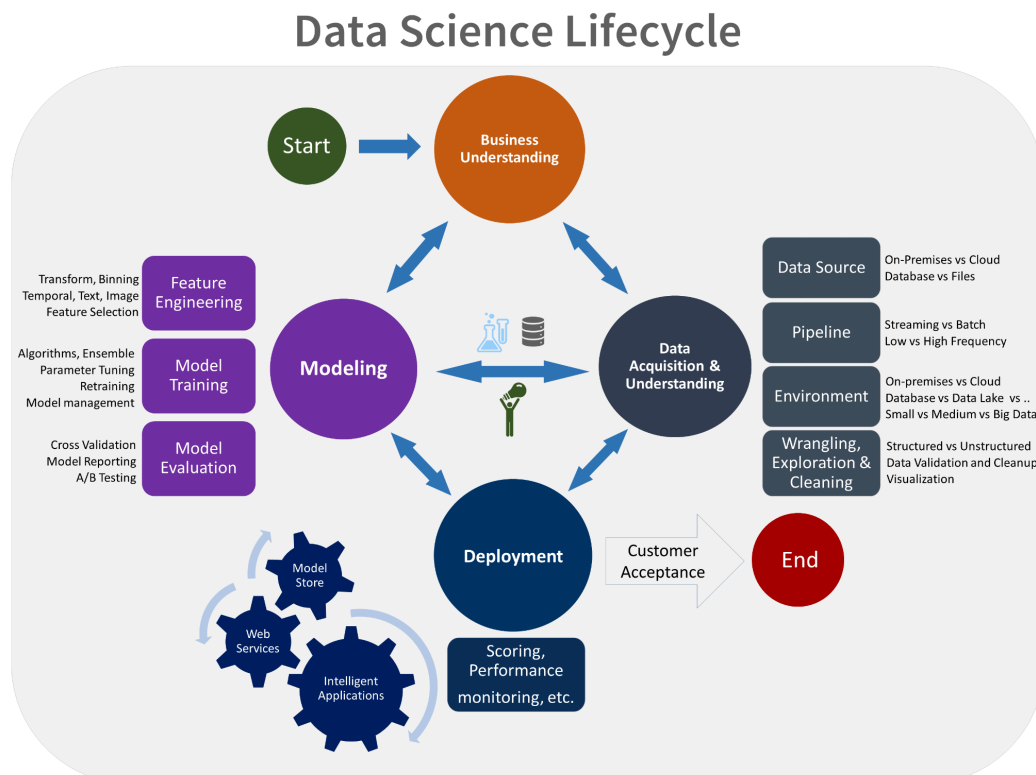


Figura 3: Diagrama TDSP

Fonte: [2]

As fases do TDSP são [2]:

- Entendimento do Negócio (Business Understanding): Define os objetivos de negócio e as variáveis-chave a serem previstas, formulando "perguntas afiadas" (sharp questions) para guiar o projeto. Identifica as fontes de dados necessárias e estabelece métricas de sucesso (SMART) para medir os resultados.
- Aquisição e Entendimento dos Dados (Data Acquisition and Understanding): Insere os dados no ambiente de análise e desenvolve um profundo entendimento sobre eles. Avalia a qualidade dos dados por meio de sumarização e visualização, verificando se são adequados para responder à pergunta do negócio. Desenvolve um pipeline de dados para processar novos dados ou atualizar os existentes.
- Modelagem (Modeling): Realiza engenharia de features, divide os dados em conjuntos de treino e teste, constrói modelos usando os dados de treino e avalia o desempenho de vários algoritmos concorrentes. Compara os modelos para determinar a "melhor" solução para a questão do negócio.
- Implantação (Deployment): Os modelos de bom desempenho, são expostos por meio de uma interface de API para que possam ser consumidos por outras aplicações. As previsões podem ser feitas em tempo real ou em lote, dependendo dos requisitos do negócio. Constrói um sistema de telemetria e monitoramento para acompanhar o status e o desempenho do modelo em produção.
- Aceitação do Cliente (Customer Acceptance): Finaliza as entregas do projeto, com o cliente validando se o sistema atende às suas necessidades de negócio e se o modelo responde às perguntas com precisão aceitável. A documentação é concluída e revisada, e o projeto é formalmente entregue à equipe responsável pela operação e manutenção do sistema em produção.

5.3.3. SEMMA

O SEMMA, um acrônimo para Sample, Explore, Modify, Model, and Assess, é uma metodologia de mineração de dados desenvolvida pelo SAS Institute. Ele funciona como a estrutura principal da ferramenta SAS Enterprise Miner, de modo que a própria interface do software é organizada em torno dessas cinco fases. A metodologia orienta o analista por uma sequência lógica, onde cada fase agrupa um conjunto de ferramentas específicas (chamadas de

"nós") que o software oferece para executar a tarefa correspondente. Por ter um foco maior na execução técnica dentro da ferramenta, o SEMMA se concentra mais nas etapas de modelagem do que no ciclo de vida completo do projeto, como o entendimento do negócio [19].

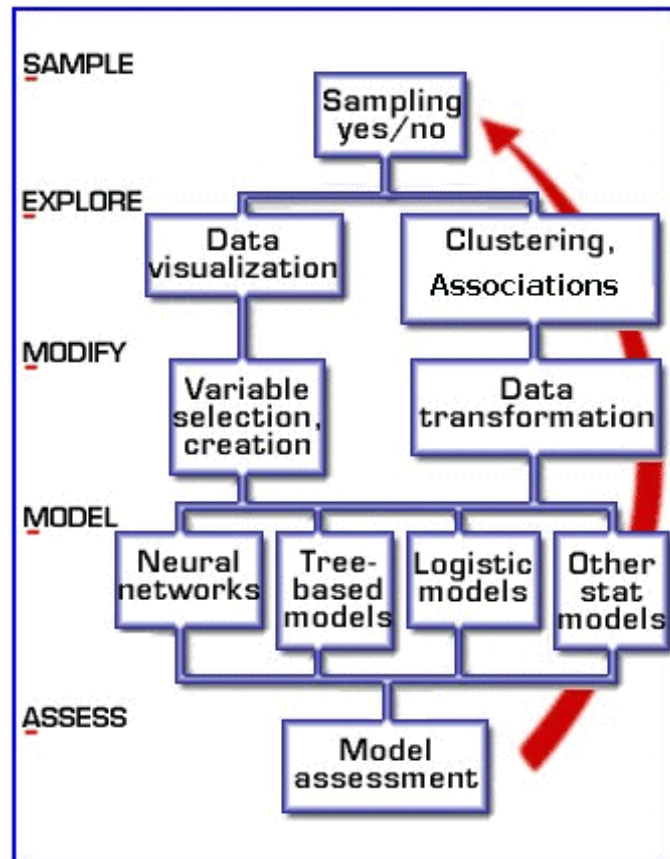


Figura 4: Diagrama SEMMA

Fonte: [19]

As fases do SEMMA são [19]:

- Amostragem (Sample): Seleciona uma amostra representativa dos dados disponíveis, grande o suficiente para conter padrões significativos, mas pequena o bastante para ser manipulada de forma eficiente.
- Exploração (Explore): Analisa estatisticamente os dados, utilizando visualizações e medidas descritivas para identificar tendências, anomalias e relações relevantes.
- Modificação (Modify): Realiza transformações nos dados, como criação de variáveis derivadas, tratamento de valores ausentes e normalização, ajustando o conjunto para a modelagem.

- Modelagem (Model): Modela os dados usando as ferramentas analíticas para procurar uma combinação de dados que preveja de forma confiável um resultado desejado.
- Avaliação (Assess): Avalia o desempenho e a robustez dos modelos gerados com base em métricas apropriadas

5.3.4. KDD Process

O KDD (Knowledge Discovery in Databases) Process é uma estrutura ampla que descreve o processo de extração de conhecimento útil a partir de dados. É um processo interativo e iterativo, envolvendo múltiplos passos e decisões por parte do usuário. Ao contrário de uma visão focada apenas na tecnologia, o KDD enfatiza que a etapa de mineração de dados (a busca por padrões) é apenas uma parte de um ciclo maior, e que etapas como o entendimento do domínio e a preparação dos dados são igualmente, ou até mais, importantes para o sucesso de um projeto na prática [20].

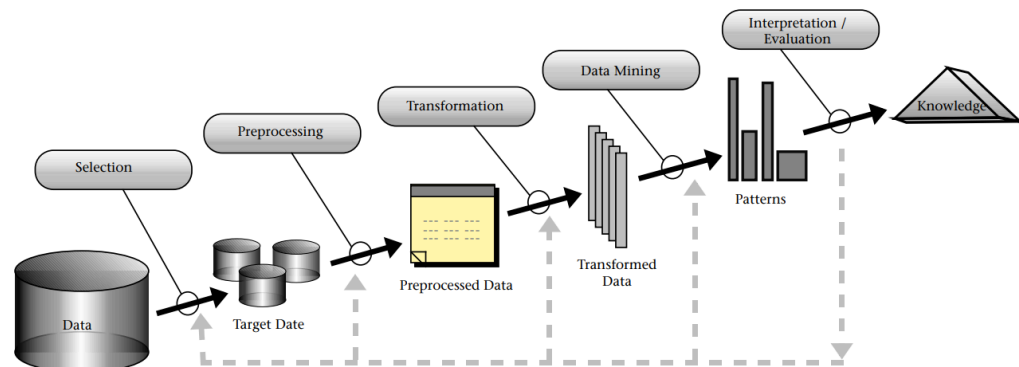


Figura 5: Diagrama KDD Process

Fonte: [20]

As fases do KDD são [20]:

- Entendimento do Domínio e do Objetivo: Desenvolver uma compreensão da área de aplicação e do conhecimento prévio, e identificar o objetivo do processo KDD sob a perspectiva do cliente.
- Criação do Conjunto de Dados-alvo: Selecionar o conjunto de dados ou focar em um subconjunto de variáveis ou amostras de dados sobre os quais a descoberta será realizada.

- Limpeza e Pré-processamento: Executar operações para tratar os dados brutos, como remover ruídos, definir estratégias para lidar com dados ausentes e considerar informações de sequência temporal.
- Redução e Projeção dos Dados: Encontrar características úteis (features) para representar os dados, dependendo do objetivo. Isso pode envolver a redução da dimensionalidade ou a transformação de variáveis para encontrar representações mais eficientes.
- Definição da Tarefa de Mineração: Conectar os objetivos do projeto (definidos na primeira etapa) a uma tarefa específica de mineração de dados, como classificação, regressão, sumarização ou clusterização.
- Seleção do Algoritmo e Hipótese: Realizar a análise exploratória, escolher os algoritmos e métodos que serão usados para buscar os padrões e decidir quais modelos e parâmetros são mais apropriados para os dados e os critérios do projeto.
- Mineração de Dados: Executar efetivamente o algoritmo escolhido para buscar padrões de interesse nos dados, gerando representações como regras de classificação, árvores de decisão ou clusters.
- Interpretação dos Padrões Extraídos: Analisar os padrões e modelos extraídos, muitas vezes com o auxílio de técnicas de visualização. Esta etapa é iterativa e pode levar ao retorno a qualquer uma das etapas anteriores para refinamento.
- Uso do Conhecimento Descoberto: Aplicar o conhecimento adquirido, seja usando-o diretamente, incorporando-o em outro sistema, ou simplesmente documentando-o e reportando-o às partes interessadas.

5.3.5. Domino Data Science Lifecycle

O Domino Data Science Lifecycle é uma metodologia de ciclo de vida desenvolvida pela Domino Data Lab para gerenciar projetos de ciência de dados em escala. Inspirado em frameworks como o CRISP-DM, o modelo se baseia em práticas e realidades observadas em grandes organizações para abranger todo o processo, desde a ideação até o monitoramento contínuo [21].

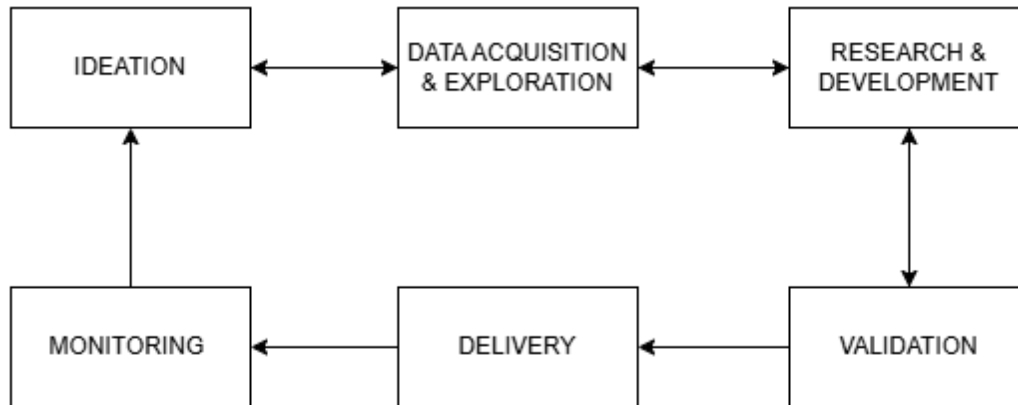


Figura 6: Diagrama Domino Data Science Lifecycle

Adaptado de [21]

As fases do Domino Data Science Lifecycle são [21]:

- **Ideação (Ideation):** Definição do problema de negócio, levantamento de requisitos, análise de viabilidade e cálculo preliminar de retorno sobre investimento (ROI). Também envolve revisão de trabalhos anteriores e seleção do artefato de entrega esperado.
- **Aquisição e Exploração de Dados (Data Acquisition & Exploration):** Se tiver dados disponíveis identifica fontes de dados, com apoio dos stakeholders, senão, toma decisões sobre captura ou compra de dados externos. Então, realiza análises exploratórias e prepara os dados tanto para o projeto atual quanto para reuso em iniciativas futuras.
- **Pesquisa e Desenvolvimento (Research & Development):** Desenvolvimento de modelos em ciclos iterativos, com foco em hipóteses, experimentação, compartilhamento de insights e verificação da necessidade de novos dados ou ferramentas.
- **Validação (Validation):** Avaliação técnica e de negócio dos modelos desenvolvidos. Inclui verificação da reprodutibilidade, revisão de hipóteses, discussão com stakeholders e documentação dos resultados, mesmo em casos que o projeto não vá para produção.
- **Entrega (Delivery):** Implantação do modelo em ambiente de produção. Pode assumir formas diversas, como relatórios, APIs, dashboards ou integração em sistemas. Envolve testes e capacitação dos usuários finais.
- **Monitoramento (Monitoring):** Acompanhamento do desempenho do modelo em uso real, com métricas de valor gerado, performance e uso.

5.4. Metodologias descartadas

5.4.1. Agile Data Science Lifecycle

Embora amplamente citado, não é um ciclo de vida em si, mas sim uma adaptação de frameworks ágeis ao contexto dos dados, sem convergência clara sobre fases de projeto completas.

5.5. Modelo de referência para ciclo de vida de Ciência de Dados

Com base nas metodologias analisadas, propõe-se um modelo de referência para o ciclo de vida de projetos de Ciência de Dados, sintetizando os analisados. É importante ressaltar que este modelo não tem a pretensão de substituir as metodologias de mercado existentes, nem de propor um novo padrão para a indústria. Trata-se de um artefato construído especificamente para este trabalho, com o objetivo de sintetizar as etapas comuns da literatura e fornecer uma estrutura padronizada para que as assertivas sobre o uso de ontologias pudessem ser elaboradas e validadas. O modelo reconhece o caráter iterativo do processo, permitindo o retorno a fases anteriores sempre que necessário, e admite que nem todas as etapas se aplicam a todos os projetos, dependendo do tipo de entrega e do contexto.

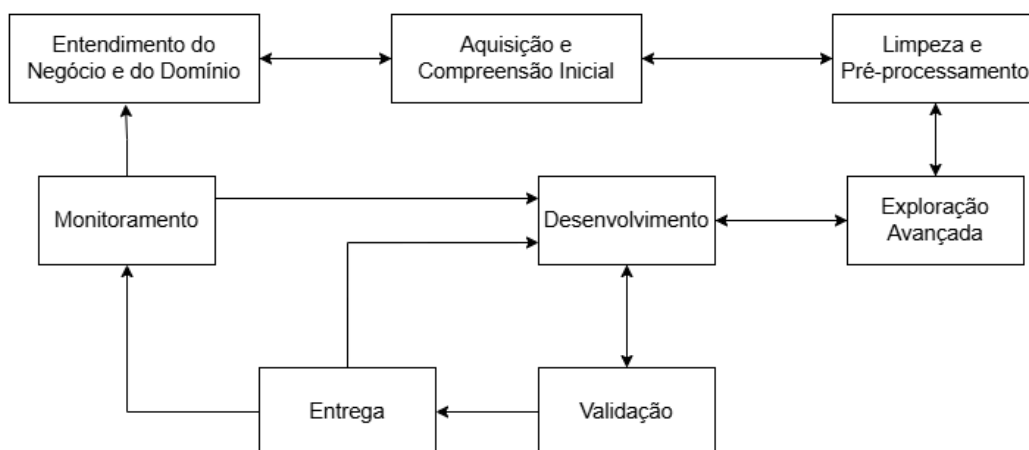


Figura 7: Diagrama modelo de referência

As fases do modelo de referência são:

- Entendimento do Negócio e do Domínio: Etapa inicial focada em compreender as necessidades do negócio e traduzi-las em objetivos claros de Ciência de Dados. Envolve definição de escopo, hipóteses iniciais e métricas de sucesso, além do alinhamento das expectativas com as partes interessadas.
- Aquisição e Compreensão Inicial: Identificação de fontes, coleta e carga dos dados brutos. Inclui uma análise inicial do formato, tipos, estatísticas básicas e distribuição de valores faltantes/anômalos.
- Limpeza e Pré-processamento: Tratamento de inconsistências, valores ausentes, outliers e integração de diferentes fontes. Engloba também transformações iniciais como normalização.
- Exploração Avançada: Com os dados limpos, inicia-se uma análise mais aprofundada para investigar padrões, comportamentos e possíveis relações entre variáveis. Essa etapa ajuda a gerar novas hipóteses, confirmar ou ajustar as hipóteses iniciais e aprofundar o entendimento do contexto, orientando as decisões nas fases seguintes.
- Desenvolvimento: A solução é construída com base nos objetivos do projeto. Isso pode envolver desde análises estatísticas e dashboards até a aplicação de algoritmos de aprendizado de máquina. Também pode incluir a criação de novas variáveis que melhorem a representação da informação e a qualidade dos resultados.
- Validação: Verificação dos resultados com base nas métricas definidas na fase inicial e validação qualitativa dos resultados com as partes interessadas.
- Entrega: A solução é disponibilizada no formato mais apropriado ao seu uso final, como relatórios, dashboards, APIs, aplicações ou integrações com sistemas existentes. Em seguida, realizam-se testes finais de aceitação com os usuários finais para verificar se tudo funciona conforme o esperado. Por fim, são fornecidas instruções e treinamentos aos usuários finais para garantir o uso correto da solução e sua adoção eficaz.
- Monitoramento: Uma solução de Ciência de Dados só é bem-sucedida se continuar gerando impacto depois de entregue. Por isso, é essencial acompanhar continuamente seus efeitos, combinando medidas técnicas, como estabilidade e desempenho, com indicadores que reflitam a qualidade e o sentido prático dos resultados ao longo do tempo. Um bom monitoramento permite identificar desvios, evitar usos inadequados e

ajustar a solução sempre que necessário, garantindo que ela permaneça útil, confiável e alinhada ao propósito original.

6. Ontologias

Ontologia é uma especificação explícita e formal de uma conceitualização do mundo real, isto é, um vocabulário organizado que define conceitos, relações e restrições usadas para representar um domínio de conhecimento. A definição clássica de Gruber é: “ontologia é uma especificação de uma conceitualização” [6]. Na prática, torna o significado dos dados explícito, compartilhável e utilizável por humanos e máquinas, reduzindo ambiguidade e promovendo a interoperabilidade entre sistemas.

Distinguem-se dois níveis importantes de ontologia: as ontologias de domínio, que descrevem conceitos específicos e detalhes operacionais de uma área (por exemplo, medicina ou finanças), e as ontologias fundamentais (foundational ontologies), que oferecem categorias gerais e estruturais, independentes de domínio, usadas como base conceitual para outras ontologias.

Neste trabalho, optou-se por focar exclusivamente nas ontologias de domínio, não sendo abordada a aplicação de ontologias fundamentais.

6.1. A Estrutura de uma Ontologia

Toda ontologia é construída a partir de um conjunto de elementos fundamentais que, juntos, modelam um domínio de conhecimento [6]:

- **Classes (Conceitos):** Categorias abstratas que representam os tipos de coisas que existem no domínio (exemplo: Pessoa, Evento, Doença).
- **Instâncias (Indivíduos):** Ocorrências concretas ou exemplos dessas classes (exemplo: Joana como uma instância de Pessoa).
- **Propriedades (Relações/Atributos):** Conectam instâncias e classes, descrevendo seus atributos e como se relacionam entre si (exemplo: temData, ocorreuEm).
- **Axiomas e Restrições:** regras lógicas que governam classes e propriedades, definindo o que é possível e o que não é.(exemplo: um Carro deve ter exatamente 4 rodas (cardinalidade), uma Pessoa não pode ser uma Doença (disjunção))
- **Vocabulário e documentação:** rótulos e descrições em linguagem natural que facilitam entendimento humano.

6.2. Tecnologias para Ontologias

Para passar a estrutura conceitual de uma ontologia em um modelo computacional são usadas diversas tecnologias:

- RDF (Resource Description Framework): A base para representar dados em formato de grafo, utilizando triplas (sujeito-predicado-objeto) [22].
- RDFS (RDF Schema) e OWL (Web Ontology Language): Linguagens que permitem definir a lógica da ontologia. Enquanto o RDFS cria hierarquias básicas, o OWL, baseado em lógicas descritivas, permite a criação de axiomas mais expressivos [23].
- Raciocinadores (Reasoners): Ferramentas como Pellet ou Hermit processam a lógica da ontologia. Eles verificam a consistência do modelo e, mais importante, inferem novos fatos que não foram explicitamente declarados. Por exemplo, se "Pessoa \equiv SerHumano" e "Maria é Pessoa", o raciocinador automaticamente conclui que "Maria é SerHumano".
- SPARQL: Uma linguagem de consulta para extrair dados do grafo de conhecimento, incluindo as inferências geradas pelo raciocinador [24].
- Protégé: Um ambiente gráfico padrão para criar, editar e visualizar ontologias OWL [25].

6.3. Web Semântica

Web Semântica é um conjunto de padrões, formatos e boas práticas que permite publicar e conectar dados na web de forma que máquinas entendam seu significado, não apenas sua sintaxe [26]. Ela fornece formatos padronizados para representar afirmações sobre recursos (RDF), esquemas e vocabulários para organizar esses termos (RDFS, OWL) e uma linguagem de consulta para extrair informação e inferências (SPARQL).

A Web Semântica não é sinônimo de ontologia, mas é a infraestrutura prática que torna as ontologias efetivas em escala distribuída. Ontologias formalizam vocabulários e axiomas, a Web Semântica oferece meios padronizados para codificá-los, publicá-los e ligá-los via URIs⁵, além de ferramentas que executam raciocínio automático e consultas sobre esses grafos.

⁵ URI é um identificador único utilizado para nomear e localizar recursos na web.

6.4. Exemplo de Ontologia de Domínio: OWL-Time

Um exemplo prático de ontologia de domínio é a Time Ontology in OWL (OWL-Time), um padrão mantido pelo W3C [27]. Seu objetivo é fornecer um vocabulário formal e explícito para descrever as propriedades temporais de recursos, permitindo a interoperabilidade entre sistemas que precisam representar e raciocinar sobre o tempo.

Em sua essência, a ontologia é construída em torno de classes centrais que definem entidades temporais. As classes principais, como `time:Instant` (um ponto no tempo sem duração) e `time:Interval` (um período com duração), herdam de uma superclasse `TemporalEntity`. A Figura 8 ilustra essa hierarquia e as propriedades que conectam esses conceitos.

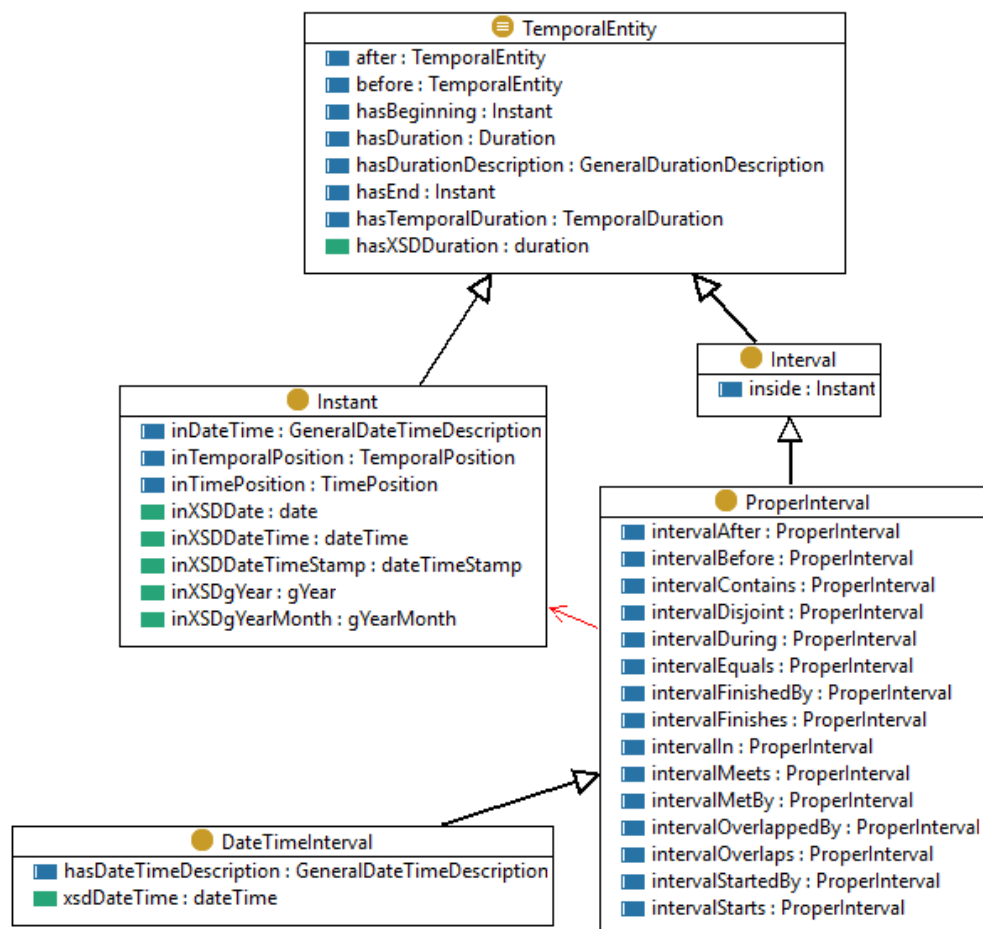


Figura 8: Modelo central de entidades temporais

Fonte: [27]

Como detalhado no diagrama, a OWL-Time oferece um rico conjunto de propriedades para conectar essas entidades. Isso inclui não apenas propriedades básicas de data e hora (como `inXSDDateTime`), mas

principalmente propriedades de relação temporal. A classe `ProperInterval`, por exemplo, define um vocabulário completo para as 13 relações de Allen, um conjunto de relações temporais fundamentais proposto por James F. Allen [28] para descrever como dois intervalos de tempo podem se relacionar. Essas relações incluem, entre outras, `intervalBefore` (um intervalo ocorre antes do outro), `intervalAfter` (um intervalo ocorre depois do outro), `intervalContains` (um intervalo contém o outro) e `intervalMeets` (um intervalo termina exatamente quando o outro começa). Elas permitem modelar com precisão interações temporais entre eventos.

A estrutura formal da OWL-Time a torna fundamental para resolver desafios práticos. Sua lógica permite que sistemas de agendamento, por exemplo, detectem automaticamente conflitos de horários e infiram a ordem de eventos em uma linha do tempo, mesmo quando essa informação não está explicitamente declarada. Além disso, a ontologia se torna um componente chave para a interoperabilidade, permitindo modelar calendários complexos e fornecer o contexto temporal necessário para serviços da Web Semântica.

7. Aplicação de Ontologias no Ciclo de Vida de Ciência de Dados

Nesta seção, é explorado como as ontologias podem beneficiar cada fase do modelo de referência de Ciência de Dados proposto (Seção 5.5). O foco está na aplicação de ontologias de domínio, enquanto [29] discute os benefícios de se utilizar ontologias fundamentais para a mineração de dados.

Para habilitar esses benefícios, duas ações principais são executadas nas fases iniciais do projeto. Primeiro, na fase de Entendimento do Negócio, a ação é construir ou reutilizar uma ontologia de domínio. A equipe cria ou adapta um modelo formal que representa o conhecimento do negócio, podendo reutilizar ontologias existentes. Em seguida, na fase de Aquisição e Compreensão Inicial, a ação principal é criar um mapeamento da fonte para a ontologia. Este trabalho consiste em documentar formalmente como os dados brutos se conectam aos conceitos e propriedades da ontologia (ex: `fonte.reviews.rating` → `Avaliacao.nota`)

Entendimento do Negócio e do Domínio

- Uma ontologia força as partes interessadas a definir formalmente os termos de negócio. Isso elimina ambiguidades desde o início.
 - O processo de criação da ontologia exige que todos cheguem a um consenso sobre o significado exato dos conceitos, eliminando ambiguidades que poderiam levar a erros de interpretação.
 - Exemplo: O time de marketing fala em "produto popular". A ontologia formaliza Cliente e Produto e define ProdutoPopular como subclasse de Produto com a regra (axioma) de ter `numeroDeVendas > 1000` e `avaliacaoMedia > 4.5`, assim, a ambiguidade acaba.
- Uma ontologia alinha vocabulário entre analistas, especialistas e stakeholders.
 - A ontologia centraliza e padroniza a terminologia do projeto. Termos que diferentes áreas usam de formas distintas são mapeados para um único conceito formal.
 - Exemplo: A equipe de logística chama de "SKU", o desenvolvedor de `product_id`, e o analista de item. A ontologia unifica todos sob o conceito `Produto.identificadorUnico`, com uma descrição clara e semântica.
- Uma ontologia oferece um modelo visual das entidades chave do negócio e suas relações, ajudando a definir o escopo e a identificar os fatores que dão valor ao projeto.
 - Ao representar o conhecimento de negócio em um mapa visual, a ontologia define o domínio de forma clara e objetiva, destacando as relações mais críticas entre as entidades. Isso permite concentrar os esforços do projeto nos fatores de maior valor.
 - Exemplo: O diagrama mostra que Cliente pode fazer múltiplos Pedidos, Pedido contém Produtos, e Produto pertence a Categoria e Marca. Com isso fica claro que uma recomendação de produto exige integrar dados de Pedidos, Histórico do Cliente e Avaliação do Produto.
- As relações e axiomas numa ontologia ajudam a formular hipóteses mais ricas e estruturadas.
 - As conexões lógicas definidas no modelo sugerem caminhos de investigação e hipóteses de negócio que podem ser testadas com dados.

- Exemplo: A ontologia define o axioma $\text{ClienteVIP} \equiv \text{Cliente} \text{ and } (\text{valorTotalPedidos} > 10000)$. A partir disso, a equipe formula a hipótese: “Clientes VIP realizam mais que o triplo de compras que Clientes não VIP?”.

Aquisição e Compreensão Inicial

- Uma ontologia define os conceitos importantes, transformando uma busca cega em uma investigação direcionada.
 - A ontologia funciona como uma "lista de compras" para o cientista de dados. Em vez de explorar centenas de tabelas e arquivos aleatoriamente, ele busca ativamente nas fontes brutas os dados que correspondem aos conceitos já definidos como essenciais para o projeto.
 - Exemplo: A ontologia requer os conceitos Cliente, Produto, Pedido e Avaliacao. Com essa lista, o cientista de dados foca sua busca nos bancos de dados por tabelas com nomes como users, products, orders e reviews, economizando tempo e esforço.

Limpeza e Pré-processamento

- Uma ontologia funciona como o modelo canônico para a integração dos dados.
 - Aqui o mapeamento já serve como guia para a transformação prática, colunas são renomeadas, tipos ajustados e tabelas unificadas de acordo com a ontologia, gerando um dataset consistente.
 - Exemplo: Dados de clientes vêm de dois sistemas: o de vendas (com a coluna sales.client_id) e o de suporte (com a coluna tickets.customer_num). O script de limpeza, então, usa a ontologia como guia para integrar essas três fontes, criando para cada cliente um perfil unificado.
- Os axiomas de uma ontologia servem como regras para a validação semântica.
 - As regras lógicas da ontologia são usadas para validar dados antes da análise, sinalizando quando algo não faz sentido conceitual ou viola a lógica do domínio.
 - Exemplo: Um axioma estabelece que a dataDeCriacao de um Pedido deve ser posterior ou igual à dataDeCadastro do Cliente

que o realizou. Durante a limpeza, o sistema identifica pedidos com data anterior ao cadastro do cliente, o que viola o axioma e indica erro de ingestão ou inconsistência no histórico.

Exploração Avançada

- Uma ontologia contextualiza a análise, ajudando a separar correlações enganosas de relações de negócio significativas.
 - Ao fornecer o contexto semântico do domínio, a ontologia permite que o cientista de dados interprete as correlações estatísticas dentro de um quadro de conhecimento de negócio, distinguindo entre associações espúrias e relações causalmente ou logicamente relevantes.
 - Exemplo: Uma análise exploratória pode apontar correlação entre o último dígito do id do cliente e o valor médio dos pedidos. A ontologia mostra que não existe nenhuma relação semântica entre identificadores técnicos e comportamento de compra, classificando essa associação como espúria. Em contrapartida, uma correlação entre Categoria do Produto e frequência de Pedido é validada como significativa, já que a ontologia modela explicitamente que todo Pedido contém Produtos que pertencem a uma Categoria.
- Uma ontologia permite explorar conexões que não seriam visíveis apenas pela estatística, orientando a análise para padrões relevantes de negócio.
 - O modelo de dados tradicional (tabelas) dificulta a descoberta de relações indiretas. A representação em grafo da ontologia permite que o analista caminhe pelas conexões.
 - Exemplo: Explorando o grafo, um analista descobre um caminho: um Cliente faz um Pedido que contém um Produto de uma determinada Marca. O analista percebe que muitos desses clientes, em Pedidos futuros, compram outros Produtos da mesma Marca, mesmo que de Categorias diferentes. Isso revela um padrão de "lealdade à marca" que pode ser usado para enriquecer as recomendações.

Desenvolvimento

- Uma ontologia orienta a engenharia de features, ajudando a selecionar variáveis de forma inteligente.

- A ontologia destaca quais conceitos e relações são mais importantes. Isso foca o esforço na criação de features que carregam significado para o negócio, em vez de testar combinações aleatórias de colunas.
- Exemplo: A ontologia modela a relação entre um Cliente e seus Pedidos. Isso inspira o cientista de dados a criar features como `frequencia_de_pedidos_mensal`, `valor_medio_por_pedido` e `numero_de_categorias_distintas_compradas`, todas baseadas diretamente na estrutura do modelo.
- Uma ontologia enriquece os dados para modelos de machine learning ao permitir a inferência automática de novas variáveis a partir de seus axiomas.
 - Ferramentas de inferência aplicam axiomas para derivar atributos ausentes ou classificações úteis como novas colunas de entrada.
 - Exemplo: A ontologia define o axioma `ClienteVIP ≡ Cliente and (valorTotalPedidos > 10000)`, o raciocinador marca os Clientes que cumprem essa condição. Assim, nos dados de entrada do modelo é criada a variável booleana `ClienteVIP`.

Validação

- Uma ontologia ajuda a verificar se os resultados respeitam restrições de negócio e se fazem sentido conceitual.
 - Na avaliação dos modelos, confronta-se os resultados com os axiomas da ontologia para checar a coerência.
 - Exemplo: O modelo de machine learning sugere vendas de um produto em uma categoria que não existe para esse produto, se a ontologia diz que aquele produto só pertence a uma categoria fixa, essa previsão é inconsistente.
- Uma ontologia permite validação cruzada com especialistas, pois os resultados podem ser mapeados diretamente em conceitos da ontologia.
 - Como os insights podem ser expressos em termos bem definidos, é mais fácil discutir conclusões com os especialistas.
 - Exemplo: O cientista de dados se reúne com o gerente de produtos. Em vez de perguntar "O que você acha da recomendação do produto 321 para o cliente 123?", ele usa a ontologia para agregar os resultados e pergunta: "Nosso modelo está recomendando massivamente Produtos da Marca Y para

Clientes que foram inferidos como CompradoresTecnologicos. Isso faz sentido com a nossa estratégia de marca ou o modelo está com um viés indesejado?". Essa pergunta, baseada nos conceitos da ontologia, permite uma validação de negócio muito mais rica.

- Uma ontologia fornece uma camada de validação lógica para as previsões de um modelo, identificando resultados que violam suas regras de negócio e auxiliando na explicabilidade.
 - Ao comparar previsões com as restrições definidas na ontologia, é possível identificar saídas incoerentes com o domínio de negócio. Isso aumenta a confiabilidade do sistema e fornece justificativas claras sobre por que certas previsões não devem ser aceitas.
 - Exemplo: Um modelo de recomendação sugere que um Pedido seja criado sem Produto associado. A ontologia possui o axioma “todo Pedido deve conter pelo menos um Produto”. A violação desse axioma sinaliza que a previsão é incorreta, permitindo que a equipe descarte o resultado e registre a justificativa baseada em regra de negócio.

Entrega

- Uma ontologia atua como documentação, explicando de forma formal e acessível os conceitos usados no projeto.
 - Ela atua como documentação viva que registra entidades, propriedades e restrições de forma clara, servindo como referência tanto para especialistas técnicos quanto para as equipes de negócio.
 - Exemplo: O diagrama ontológico mostra como Cliente, Pedido, Produto, Categoria e Marca se relacionam. Esse material é anexado à entrega final do projeto para que novos membros entendam rapidamente o domínio.
- Uma ontologia garante um vocabulário consistente em todos os produtos finais (relatórios, dashboards, APIs).
 - A ontologia padroniza os termos usados em relatórios e sistemas, evitando ambiguidades.
 - Exemplo: Todos os relatórios e dashboards usam Categoria para agrupar produtos, em vez de alternar entre “tipo de produto”, “linha” ou “segmento”.

- Uma ontologia facilita integração com outros sistemas, pois o modelo ontológico serve como referência semântica.
 - Ao expor conceitos bem definidos, a ontologia torna os dados auto-descritivos e padronizados. Isso reduz o esforço de integração, elimina ambiguidades e permite que sistemas externos entendam o significado dos dados.
 - Exemplo: Um sistema de gestão de estoque usa `item_code`, que é mapeado diretamente para `Produto.identificadorUnico` na ontologia, garantindo consistência. Já a API de recomendação retorna uma estrutura de dados informando que o item é instância da classe `Produto` e pertence a uma `Marca`. Assim, o sistema de estoque pode consumir essa API e entender automaticamente do que se trata, sem precisar de uma reunião para alinhar o significado dos campos.

Monitoramento

- Uma ontologia funciona como referência contínua para detectar desvios conceituais.
 - Durante a operação, novos dados podem não se encaixar no modelo definido. Isso sinaliza possíveis problemas de qualidade ou mudanças no negócio.
 - Exemplo: A ontologia tem um axioma que diz que todo `Pedido` deve conter pelo menos um `Produto`. Um novo sistema de "assinaturas de caixas misteriosas" é lançado e começa a gerar `Pedidos` sem uma lista de produtos explícita. O sistema de monitoramento, que valida os dados contra a ontologia, imediatamente sinaliza essa violação, alertando que o conceito de `Pedido` evoluiu e a ontologia precisa ser atualizada.
- Uma ontologia permite auditoria semântica, explicando decisões do modelo com base em axiomas e relações.
 - A ontologia registra as regras que orientam inferências e previsões. Isso possibilita rastrear a lógica de cada decisão e justificar saídas para clientes ou funcionários.
 - Exemplo: Um cliente reclama de uma recomendação de roupa feminina que parece aleatória. A auditoria, no entanto, usa a ontologia para explicar a decisão. Ela mostra que o sistema inferiu o perfil do cliente como `CompradorDePresentes`, com base em

um axioma que analisa seu histórico de entregas para diferentes endereços. O item recomendado, por sua vez, estava classificado como PresentePopular. A recomendação, portanto, não foi um erro, mas o resultado de conexão entre o perfil "comprador de presentes" e itens populares para essa finalidade.

- Uma ontologia ajuda a identificar o surgimento de novas categorias, entidades ou relações que não se encaixam no modelo de conhecimento atual, sinalizando como o sistema precisa evoluir.
 - A ontologia representa o conhecimento atual. Ao analisar dados que "não se encaixam" no modelo existente, é possível descobrir novas tendências e conceitos de negócio.
 - Exemplo: Um modelo de análise de texto que lê as avaliações dos clientes começa a identificar um cluster frequente de termos como "sustentável", "ecológico", "reciclado". Esses termos não se mapeiam a nenhuma Categoria existente na ontologia. O sistema de monitoramento sinaliza isso como a provável emergência de um novo conceito, levando a equipe a criar a Categoria "Produtos Sustentáveis" e a retrainar o modelo de recomendação para alavancar essa nova tendência.

8. Entrevistas

Para a validação das assertivas, buscou-se identificar profissionais com experiência comprovada em Ciência de Dados. O processo de validação consistiu no envio de um email convite contendo o formulário para avaliação das assertivas propostas. Após o recebimento das respostas, foi marcada uma entrevista com o objetivo de aprofundar a identificação de pontos fortes, fracos e facilitadores da abordagem. Essa etapa foi realizada de forma síncrona com três participantes (E1, E2 e E3), enquanto o especialista E4 contribuiu com respostas estruturadas e detalhadas enviadas por email.

8.1. Participantes

A Tabela 1 apresenta o perfil dos especialistas selecionados, detalhando suas áreas de atuação acadêmica e profissional, bem como o tempo de experiência.

Especialista	Perfil Principal	Anos de Experiência
E1	Mestrando Ciência de Dados	3
E2	Engenheiro de Software e Cientista de Dados	9
E3	Professora e Pesquisadora em Ontologias e Ciência de Dados	23
E4	Arquiteto de Enterprise com foco em Cloud Data Analytics e Engenharia de Dados	10

Tabela 1: Especialistas entrevistados

8.2. Formulário

Para a validação das assertivas, foi elaborado um formulário estruturado, disponível no Apêndice 1. Ele foi desenhado para coletar a percepção de especialistas sobre os benefícios de cada assertiva que conecta ontologias a cada etapa do ciclo de vida de Ciência de Dados.

8.3. Email

O convite formal para a validação foi por meio de um email, disponível no Apêndice 2. O texto apresenta ao especialista, os objetivos do trabalho e o formato da validação, solicitando a participação do profissional e sugerindo os próximos passos, como o preenchimento do formulário e a disponibilidade para uma entrevista.

9. Resultados

Esta seção apresenta a consolidação dos dados obtidos durante a validação. A Tabela 2 exibe a síntese quantitativa das respostas objetivas coletadas através do formulário. As seções subsequentes detalham a análise qualitativa referente aos pontos fortes, fracos e facilitadores apontados pelos participantes. Os comentários detalhados apresentados pelos especialistas encontram-se no Apêndice 3.

9.1. Síntese das respostas objetivas

A seguir, a Tabela mostra a distribuição dos níveis de concordância dos especialistas em relação às vinte assertivas propostas, permitindo uma visualização clara da aderência do modelo teórico à visão prática dos profissionais.

Assertiva	Concordo totalmente	Concordo	Nem concordo nem discordo	Discordo	Discordo totalmente
1	25%	75%			
2	75%	25%			
3	25%	50%	25%		
4	25%	75%			
5	25%	50%		25%	
6	75%	25%			
7	50%	50%			
8	50%	25%	25%		
9	25%	50%	25%		
10	75%	25%			
11	25%	75%			

12	75%	25%			
13	25%	50%		25%	
14	50%	50%			
15	50%	50%			
16	50%		25%	25%	
17	25%	75%			
18	75%	25%			
19	25%	75%			
20	50%	50%			

Tabela 2: síntese das respostas objetivas

9.2. Pontos mais fortes

Houve consenso entre os especialistas de que o principal benefício da abordagem é a especificação formal do conhecimento. Um dos participantes (E2) contrastou essa característica com a realidade comum dos projetos de Ciência de Dados, onde grande parte do conhecimento do domínio permanece implícita ou distribuída entre stakeholders. A ontologia foi vista como um artefato capaz de transformar esse conhecimento disperso em documentação estruturada, apoiando o alinhamento de expectativas, que é um dos maiores desafios da área.

Além da formalização, a padronização terminológica foi amplamente citada como um ponto forte. O especialista E1 destacou que a fase de entendimento do domínio é a mais beneficiada pelo uso de ontologias, pois a ontologia estabelece um padrão que facilita a comunicação entre equipes distintas reduzindo significativamente o risco de retrabalho futuro. Comentários semelhantes apareceram em outras respostas, enfatizando que a ontologia promove alinhamento entre equipes técnicas e de negócio e evita inconsistências conceituais que surgem quando cada área utiliza vocabulário próprio.

Por fim, o especialista E4 expandiu a visão dos benefícios para além da documentação, citando a integração semântica e a capacidade de validação lógica como diferenciais técnicos. Segundo ele, a ontologia não apenas

padroniza a linguagem humana, mas habilita a inferência automática e validações avançadas, permitindo conectar sistemas heterogêneos e eliminar ambiguidades que o simples cruzamento de tabelas não resolveria.

9.3. Pontos mais fracos

A principal limitação apontada pelos especialistas foi o custo de construção e manutenção. Para E1, o esforço necessário para construir uma ontologia do zero pode ser desproporcional em projetos rápidos ou de menor complexidade, especialmente quando não há ontologias de domínio consolidadas para reutilização. O E2 questionou o trade-off entre esse esforço de modelagem formal e entregas ágeis. Em muitos cenários, modelos de machine learning fornecem valor mesmo sem o suporte de ontologias, e isso levanta a dúvida sobre quando o esforço extra se justifica.

Outro ponto crítico levantado refere-se à curva de aprendizado e à disponibilidade de competências. O especialista E4 enfatizou que o uso de ontologias exige o domínio de tecnologias complexas e ainda raras no mercado, como OWL, RDF, SPARQL e SHACL. Essa barreira técnica vai de encontro com um ponto levantado pelo E2 sobre a responsabilidade prática pelo artefato dentro das equipes: se essa função caberia ao engenheiro de requisitos, ao cientista de dados ou a um novo papel especializado, visto que muitas equipes atuais não possuem profissionais com esse perfil.

Adicionalmente, a governança foi citada como um desafio que supera a complexidade técnica. O especialista E4 alertou que a manutenção de ontologias é mais exigente do que a de modelos relacionais ou data lakes tradicionais, demandando controle rígido de alterações que quebram compatibilidade (breaking changes), versionamento semântico e sincronização constante com APIs e esquemas físicos. Em cenários de dados altamente dinâmicos ou não estruturados, essa rigidez pode se tornar um obstáculo operacional.

9.4. Facilitadores

Os especialistas indicaram que a adoção é facilitada quando existem recursos reutilizáveis. A disponibilidade de ontologias de domínio já consolidadas para macrodomínios foi apontada pelo especialista E1 como essencial para reduzir o esforço inicial de modelagem, evitando que cada projeto precise começar do zero. A presença de alguém com conhecimento do domínio e

experiência prévia com ontologias na equipe também foi apontada como fator facilitador.

Outro facilitador mencionado é a necessidade de uma comunicação clara sobre o valor da abordagem. E2 observou que, à primeira vista, um diagrama de classes pode parecer equivalente, o que torna fundamental explicar o diferencial das ontologias de forma objetiva, principalmente os recursos de restrições formais e inferência automática.

Por fim, questões de operacionalização e o uso de ferramenta adequado foram destacados como essenciais para a viabilidade cotidiana da abordagem. O especialista E4 apontou que glossários corporativos padronizados facilitam a incorporação da ontologia ao fluxo de trabalho ao alinhar conceitos entre as áreas. Ele também ressaltou que ferramentas de validação semântica, como o SHACL, permitem integrar regras de negócio diretamente aos pipelines, automatizando verificações e reduzindo a necessidade de limpeza manual dos dados, o que torna o benefício tangível já nas etapas iniciais do processo.

9.5. Adicionais

Além da avaliação das assertivas propostas, os especialistas trouxeram à tona conceitos e lacunas que não haviam sido mapeados no modelo de referência inicial, sugerindo ampliações importantes para a teoria aplicada. O especialista E4 introduziu a distinção crítica entre drift estatístico e drift semântico. Enquanto o monitoramento tradicional de Machine Learning foca na mudança das distribuições dos dados (data drift), o drift semântico ocorre quando a regra de negócio ou o significado do conceito muda, mas os dados continuam estatisticamente parecidos. Por exemplo, se a definição de “Cliente Ativo” muda de “compra em 30 dias” para “login em 7 dias”, um monitor estatístico pode não identificar a mudança. Nesse cenário, uma ontologia permite identificar discrepâncias quando indivíduos classificados operacionalmente como “Cliente Ativo” não satisfazem os axiomas que o definem, evidenciando uma divergência semântica entre a regra de negócio formalizada e o comportamento efetivamente adotado pelo sistema. Esse tipo de lacuna conceitual não é, em geral, abordado pelas ferramentas atuais de MLOps⁶, que se concentram predominantemente em métricas estatísticas e de desempenho do modelo.

⁶ Conjunto de práticas para automatizar e gerenciar o ciclo de vida de modelos de machine learning, incluindo desenvolvimento, implantação, monitoramento e governança.

Outra adição relevante foi a identificação da governança de vocabulário (stewardship) como um pré-requisito mandatário. Foi apontado que a ontologia, por si só, é apenas um artefato técnico, ela não elimina ambiguidades se não houver um processo político de gestão de termos e engajamento real dos stakeholders. O modelo proposto, portanto, carece de uma etapa explícita de governança que anteceda ou acompanhe a modelagem técnica.

Por fim, um novo caso de uso prático foi apresentado, expandindo a aplicação das ontologias para a função de "Bloqueio de Qualidade" (Gatekeeper). O especialista E4 relatou um caso onde a ontologia não foi usada apenas para exploração ou inferência, mas como uma barreira lógica na entrada do pipeline de treinamento. Nesse cenário, regras de negócio formalizadas (como "Data do Pedido deve ser posterior à Data de Cadastro") impediram a ingestão de dados inconsistentes que passariam por validações de tipo simples, demonstrando um papel ativo da ontologia na garantia da integridade lógica dos dados antes da modelagem.

10. Ajustes

A análise das entrevistas e dos questionários mostrou que algumas assertivas precisavam de refinamento para representar de forma mais fiel as condições reais de uso das ontologias em Ciência de Dados. Os especialistas destacaram que certas formulações utilizavam termos absolutos que não refletem adequadamente a dependência de fatores como governança, maturidade da equipe, qualidade da ontologia e disponibilidade de ferramentas. Esses comentários motivaram a substituição de expressões que sugeriam garantias automáticas por termos que indicam contribuição ou suporte, deixando claro que os benefícios dependem de práticas organizacionais que sustentam o artefato.

Os entrevistados também apontaram a importância de distinguir explicitamente ontologias de artefatos tradicionais, como diagramas de classes. Para evitar interpretações equivocadas, o modelo revisado enfatiza que o diferencial das ontologias não está na representação visual, mas na capacidade de aplicar axiomas, executar validações formais e habilitar inferências lógicas durante o uso, funções que diagramas estáticos não oferecem.

Outro ajuste relevante está relacionado ao custo de adoção. A construção e manutenção de ontologias envolve esforço e demanda competências

especializadas. Assim, o modelo passou a incluir uma recomendação explícita de análise de trade-off, reconhecendo que a abordagem tende a gerar mais valor em contextos de maior complexidade, integração entre múltiplas fontes e necessidade de explicabilidade. Como forma de mitigar o esforço inicial, o modelo também passa a recomendar o uso de glossários padronizados e ontologias de domínio reutilizáveis nas etapas iniciais do processo.

A seguir são apresentadas as assertivas na sua forma final, já incorporando todos os ajustes identificados pelos especialistas.

Assertiva 1 - Uma ontologia contribui para reduzir ambiguidades ao apoiar a definição formal dos termos de negócio, desde que haja participação efetiva dos stakeholders.

Assertiva 2 - Uma ontologia auxilia no alinhamento do vocabulário entre analistas, especialistas e stakeholders, especialmente quando integrada a processos formais de gestão de termos.

Assertiva 3 - Uma ontologia fornece uma representação conceitual das entidades do domínio e suas relações, ajudando a esclarecer escopo e dependências de negócio de maneira mais rica que modelos estáticos.

Assertiva 4 - As relações e axiomas de uma ontologia ajudam a formular hipóteses mais estruturadas, desde que o nível de granularidade seja adequado ao domínio.

Assertiva 5 - Uma ontologia orienta a investigação dos dados ao destacar conceitos relevantes e seus vínculos, direcionando a análise para aspectos significativos do problema.

Assertiva 6 - Uma ontologia pode funcionar como modelo canônico para integração de dados, reduzindo mapeamentos ad hoc e promovendo consistência semântica entre fontes.

Assertiva 7 - Os axiomas de uma ontologia podem ser usados como base para validação semântica dos dados, ampliando o escopo de verificação além de regras técnicas.

Assertiva 8 - Uma ontologia oferece contexto conceitual que auxilia na interpretação dos resultados exploratórios, reduzindo o risco de conclusões baseadas em correlações enganosas.

Assertiva 9 - Uma ontologia permite explorar relações conceituais e conexões de negócio que podem não ser evidentes apenas pela análise estatística tradicional.

Assertiva 10 - Uma ontologia apoia a engenharia de features ao destacar variáveis conceituais relevantes e suas relações, favorecendo escolhas mais fundamentadas.

Assertiva 11 - Uma ontologia pode enriquecer o conjunto de dados ao permitir a derivação de novas variáveis via inferência, quando há motores de raciocínio compatíveis com o volume e a complexidade dos dados.

Assertiva 12 - Uma ontologia contribui para verificar se os resultados respeitam restrições conceituais e regras de negócio, fortalecendo a coerência dos modelos produzidos.

Assertiva 13 - Uma ontologia facilita a revisão com especialistas, pois os resultados podem ser analisados em termos de conceitos formais e bem definidos.

Assertiva 14 - Uma ontologia fornece uma camada adicional de validação lógica que ajuda a identificar previsões incompatíveis com regras de negócio e apoia a explicabilidade do modelo.

Assertiva 15 - Uma ontologia funciona como documentação formal e viva dos conceitos utilizados no projeto, desde que mantida de forma contínua.

Assertiva 16 - Uma ontologia facilita a consistência terminológica entre diferentes produtos finais, apoiando a padronização do vocabulário ao longo do ciclo de vida.

Assertiva 17 - Uma ontologia atua como referência semântica que facilita integrações entre sistemas, quando esses sistemas conseguem adotar ou mapear seus conceitos ao modelo ontológico.

Assertiva 18 - Uma ontologia funciona como referência contínua para detectar mudanças conceituais no domínio, ajudando a identificar possíveis desvios semânticos.

Assertiva 19 - Uma ontologia apoia auditoria semântica ao permitir rastrear decisões de modelos com base em axiomas, relações e restrições conceituais formalizadas.

Assertiva 20 - Uma ontologia contribui para identificar categorias, entidades ou relações emergentes que não se encaixam no modelo existente, indicando pontos de evolução do conhecimento.

11. Conclusão

Nessa última sessão será comentado o quanto os objetivos iniciais foram atingidos e apresentadas sugestões de trabalhos futuros.

11.1. Objetivos x Realidade

A meta deste trabalho foi investigar a aplicação de ontologias ao ciclo de vida da Ciência de Dados com o objetivo de aprimorar consistência, interoperabilidade e explicabilidade. Essa meta foi atingida. A revisão teórica resultou em um conjunto de vinte assertivas que relacionam benefícios do uso de ontologias a cada etapa do processo, e a validação realizada com especialistas mostrou concordância ampla com esses fundamentos. A especificação formal do domínio e o alinhamento da comunicação entre áreas surgiram como os aspectos mais valorizados.

A validação também evidenciou a distância entre os benefícios teóricos e a adoção prática. Os especialistas apontaram barreiras importantes, como o custo inicial percebido, a falta de clareza sobre responsabilidades dentro da equipe e a ausência de ferramentas que integrem a ontologia ao fluxo de trabalho existente.

Esses fatores explicam por que a abordagem, embora promissora, ainda é pouco empregada no mercado.

11.2. Trabalhos futuros

Trabalhos futuros podem aprofundar e ampliar os resultados alcançados neste estudo. Uma primeira direção é a elaboração de um guia prático que descreva, de forma passo a passo, como conduzir a construção incremental de ontologias em projetos de Ciência de Dados. Esse material auxiliaria equipes a incorporar a abordagem de maneira progressiva, reduzindo o esforço inicial e tornando a técnica mais acessível.

Outra possibilidade é a criação de um repositório de ontologias de domínio reutilizáveis para macrodomínios de maior demanda. A disponibilização de modelos prontos e adaptáveis reduziria a necessidade de começar cada projeto do zero, endereçando um dos principais pontos fracos identificados pelos especialistas.

Por fim, sugere-se o desenvolvimento de um estudo de caso aplicado que compare cenários com e sem o uso de ontologias, apresentando métricas sobre redução de retrabalho, inconsistências conceituais e qualidade dos dados. Um estudo dessa natureza permitiria evidenciar o impacto concreto da abordagem na prática profissional, fortalecendo sua adoção em ambientes organizacionais.

12. Referências

- [1] CHAPMAN, Pete et al. **CRISP-DM 1.0: Step-by-step data mining guide**. SPSS inc, v. 9, n. 13, p. 1-73, 2000.
- [2] MICROSOFT. **The Team Data Science Process (TDSP) lifecycle**. Azure/Microsoft-TDSP. GitHub, 2017. Disponível em: <https://github.com/Azure/Microsoft-TDSP/blob/master/Docs/lifecycle-detail.md>. Acesso em: 23 maio 2025.
- [3] MARTINEZ, Iñigo; VILES, Elisabeth; OLAIZOLA, Igor G. Data science methodologies: Current challenges and future approaches. **Big Data Research**, v. 24, p. 100183, 2021.
- [4] DATA SCIENCE PM. **What is a Data Science Workflow?** Disponível em: <https://www.datascience-pm.com/data-science-workflow/>. Acesso em: 12 maio 2025.
- [5] DATA SCIENCE PM. **What is a Data Science Life Cycle?** Disponível em: <https://www.datascience-pm.com/data-science-life-cycle/>. Acesso em: 12 maio 2025.
- [6] GRUBER, Thomas R. A translation approach to portable ontology specifications. **Knowledge acquisition**, v. 5, n. 2, p. 199-220, 1993.
- [7] GHIDALIA, Sarah et al. Combining machine learning and ontology: A systematic literature review. **arXiv preprint arXiv:2401.07744**, 2024.
- [8] CALVANESE, Diego et al. Ontop: Answering SPARQL queries over relational databases. **Semantic Web**, v. 8, n. 3, p. 471-487, 2016.
- [9] W3C. **Achieving Semantic Interoperability Using RDF and OWL**. W3C Best Practices, 2005. Disponível em: <https://www.w3.org/2001/sw/BestPractices/OEP/SemInt/>. Acesso em: 18 maio 2025.

- [10] HOVLAND, Dag et al. OBDA constraints for effective query answering (extended version). **arXiv preprint arXiv:1605.04263**, 2016.
- [11] W3C. **Shapes Constraint Language (SHACL)**. Disponível em: <https://www.w3.org/TR/shacl/>. Acesso em: 12 maio 2025.
- [12] TOPQUADRANT. **Why I Use SHACL For Defining Ontology Models**. Disponível em: <https://www.topquadrant.com/resources/why-i-use-shacl-for-defining-ontology-models/>. Acesso em: 12 maio 2025.
- [13] ABDELAZIZ, Ibrahim et al. A survey and experimental comparison of distributed SPARQL engines for very large RDF data. **Proceedings of the VLDB Endowment**, v. 10, n. 13, p. 2049–2060, set. 2017.
- [14] CONFALONIERI, Roberto; GUIZZARDI, Giancarlo. On the multiple roles of ontologies in explanations for neuro-symbolic AI. **Neurosymbolic Artificial Intelligence**, v. 1, p. NAI-240754, 2025.
- [15] SAHOO, Satya S. et al. Ontology-based feature engineering in machine learning workflows for heterogeneous epilepsy patient records. **Scientific reports**, v. 12, n. 1, p. 19430, 2022.
- [16] PIATETSKY, Gregory. CRISP-DM, still the top methodology for analytics, data mining, or data science projects. **KDD News**, 2014.
- [17] MARTINEZ, Iñigo; VILES, Elisabeth; OLAIZOLA, Igor G. A survey study of success factors in data science projects. In: **2021 IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA)**. IEEE, 2021. p. 2313-2318.
- [18] SCHRÖER, Christoph; KRUSE, Felix; GÓMEZ, Jorge Marx. A systematic literature review on applying CRISP-DM process model. **Procedia computer science**, v. 181, p. 526-534, 2021.
- [19] SAS INSTITUTE. **Introduction to SEMMA. SAS Enterprise Miner 14.3 Documentation**. Disponível em:

<https://documentation.sas.com/doc/en/emref/14.3/n061bzurmej4j3n1jnj8bbj1a2.htm>. Acesso em: 09 jun. 2025.

[20] FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37-37, 1996.

[21] DOMINO DATA LAB. **The Practical Guide to Managing Data Science at Scale**. 2017. Disponível em: <https://domino.ai/resources/managing-data-science>. Acesso em: 02 jun. 2025.

[22] W3C. **RDF 1.1 Concepts and Abstract Syntax**. Disponível em: <https://www.w3.org/TR/rdf11-concepts/>. Acesso em: 03 ago. 2025.

[23] W3C. **OWL 2 Web Ontology Language Document Overview (Second Edition)**. Disponível em: <https://www.w3.org/TR/owl2-overview/>. Acesso em: 03 ago. 2025.

[24] W3C. **SPARQL 1.1 Overview**. Disponível em: <https://www.w3.org/TR/sparql11-overview/>. Acesso em: 03 ago. 2025.

[25] Stanford Center for Biomedical Informatics Research. **Protégé**: A free, open-source ontology editor and framework for building intelligent systems. Disponível em: <https://protege.stanford.edu/>. Acesso em: 16 nov. 2025.

[26] SHADBOLT, Nigel; BERNERS-LEE, Tim; HALL, Wendy. The semantic web revisited. **IEEE intelligent systems**, v. 21, n. 3, p. 96-101, 2006.

[27] W3C. **Time Ontology in OWL**. Disponível em: <https://www.w3.org/TR/owl-time/>. Acesso em: 09 ago. 2025.

[28] ALLEN, James F. Maintaining knowledge about temporal intervals. **Communications of the ACM**, v. 26, n. 11, p. 832-843, 1983.

[29] AMARAL, Glenda; BAIÃO, Fernanda; GUIZZARDI, Giancarlo. Foundational ontologies, ontology-driven conceptual modeling, and their multiple benefits to

data mining. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 11, n. 4, p. e1408, 2021.

13. Apêndices

13.1. Apêndice 1: Formulário

A seguir, apresento uma série de assertivas sobre os benefícios do uso de ontologias em cada uma das fases do modelo de referência. Peço, por favor, que avalie o seu grau de concordância com cada uma delas.

Entendimento do Negócio e do Domínio

ASSERTIVA 1: Uma ontologia força as partes interessadas a definir formalmente os termos de negócio. Isso elimina ambiguidades desde o início.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>O processo de criação da ontologia exige que todos cheguem a um consenso sobre o significado exato dos conceitos, eliminando ambiguidades que poderiam levar a erros de interpretação.</p>	<p>O time de marketing fala em "produto popular". A ontologia formaliza Cliente e Produto e define ProdutoPopular como subclasse de Produto com a regra (axioma) de ter numeroDeVendas > 1000 e avaliacaoMedia > 4.5, assim, a ambiguidade acaba.</p>
<p>COMENTÁRIOS?</p>	

ASSERTIVA 2: Uma ontologia alinha vocabulário entre analistas, especialistas e stakeholders.

- Concordo totalmente
 Concordo
 Nem concordo nem discordo
 Discordo
 Discordo totalmente

Explicação	Exemplo
A ontologia centraliza e padroniza a terminologia do projeto. Termos que diferentes áreas usam de formas distintas são mapeados para um único conceito formal.	A equipe de logística chama de "SKU", o desenvolvedor de product_id, e o analista de item. A ontologia unifica todos sob o conceito Produto.identificadorUnico, com uma descrição clara e semântica.
COMENTÁRIOS?	

ASSERTIVA 3: Uma ontologia oferece um modelo visual das entidades chave do negócio e suas relações, ajudando a definir o escopo e a identificar os fatores que dão valor ao projeto.

- Concordo totalmente
- Concordo
- Nem concordo nem discordo
- Discordo
- Discordo totalmente

Explicação	Exemplo
Ao representar o conhecimento de negócio em um mapa visual, a ontologia define o domínio de forma clara e objetiva, destacando as relações mais críticas entre as entidades. Isso permite concentrar os esforços do projeto nos fatores de maior valor.	O diagrama mostra que Cliente pode fazer múltiplos Pedidos, Pedido contém Produtos, e Produto pertence a Categoria e Marca. Com isso fica claro que uma recomendação de produto exige integrar dados de Pedidos, Histórico do Cliente e Avaliação do Produto.
COMENTÁRIOS?	

ASSERTIVA 4: As relações e axiomas numa ontologia ajudam a formular hipóteses mais ricas e estruturadas.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>As conexões lógicas definidas no modelo sugerem caminhos de investigação e hipóteses de negócio que podem ser testadas com dados.</p>	<p>A ontologia define o axioma $\text{ClienteVIP} \equiv \text{Cliente and } (\text{valorTotalPedidos} > 10000)$. A partir disso, a equipe formula a hipótese: "Clientes VIP realizam mais que o triplo de compras que Clientes não VIP?".</p>
<p>COMENTÁRIOS?</p>	

Aquisição e Compreensão Inicial

ASSERTIVA 5: Uma ontologia define os conceitos importantes, transformando uma busca cega em uma investigação direcionada.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>A ontologia funciona como uma "lista de compras" para o cientista de dados. Em vez de explorar centenas de tabelas e arquivos aleatoriamente, ele busca ativamente nas fontes brutas os dados que correspondem aos conceitos já definidos como essenciais para o projeto.</p>	<p>A ontologia requer os conceitos Cliente, Produto, Pedido e Avaliação. Com essa lista, o cientista de dados foca sua busca nos bancos de dados por tabelas com nomes como users, products, orders e reviews, economizando tempo e esforço.</p>
<p>COMENTÁRIOS?</p>	

Limpeza e Pré-processamento

ASSERTIVA 6: Uma ontologia funciona como o modelo canônico para a integração dos dados.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>Aqui o mapeamento já serve como guia para a transformação prática, colunas são renomeadas, tipos ajustados e tabelas unificadas de acordo com a ontologia, gerando um dataset consistente.</p>	<p>Dados de clientes vêm de dois sistemas: o de vendas (com a coluna sales.client_id) e o de suporte (com a coluna tickets.customer_num). O script de limpeza, então, usa a ontologia como guia para integrar essas três fontes, criando para cada cliente um perfil unificado.</p>
<p>COMENTÁRIOS?</p>	

Exploração Avançada**ASSERTIVA 7: Os axiomas de uma ontologia servem como regras para a validação semântica.**

- Concordo totalmente
 Concordo
 Nem concordo nem discordo
 Discordo
 Discordo totalmente

Explicação	Exemplo
<p>As regras lógicas da ontologia são usadas para validar dados antes da análise, sinalizando quando algo não faz sentido conceitual ou viola a lógica do domínio.</p>	<p>Um axioma estabelece que a dataDeCriacao de um Pedido deve ser posterior ou igual à dataDeCadastro do Cliente que o realizou. Durante a limpeza, o sistema identifica pedidos com data anterior ao cadastro do cliente, o que viola o axioma e indica erro de ingestão ou inconsistência no histórico.</p>
<p>COMENTÁRIOS?</p>	

ASSERTIVA 8: Uma ontologia contextualiza a análise, ajudando a separar correlações enganosas de relações de negócio significativas.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>Ao fornecer o contexto semântico do domínio, a ontologia permite que o cientista de dados interprete as correlações estatísticas dentro de um quadro de conhecimento de negócio, distinguindo entre associações espúrias e relações causalmente ou logicamente relevantes.</p>	<p>Uma análise exploratória pode apontar correlação entre o último dígito do id do cliente e o valor médio dos pedidos. A ontologia mostra que não existe nenhuma relação semântica entre identificadores técnicos e comportamento de compra, classificando essa associação como espúria. Em contrapartida, uma correlação entre Categoria do Produto e frequência de Pedido é validada como significativa, já que a ontologia modela explicitamente que todo Pedido contém Produtos que pertencem a uma Categoria.</p>
<p>COMENTÁRIOS?</p>	

ASSERTIVA 9: Uma ontologia permite explorar conexões que não seriam visíveis apenas pela estatística, orientando a análise para padrões relevantes de negócio.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>O modelo de dados tradicional (tabelas) dificulta a descoberta de relações indiretas. A representação em grafo da ontologia permite que o analista caminhe pelas conexões.</p>	<p>Explorando o grafo, um analista descobre um caminho: um Cliente faz um Pedido que contém um Produto de uma determinada Marca. O analista percebe que muitos desses clientes, em Pedidos futuros, compram outros Produtos da mesma Marca, mesmo que de Categorias diferentes. Isso revela um padrão de "lealdade à marca" que pode ser usado para enriquecer as recomendações.</p>
<p>COMENTÁRIOS?</p>	

Desenvolvimento

ASSERTIVA 10: Uma ontologia orienta a engenharia de features, ajudando a selecionar variáveis de forma inteligente.

- Concordo totalmente
 Concordo
 Nem concordo nem discordo
 Discordo
 Discordo totalmente

Explicação	Exemplo
<p>A ontologia destaca quais conceitos e relações são mais importantes. Isso foca o esforço na criação de features que carregam significado para o negócio, em vez de testar combinações aleatórias de colunas.</p>	<p>A ontologia modela a relação entre um Cliente e seus Pedidos. Isso inspira o cientista de dados a criar features como <code>frequencia_de_pedidos_mensal</code>, <code>valor_medio_por_pedido</code> e <code>numero_de_categorias_distintas_compradas</code>, todas baseadas diretamente na estrutura do modelo.</p>
<p>COMENTÁRIOS?</p>	

ASSERTIVA 11: Uma ontologia enriquece os dados para modelos de machine learning ao permitir a inferência automática de novas variáveis a partir de seus axiomas.

- Concordo totalmente
- Concordo
- Nem concordo nem discordo
- Discordo
- Discordo totalmente

Explicação	Exemplo
Ferramentas de inferência aplicam axiomas para derivar atributos ausentes ou classificações úteis como novas colunas de entrada.	A ontologia define o axioma $\text{ClienteVIP} \equiv \text{Cliente and } (\text{valorTotalPedidos} > 10000)$, o raciocinador marca os Clientes que cumprem essa condição. Assim, nos dados de entrada do modelo é criada a variável booleana ClienteVIP.
COMENTÁRIOS?	

Validação

ASSERTIVA 12: Uma ontologia ajuda a verificar se os resultados respeitam restrições de negócio e se fazem sentido conceitual.

- () Concordo totalmente
- () Concordo
- () Nem concordo nem discordo
- () Discordo
- () Discordo totalmente

Explicação	Exemplo
Na avaliação dos modelos, confronta-se os resultados com os axiomas da ontologia para checar a coerência.	O modelo de machine learning sugere vendas de um produto em uma categoria que não existe para esse produto, se a ontologia diz que aquele produto só pertence a uma categoria fixa, essa previsão é inconsistente.
COMENTÁRIOS?	

ASSERTIVA 13: Uma ontologia permite validação cruzada com especialistas, pois os resultados podem ser mapeados diretamente em conceitos da ontologia.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>Como os insights podem ser expressos em termos bem definidos, é mais fácil discutir conclusões com os especialistas.</p>	<p>O cientista de dados se reúne com o gerente de produtos. Em vez de perguntar "O que você acha da recomendação do produto 321 para o cliente 123?", ele usa a ontologia para agregar os resultados e pergunta: "Nosso modelo está recomendando massivamente Produtos da Marca Y para Clientes que foram inferidos como CompradoresTecnologicos. Isso faz sentido com a nossa estratégia de marca ou o modelo está com um viés indesejado?". Essa pergunta, baseada nos conceitos da ontologia, permite uma validação de negócio muito mais rica.</p>
<p>COMENTÁRIOS?</p>	

ASSERTIVA 14: Uma ontologia fornece uma camada de validação lógica para as previsões de um modelo, identificando resultados que violam suas regras de negócio e auxiliando na explicabilidade.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>Ao comparar previsões com as restrições definidas na ontologia, é possível identificar saídas incoerentes com o domínio de negócio. Isso aumenta a confiabilidade do sistema e fornece justificativas claras sobre por que certas previsões não devem ser aceitas.</p>	<p>Um modelo de recomendação sugere que um Pedido seja criado sem Produto associado. A ontologia possui o axioma “todo Pedido deve conter pelo menos um Produto”. A violação desse axioma sinaliza que a previsão é incorreta, permitindo que a equipe descarte o resultado e registre a justificativa baseada em regra de negócio.</p>
<p>COMENTÁRIOS?</p>	

Entrega

ASSERTIVA 15: Uma ontologia atua como documentação, explicando de forma formal e acessível os conceitos usados no projeto.

- () Concordo totalmente
- () Concordo
- () Nem concordo nem discordo
- () Discordo
- () Discordo totalmente

Explicação	Exemplo
Ela atua como documentação viva que registra entidades, propriedades e restrições de forma clara, servindo como referência tanto para especialistas técnicos quanto para as equipes de negócio.	O diagrama ontológico mostra como Cliente, Pedido, Produto, Categoria e Marca se relacionam. Esse material é anexado à entrega final do projeto para que novos membros entendam rapidamente o domínio.
COMENTÁRIOS?	

ASSERTIVA 16: Uma ontologia garante um vocabulário consistente em todos os produtos finais (relatórios, dashboards, APIs).

- Concordo totalmente
 Concordo
 Nem concordo nem discordo
 Discordo
 Discordo totalmente

Explicação	Exemplo
A ontologia padroniza os termos usados em relatórios e sistemas, evitando ambiguidades.	Todos os relatórios e dashboards usam Categoria para agrupar produtos, em vez de alternar entre “tipo de produto”, “linha” ou “segmento”.
COMENTÁRIOS?	

ASSERTIVA 17: Uma ontologia facilita integração com outros sistemas, pois o modelo ontológico serve como referência semântica.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>Ao expor conceitos bem definidos, a ontologia torna os dados auto-descritivos e padronizados. Isso reduz o esforço de integração, elimina ambiguidades e permite que sistemas externos entendam o significado dos dados.</p>	<p>Um sistema de gestão de estoque usa <code>item_code</code>, que é mapeado diretamente para <code>Produto.identificadorUnico</code> na ontologia, garantindo consistência. Já a API de recomendação retorna uma estrutura de dados informando que o item é instância da classe <code>Produto</code> e pertence a uma <code>Marca</code>. Assim, o sistema de estoque pode consumir essa API e entender automaticamente do que se trata, sem precisar de uma reunião para alinhar o significado dos campos.</p>
<p>COMENTÁRIOS?</p>	

Monitoramento

ASSERTIVA 18: Uma ontologia funciona como referência contínua para detectar desvios conceituais.

- Concordo totalmente
 Concordo
 Nem concordo nem discordo
 Discordo
 Discordo totalmente

Explicação	Exemplo
<p>Durante a operação, novos dados podem não se encaixar no modelo definido. Isso sinaliza possíveis problemas de qualidade ou mudanças no negócio.</p>	<p>A ontologia tem um axioma que diz que todo Pedido deve conter pelo menos um Produto. Um novo sistema de "assinaturas de caixas misteriosas" é lançado e começa a gerar Pedidos sem uma lista de produtos explícita. O sistema de monitoramento, que valida os dados contra a ontologia, imediatamente sinaliza essa violação, alertando que o conceito de Pedido evoluiu e a ontologia precisa ser atualizada.</p>
<p>COMENTÁRIOS?</p>	

ASSERTIVA 19: Uma ontologia permite auditoria semântica, explicando decisões do modelo com base em axiomas e relações.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>A ontologia registra as regras que orientam inferências e previsões. Isso possibilita rastrear a lógica de cada decisão e justificar saídas para clientes ou funcionários.</p>	<p>Um cliente reclama de uma recomendação de roupa feminina que parece aleatória. A auditoria, no entanto, usa a ontologia para explicar a decisão. Ela mostra que o sistema inferiu o perfil do cliente como CompradorDePresentes, com base em um axioma que analisa seu histórico de entregas para diferentes endereços. O item recomendado, por sua vez, estava classificado como PresentePopular. A recomendação, portanto, não foi um erro, mas o resultado de conexão entre o perfil "comprador de presentes" e itens populares para essa finalidade.</p>
<p>COMENTÁRIOS?</p>	

ASSERTIVA 20: Uma ontologia ajuda a identificar o surgimento de novas categorias, entidades ou relações que não se encaixam no modelo de conhecimento atual, sinalizando como o sistema precisa evoluir.

- () Concordo totalmente
 () Concordo
 () Nem concordo nem discordo
 () Discordo
 () Discordo totalmente

Explicação	Exemplo
<p>A ontologia representa o conhecimento atual. Ao analisar dados que "não se encaixam" no modelo existente, é possível descobrir novas tendências e conceitos de negócio.</p>	<p>Um modelo de análise de texto que lê as avaliações dos clientes começa a identificar um cluster frequente de termos como "sustentável", "ecológico", "reciclado". Esses termos não se mapeiam a nenhuma Categoria existente na ontologia. O sistema de monitoramento sinaliza isso como a provável emergência de um novo conceito, levando a equipe a criar a Categoria "Produtos Sustentáveis" e a retreinar o modelo de recomendação para alavancar essa nova tendência.</p>
<p>COMENTÁRIOS?</p>	

Obrigado por sua participação!

13.2. Apêndice 2: Email

Assunto: Ciência de Dados e Ontologias (Convite)

Prezado(a) <nome>,

Meu nome é Luiz Augusto Masset Lacombe Papa e sou estudante concluinte do curso de Ciência da Computação na PUC-Rio.

Estou finalizando o meu Projeto Final de Graduação sob a orientação do Prof. Marcos Villas, com o tema "Ontologias aplicadas à Ciência de Dados". O trabalho investiga como o uso de ontologias pode aprimorar a consistência, a interoperabilidade e a explicabilidade em cada etapa do ciclo de vida de projetos de Ciência de Dados.

Para validar a aplicabilidade e a relevância das propostas deste trabalho, elaborei um formulário que busca a avaliação de especialistas experientes na área. A sua participação seria de imenso valor.

O formulário está estruturado da seguinte forma:

1. Contextualização: Apresenta brevemente alguns conceitos importantes e o modelo de referência do ciclo de vida de Ciência de Dados.
2. Avaliação das Assertivas: Lista uma série de afirmações sobre os benefícios da aplicação de ontologias em cada uma das fases do ciclo de vida, para que você avalie seu grau de concordância.

Anexei a este e-mail o formulário. Para preenchê-lo, basta indicar seu nível de concordância para cada assertiva e, caso deseje, adicionar seus comentários nos campos designados. O preenchimento leva cerca de 20 minutos.

Adicionalmente, após a análise das respostas, eu ficaria grato se tivéssemos a oportunidade de agendar uma breve entrevista de no máximo meia hora para aprofundar alguns pontos e colher suas percepções de forma mais detalhada.

Gostaria de saber se um prazo de uma semana para a sua análise seria razoável. Agradeço antecipadamente pela sua consideração e aguardo um retorno para saber se posso contar com sua ajuda.

Atenciosamente,

Luiz Augusto Masset Lacombe Papa

13.3. Apêndice 3: Comentários formulário

ASSERTIVA 1: Uma ontologia força as partes interessadas a definir formalmente os termos de negócio. Isso elimina ambiguidades desde o início.

E1 - Acredito que essa é a etapa em que é mais proveitoso o uso de Ontologias, visando a definir um padrão para os termos ao longo do processo. Ter esse padrão no início o processo é bastante útil e facilita a comunicação com clientes, desenvolvedores, POs e gerentes, além de evitar retrabalho.

E4 - Embora ontologias promovam clareza terminológica, elas não eliminam ambiguidades por si só apenas quando existe envolvimento real dos stakeholders na sua construção. Em muitas organizações, a dificuldade não é técnica, mas política ou cultural. A ontologia ajuda, mas não garante consenso sem governança.

ASSERTIVA 2: Uma ontologia alinha vocabulário entre analistas, especialistas e stakeholders.

E4 - A padronização semântica é um dos benefícios mais claros e imediatos do uso de ontologias. No entanto, esse alinhamento só se sustenta ao longo do tempo quando há um processo formal de gestão de termos (glossário corporativo, stewardship).

ASSERTIVA 3: Uma ontologia oferece um modelo visual das entidades chave do negócio e suas relações, ajudando a definir o escopo e a identificar os fatores que dão valor ao projeto.

E1 - Acredito que a ontologia pode sim auxiliar ao desenvolver esquemas e diagramas que representem relacionamentos entre entidades e fluxos de informações.

E2 - Entendo que nesse caso específico teria a mesma função de um diagrama de classes, correto? No caso da ontologia, o que vai além quando comparado com um diagrama de classes tradicional?

E4 - Representações ontológicas fornecem uma visão mais rica que simples ER diagrams, pois capturam relações conceituais e restrições. Isso ajuda sim a clarear escopo e identificar dependências. Porém, se o modelo ficar complexo demais, pode virar um obstáculo para times não técnicos.

ASSERTIVA 4: As relações e axiomas numa ontologia ajudam a formular hipóteses mais ricas e estruturadas.

E4 - A ontologia de fato expõe relações que inspiram hipóteses melhores, mas isso depende do nível de granularidade. Ontologias muito abstratas ajudam pouco; muito específicas tornam-se rígidas. Há um equilíbrio crítico entre capacidade explicativa e flexibilidade analítica

ASSERTIVA 5: Uma ontologia define os conceitos importantes, transformando uma busca cega em uma investigação direcionada.

E2 - Nesse caso parece meio óbvio, mas vejo a ontologia nesse caso auxiliando na construção e otimização de queries SQL mais complexas (e.g., JOINS).

E4 - A ontologia direciona a busca, mas ela não substitui a descoberta exploratória, especialmente em ambientes com múltiplos legados. Ela reduz dispersão, mas o cientista de dados ainda precisa validar empiricamente se as fontes realmente representam os conceitos modelados.

ASSERTIVA 6: Uma ontologia funciona como o modelo canônico para a integração dos dados.

E4 - Modelos canônicos são particularmente úteis em cenários com múltiplos sistemas. A ontologia evita mapeamentos ad hoc e consolida padrões. A única ressalva é que, se a ontologia mudar, o pipeline de integração precisa evoluir junto.

ASSERTIVA 7: Os axiomas de uma ontologia servem como regras para a validação semântica.

E4 - Validação semântica é extremamente valiosa, mas exige ontologias com axiomas realmente úteis o que nem sempre ocorre. Muitas ontologias corporativas focam apenas em vocabulário, e não em restrições formais.

ASSERTIVA 8: Uma ontologia contextualiza a análise, ajudando a separar correlações enganosas de relações de negócio significativas.

E2 - Nem sempre eu acho. Para isso a ontologia precisaria de boa qualidade e completude do conhecimento específico daquele domínio/problema. Fico pensando, até que ponto (trade-off), o desenvolvimento da ontologia pode ser inclusive mais complexo que o desenvolvimento de um modelo de machine learning 'OK' que já agrega valor.

E4 - Um dos principais desafios da EDA é interpretar padrões sem contexto. Ontologias reduzem esse risco ao fornecerem uma estrutura semântica que orienta a análise indicando o que realmente faz sentido comparar. Assim, o analista evita conclusões equivocadas e foca em relações que são relevantes para o negócio.

ASSERTIVA 9: Uma ontologia permite explorar conexões que não seriam visíveis apenas pela estatística, orientando a análise para padrões relevantes de negócio.

E4 - A exploração orientada por grafos e relações conceituais realmente revela padrões difíceis de perceber apenas com tabelas. Porém, nem todas as empresas têm ferramentas de raciocínio semântico maduras para aproveitar esse potencial.

ASSERTIVA 10: Uma ontologia orienta a engenharia de features, ajudando a selecionar variáveis de forma inteligente.

E4 - Ontologias são extremamente úteis para feature engineering, especialmente em domínios onde o conhecimento de negócio é chave. Elas reduzem a

experimentação cega e ajudam a gerar features com significado real e não apenas estatístico.

ASSERTIVA 11: Uma ontologia enriquece os dados para modelos de machine learning ao permitir a inferência automática de novas variáveis a partir de seus axiomas.

E4 - Inferência ontológica adiciona valor, mas depende de motores de raciocínio que nem sempre escalam bem para grandes volumes de dados. Em cenários de Big Data, é mais comum gerar features derivadas manualmente

ASSERTIVA 12: Uma ontologia ajuda a verificar se os resultados respeitam restrições de negócio e se fazem sentido conceitual.

E4 - Modelos que violam regras de negócio podem gerar custos enormes. A ontologia funciona como uma camada adicional de “sanidade conceitual”. Porém, é essencial que a ontologia esteja alinhada à realidade operacional atual o que nem sempre ocorre.

ASSERTIVA 13: Uma ontologia permite validação cruzada com especialistas, pois os resultados podem ser mapeados diretamente em conceitos da ontologia.

E4 - Expressar resultados em termos conceituais torna discussões mais claras, mas a maturidade dos especialistas determina o valor dessa etapa. Se a ontologia for mal compreendida, pode gerar barreiras em vez de facilitar.

ASSERTIVA 14: Uma ontologia fornece uma camada de validação lógica para as previsões de um modelo, identificando resultados que violam suas regras de negócio e auxiliando na explicabilidade.

E4 - A capacidade de rastrear erros conceituais e explicar previsões é extremamente valiosa, especialmente em modelos com obrigação regulatória ou de transparência (bancos, saúde, telecom). Ontologias são excelentes nesse papel.

ASSERTIVA 15: Uma ontologia atua como documentação, explicando de forma formal e acessível os conceitos usados no projeto.

E4 - Ontologias funcionam como documentação viva, mais eficaz que documentos estáticos. Porém, manutenção contínua é essencial uma ontologia desatualizada é tão ruim quanto nenhuma.

ASSERTIVA 16: Uma ontologia garante um vocabulário consistente em todos os produtos finais (relatórios, dashboards, APIs).

E2 - Depende, não sei se garante, mas concordo que ajuda!

E4 - Consistência terminológica é um dos maiores fatores de qualidade no consumo de dados. Ontologias são ideais para resolver esse problema de forma duradoura.

ASSERTIVA 17: Uma ontologia facilita integração com outros sistemas, pois o modelo ontológico serve como referência semântica.

E4 - A ontologia facilita integrações, mas isso só se materializa quando todos os sistemas adotam o modelo semântico. Em empresas com legado pesado, a transição pode ser lenta.

ASSERTIVA 18: Uma ontologia funciona como referência contínua para detectar desvios conceituais.

E4 - Detectar drift semântico é tão importante quanto detectar drift estatístico. Ontologias oferecem uma referência estrutural que ajuda a captar mudanças no próprio negócio.

ASSERTIVA 19: Uma ontologia permite auditoria semântica, explicando decisões do modelo com base em axiomas e relações.

E4 - A auditoria semântica é poderosa, mas depende de axiomas claros. Muitas ontologias não possuem regras suficientes para permitir rastreabilidade rica.

ASSERTIVA 20: Uma ontologia ajuda a identificar o surgimento de novas categorias, entidades ou relações que não se encaixam no modelo de

conhecimento atual, sinalizando como o sistema precisa evoluir.

E4 - Ontologias ajudam a identificar novos conceitos emergentes, algo crucial em mercados dinâmicos. Isso permite evolução contínua do modelo de conhecimento e maior adaptação do pipeline analítico.