

UNIVERSIDADE FEDERAL DO PARANÁ

RÚBIA MARIA PEREIRA

APRIMORAMENTO DO MODELO BERT PARA ANÁLISE DE INTENÇÃO EM
TEXTOS: APLICAÇÃO EM CHATBOT NO SETOR DE ÓLEO E GÁS

CURITIBA

2025

RÚBIA MARIA PEREIRA

APRIMORAMENTO DO MODELO BERT PARA ANÁLISE DE INTENÇÃO EM
TEXTOS: APLICAÇÃO EM CHATBOT NO SETOR DE ÓLEO E GÁS

Tese apresentada ao curso de Pós-Graduação em
Métodos Numéricos em Engenharia, Setor de
Exatas e Tecnológica, Universidade Federal do
Paraná, como requisito parcial à obtenção do título
de Doutora em Métodos Numéricos em Engenharia

Orientador: Prof. Dr. Cassius Tadeu Scapin
Coorientador: Prof. Dr. Pablo Deivid Valle

CURITIBA

2025

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

Pereira, Rubia Maria

Aprimoramento do modelo BERT para análise de intenção em textos:
aplicação em chatbot no setor de óleo e gás / Rubia Maria Pereira. –
Curitiba, 2025.

1 recurso on-line : PDF.

Tese (Doutorado) - Universidade Federal do Paraná, Setor de Ciências
Exatas, Programa de Pós-Graduação em Métodos Numéricos em
Engenharia.

Orientador: Cassius Tadeu Scapin

Coorientador: Pablo Deivid Valle

1. Referência (Linguística). 2. Textos – Análise. 3. Textos – Classificação.
4. Petróleo. I. Universidade Federal do Paraná. II. Programa de Pós-
Graduação em Métodos Numéricos em Engenharia. III. Scapin, Cassius
Tadeu. IV. Valle, Pablo Deivid. V. Título.

Bibliotecário: Douglas Lenon da Silva CRB-9/1892



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MÉTODOS NUMÉRICOS
EM ENGENHARIA - 40001016030P0

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **RÚBIA MARIA PEREIRA**, intitulada: **APRIMORAMENTO DO MODELO BERT PARA ANÁLISE DE INTENÇÃO EM TEXTOS: APLICAÇÃO EM CHATBOT NO SETOR DE ÓLEO E GÁS**, sob orientação do Prof. Dr. CASSIUS TADEU SCARPIN, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutora está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 18 de Dezembro de 2025.

Assinatura Eletrônica
18/12/2025 19:17:33.0
CASSIUS TADEU SCARPIN
Presidente da Banca Examinadora

Assinatura Eletrônica
19/12/2025 15:32:58.0
NELSON HEIN
Avaliador Externo (FUNDAÇÃO UNIVERSIDADE REGIONAL DE
BLUMENAU)

Assinatura Eletrônica
19/12/2025 09:02:03.0
ROBERTO ZANETTI FREIRE
Avaliador Externo (UNIVERSIDADE TECNOLÓGICA FEDERAL DO
PARANÁ)

Assinatura Eletrônica
19/12/2025 15:36:03.0
ADRIANA KROENKE
Avaliador Externo (UNIVERSIDADE REGIONAL DE BLUMENAU)

À minha família, que sempre me sustentou com amor.

Ao meu companheiro, pelo apoio diário.

Ao Prof. Dr. Cassius, pela orientação atenciosa ao longo de toda a pesquisa.

AGRADECIMENTOS

A Deus, por fortalecer meus passos e iluminar meu caminho ao longo desta jornada.

À minha família, em especial aos meus pais, expresso minha gratidão pelo apoio constante, pela compreensão nos momentos de ausência e pelo incentivo incondicional ao meu crescimento pessoal e acadêmico.

Ao meu companheiro, Gilson, agradeço a paciência, apoio diário e palavras de encorajamento que me acompanharam nos momentos mais desafiadores. Obrigada por sempre me permitir correr atrás dos meus sonhos e por caminhar ao meu lado enquanto eles ganhavam forma.

Aos amigos que estiveram próximos, especialmente à Lais, meu sincero agradecimento pela companhia, pelas conversas e pela leveza que ajudaram a equilibrar esta caminhada.

Aos meus queridos professores Adriana e Nelson, por me acompanharem e me apoiarem desde a graduação até o Doutorado. Vocês são exemplos que levo para minha vida, muito obrigada!

Registro meu profundo agradecimento ao meu orientador, Prof. Cassius, por sua orientação competente, pelas discussões técnicas, pela confiança depositada em meu trabalho e pela serenidade com que conduziu cada etapa desta pesquisa. Sua dedicação foi fundamental para que este estudo se concretizasse.

Aos meus colegas do projeto AKHub, em especial a Fernanda e Andressa, que sempre me apoiaram e me motivaram a chegar até aqui. Sou eternamente grata a vocês, por dividir esses momentos comigo.

À ExxonMobil, pela colaboração estabelecida, que possibilitou o desenvolvimento de pesquisas aplicadas em um contexto real de engenharia e análise de dados.

À Universidade Federal do Paraná e ao Programa de Pós-Graduação em Métodos Numéricos em Engenharia, pela oportunidade de desenvolvimento acadêmico e científico. Estendo meus agradecimentos à coordenação, representada pelo Prof. Portela e pela nossa querida Mika, pelo apoio institucional ao longo desta jornada, e à banca examinadora, pela disponibilidade em avaliar este trabalho e contribuir com sugestões valiosas.

“Tudo parece impossível até que seja feito.”
— Nelson Mandela

RESUMO

Esta tese investiga a classificação de intenções textuais no setor de petróleo e gás, considerando cenários caracterizados por conjuntos de dados rotulados pequenos e linguagem técnico-financeira especializada. A motivação do estudo decorre das limitações observadas na aplicação direta de modelos de linguagem pré-treinados em domínios específicos, nos quais a escassez de dados anotados e a proximidade semântica entre categorias de intenção dificultam a obtenção de classificações consistentes. Nesse contexto, o objetivo central foi avaliar a eficácia de uma abordagem de análise de intenção adaptada a esse domínio. Para atender a esse objetivo, foi proposto o modelo OilFinBERT, baseado na arquitetura BERT e adaptado ao setor de petróleo e gás por meio de pré-treinamento contínuo e estratégias de ajuste supervisionado. O estudo considerou cinco categorias de intenção relevantes ao domínio técnico-financeiro (sendo elas: previsão de preço, análise de sentimento, conceito, outros e preço atual), organizadas na base de dados OilFinIntent, construída especificamente para este trabalho. Além disso, foi utilizado um conjunto de mais de 400 mil títulos de notícias do setor como dados não rotulados, empregado na etapa de especialização do modelo ao vocabulário do domínio. A metodologia incluiu experimentos de *fine-tuning* com diferentes otimizadores, técnicas de regularização e duas estratégias de modelagem para a tarefa de classificação de intenções: multiclasse plana e hierárquica. O desempenho do OilFinBERT foi comparado a modelos de referência amplamente utilizados na literatura, incluindo *Naive Bayes*, *Convolutional Neural Network (CNN)*, *Recurrent Neural Network (RNN)*, *Long Short-Term Memory (LSTM)* e *BERT-base*. A avaliação foi conduzida por meio de métricas adequadas a cenários multiclasse e com distribuição equilibrada entre classes, com destaque para o F1-macro e a PR-AUC. Os resultados indicaram que a abordagem hierárquica do OilFinBERT apresentou desempenho superior em relação às demais configurações avaliadas, evidenciando maior equilíbrio entre precisão e *recall* ao longo das diferentes classes de intenção. Observou-se ainda que estratégias de regularização e otimizadores adaptativos contribuem para maior estabilidade no processo de ajuste supervisionado em conjuntos de dados reduzidos. A comparação com os modelos de referência reforçou a importância da adaptação explícita ao domínio para alcançar desempenho consistente em tarefas de análise de intenção com textos especializados. Conclui-se que a combinação de pré-treinamento contínuo, técnicas de regularização e modelagem hierárquica constitui uma abordagem adequada para a classificação de intenções em contextos técnicos com disponibilidade limitada de dados rotulados. Como principal contribuição, esta tese demonstra a eficácia do OilFinBERT hierárquico e disponibiliza uma base de dados específica para o setor de petróleo e gás, contribuindo para a reprodutibilidade e o avanço de pesquisas futuras em Processamento de Linguagem Natural (PLN) aplicado a domínios especializados.

Palavras-chave: Análise de Intenção; classificação de texto; BERT; petróleo e gás; OilFinBERT.

ABSTRACT

This thesis investigates textual intent classification in the oil and gas sector, considering scenarios characterized by small labeled datasets and specialized technical-financial language. The motivation for this study arises from the limitations observed in the direct application of pre-trained language models to domain-specific contexts, in which the scarcity of annotated data and the semantic proximity between intent categories hinder the achievement of consistent classifications. Within this context, the main objective was to evaluate the effectiveness of an intent analysis approach adapted to this domain. To achieve this objective, the OilFinBERT model was proposed, based on the BERT architecture and adapted to the oil and gas sector through continued pre-training and supervised fine-tuning strategies. The study considered five intent categories relevant to the technical-financial domain—namely price forecasting, sentiment analysis, concept, others, and current price—organized in the OilFinIntent dataset, which was specifically constructed for this research. In addition, a collection of more than 400,000 news headlines from the sector was used as unlabeled data during the domain specialization stage of the model. The methodology included fine-tuning experiments with different optimizers, regularization techniques, and two modeling strategies for the intent classification task: flat multiclass and hierarchical classification. The performance of OilFinBERT was compared with widely used baseline models in the literature, including Naive Bayes, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), and BERT-base. The evaluation was conducted using metrics suitable for multiclass scenarios with balanced class distributions, with emphasis on F1-macro and PR-AUC. The results indicated that the hierarchical OilFinBERT approach achieved superior performance compared to the other evaluated configurations, demonstrating a more balanced trade-off between precision and recall across the different intent classes. It was also observed that regularization strategies and adaptive optimizers contributed to greater stability in the supervised fine-tuning process when applied to small datasets. The comparison with baseline models reinforced the importance of explicit domain adaptation to achieve consistent performance in intent analysis tasks involving specialized texts. It is concluded that the combination of continued pre-training, regularization techniques, and hierarchical modeling constitutes an appropriate approach for intent classification in technical contexts with limited availability of labeled data. As its main contribution, this thesis demonstrates the effectiveness of the hierarchical OilFinBERT model and provides a domain-specific dataset for the oil and gas sector, contributing to reproducibility and to future research in Natural Language Processing (NLP) applied to specialized domains.

Keywords: Intent Analysis; text Classification; BERT; oil and gas; OilFinBERT.

LISTA DE FIGURAS

Figura 1 - Fluxo conceitual adotado para representar o problema	30
Figura 2 - Fluxo de uma <i>Convolutional Neural Network</i> (CNN) aplicada a texto.....	42
Figura 3 - Ilustração de <i>Mixout</i>	56
Figura 4 - Fluxo de Trabalho	64

LISTA DE TABELAS

Tabela 1 - Vantagens e Desvantagens do NB.....	38
Tabela 2 - Resumo dos Trabalhos Relacionados.....	58
Tabela 3 - Amostra do conjunto de dados OilFinIntent.....	66
Tabela 4 - Resultados dos Modelos	73

LISTA DE ABREVIATURAS OU SIGLAS

BERT	- Bidirectional Encoder Representations from Transformers
BoW	- Bag of Words
CNN	- Convolutional Neural Network
CNB	- Complement Naïve Bayes
DL	- Deep Learning
GAN	- Generative Adversarial Network
GPT	- Generative Pre-trained Transformer
GRU	- Gated Recurrent Unit
GloVe	- Global Vectors for Word Representation
KNN	- K-Nearest Neighbors
LLM	- Large Language Model
LLRD	- Layer-wise Learning Rate Decay
LSTM	- Long Short-Term Memory
MLM	- Masked Language Modeling
ML	- Machine Learning
MNB	- Multinomial Naïve Bayes
NB	- Naive Bayes
PLN	- Processamento de Linguagem Natural
NSP	- Next Sentence Prediction —
OVA	- One-Versus-All —
PR-AUC	- Precision-Recall Area Under Curve
RNN	- Recurrent Neural Network
SGD	- Stochastic Gradient Descent
SWA	- Stochastic Weight Averaging
SVM	- Support Vector Machines
TF-IDF	- Term Frequency-Inverse Document Frequency

SUMÁRIO

1 INTRODUÇÃO	16
1.1 OBJETIVO	19
1.1.1 Objetivo Geral	19
1.1.2 Objetivos Específicos	19
1.2 JUSTIFICATIVA	20
1.3 CONTRIBUIÇÕES	21
1.4 ESTRUTURA DO TRABALHO	22
2 DESCRIÇÃO DO PROBLEMA	24
2.1 ENTRADA.....	24
2.2 REPRESENTAÇÃO DE ENTRADA	25
2.3 ESPAÇO DE RÓTULOS	26
2.4 MODELAGEM.....	26
2.5 FUNÇÃO DE CLASSIFICAÇÃO	27
2.6 OBJETIVO DE OTIMIZAÇÃO	27
2.7 SAÍDAS E MÉTRICAS	28
2.8 RESTRIÇÕES DE CENÁRIO.....	30
3 REVISÃO DE LITERATURA	31
3.1 PROCESSAMENTO DE LINGUAGEM NATURAL.....	31
3.1.1 Representação de texto	32
3.2 ANÁLISE DE INTENÇÃO.....	34
3.2.1 Machine Learning.....	35
3.2.1.1 NAIVE BAYES.....	36
3.2.2 Deep Learning.....	39
3.2.2.1 <i>Convolutional Neural Networks</i> - CNN.....	41
3.2.2.2 <i>Recurrent Neural Networks</i> - RNN.....	44
3.2.2.3 <i>Long Short-Term Memory</i> - LSTM	46
3.2.2.4 <i>Bidirectional Encoder Representations from Transformers</i> - BERT	48
3.2.3 Técnicas aplicadas ao <i>fine-tuning</i>	52
3.2.3.1 Otimizadores	52
3.2.3.2 Regularização e ajustes	54
3.2.3.3 Modelo Hierárquico	57
3.3 TRABALHOS RELACIONADOS	57

4 PROPOSTA DE MODELO	64
4.1 SELEÇÃO DO MODELO	65
4.2 DEFINIÇÃO DA BASE DE DADOS NÃO ROTULADA PARA PRÉ- TREINAMENTO	65
4.3 DEFINIÇÃO DOS MODELOS DE REFERÊNCIA.....	66
4.4 CONSTRUÇÃO DO CONJUNTO DE DADOS DE INTENÇÃO	66
4.5 PRÉ-TREINAMENTO DO OILFINBERT	67
4.6 <i>FINE-TUNING</i> SUPERVISIONADO	67
4.7 ESTRATÉGIAS DE AJUSTE NO <i>FINE-TUNING</i>	68
4.7.1 Treinamento padrão	68
4.7.2 Reinicialização das camadas superiores.....	68
4.7.3 LLRD.....	69
4.7.4 Estratégia combinada.....	69
4.8 MODELAGEM HIERÁRQUICA	69
4.9 TÉCNICAS DE REGULARIZAÇÃO E ESTABILIDADE	69
4.9.1 Mixout	70
4.9.2 SWA.....	70
4.10 MÉTRICAS DE AVALIAÇÃO.....	70
4.11 PROCEDIMENTOS EXPERIMENTAIS	70
5 RESULTADO E DISCUSSÃO	71
5.1 PROTOTIPAGEM E FERRAMENTAS COMPUTACIONAIS UTILIZADAS	71
5.2 RESULTADOS.....	73
5.3 ANÁLISE DE FALHA	76
5.4 CONSIDERAÇÕES FINAIS	78
6 CONCLUSÃO	79
6.1 LIMITAÇÕES	80
6.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	80
6.3 PUBLICAÇÕES.....	81
REFERÊNCIAS	82
APÊNDICE A – DATASET OINFININTENT	88

1 INTRODUÇÃO

A maioria dos dados, embora grandes volumes textuais sejam gerados continuamente devido ao rápido avanço da internet, não está rotulado. Esse fenômeno limita o seu uso imediato em aplicações de aprendizado de máquina. (Liu; Xu, 2021). A necessidade de classificar e gerenciar esses dados de forma eficaz, por meio de tecnologias de Processamento de Linguagem Natural (PLN), tornou-se um tema de pesquisa cada vez mais relevante (Xu et al., 2020).

O PLN é um campo interdisciplinar que envolve a interação entre a computação e a linguística, com o objetivo de permitir que as máquinas processem e gerem linguagem humana de maneira útil e significativa (Aslam, 2023). Desta forma, torna-se importante para a compreensão da linguagem humana por parte dos programas de computador, tanto em sua forma falada quanto escrita (Sudhir; Suresh, 2021). Aplicações como *chatbots*, atendimento ao cliente e assistentes de voz dependem do PLN para facilitar a interação entre humanos e máquinas (Wu et al., 2024).

A análise de intenção é o primeiro elemento do fluxo de um *chatbot*, sendo responsável por interpretar a mensagem enviada pelo usuário. Quando uma mensagem é recebida, o sistema identifica a intenção associada e determina o próximo passo (Pérez-Soler; Guerra; De Lara, 2026). Assim, a identificação de intenções desempenha um papel fundamental nessas interações, permitindo ao sistema compreender as metas dos usuários e fornecer respostas ou ações adequadas (Huang et al., 2023).

Neste trabalho, o termo análise de intenção refere-se ao processo de identificação da finalidade comunicativa expressa em um texto, isto é, à determinação do objetivo implícito do autor ao produzir uma determinada mensagem. Diferentemente de tarefas voltadas exclusivamente à classificação temática ou à extração de informações pontuais, a análise de intenção busca capturar o propósito semântico subjacente ao texto, considerando o contexto linguístico no qual a mensagem está inserida. Assim, a análise de intenção é tratada como um problema de classificação supervisionada, no qual cada instância textual é associada a uma classe de intenção previamente definida, conforme discutido em Huang et al. (2023).

Para realizar a análise de intenções, utilizamos a técnica de classificação de texto, cujo objetivo é categorizar as expressões textuais do usuário em classes

predefinidas e que melhor correspondam às suas intenções (Fernández-Martínez et al., 2022). Para isso, é possível a utilização de diversas opções de algoritmos, desde técnicas de *Machine Learning* (ML), que consistem em métodos capazes de aprender padrões a partir de atributos previamente extraídos dos dados — como Naive Bayes (NB) e Support Vector Machine (SVM) — até abordagens de *Deep Learning* (DL), uma subárea do ML que permite capturar relações mais complexas e extrair automaticamente características relevantes do texto. Entre os modelos, destacam-se arquiteturas como Recurrent Neural Networks (RNN) e Convolutional Neural Networks (CNN), além de modelos pré-treinados como BERT¹ e GPT² (Ayadi; Lachiri, 2023; LeCun; Bengio; Hinton, 2015).

A precisão na identificação das intenções dos usuários é importante para melhorar a qualidade das respostas e apoiar a tomada de decisões estratégicas (Singh; Jia; Nalagatla, 2023). Esse desafio torna-se mais evidente no setor de petróleo e gás, que, com a implantação de conceitos da Indústria 4.0 — caracterizada pela integração de tecnologias digitais, automação inteligente e sistemas conectados que tornam os processos mais eficientes e autônomos (Ghobakhloo, 2020; Cañas et al., 2021) —, passou a explorar essas tecnologias para aumentar a produtividade, segurança e eficiência, enquanto minimiza custos e riscos ambientais e à saúde (Wanasinghe et al., 2020).

O setor de petróleo e gás foi adotado como domínio de estudo a partir do contexto de desenvolvimento do projeto AKHub³, no qual se evidenciou a necessidade de analisar intenções em textos técnico-financeiros sob condições de linguagem especializada e disponibilidade limitada de dados rotulados. As limitações observadas nesse contexto não se restringem ao projeto em si, mas refletem características recorrentes de aplicações reais em setores técnicos, nos quais a adaptação de modelos de análise de intenção ainda representa um desafio. Dessa forma, o domínio selecionado constitui um cenário adequado para investigar o problema proposto nesta

¹ *BERT (Bidirectional Encoder Representations from Transformers)* é um modelo de linguagem desenvolvido pelo Google AI, que entende o contexto das palavras em ambas as direções numa frase. Ele é amplamente utilizado para diversas tarefas de PNL. Para mais detalhes, ver tópico 3.2.4.

² *GPT (Generative Pre-trained Transformer)* é uma série de modelos de linguagem da OpenAI que gera texto prevendo a próxima palavra em uma sequência. Ele é utilizado para várias tarefas de PLN.

³ Projeto AKHub: Projeto desenvolvido através de uma parceria entre UFPR e ExxonMobil, com o enfoque com enfoque na análise de informações técnico-financeiras do setor de petróleo e gás.

tese, sem prejuízo da generalização conceitual dos resultados para outros contextos semelhantes.

Sendo assim, a colaboração entre engenheiros, cientistas de dados e empresas é essencial para integrar a IA aos processos e melhorar a tomada de decisões (Sircar et al., 2021). O uso de algoritmos de análise de intenção pode contribuir para resultados mais precisos e facilitar a comunicação entre os diferentes setores da empresa (Huang et al., 2023; Lu et al., 2019).

No entanto, a principal limitação enfrentada, em setores especializados como o de petróleo e gás, é a escassez desses grandes volumes de dados rotulados. Esses dados consistem em exemplos anotados manualmente por especialistas, que definem a classe ou a intenção correta associada a cada instância textual. Apesar da ampla disponibilidade, a rotulagem desses dados para treinamento de modelos de ML e DL é frequentemente limitada devido ao custo e ao tempo necessários para a anotação precisa (Ezen-Can, 2020). Essa limitação impacta diretamente o desenvolvimento e a aplicação eficaz de algoritmos de PLN, que necessitam de grandes quantidades de dados rotulados para alcançar um desempenho ideal (Wu et al., 2024).

Em cenários como esse, torna-se comum trabalhar com bases consideradas pequenas, já que a quantidade de exemplos anotados é insuficiente diante da complexidade da tarefa ou do número de variáveis envolvidas (Kokol; Kokol; Zagoranski, 2022). A distinção entre *big data* e bases pequenas não é absoluta e sim, relativa. Além disso, depende tanto do volume de informações disponíveis quanto da forma como elas são geradas, enquanto alguns domínios produzem dados de maneira automatizada e em grande escala, outros dependem de processos manuais ou especializados, o que reduz significativamente o número de amostras rotuladas (Xu et al., 2023). Como resultado, modelos de ML e DL encontram mais dificuldade para identificar padrões e generalizar adequadamente quando treinados com conjuntos de dados reduzidos (Kokol; Kokol; Zagoranski, 2022).

Esse contexto se manifesta de forma particularmente evidente em tarefas de análise de intenção textual em domínios especializados, como o setor de petróleo e gás. Nesse caso, o problema abordado neste estudo está relacionado à identificação da intenção presente em mensagens curtas, tais como consultas, que podem envolver preços, acontecimentos de mercado ou conceitos técnicos. A presença de linguagem técnico-financeira específica, aliada à similaridade semântica entre diferentes tipos de mensagens, faz com que intenções distintas sejam frequentemente expressas de

forma semelhante, o que dificulta sua correta identificação, especialmente quando se dispõe de conjuntos de dados rotulados limitados.

Além disso, a quantidade reduzida de textos previamente classificados nesse domínio representa uma limitação adicional. A rotulagem dessas mensagens depende de conhecimento especializado e demanda tempo, o que restringe o volume de dados disponível para o treinamento dos modelos. Como consequência, métodos de classificação tendem a apresentar dificuldades para generalizar adequadamente, especialmente em cenários com poucos dados e vocabulário técnico específico. Essas características tornam o problema de classificação de intenções no setor de petróleo e gás particularmente desafiador e evidenciam a necessidade de investigar abordagens mais adequadas a esse contexto.

A natureza técnica e específica dos dados de previsão de preços do petróleo requer abordagens adaptadas para lidar com a complexidade e a especialização do setor (Gomes et al., 2021). Superar esses desafios requer o desenvolvimento contínuo de metodologias robustas e eficazes, combinando conhecimentos de linguística, ciência da computação e outras disciplinas relacionadas (Wu et al., 2024). Investir nesse desenvolvimento é importante para avançar no campo do PLN e melhorar a capacidade dos sistemas de entender e interagir com a linguagem humana de maneira precisa e eficaz.

1.1 OBJETIVO

1.1.1 Objetivo Geral

Avaliar a eficácia de uma abordagem de análise de intenção de textos aplicada ao setor de petróleo e gás, considerando cenários com conjuntos de dados pequenos.

1.1.2 Objetivos Específicos

Para atingir o objetivo geral, tem-se como objetivos específicos:

- Desenvolver um modelo de análise de intenção de textos baseado no modelo BERT, voltado ao domínio técnico-financeiro do setor de petróleo e gás;

- Construir uma base de dados rotulada de intenções textuais voltada ao setor de petróleo e gás, adequada ao treinamento e à avaliação de modelos de análise de intenção em cenários com dados limitados.

1.2 JUSTIFICATIVA

Na indústria de óleo e gás, a análise de intenções textuais tornou-se uma necessidade devido à complexidade e ao aumento do volume de dados técnicos disponíveis. Neste contexto, o reconhecimento de intenções textuais desempenha um papel importante em sistemas de pergunta e resposta (Liu; Xu, 2021). Para esse fim, técnicas de classificação de texto, usadas para classificação de intenções, são frequentemente utilizadas, incluindo métodos baseados em regras, métodos estatísticos e abordagens de DL, cada uma com seus próprios resultados e limitações (Wu et al., 2024).

Métodos baseados em regras e estatísticas enfrentam desafios como a falta de dados e a dificuldade em identificar padrões e relações contextuais em textos curtos (Liu; Xu, 2021). Em contraste, as abordagens de DL, como o BERT, têm mostrado avanços ao oferecer representações automáticas e semanticamente ricas (Wu et al., 2024). No entanto, o BERT também enfrenta desafios ao lidar com poucos dados e dados especializados, o que indica a necessidade de aprimoramentos para melhorar a interpretação em contextos específicos (Ezen-Can, 2020).

Nesse contexto, iniciativas de pesquisa aplicada permitem observar de forma concreta como essas limitações se manifestam em aplicações reais. O desenvolvimento desta tese está associado ao projeto *Analytics Knowledge Hub* (AKHub), um projeto de Pesquisa, Desenvolvimento e Inovação conduzido em parceria entre a Universidade Federal do Paraná (UFPR) e a ExxonMobil, no âmbito das cláusulas de investimento da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). O projeto teve como objetivo a análise integrada de informações técnico-financeiras do setor de petróleo e gás para apoio à tomada de decisão.

No escopo do AKHub, foram desenvolvidas soluções computacionais voltadas à previsão de preços, à análise de sentimento em textos noticiosos e à integração de séries temporais de mercado com informações textuais provenientes de notícias especializadas. Esses resultados foram consolidados em dashboards analíticos,

permitindo a visualização conjunta de indicadores econômicos, séries temporais e informações textuais. Como etapa complementar, o projeto previu o desenvolvimento de um *chatbot* como ferramenta de apoio, no qual a análise de intenção se mostrou necessária para a correta interpretação de consultas textuais curtas relacionadas a preços, notícias e conceitos técnicos do domínio.

Durante a execução do projeto, observou-se que a aplicação de modelos de análise de intenção em textos técnico-financeiros enfrenta limitações relevantes, especialmente em cenários com dados rotulados reduzidos e linguagem especializada. Evidências empíricas reforçam esse comportamento: Ezen-Can (2020), ao comparar o desempenho do BERT e do LSTM em um conjunto de dados pequeno (contendo 150 classes de intenção, cada uma com 100 exemplos de treinamento, além de 20 exemplos de validação e 30 de teste por classe) para classificação de intenções, observou que o LSTM apresentou maior precisão e menor custo computacional, indicando limitações do BERT nesse tipo de cenário. De forma geral, a literatura recente aponta para a necessidade de metodologias capazes de aprimorar o desempenho de modelos baseados em BERT em contextos com dados limitados e domínios especializados (Ezen-Can, 2020; Wu et al., 2024).

Nesse sentido, a adaptação de modelos de linguagem a domínios específicos tem sido apontada como uma estratégia relevante para aumentar a confiabilidade das análises textuais e apoiar processos de tomada de decisão em aplicações reais (Huang et al., 2023; Ayadi; Lachiri, 2023). Assim, esta tese busca abordar essa lacuna por meio da investigação e adaptação de modelos baseados em BERT para a classificação de intenções em textos técnico-financeiros do setor de petróleo e gás.

O estudo justifica-se também pela sua contribuição à linha de otimização teórica e aplicada do Programa de Pós-graduação em Métodos Numéricos em Engenharia da Universidade Federal do Paraná (UFPR), ao grupo de tecnologia aplicada à otimização (GTAO) e ao projeto de pesquisa *Analytics Knowledge Hub* (AKHub).

1.3 CONTRIBUIÇÕES

As principais contribuições deste trabalho podem ser sintetizadas em quatro pontos:

1. Desenvolvimento do modelo OilFinBERT, uma adaptação do BERT para o domínio técnico-financeiro do setor de petróleo e gás, com pré-treinamento contínuo em conjunto de dados especializado.
2. Construção de um base de dados balanceado de intenções em cinco classes, especificamente desenhado para cenários de chatbot no setor.
3. Avaliação sistemática de técnicas de fine-tuning (reinicialização de camadas, Mixout, SWA e LLRD) e otimizadores (AdamW, AdamP e SGD), gerando evidências sobre suas vantagens e limitações.
4. Comparação abrangente com modelos clássicos (NB, CNN, RNN, LSTM e BERT base), situando a proposta no contexto das alternativas disponíveis.

1.4 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em seis capítulos principais.

No Capítulo 2 apresenta a descrição formal do problema de classificação de intenções no domínio técnico-financeiro do petróleo e gás. São definidos os elementos fundamentais do modelo supervisionado — representação de entrada, espaço de rótulos, funções de classificação, objetivo de otimização e métricas.

O Capítulo 3 é dividido em três subcapítulos: PLN, Análise de intenção e Trabalhos Relacionados. No primeiro é abordado o PLN, apresentando uma visão geral das técnicas utilizadas para representar texto de forma que possa ser processado por máquinas. Na segunda seção foca-se na análise de intenção no contexto de PLN, explorando duas abordagens principais: aprendizado de máquina e aprendizado profundo. Na terceira seção são revisados os trabalhos relacionados ao tema, com uma análise da literatura existente e das contribuições práticas das técnicas discutidas anteriormente.

No Capítulo 4 apresenta-se a proposta metodológica adotada nesta pesquisa. Detalha-se o fluxo de trabalho da pesquisa, dividido em etapas que incluem a seleção do modelo, definição do *dataset* para pré-treinamento, construção do *dataset* de classificações de intenções (proposto pela tese, denominado OinFinInten),

implementação dos modelos comparativos e, por fim, o pré-treinamento e *fine-tuning*⁴ supervisionado do modelo proposto pela tese, OilFinBERT.

O Capítulo 5 é dedicado à apresentação e análise dos resultados. Inicialmente, descrevem-se as ferramentas computacionais utilizadas e o processo de prototipagem do *pipeline*⁵. Em seguida, são discutidos os experimentos conduzidos, comparando o desempenho dos diferentes modelos e das variações de regularização e ajustes aplicadas. Além dos resultados quantitativos, são apresentadas análises críticas, incluindo a avaliação das métricas de desempenho, a robustez dos modelos frente a dados escassos e as limitações observadas. Por fim, o capítulo contempla uma análise de falhas, destacando cenários em que os modelos apresentaram dificuldades e propõe recomendações para ajustes futuros, alinhando as descobertas práticas às hipóteses iniciais.

Por fim, o Capítulo 6 apresenta as conclusões do estudo, destacando as contribuições obtidas, as limitações encontradas e as recomendações para trabalhos futuros. Além disso, são listadas as publicações decorrentes desta pesquisa, evidenciando o impacto científico e prático dos resultados alcançados.

⁴ *Fine-tuning* é o processo de ajustar um modelo pré-treinado em uma nova tarefa específica, atualizando seus pesos com base em um conjunto de dados rotulado menor e especializado.

⁵ Um *pipeline* é uma sequência organizada de etapas pelas quais os dados passam até gerar uma saída final.

2 DESCRIÇÃO DO PROBLEMA

A tarefa de classificação de intenções em textos técnico-financeiros consiste em mapear mensagens curtas para um conjunto de categorias previamente definidas. Para isso, é necessário caracterizar os elementos envolvidos no processo de modelagem — desde a natureza dos textos, passando pela representação da entrada, pelo espaço de rótulos utilizado, até as funções de classificação e as métricas que permitem avaliar o desempenho dos modelos.

Nesse estudo, foi proposto um classificador de intenção, isto é, um modelo supervisionado cujo objetivo é identificar a intenção expressa em uma mensagem e atribuí-la a uma classe predefinida. Esse classificador é implementado, no modelo principal deste trabalho, pelo OilFinBERT, um *Transformer*⁶ especializado para o domínio técnico-financeiro do setor de petróleo e gás.

2.1 ENTRADA

As entradas do sistema são mensagens curtas, frequentemente compostas por poucas palavras ou pequenas frases. Essas mensagens reproduzem consultas típicas de usuários interessadas em informações sobre o setor de petróleo e gás, com menções a preços, notícias, eventos de mercado e conceitos técnicos especializados.

As mensagens utilizadas no treinamento supervisionado foram organizadas no conjunto OilFinIntent, criado pela autora, o qual formaliza o tipo de entrada considerado no problema.

Formalmente, temos x o espaço de texto, que representa todas as possíveis entradas aceitas pelo modelo, $y = \{\text{previsao_preço, sentimento_noticia, preço_atual, conceito, outros}\}$ o conjunto de rótulos possíveis e $D = \{(x_i, y_i)\}_{i=1}^N$, em que cada texto recebeu um rótulo atribuído por um anotador humano. O objetivo é aprender uma função de decisão (Equação 1) que mapeia elementos de X em Y , estimando a classe mais provável para cada nova mensagem.

⁶ *Transformers*: Arquitetura de redes neurais proposta por Vaswani et al. (2017), baseada em mecanismos de atenção, capaz de modelar relações de dependência entre *tokens* independentemente da distância na sequência.

$$F_{\theta}: X \rightarrow Y \quad (1)$$

parametrizada por θ , que modela $\Pr(y | x; \theta)$ e, durante o treinamento, é ajustada para minimizar uma função de perda supervisionada sobre D . A função retorna a classe mais provável $\hat{y} = \arg \max P(y | x; \theta)$. O desempenho é avaliado por meio de quatro métricas: F1-score, recall, precisão e PR-AUC.

2.2 REPRESENTAÇÃO DE ENTRADA

Cada texto de entrada x passa, inicialmente, pelo processo de tokenização, etapa na qual a mensagem é segmentada em unidades menores chamadas *tokens*, que podem corresponder a palavras, subpalavras ou símbolos especiais utilizados pelo modelo *Transformer* (Equação 2).

$$x = (t_1, t_2, \dots, t_T) \quad (2)$$

Em que T é o número total de *tokens*.

Cada *token* t_i é então convertido em um vetor denso, denominado *embedding*⁷, que projeta o *token* em um espaço vetorial de dimensão d , preservando relações semânticas relevantes (Equação 3).

$$e_i = E(t_i), \quad \text{com } e_i \in \mathbb{R}^d \quad (3)$$

Esses vetores iniciais são processados pelo *Encoder Transformer*⁸, conjunto de camadas responsáveis por modelar dependências de longo alcance na sequência por meio do mecanismo de atenção. O resultado são representações contextuais h_i , que incorporam informações sobre a posição do *token* na frase e sua relação com os demais elementos do texto (Equação 4).

$$H = (h_1, h_2, \dots, h_T), \quad \text{com } h_i \in \mathbb{R}^d \quad (4)$$

⁷ *Embeddings*: vetores densos que representam cada *token* em um espaço n-dimensional, codificando relações semânticas e sintáticas

⁸ *Encoder Transformer*: Arquitetura baseada em mecanismos de atenção, responsável por processar a sequência de *embeddings* e capturar dependências entre tokens, independentemente de sua posição no texto.

Para obter uma representação única da mensagem, aplica-se uma operação de agregação (*pooling*⁹). Neste trabalho, foram consideradas duas alternativas:

- O vetor especial [CLS]¹⁰: $z = h_{[CLS]}$
- A média dos vetores contextuais: $z = \frac{1}{T} \sum_{i=1}^T h_i$

O vetor $z \in \mathbb{R}^d$ representa a mensagem de forma compacta e é utilizado pela camada de classificação para prever a intenção.

2.3 ESPAÇO DE RÓTULOS

O espaço de saída é composto pelas cinco categorias definidas no OilFinIntent, as quais refletem intenções típicas do domínio técnico-financeiro do petróleo e gás. Essas categorias apresentam proximidade semântica, o que afeta a dificuldade da tarefa e motiva o uso de estruturas de classificação plana ou hierárquica.

A definição do espaço de rótulos apresenta particularidades específicas do domínio estudado. As intenções consideradas estão diretamente relacionadas ao comportamento do preço do petróleo, à interpretação de notícias de impacto financeiro e à consulta de conceitos técnicos utilizados no mercado de energia. Essa estrutura difere de cenários gerais de detecção de intenção, nos quais as categorias são mais amplas e menos dependentes de vocabulário especializado. Essas características definem a especificidade do problema analisado.

2.4 MODELAGEM

Dois esquemas são considerados nesse estudo:

(i) Multiclasse “plana”

Nesse caso, a distribuição condicional $P(y|x)$ é estimada diretamente sobre as cinco classes em y , definido como o conjunto de categorias possíveis de intenção.

⁹ *Pooling*: operação que reduz uma sequência de vetores para uma única representação, por meio de procedimentos como média, soma ou seleção de um token especial.

¹⁰ [CLS]: *Token* especial introduzido no início da sequência em modelos *BERT-like*, usado como representação agregada da mensagem para tarefas de classificação

(ii) Hierárquica em quatro níveis

Aqui, adota-se uma estrutura em árvore, na qual as decisões são tomadas de forma sequencial. Cada nó $g_n(x)$ corresponde a um classificador binário responsável por avaliar uma categoria específica. A árvore segue a seguinte organização:

- Nó $g_1(x)$: verifica se a mensagem pertence à classe `previsao_preco`.
- Nó $g_2(x)$: se não for previsão de preço, verifica se pertence à classe `preco_atual`.
- Nó $g_3(x)$: se não for preço atual, verifica se é `sentimento_noticia`.
- Nó $g_4(x)$: classifica entre as duas classes restantes: {conceito, outros}.

A predição final é dada pelo caminho de maior probabilidade ao longo da árvore (Equação 5).

$$\hat{y} = \arg \max \prod_{n \in \text{caminho}(y)} P_n(\text{decisão} | x) \quad (5)$$

2.5 FUNÇÃO DE CLASSIFICAÇÃO

Na abordagem multiclasse plana, utiliza-se uma camada linear seguida da função *softmax*, que transforma o vetor de pontuações $Wz + b$ em uma distribuição de probabilidades sobre as classes (Equação 6).

$$P(y | x; 0) = \text{softmax}(Wz + b) \quad (6)$$

Na abordagem hierárquica, a classificação é estruturada como uma árvore de decisão, na qual cada nó aplica uma função de decisão binária para determinar se a mensagem pertence ou não a um grupo específico de classes. Assim, a predição final é obtida pela combinação sequencial das probabilidades estimadas ao longo do caminho correspondente à classe prevista, resultando em um processo de decisão mais estruturado e adequado a relações internas entre as categorias.

2.6 OBJETIVO DE OTIMIZAÇÃO

O treinamento supervisionado busca minimizar a *entropia cruzada ponderada*, utilizada para lidar com o desbalanceamento entre classes (Equação 7). Essa função de perda penaliza mais fortemente erros em classes menos frequentes, por meio dos pesos w_c , permitindo que o modelo aprenda de forma mais equilibrada.

$$\mathcal{L}(\theta) = - \sum_{i=1}^N \sum_{c \in Y} w_c q_i(c) \log P(c | x_i; \theta) \quad (7)$$

Em que w_c representa o peso atribuído a cada classe, ajustado de acordo com sua frequência no conjunto de dados; e q_i é a distribuição alvo, podendo assumir forma *one-hot*¹¹ ou incorporar *label smoothing*¹², técnica que reduz *overfitting*¹³ ao suavizar a probabilidade do rótulo correto.

Além disso, quando é empregado um pré-treinamento adicional utilizando uma base de dados não rotulada do setor de petróleo e gás, incorpora-se a tarefa de modelagem de linguagem mascarada (*Masked Language Modeling* — MLM), em que o modelo aprende a prever *tokens* ocultos no texto. Nesses casos, o ajuste ocorre em duas fases distintas:

1. Pré-treinamento contínuo (*continued pretraining*), no qual o modelo aprimora suas representações linguísticas específicas do domínio;
2. *Fine-tuning* supervisionado, etapa em que os parâmetros são refinados para a tarefa de classificação de intenções.

Essa combinação permite que o modelo obtenha representações mais adequadas ao domínio técnico-financeiro, resultando em melhor desempenho quando treinado com bases rotuladas reduzidas.

2.7 SAÍDAS E MÉTRICAS

¹¹ *One-hot encoding*: Representação vetorial em que cada classe é codificada como um vetor binário com valor 1 apenas na posição correspondente à classe correta.

¹² *Label smoothing*: Técnica de regularização que reduz o *overfitting* suavizando a distribuição-alvo, substituindo o rótulo “duro” (1.0) por um valor próximo, como 0.9, distribuindo o restante entre as classes incorretas.

¹³ *Overfitting*: Fenômeno em que o modelo aprende padrões específicos do conjunto de treinamento, perdendo capacidade de generalização para dados novos.

O sistema do sistema, para cada mensagem x , a classe prevista \hat{y} e a distribuição de probabilidades $P_{\theta}(x | y)$. A avaliação é feita com quatro métricas, amplamente empregadas na literatura de aprendizado de máquina. A precisão (*Precision*) mede a proporção de instâncias corretamente classificadas como positivas em relação ao total de predições positivas (Equação 8), em que TP (*True Positives*) corresponde aos verdadeiros positivos e FP (*False Positives*) aos falsos positivos.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

O *Recall* (ou sensibilidade) quantifica a proporção de exemplos positivos corretamente recuperados pelo modelo (Equação 9), onde FN (*False Negatives*) representa os falsos negativos.

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

A partir dessas duas métricas, calcula-se o *F1-score* (Equação 10), definido como a média harmônica entre precisão e *Recall*, penalizando mais fortemente situações em que há desequilíbrio entre ambas. Essa métrica foi empregada como principal indicador de comparação, tanto em sua forma *macro* — média simples entre as classes — quanto *weighted*, ponderada pela frequência de cada classe no conjunto de dados.

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (10)$$

Além disso, utilizou-se a *Precision-Recall Area Under Curve* (PR-AUC), calculada pela integral da curva *Precision-Recall* ao longo de diferentes limiares de decisão. Na prática, essa integral é aproximada pela métrica de *Average Precision* (AP), conforme recomendado por Saito e Rehmsmeier (2015). A PR-AUC é especialmente adequada em cenários desbalanceados¹⁴, pois enfatiza a relação entre recuperar exemplos positivos reais e manter o nível de precisão das predições.

¹⁴ Desbalanceamento entre classes: condição em que algumas classes possuem número menor de exemplos do que outras, afetando a capacidade do modelo de aprender padrões representativos

Assim, a combinação de precisão, *Recall*, *F1-score* e *PR-AUC* permite avaliar o desempenho dos modelos sob múltiplas perspectivas, contemplando tanto a qualidade global da classificação quanto o comportamento em bases desbalanceadas e classes específicas.

2.8 RESTRIÇÕES DE CENÁRIO

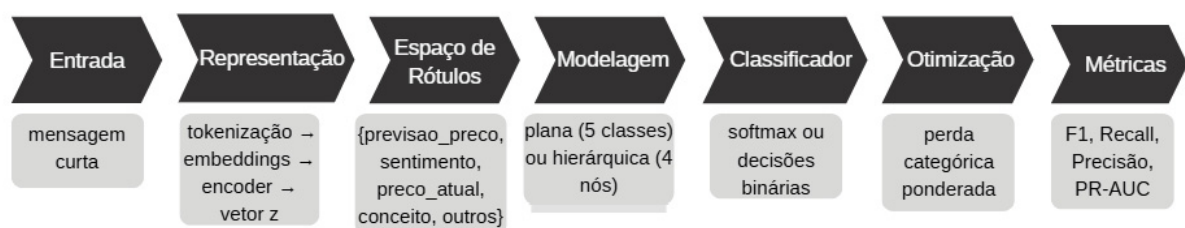
O problema de classificação de intenções apresenta algumas condições específicas:

- Dados escassos: embora as classes tenham a mesma quantidade de exemplos, o conjunto de dados é reduzido em tamanho total, o que pode limitar a capacidade de generalização do modelo.
- Mensagens curtas: frases pequenas podem gerar incerteza entre classes próximas.
- Vocabulário especializado: termos financeiros e técnicos do setor de petróleo e gás concentram significados específicos, o que aumenta a necessidade de representação adequada desse domínio.
- Limite de contexto: há um comprimento máximo de entrada, sendo necessário truncar mensagens longas.

Essas restrições precisam ser consideradas no treinamento e na avaliação do modelo.

As etapas descritas ao longo deste capítulo podem ser integradas em um esquema único que representa o problema de forma consolidada. A Figura 1 apresenta esse resumo, ilustrando a organização lógica da tarefa de classificação de intenções utilizada neste estudo.

Figura 1 - Fluxo conceitual adotado para representar o problema



Fonte: A autora (2025)

3 REVISÃO DE LITERATURA

Neste capítulo explora-se conceitos e técnicas para o desenvolvimento de sistemas de análise de intenção baseados em PLN. Inicialmente, apresenta-se uma discussão a representação de texto, um componente que visa converter textos em formatos adequados para processamento computacional, explorando métodos como *Bag of Words* (BoW) e *Term Frequency-Inverse Document Frequency* (TF-IDF), bem como abordagens mais recentes como *Word Embeddings* e representações contextuais baseadas em BERT.

Na sequência, abordar-se a análise de intenção, uma tarefa em aplicações que envolvem a compreensão e categorização de intenções humanas a partir de textos, como assistentes virtuais e *chatbots*. Esse tópico foi desdobrado em duas abordagens: ML e DL. No contexto de ML, serão exploradas técnicas como NB, SVM e *k-Nearest Neighbor* (k-NN). Em contrapartida, no âmbito do DL, daremos ênfase a técnicas como CNN, RNN, LSTM, e ao modelo BERT.

Cada um desses tópicos foi abordado nas seções subsequentes, proporcionando uma compreensão das técnicas e metodologias empregadas no desenvolvimento de sistemas de PLN e análise de intenção.

3.1 PROCESSAMENTO DE LINGUAGEM NATURAL

O PLN é uma área que combina computação e linguística para permitir que máquinas compreendam e produzam linguagem humana de maneira eficaz (Aslam, 2023). A abrangência do PLN vai desde tarefas simples, como a correção ortográfica, até aplicações mais complexas, como tradução automática, geração de texto, análise de sentimentos e sistema de diálogo (Kowsari et al., 2019).

A análise de textos pode oferecer ideias valiosas para otimizar a eficiência operacional, melhorar a segurança e fortalecer o relacionamento com os clientes (Liu; Qiu; Huang, 2016; Wu et al., 2024). A capacidade de interpretar corretamente as intenções expressas em perguntas enviadas a sistema de diálogo, por exemplo, pode transformar a maneira como as empresas interagem com seus *stakeholders*, oferecendo respostas mais precisas e relevantes.

Para alcançar esses objetivos, o PLN faz uso de um conjunto de técnicas avançadas. Entre elas, destacam-se os algoritmos de ML, as *deep neural networks* e

os procedimentos de pré-processamento que preparam os dados brutos para análise. *Deep neural networks* são caracterizadas pela presença de múltiplas camadas ocultas, o que lhes permite aprender representações hierárquicas e identificar relações linguísticas complexas, constituindo a base dos modelos modernos como BERT e GPT (Kowsari et al., 2019).

Essas técnicas se articulam com um conjunto fundamental de operações de pré-processamento de texto, etapa responsável por transformar o texto bruto — isto é, o conteúdo original coletado antes de qualquer tratamento computacional — em dados consistentes e adequados para modelagem. Entre as etapas mais utilizadas, destacam-se: a limpeza do texto, que remove ruídos como caracteres especiais, múltiplos espaços, URLs e símbolos desnecessários; a normalização, que padroniza o texto por meio de ações como converter caracteres para minúsculas ou unificar variações de escrita; a tokenização, que segmenta o texto em unidades menores (palavras, subpalavras ou sentenças); a remoção seletiva de *stopwords*, quando adequada à tarefa; e procedimentos de redução morfológica, como o *stemming*, que reduz palavras a suas raízes (Kowsari et al., 2019; Anandarajan; Hill; Nolan, 2019).

A integração entre essas tecnologias — *deep neural networks*, algoritmos de aprendizado e pré-processamento estruturado — permite que sistemas de PLN identifiquem padrões, extraiam informações contextuais e ofereçam respostas mais precisas e alinhadas às necessidades dos usuários, mesmo em domínios de linguagem altamente especializada (Lee et al., 2024).

3.1.1 Representação de texto

A etapa de pré-processamento é o ponto de partida do PLN, pois prepara o texto para ser transformado em representações numéricas. O texto bruto contém ruídos como caracteres especiais, URLs, múltiplos espaços ou variações inconsistentes de capitalização que não contribuem para a interpretação semântica. A remoção ou padronização desses elementos melhora a qualidade dos dados (Anandarajan; Hill; Nolan, 2019).

Outras operações, como tokenização e normalização, garantem consistência ao longo do conjunto de texto. A tokenização consiste em dividir o texto em unidades menores — como palavras, subpalavras ou sentenças — que servirão como elementos básicos para análise. Já a normalização envolve procedimentos que

padronizam essas unidades, como converter todos os caracteres para minúsculas, corrigir variações ortográficas, remover acentuação ou unificar formas equivalentes de um mesmo termo. Além disso, a remoção seletiva de *stopwords* ajuda a controlar termos que, em muitos casos, pouco agregam à análise. Contudo, em tarefas como detecção de intenção, certas *stopwords* devem ser preservadas, especialmente quando indicam negação (*não, nunca*), estruturalmente definem o tipo da pergunta (*o que, quando, como*), alteram relações semânticas por meio de preposições (*no, do*) ou expressam modalidade (*pode, deve, vai*). Por exemplo, em consultas como “quando o preço do Brent deve subir?”, a palavra quando indica que o usuário não busca o valor atual, mas sim uma estimativa temporal, o que diferencia essa intenção de perguntas como “qual é o preço do Brent?”. A remoção desse termo poderia levar o modelo a classificar incorretamente a consulta (Kowsari et al., 2019).

Após o pré-processamento, inicia-se a representação vetorial do texto, etapa central no PLN, pois converte informação linguística em estruturas matemáticas adequadas para modelos de classificação e análise de intenção (Qiu et al., 2020). Técnicas tradicionais, como BoW e TF-IDF, representam documentos por meio da contagem de termos: BoW utiliza apenas a frequência das palavras, enquanto TF-IDF pondera essa frequência considerando a raridade do termo no corpus (Li; Shang; Wang, 2019). Embora simples, tais métodos geram vetores esparsos e desconsideram tanto a ordem quanto o contexto semântico, o que limita sua capacidade de capturar significados mais complexos (Zhang et al., 2024).

Com a evolução do PLN, surgiram métodos distribuídos que aprendem relações semânticas entre palavras, gerando *embeddings*, isto é, vetores densos contínuos que representam termos preservando significados. No Word2Vec (Mikolov et al., 2013), a representação é aprendida otimizando uma função de probabilidade: no modelo Skip-gram, maximiza-se a probabilidade de observar palavras de contexto dadas a palavra central, enquanto no CBOW maximiza-se a probabilidade inversa. Essa otimização é implementada por meio de *negative sampling*, que ensina o modelo a distinguir palavras corretas de palavras negativas selecionadas aleatoriamente, ou pelo *hierarchical softmax*, que utiliza uma árvore binária para acelerar o cálculo das probabilidades.

O GloVe (Pennington et al., 2014) segue outro princípio matemático: ele fatoriza a matriz global de coocorrência de palavras e minimiza o erro entre as razões

observadas e previstas dessas coocorrências. Assim, palavras que aparecem em contextos semelhantes recebem vetores próximos no espaço vetorial.

Já o FastText (Bojanowski et al., 2017) representa cada palavra como a soma dos vetores de seus n-gramas de caracteres. Isso permite capturar regularidades morfológicas e gerar representações mais robustas para palavras raras ou desconhecidas.

Diferentemente de BoW e TF-IDF, que produzem vetores esparsos baseados apenas em contagem de termos, essas abordagens geram representações contínuas que refletem propriedades semânticas e sintáticas, permitindo modelos de PLN mais precisos e generalizáveis.

Mais recentemente, modelos baseados em redes neurais profundas — arquiteturas compostas por múltiplas camadas capazes de aprender representações hierárquicas — revolucionaram o campo. Modelos como BERT e GPT utilizam mecanismos de atenção para gerar representações contextuais dinâmicas, nas quais o significado de cada palavra é ajustado conforme o contexto em que aparece (Qiu et al., 2020).

No setor de petróleo, representações textuais robustas são fundamentais para tarefas como a análise de intenção em *chatbots* corporativos, que precisam identificar nuances das perguntas enviadas pelos usuários. A escolha adequada da técnica de representação impacta diretamente a precisão das respostas e a interpretação de comandos, sobretudo em ambientes de linguagem altamente especializada (Tariq et al., 2021).

3.2 ANÁLISE DE INTENÇÃO

Na indústria de petróleo, a previsão de preços é uma tarefa sensível que depende de diversos fatores, incluindo políticas econômicas, geopolítica, e mudanças na oferta e demanda. Para apoiar stakeholders nesse contexto, podem ser empregadas técnicas avançadas de previsão de preços, enquanto um *chatbot* inteligente atua como interface interpretando as intenções dos usuários e fornecendo acesso a informações e estimativas de forma rápida e contextualizada (Hasan; Ibna Basher; Shawon, 2023).

Por exemplo, um *chatbot* projetado para interagir com *stakeholders* que buscam previsões ou ideias sobre os preços futuros do petróleo precisa entender a

intenção da pergunta do usuário. Quando um usuário faz uma consulta, como "Qual é a previsão para o preço do produto X para os próximos 180 dias?", o *chatbot* precisa identificar a intenção por trás da pergunta para fornecer uma resposta relevante e acionável.

Para reconhecer diferentes intenções nas consultas sobre o preço do petróleo, o sistema emprega um modelo de análise baseado em técnicas avançadas de PLN. Uma vez identificada a intenção, a solicitação é tratada e a resposta apropriada é apresentada. No caso de previsões de preço, o sistema apenas apresenta estimativas provenientes de modelos especializados, em vez de gerar previsões diretamente.

Diversos estudos sobre classificação de intenções têm investigado métodos para aprimorar o reconhecimento das consultas formuladas por usuários em sistemas conversacionais. A classificação de intenção consiste em categorizar entradas textuais em classes previamente definidas, permitindo que o sistema identifique o propósito da mensagem (Hasan; Ibna Basher; Shawon, 2023; Chen; Zhuo; Wang, 2019).

Nesse sentido, o reconhecimento de intenção ocorre pela análise de textos em linguagem natural dos usuários para determinar o propósito por trás de suas mensagens, resultando em respostas ou ações adequadas (Wu et al., 2024; Huang et al., 2023). Esses textos são contextualmente dependentes, influenciados por diálogos anteriores ou subsequentes, exigindo consideração e utilização eficiente das informações contextuais durante o processo de reconhecimento de intenção (Wu et al., 2024).

A abordagem para classificar intenções segue os princípios da classificação de texto, integrando as intenções identificadas ao sistema de diálogo para gerar respostas adequadas pelo *chatbot* (Hasan; Ibna Basher; Shawon, 2023).

Há diversos métodos existentes para a classificação de intenção, como os modelos de ML: NB, SVM, Regressão Logística; e os modelos de DL: CNN, RNN, BERT (Otter; Medina; Kalita, 2021). Alguns desses modelos são apresentados nas próximas sessões.

3.2.1 Machine Learning

O ML refere-se à habilidade dos sistemas de utilizar dados de treinamento específicos do problema para automatizar a criação de modelos analíticos e solucionar tarefas relacionadas (Janiesch; Zschech; Heinrich, 2021).

Os algoritmos de ML podem ser classificados em supervisionados, não supervisionados e semissupervisionados (Luo, 2021). No aprendizado supervisionado, a classificação é realizada com base em dados rotulados; no aprendizado não supervisionado, são utilizados dados não rotulados; e no aprendizado semissupervisionado, ambos os tipos de dados são combinados para tratar grandes volumes de dados complexos (Thangaraj; Sivakami, 2018; Ayadi; Lachiri, 2023).

Dentro dessas categorias, técnicas como SVM, NB e k-NN são empregadas na classificação de textos em cenários de aprendizado supervisionado, nos quais a disponibilidade de dados rotulados é necessária para o processo de treinamento dos modelos (Palanivinayagam; El-Bayeh; Damaševičius, 2023).

Entre essas técnicas, o SVM é um método de aprendizado de máquina aplicado em tarefas de classificação e regressão, baseado no conceito de *Maximum Margin Hyperplane* (MMH) para separação linear e no uso de funções *kernel* para casos não lineares. O modelo é amplamente empregado em cenários de classificação binária (Hasan; Ibna Basher; Shawon, 2023).

Por outro lado, o k-NN classifica um documento de teste encontrando os k documentos mais próximos no conjunto de treinamento e atribuindo a categoria mais frequente entre esses vizinhos como a classe do documento de teste. A similaridade entre o documento de teste e cada vizinho determina a pontuação de cada categoria, sendo a classe com a maior pontuação atribuída ao documento de teste (Kowsari et al., 2019).

Neste trabalho, será utilizada a técnica NB devido à sua eficácia na classificação de texto, que será abordada a seguir.

3.2.1.1 NAIVE BAYES

O modelo NB é popularmente conhecido como um método elementar e relativamente eficiente e eficaz para algoritmos de mineração de dados, mais especificamente, a classificação de texto (Chen; Zhuo; Wang, 2019; Jiang et al., 2016). O modelo fundamenta-se no teorema bayesiano, adotando a suposição de

independência condicional entre os preditores (Diab; El Hindi, 2017). Na prática, essa suposição é uma aproximação: embora raramente verdadeira em dados textuais, ela costuma produzir estimativas razoáveis o suficiente para resultar em bom desempenho preditivo, especialmente quando a representação do texto e o alisamento são adequados (Mccallum; Nigam, 1998)

Conforme Mccallum e Nigam (1998), a decisão do classificador NB pode ser expressa pela Equação 1.

$$\hat{c} = \mathit{arg} \max_{c \in C} P(c) \prod_{i=1}^n P(x_i | c) \quad (1)$$

Em que \hat{c} é a classe escolhida, $P(c)$ e a probabilidade a priori da classe c , $P(x_i | c)$ é a probabilidade condicional do atributo x_i dado a classe c e n é o número de atributos ou termos do documento.

Essa formulação assume que os atributos x_i são condicionalmente independentes entre si, dado c . Embora essa hipótese raramente seja verdadeira em textos (onde há forte correlação entre palavras), o modelo ainda apresenta resultados satisfatórios em muitas tarefas.

Mccallum e Nigam (1998) também distinguem dois modelos de evento para textos:

- Modelo Multinomial (MNB)

Neste caso, considera-se a frequência das palavras em um documento. A probabilidade de um documento d , dado que ele pertence a uma classe c , é calculada como mostra a Equação 2.

$$P(d | c) = P(c) \prod_{k=1}^{|V|} P(w_k | c)^{n_{k,d}} \quad (2)$$

Em que $|V|$ é o tamanho do vocabulário, w_k é a k -ésima palavra do vocabulário e $n_{k,d}$ é o número de ocorrências da palavra w_k no documento d . Esse modelo é amplamente utilizado para classificação de texto, pois leva em conta a frequência de termos e não apenas sua presença.

- Modelo Multivariado de Bernoulli (BNB)

Considera apenas a presença ou ausência da palavra, ignorando sua frequência. A probabilidade é calculada pela Equação 3.

$$P(d | c) = P(c) \prod_{k=1}^{|V|} P(w_k | c)^{I(w_k \in d)} (1 - P(w_k | c))^{1-I(w_k \in d)} \quad (3)$$

Onde $I(w_k \in d)$ é uma função indicadora, que vale 1 se a palavra w_k estiver presente no documento d e 0 caso contrário. Esse modelo é mais simples, mas pode ser menos eficaz em textos longos, pois ignora a frequência das palavras.

Palanivinayagam, El-Bayeh e Damaševičius (2023) apresentam uma série de vantagens e desvantagens, ao utilizar o modelo NB para classificação de texto, conforme apresentado na Tabela 1.

Tabela 1 - Vantagens e Desvantagens do NB

Vantagens	Desvantagens
<ul style="list-style-type: none"> - Se o conjunto de treinamento for pequeno, então o NB pode produzir um bom desempenho; - Pode ser usado para classificação multiclasse; - Não requer muitos dados de treinamento. 	<ul style="list-style-type: none"> - Não é adequado para pequenos conjuntos de dados; - Se as características não forem independentes, NB não é a melhor escolha para classificação.

Fonte: Adaptado de Palanivinayagam *et al.* (2023)

No entanto, há uma contradição aparente na literatura sobre o desempenho do NB em conjuntos de dados pequenos. Enquanto Palanivinayagam, El-Bayeh e Damaševičius (2023) sugerem que o NB pode produzir bons resultados com um número reduzido de amostras, essa vantagem é limitada a cenários onde as características são independentes. Quando essa condição não é satisfeita, o desempenho do NB pode ser comprometido, conforme observado em Milosevic, Dehghantanha e Choo (2017). Portanto, a adequação do NB para conjuntos de dados pequenos depende fortemente das características específicas do problema.

Mccallum e Nigam (1998) destacam o uso de técnicas de suavização, como Laplace ou Lidstone. Sem esse ajuste, a probabilidade atribuída a uma palavra não observada no conjunto de treinamento é igual a zero, o que resulta na anulação da probabilidade do documento inteiro. A suavização atribui uma probabilidade mínima a

termos raros ou ausentes, evitando esse efeito e permitindo o cálculo da distribuição probabilística pelo classificador.

O modelo NB utiliza um conjunto de dados de treinamento para calcular os termos de probabilidade necessários para realizar classificações. A eficácia do algoritmo depende da obtenção precisa desses termos de probabilidade, o que pode se tornar desafiador, especialmente em situações em que os dados de treinamento são limitados, como geralmente ocorre (Diab; El Hindi, 2017; Ayadi; Lachiri, 2023).

Na literatura de PLN, o NB é frequentemente utilizado como método de referência em experimentos, justamente por sua simplicidade, baixo custo computacional e facilidade de implementação. Esse papel de referência é importante, pois permite comparar o ganho real proporcionado por modelos mais sofisticados, como CNN, RNN ou BERT (Hasan; Ibna Basher; Shawon, 2023).

Além disso, em contextos de textos curtos e especializados, como no domínio técnico-financeiro do setor de petróleo e gás, a suposição de independência entre palavras pode ser ainda mais restritiva. Termos como “preço do barril” ou “*spread* do *Brent*” aparecem de forma conjunta e correlacionada, o que limita a capacidade do NB de capturar essas relações semânticas. Nessas situações, o modelo tende a ser superado por abordagens mais avançadas, mas ainda assim fornece uma base inicial de comparação confiável (Kowsari et al., 2019).

Além do modelo original NB, há diversas variações do modelo, como NB Multinomiais (MNB), NB Multivariado de Bernoulli (BNB), NB Complementar (CNB) e NB um-contra-todos-menos-um (OVA) (Palanivinayagam; El-Bayeh; Damaševičius, 2023).

3.2.2 Deep Learning

O DL é uma abordagem de ML que utiliza redes neurais com múltiplas camadas, sendo formalmente consolidada por LeCun, Bengio e Hinton (2015). Essas redes empregam técnicas de aprendizado supervisionado e não supervisionado para automaticamente descobrir representações latentes em diferentes níveis de complexidade (Selmy; Mohamed; Medhat, 2024).

Modelos de DL frequentemente adotam a arquitetura sequência a sequência, onde um codificador transforma a entrada em vetores compactos e um decodificador

gera o texto final com base nesses vetores, proporcionando melhorias notáveis em comparação com abordagens estatísticas anteriores (Li et al., 2024).

Uma das principais vantagens do DL em PLN é sua capacidade de aprender diretamente a partir dos dados brutos, minimizando a necessidade de pré-processamento extensivo (Selmy; Mohamed; Medhat, 2024). Nesse contexto, modelos baseados na arquitetura Transformers utilizam mecanismos de atenção para processar sequências textuais sem recorrer a recorrência ou convoluções, permitindo que cada *token* seja contextualizado em relação aos demais elementos da sequência (Vaswani et al., 2017). Essa estrutura possibilita a modelagem de relações contextuais e semânticas em tarefas como classificação de textos, sendo amplamente empregada em estudos recentes (Wu et al., 2024; Li et al., 2024).

A evolução das arquiteturas de redes neurais, como as CNNs, RNNs, redes LSTM, e modelos baseados em *transformers* como o BERT, tem ampliado significativamente o potencial do DL para análise de intenção (Lee et al., 2024). Cada uma dessas arquiteturas tem suas próprias características e vantagens, tornando-as adequadas para diferentes aspectos da classificação de textos.

As CNNs são particularmente eficazes na captura de padrões locais e n-gramas, as RNNs são úteis para dados sequenciais e dependências temporais, enquanto as LSTMs são projetadas para preservar informações ao longo de longas sequências, superando limitações das RNNs tradicionais (Otter; Medina; Kalita, 2021). O BERT, por outro lado, utiliza uma abordagem bidirecional que considera o contexto de todas as palavras na sentença simultaneamente, resultando em uma compreensão mais rica e precisa do texto (Devlin et al., 2019).

Modelos de *DL* têm sido aplicados em tarefas de reconhecimento de intenções por aprenderem padrões presentes em sequências textuais (Lin; Huang; Yang, 2023). Essas abordagens empregam aprendizado por transferência, mecanismos de atenção e arquiteturas como LSTM e BERT para gerar representações contextuais sem dependência de regras manuais. Entretanto, o desempenho permanece associado ao tamanho e à qualidade do conjunto de dados de treinamento, o que pode restringir sua utilização em bases reduzidas (Deng et al., 2024).

Embora o DL ofereça vantagens significativas, como maior precisão e capacidade de identificar relações complexas em textos, ele também demanda alto custo computacional e grandes volumes de dados rotulados. Esse cenário representa

uma limitação em domínios especializados, como o setor de petróleo e gás, onde os conjuntos de dados são frequentemente reduzidos. Assim, o uso de DL precisa ser complementado por estratégias de adaptação, pré-treinamento e regularização, o que será explorado nos próximos tópicos (Palanivinayagam; El-Bayeh; Damaševičius, 2023).

3.2.2.1 Convolutional Neural Networks - CNN

As CNNs foram inicialmente desenvolvidas para a classificação de imagens, empregando filtros convolucionais para extrair características essenciais (LeCun; Bengio; Hinton, 2015). Embora tenham sido projetadas originalmente para visão computacional, as CNNs também demonstraram eficácia em tarefas de Processamento de Linguagem Natural, como a classificação de sentenças e a análise de intenções (Kim, 2014; Tun; Soe, 2020). Essa adaptação ocorre porque, apesar das diferenças entre imagens e textos, ambos podem ser tratados como arranjos de dados estruturados, nos quais operações convolucionais permitem capturar padrões locais.

A estrutura típica das CNNs inclui uma camada de entrada, várias camadas convolucionais e de *pooling*, seguidas por camadas totalmente conectadas e uma camada de saída. No caso dos textos, a entrada corresponde a uma matriz de *embeddings*, em que cada linha representa uma palavra e cada coluna corresponde a uma dimensão vetorial. Esse arranjo matricial possibilita a aplicação direta de filtros convolucionais, assim como ocorre com *pixels* em imagens.

No trabalho de Kim (2014) o processo de convolução em uma janela de tamanho h é definido pela Equação 4.

$$c_i = f(w \cdot x_{i:i+h-1} + b) \quad (4)$$

Onde, $x_{i:i+h-1}$ representa a concatenação dos vetores de palavras do i -ésimo até o $(i + h - 1)$ -ésimo termo, w é o filtro convolucional, b é o viés e f é a função de ativação (comumente ReLU).

Esse cálculo gera uma sequência de valores c_i , cada um correspondente a um padrão detectado em uma janela do texto. Em seguida, aplica-se uma operação de *max-pooling*, que seleciona o valor mais representativo de cada filtro, apresentado na Equação 5.

$$\hat{c} = \max_i c_i \quad (5)$$

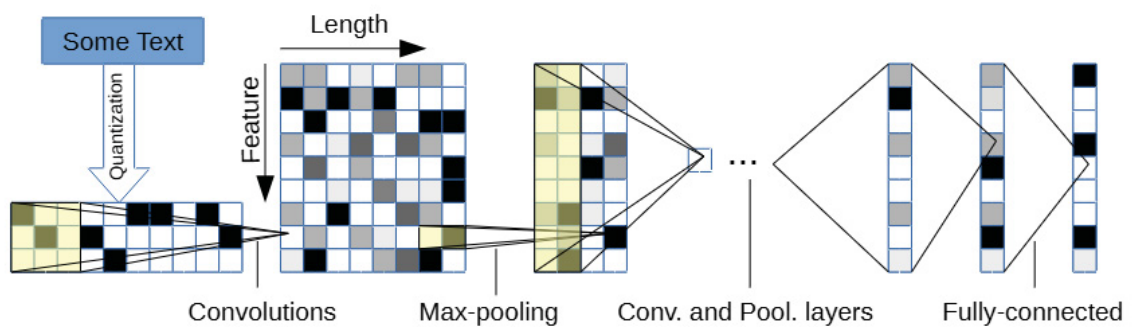
Esse valor \hat{c} representa a característica dominante para aquela janela. A utilização de múltiplos filtros, com diferentes larguras de janela, permite capturar padrões variados, desde bigramas até sequências mais longas de palavras. O vetor resultante de características é, então, conectado a uma camada totalmente conectada e normalizado por *softmax* para realizar a classificação da sentença, conforme a Equação 6.

$$P(y|x) = \text{softmax}(W\hat{c} + b) \quad (6)$$

Onde W e b representam os parâmetros da camada final de decisão.

Esse fluxo — entrada → convolução → pooling → vetor de características → classificação — resume a lógica das CNNs aplicadas a texto. A cada etapa, a rede reduz a dimensionalidade e concentra a informação em características mais abstratas, permitindo que mesmo frases curtas sejam transformadas em representações discriminativas, como ilustrado na Figura 2.

Figura 2 - Fluxo de uma *Convolutional Neural Network* (CNN) aplicada a texto.



Fonte: Zhang, Zhao e LeCun (2015)

Avanços recentes expandiram essa formulação clássica. Wu et al. (2024), por exemplo, descrevem a utilização de três camadas convolucionais unidimensionais,

cada uma com diferentes tamanhos de *kernel*¹⁵ (2, 3 e 4) e canais de saída (50, 100 e 150). *Kernels* menores capturam padrões locais mais específicos, enquanto *kernels* maiores extraem informações semânticas mais amplas. Essa configuração aumenta a capacidade do modelo de representar os dados de entrada em diferentes níveis de granularidade. A concatenação das saídas é formalizada como mostrado na Equação 7.

$$h = [h_2; h_3; h_4] \quad (7)$$

Em que cada h_k corresponde à saída da convolução com *kernel* de tamanho k . O vetor concatenado amplia a diversidade de padrões detectados e fornece uma base robusta para a etapa final de classificação.

Do ponto de vista empírico, CNNs têm demonstrado resultados competitivos. O estudo de Bilah, Adji e Setiawan (2022), por exemplo, aplicou CNN com *embeddings* GloVe no conjunto ATIS¹⁶ (*Airline Travel Information System*) e alcançou uma precisão de 95,84%. Esse resultado mostra que, quando bem ajustadas, CNNs podem superar métodos tradicionais, oferecendo modelos mais precisos e eficientes para classificação de intenções.

Entretanto, é importante ressaltar algumas limitações. As CNNs capturam padrões locais (n -gramas) com grande eficácia, mas apresentam dificuldades em modelar dependências de longo prazo, que exigem considerar relações entre palavras distantes. Esse é um ponto em que arquiteturas como RNN e LSTM apresentam vantagens, por incorporarem mecanismos de memória que preservam o contexto sequencial (Otter; Medina; Kalita, 2021). Já os *Transformers*, como o BERT, avançam ainda mais ao modelar dependências globais por meio de mecanismos de atenção.

No setor de petróleo e gás, essas limitações tornam-se ainda mais relevantes. Expressões como “*forecast oil demand*” ou “*Brent spread narrowing*” carregam significados que dependem não apenas da concorrência local de palavras, mas também de seu contexto econômico e temporal. Uma CNN pode reconhecer padrões como “*oil price*” ou “*market forecast*”, mas pode falhar em capturar nuances

¹⁵ Em CNN, um *kernel* (também chamado de filtro) é uma matriz com pesos treináveis que percorre a entrada — como uma sequência de texto ou uma imagem — realizando uma operação de convolução.

¹⁶ ATIS (*Airline Travel Information System*): conjunto de dados amplamente utilizado em PLN, composto por consultas de usuários sobre informações de voos, como reservas, horários e preços.

semânticas mais complexas, como a distinção entre uma previsão futura e uma análise retrospectiva. Estudos indicam que CNNs são empregadas como modelos de referência em tarefas de classificação de textos curtos, especialmente em experimentos comparativos com outras arquiteturas (Kim, 2014; Ayadi; Lachiri, 2023).

Por fim, estudos de revisão também destacam o papel das CNNs no desenvolvimento de *chatbots* e sistemas de diálogo. Zhang et al. (2024) apontam que, embora as CNNs tenham contribuído para avanços significativos, sua adoção em aplicações industriais está cada vez mais limitada diante do crescimento de modelos baseados em *Transformers*. Ainda assim, as CNNs permanecem relevantes como modelos intermediários, oferecendo uma boa relação entre desempenho e custo computacional, especialmente em cenários onde recursos de *hardware* são restritos.

3.2.2.2 Recurrent Neural Networks - RNN

As RNNs surgiram como uma evolução em relação às CNNs, trazendo a capacidade de modelar dados sequenciais. Enquanto as CNNs capturam padrões locais sem considerar a ordem dos elementos, as RNNs introduzem conexões recorrentes que permitem ao modelo manter um estado interno, ou “memória”, responsável por armazenar informações de entradas anteriores (Williams; Zipser, 1989) . Esse mecanismo possibilita representar dependências temporais, o que sustenta sua aplicação em tarefas como tradução automática, modelagem de linguagem e análise de séries temporais (Otter; Medina; Kalita, 2021).

Formalmente, o funcionamento de uma RNN pode ser descrito pelas clássicas equações 8 e 9 (Williams; Zipser, 1989):

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (8)$$

$$y_t = g(W_{hy}h_t + b_y) \quad (9)$$

Em que, h_t representa o estado oculto no instante t , x_t é a entrada no tempo t , y_t é a saída da rede, W_{xh} , W_{hh} e W_{hy} são matrizes de pesos aprendidas durante o treinamento, f e g correspondem a funções de ativação (tipicamente *tanh*, *sigmoid* ou *ReLU*).

Esse mecanismo recorrente possibilita que a rede capture relações entre elementos sequenciais. Em um texto, por exemplo, o termo “*previsão de preços*” só

adquire sentido completo se as palavras forem analisadas em sequência, preservando a relação entre “previsão” e “preços”. Enquanto uma CNN pode detectar apenas a coocorrência local dessas palavras, a RNN consegue modelar explicitamente a dependência entre elas ao longo da sequência (Li et al., 2022).

As RNNs foram aplicadas em diversas tarefas de PLN. Em tradução automática, por exemplo, o modelo lê uma frase palavra por palavra, armazenando no estado oculto a representação acumulada do significado, e depois gera a tradução correspondente na língua alvo (Cho et al., 2014). Em análise de sentimentos, o modelo percorre a sentença para avaliar o tom geral, atribuindo mais peso a palavras como “ótimo” ou “péssimo” dependendo do contexto (Otter; Medina; Kalita, 2021). Em sistemas de diálogo, a RNN pode ajudar a manter a coerência da conversa, lembrando informações passadas pelo usuário em turnos anteriores (Zhang et al., 2024).

Apesar de sua importância histórica, as RNNs apresentam limitações significativas. O principal desafio é o problema do desaparecimento e explosão do gradiente, que dificulta o aprendizado de dependências de longo prazo. Isso ocorre porque, durante o processo de retropropagação no tempo (*Backpropagation Through Time* – BPTT), os gradientes associados aos estados ocultos podem se tornar extremamente pequenos ou grandes, comprometendo a atualização adequada dos parâmetros. Como consequência, as RNNs simples tendem a capturar apenas dependências de curto alcance, sendo menos eficazes em contextos mais complexos.

Outro fator limitante é o custo computacional. Como as RNNs processam os dados de forma sequencial, cada etapa depende da anterior, o que reduz a possibilidade de paralelização eficiente em *hardware* como GPUs. Isso torna o treinamento mais lento quando comparado a arquiteturas como CNNs e *Transformers*, que permitem operações mais paralelas. Além disso, a complexidade aumenta à medida que se deseja treinar redes mais profundas, ampliando o risco de instabilidade durante o aprendizado (Li et al., 2022).

No domínio técnico-financeiro, como o setor de petróleo e gás, essas limitações são particularmente relevantes. Consultas relacionadas à previsão de preços ou à análise de séries temporais exigem a captura de padrões que se estendem por várias etapas, o que não é bem resolvido pelas RNNs clássicas. Um usuário pode, por exemplo, perguntar: “Qual é a previsão para o preço do barril do Brent no próximo trimestre?”. Para responder adequadamente, é necessário considerar não apenas as palavras próximas, mas também o contexto mais amplo da

consulta. A RNN, sozinha, tem dificuldade em preservar essas dependências quando a sequência ultrapassa janelas curtas, resultando em perda de informação.

Ainda assim, as RNNs representam um marco na evolução do *Deep Learning* aplicado ao PLN. Elas introduziram pela primeira vez a noção explícita de memória sequencial, superando as limitações de modelos puramente estatísticos ou baseados em padrões locais. Por isso, mesmo com suas restrições práticas, são fundamentais para compreender o desenvolvimento das arquiteturas posteriores. Essas limitações, aliás, motivaram o surgimento de variantes mais robustas, como a Long Short-Term Memory (LSTM) e a Gated Recurrent Unit (GRU), projetadas para mitigar o desaparecimento de gradiente e lidar melhor com dependências de longo prazo (Hochreiter; Schmidhuber, 1997; Cho et al., 2014; Wu et al., 2024).

3.2.2.3 Long Short-Term Memory - LSTM

As *Long Short-Term Memory networks* (LSTMs) foram propostas por Hochreiter e Schmidhuber (1997) como uma solução para as limitações das RNNs tradicionais, em especial o problema do desaparecimento e da explosão do gradiente. Enquanto as RNNs simples apresentam dificuldades para aprender dependências de longo prazo, as LSTMs introduzem um mecanismo de portas de controle que regula o fluxo de informações, permitindo preservar ou descartar estados conforme a relevância da sequência (Wu et al., 2024).

A célula LSTM é composta principalmente por três portas: porta de esquecimento, porta de entrada e porta de saída. Cada uma desempenha uma função distinta na atualização do estado da célula e do estado oculto.

A porta de esquecimento (f_t) é responsável por decidir quais informações do estado anterior devem ser descartadas, conforme mostrado na Equação 9.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

Em seguida, a porta de entrada (i_t) define quais novas informações devem ser incorporadas, de acordo com a Equação 10.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (10)$$

Os valores candidatos que podem ser adicionados ao estado da célula são calculados pela Equação 11.

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (11)$$

Combinando o que deve ser esquecido e o que deve ser adicionado, obtém-se a atualização do estado da célula (c_t), conforme a Equação 12.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (12)$$

A saída da célula é regulada pela porta de saída (o_t), apresentada na Equação 13.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (13)$$

Por fim, o estado oculto atualizado (h_t) é obtido multiplicando a saída da porta com a transformação não linear do estado da célula, como mostra a Equação 14.

$$h_t = o_t \cdot \tanh(c_t) \quad (14)$$

Essa estrutura em camadas garante que a LSTM seja capaz de manter informações relevantes por longos intervalos, superando as RNNs tradicionais na modelagem de dependências de longo prazo (Otter; Medina; Kalita, 2021).

Na prática, as LSTMs apresentaram avanços importantes em tarefas de PLN, como tradução automática, análise de sentimentos e reconhecimento de fala. Em sistemas de diálogo, permitem que um chatbot mantenha o contexto da conversa por múltiplos turnos, resultando em respostas mais coerentes (Zhang et al., 2024).

Apesar dessas vantagens, as LSTMs ainda possuem alto custo computacional e processam as entradas de maneira sequencial, o que limita sua paralelização em GPUs. Nos últimos anos, modelos baseados em atenção, como os *Transformers*, têm se mostrado mais eficientes e precisos (Wu et al., 2024).

No contexto do setor de petróleo e gás, as LSTMs apresentam potencial para capturar dependências temporais em consultas relacionadas a previsões de preços, tendências de mercado e séries históricas. Contudo, a necessidade de grandes quantidades de dados para treinamento e o custo elevado restringem sua aplicação em domínios especializados com bases reduzidas, como o utilizado nesta pesquisa.

Por essa razão, a LSTM foi utilizada como arquitetura comparativa, ilustrando a transição das redes recorrentes clássicas para os modelos baseados em atenção.

3.2.2.4 *Bidirectional Encoder Representations from Transformers* - BERT

O *Bidirectional Encoder Representations from Transformers* (BERT) representa um marco na área de PLN, sendo responsável por redefinir a forma como modelos de linguagem são construídos e aplicados em tarefas de classificação de texto, incluindo a análise de intenções. Introduzido por Devlin et al. (2019) o BERT foi desenvolvido a partir da arquitetura *Transformer*, proposta por Vaswani et al. (2017). A principal adaptação está na consideração bidirecional do contexto, em que cada *token* é analisado levando em conta informações anteriores e posteriores na sequência.

No pré-treinamento, o BERT utiliza duas tarefas supervisionadas. A primeira é o *Masked Language Modeling* (MLM), na qual aproximadamente 15% das unidades textuais são substituídas por [MASK], e o modelo deve prever os itens originais com base nas demais unidades textuais não mascaradas. Essa configuração força o modelo a estimar distribuições condicionais de palavras em ambas as direções do contexto. Por exemplo, na sequência “O preço do [MASK] subiu hoje”, a unidade textual “petróleo” é substituída por [MASK], e o modelo deve estimar qual elemento completa a frase com maior probabilidade. A predição é realizada utilizando as demais unidades textuais da sequência, como “preço”, “subiu” e “hoje”, com base nos padrões estatísticos aprendidos durante o pré-treinamento. Essa tarefa expõe o modelo a relações de concorrência e dependências contextuais bidirecionais, permitindo a estimativa de termos ausentes sem necessidade de regras manuais.

A segunda tarefa é o *Next Sentence Prediction* (NSP), em que pares de sentenças são apresentados e o modelo deve determinar se a segunda sentença ocorre de forma sequencial em relação à primeira. Metade dos exemplos corresponde a sentenças adjacentes do conjunto de treinamento, enquanto a outra metade é formada por sentenças não relacionadas, permitindo ao modelo aprender dependências entre segmentos textuais. Por exemplo, no exemplo “A OPEP anunciou cortes na produção. [SEP] Isso impactou o preço do petróleo”, a resposta correta indica sequência; no exemplo “A OPEP anunciou cortes na produção. [SEP] O cachorro correu no quintal”, a resposta é negativa. Essa tarefa possibilita ao modelo

estimar relações entre sentenças, aplicáveis em contextos como pergunta e resposta ou diálogos.

Após o pré-treinamento, a entrada passa pelo processo de *tokenização WordPiece*. Cada token é representado pela soma de três componentes: o embedding do token, que representa a palavra ou subpalavra; o embedding posicional, que indica sua posição na sequência; e o embedding de segmento, que diferencia sentenças em pares, conforme a Equação 15.

$$E_i = W_{tok}(t_i) + W_{pos}(i) + W_{seg}(s_i) \quad (15)$$

Em que E_i representa o vetor final associado ao *token* i ; $W_{tok}(t_i)$ corresponde ao *embedding* da palavra ou subpalavra t_i ; $W_{pos}(i)$ é o *embedding* posicional; e $W_{seg}(s_i)$ é o *embedding* do segmento. Essa combinação garante que cada *token* carregue tanto sua representação semântica quanto sua posição estrutural no texto.

Os *embeddings* são então processados pelo mecanismo de atenção, que avalia a relação entre os tokens por meio dos vetores de consulta (Q), chave (K) e valor (V) — em que a consulta representa a informação utilizada para buscar contextos relevantes, a chave atua como identificador de cada unidade textual e o valor armazena a informação que será combinada — conforme definido pela Equação 16.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (16)$$

Nessa equação, o produto QK^T mede a similaridade entre consultas e chaves; d_k é a dimensão dos vetores de chave, usada para normalizar os resultados e evitar valores excessivos; o *softmax* converte os escores em probabilidades, que indicam a importância relativa de cada *token*; e o produto final com V gera a combinação ponderada das informações mais relevantes. Em termos práticos, isso significa que cada palavra passa a “olhar” para todas as demais da frase, atribuindo diferentes pesos a elas conforme sua contribuição para o significado do contexto.

Para ampliar a captura de relações semânticas, o BERT utiliza atenção multi-cabeça (*Multi-Head Attention*). Em vez de realizar o cálculo de atenção apenas uma vez, o processo é replicado em diversas cabeças, cada uma projetando os vetores em

subespaços diferentes e aprendendo relações distintas. A operação global é descrita pela Equação 17.

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W^O \quad (17)$$

Cada *head* de atenção é calculada conforme a Equação 18.

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (18)$$

Onde W_i^Q , W_i^K , e W_i^V são matrizes de projeção específicas para cada cabeça, enquanto W^O é a matriz que combina todas as saídas. Esse mecanismo garante que o modelo capture, de forma paralela, tanto dependências locais de palavras próximas quanto relações semânticas distantes, algo essencial para textos especializados.

Após a etapa de atenção, cada token passa por uma rede *feed-forward* aplicada de forma independente a cada posição. Essa etapa é expressa pela Equação 19.

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (19)$$

Na equação 19, W_1 e W_2 são matrizes de peso, e b_1 e b_2 são vetores de viés. Essa camada complementa a etapa de atenção ao capturar interações mais complexas entre os *tokens*, permitindo maior expressividade no modelo.

Para tarefas de classificação, como análise de intenções, o BERT utiliza o *token* especial [CLS], inserido no início da sequência de entrada. A representação final associada a essa unidade textual reúne as informações produzidas ao longo das camadas do modelo. A predição final é obtida pela Equação 20.

$$P = softmax(W_{cls} \cdot h_{[CLS]} + b) \quad (20)$$

Aqui, $h_{[CLS]}$ é o vetor final associado ao *token* [CLS], W_{cls} é a matriz de pesos da camada de classificação, e b é o viés. O *softmax* gera a distribuição de probabilidades sobre as classes de saída, permitindo identificar a intenção mais provável.

Uma vez pré-treinado em grandes conjuntos de dados, o BERT pode ser adaptado para tarefas específicas por meio de *fine-tuning*. Nesse processo, acrescenta-se uma camada adicional e o modelo é ajustado com um conjunto de dados rotulado. Esse procedimento permite aproveitar o conhecimento adquirido

durante o pré-treinamento, mas adaptando-o às particularidades da tarefa. Aplicações práticas demonstram a eficácia dessa abordagem.

No artigo de Chakraborty et al. (2024), foi desenvolvido um modelo de classificação de intenções utilizando o BERT para categorizar perguntas de usuários relacionadas a recursos e componentes de veículos, com base em um conjunto de dados extraído dos Manuais do Proprietário. O modelo, treinado com 254.412 registros distribuídos em 199 categorias distintas, alcançou uma precisão geral de 98,21%. A abordagem demonstrou resultados promissores na classificação de texto em cenários reais, evidenciando a eficácia dos agentes conversacionais em compreender intenções, mesmo diante de erros de digitação ou gramática.

De forma semelhante, Chen, Zhuo e Wang (2019) apresentaram um modelo de classificação de intenções e preenchimento de slots baseado no BERT, que superou significativamente modelos tradicionais em termos de precisão e capacidade de generalização. Avaliado nos conjuntos de dados ATIS (*Airline Travel Information System*) e Snips¹⁷, o modelo alcançou resultados superiores em relação a arquiteturas anteriores, confirmando a eficácia de combinar as duas tarefas. O estudo ainda sugere que pesquisas futuras explorem conjuntos de dados mais amplos e incorporem conhecimento externo ao BERT, de modo a ampliar seu potencial de aplicação.

No âmbito desta tese, o BERT foi escolhido como modelo central devido à sua capacidade de capturar relações contextuais complexas, algo indispensável para lidar com as consultas curtas e técnicas do setor de petróleo e gás. No entanto, o modelo também apresenta desafios importantes: seu alto custo computacional, a necessidade de grandes volumes de dados para pré-treinamento e a instabilidade observada em conjuntos pequenos. Essas limitações justificam a proposta aqui desenvolvida, que consiste em adaptar o BERT ao domínio técnico-financeiro por meio do modelo OilFinBERT, explorando estratégias de pré-treinamento adicional, técnicas de regularização e *fine-tuning*. A aplicação prática busca superar as fragilidades do BERT em bases reduzidas, garantindo maior estabilidade e precisão em cenários especializados.

Embora o BERT tenha se consolidado como referência em tarefas de PLN, sua aplicação em bases pequenas demanda abordagens complementares que

¹⁷ Snips: *dataset* de comandos em linguagem natural, criado para assistentes de voz, incluindo tarefas como tocar música, agendar eventos e obter previsões do tempo.

garantam estabilidade e capacidade de generalização. Por esse motivo, no próximo tópico serão apresentadas técnicas de pré-treinamento e *fine-tuning*, incluindo o uso de otimizadores, regularização e ajustes estruturais, que contribuem para potencializar os resultados do modelo em contextos restritos.

3.2.3 Técnicas aplicadas ao *fine-tuning*

As técnicas aplicadas ao *fine-tuning* incluem práticas voltadas ao ajuste dos modelos para tarefas específicas. A literatura descreve o uso de diferentes otimizadores, métodos de regularização e ajustes adicionais, bem como estratégias baseadas em estruturas de classificação, como a hierárquica. Em conjunto, essas abordagens são empregadas para reduzir instabilidades durante o treinamento e ajustar o modelo a cenários com conjuntos de dados reduzidos ou especializados.

3.2.3.1 Otimizadores

No treinamento de modelos baseados em DL, o otimizador é o componente responsável por atualizar os parâmetros a partir dos gradientes calculados durante a retropropagação. Em processos de *fine-tuning*, diferentes escolhas de otimizadores podem resultar em trajetórias de atualização distintas e influenciar a estabilidade do treinamento. Entre os métodos mais empregados estão o *Stochastic Gradient Descent* (SGD), o Adam e variantes como AdamW e AdamP. Estudos indicam que esses algoritmos podem apresentar comportamentos semelhantes durante o treinamento, mas resultar em soluções com diferentes características de generalização.

O SGD é considerado o método mais clássico e fundamental para a otimização de redes neurais. Ele atualiza os parâmetros a partir de gradientes estimados em *minibatches*, o que permite lidar com grandes volumes de dados de forma eficiente (Bottou, 2010; Ayadi; Lachiri, 2023). Embora simples, sua versão com *momentum* e estratégias de decaimento da taxa de aprendizado consolidou-se como prática robusta em várias áreas, sobretudo em visão computacional. No entanto, sua principal limitação é a sensibilidade à taxa de aprendizado e a convergência relativamente lenta em problemas de alta dimensionalidade.

A popularização dos métodos adaptativos ocorreu com a introdução do Adam (Kingma; Ba, 2017), que combina estimativas de primeira e segunda ordem do gradiente, aplicando correções de viés iniciais e resultando em atualizações adaptativas por parâmetro. Essa abordagem tornou o treinamento mais estável, eficiente e menos dependente de ajustes manuais da taxa de aprendizado. Contudo, estudos indicam que, apesar de facilitar o treinamento, o Adam pode levar a soluções que generalizam pior em comparação ao SGD (Wilson et al., 2018), o que estimulou a criação de variantes.

Uma dessas variantes é o AdamW, proposto por Loshchilov e Hutter (2019). O algoritmo modifica o procedimento do Adam original ao separar o termo de *weight decay*¹⁸ da atualização do gradiente, evitando que ambos sejam tratados como a mesma operação. Essa alteração permite controlar o decaimento de pesos de maneira independente, o que tem mostrado efeito consistente na estabilidade do treinamento. Em tarefas de PLN, especialmente no *fine-tuning* de modelos baseados em Transformers, o AdamW passou a ser amplamente utilizado por apresentar comportamento mais previsível em comparação ao Adam tradicional.

Mais recentemente, Heo et al. (2021) propuseram o AdamP, projetado para lidar com arquiteturas invariantes à escala, comuns em redes com camadas de normalização. Nessas condições, otimizadores com *momentum* podem induzir crescimento excessivo da norma dos pesos e reduzir prematuramente o tamanho efetivo dos passos de atualização. O AdamP introduz uma correção geométrica ao remover o componente radial do vetor de atualização, estabilizando a convergência. Avaliações em treze *benchmarks* de visão, linguagem e áudio demonstraram ganhos consistentes sobre métodos anteriores.

Comparações entre SGD, AdamW e AdamP indicam diferenças no comportamento durante o treinamento. O SGD utiliza gradientes estimados em subconjuntos dos dados e atualiza os parâmetros de forma incremental, enquanto o AdamW aplica atualizações adaptativas para cada parâmetro. Já o AdamP inclui uma etapa adicional que ajusta o vetor de atualização em situações nas quais a escala dos pesos influencia o treinamento. Em cenários de *fine-tuning* em PLN, variações nos resultados podem ocorrer devido à combinação entre o otimizador e práticas de

¹⁸ No contexto de otimização, *weight decay* é um termo que reduz gradualmente os valores dos parâmetros durante o treinamento, com o objetivo de evitar crescimento excessivo dos pesos.

treinamento, como o aquecimento gradual da taxa de aprendizado (*warmup*), a aplicação de *weight decay* e o uso de critérios para interromper o treinamento (*early stopping*) (Mosbach; Andriushchenko; Klakow, 2021).

No *fine-tuning* de modelos baseados em Transformers, diferentes inicializações aleatórias podem produzir resultados distintos, o que indica variação no processo de treinamento (Mosbach; Andriushchenko; Klakow, 2021). Nesse contexto, o SGD continua sendo utilizado em alguns cenários, o AdamW permanece como escolha frequente em modelos de linguagem, e o AdamP tem sido explorado em situações nas quais a variação na escala dos pesos interfere nas atualizações.

3.2.3.2 Regularização e ajustes

A regularização e os ajustes de treinamento referem-se ao conjunto de técnicas utilizadas durante o *fine-tuning* de modelos pré-treinados, com o objetivo de controlar a atualização dos pesos e reduzir instabilidades. Essas estratégias são empregadas por profissionais e pesquisadores que adaptam modelos como Transformers a tarefas específicas, principalmente em domínios especializados e com quantidade reduzida de dados rotulados. Nesse contexto, métodos como reinicialização de camadas, *Mixout*, *Stochastic Weight Averaging* (SWA) e *Layer-wise Learning Rate Decay* (LLRD) atuam de maneira complementar aos otimizadores (por exemplo, AdamW e AdamP): a reinicialização evita acúmulo de gradientes nas camadas superiores (Zhang et al., 2021), o *Mixout* limita a divergência em relação aos pesos do pré-treino (Lee; Cho; Kang, 2020), o SWA estabiliza a solução ao promediar pesos finais (Izmailov et al., 2019) e o LLRD preserva representações gerais ao aplicar taxas de aprendizado decrescentes ao longo das camadas (Howard; Ruder, 2018). Essas práticas são aplicadas no momento em que o modelo deixa de ser apenas pré-treinado e passa a ser ajustado para a tarefa-alvo, tornando-se especialmente relevantes quando há indícios de *overfitting*, alta variabilidade entre *seeds*¹⁹ ou discrepância entre o domínio original e o domínio de aplicação. A literatura indica que parte da instabilidade observada nesse regime decorre de dificuldades de otimização, como gradientes pequenos, e não apenas da limitação de dados ou da perda de

¹⁹ *Seed* é um controle da aleatoriedade usado para garantir reprodutibilidade e avaliar estabilidade no treinamento de modelos.

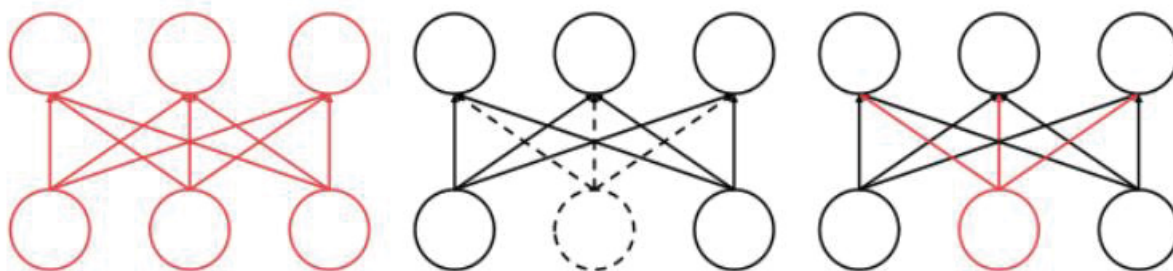
representações aprendidas, o que reforça a utilidade dessas técnicas para preservar conhecimento prévio e favorecer a generalização (Mosbach; Andriushchenko; Klakow, 2021).

A reinicialização de camadas consiste em redefinir os pesos de parte do modelo antes do início do *fine-tuning*, em vez de manter integralmente os valores do pré-treino. Em *Transformers* como o BERT, essa prática concentra-se na camada de saída, formada pela operação de agregação da sequência (geralmente o vetor do token [CLS]) e pela camada linear que projeta essa representação para o número de classes. Além disso, um pequeno conjunto de camadas superiores — isto é, as últimas camadas da pilha *Transformer*, mais próximas da saída — pode ser reinicializado para facilitar a adaptação à tarefa-alvo, enquanto as camadas inferiores permanecem intactas por armazenarem representações linguísticas mais gerais. Essa estratégia é aplicada quando o *fine-tuning* apresenta instabilidade ou degradação de métricas de validação, especialmente em bases pequenas. Estudos relatam que reinicializar a camada de saída e algumas camadas finais reduz variabilidade entre *seeds* e melhora a consistência do desempenho (Zhang et al., 2021). Do ponto de vista teórico, análises recentes indicam que o aumento do número de iterações e o uso de taxas de aprendizado menores também contribuem para reduzir essa variância durante o ajuste supervisionado (Fu; So; Collier, 2023).

O *Mixout* é uma técnica de regularização aplicada durante o *fine-tuning* de modelos de linguagem pré-treinados, com o objetivo de limitar o afastamento dos pesos em relação aos valores originais do pré-treino. A técnica atua diretamente sobre os parâmetros do modelo: em cada atualização, recebe os pesos atuais W , os pesos pré-treinados W_0 e uma probabilidade p . A operação consiste em gerar uma máscara aleatória e combinar os dois conjuntos de pesos segundo $W_{\text{mix}} = M \odot W + (1 - M) \odot W_0$, de modo que parte dos valores W permanece atualizada e o restante retorna ao estado inicial. Esse procedimento produz um conjunto de pesos regularizado, preservando informações aprendidas previamente e reduzindo oscilações quando há poucas amostras, sem modificar a arquitetura do modelo (Lee; Cho; Kang, 2020).

A Figura 3 compara o comportamento do *Dropout* e do *Mixout* durante o *fine-tuning*, destacando que o *Mixout* preserva parte dos pesos pré-treinados e tende a produzir ajustes mais estáveis (Lee; Cho; Kang, 2020).

Figura 3 - Ilustração de *Mixout*



Fonte: Extraída de Lee, Cho e Kang (2020)

A Média Estocástica dos Pesos (*Stochastic Weight Averaging — SWA*) (Izmailov et al., 2019) consiste em calcular a média de pesos obtidos em diferentes pontos da trajetória de treinamento (geralmente sob *learning rates* cíclicos ou quase constantes), produzindo soluções mais “planas” (*flat* mínima) — isto é, regiões da superfície de perda em que pequenas variações nos pesos não alteram significativamente o erro — e, em consequência, com melhor capacidade de generalização (Guo; Jin; Liu, 2023). Em *PLN*, relatos indicam que o *SWA* melhora a generalização sem custo adicional na inferência (um único modelo é mantido) (Lu et al., 2022). Também há evidências de ganhos de robustez frente a mudança entre a distribuição de dados do treinamento e a distribuição observada na inferência (*shift*), especialmente em modelos de linguagem grandes e em *fine-tuning* com conjuntos reduzidos (Talman et al., 2023; Onal et al., 2024).

O *fine-tuning* discriminativo aplica *learning rates* distintos por camada, pois cada nível do modelo retém informações de natureza diferente (Howard; Ruder, 2018). Também chamado de *layer-wise learning rate decay* (LLRD), utiliza taxas maiores nas camadas superiores e menores nas inferiores para ajustar o que é específico da tarefa sem perder o conhecimento geral do pré-treino (Zhang et al., 2021). Estudos relatam ganhos consistentes com LLRD em várias tarefas e modelos, incluindo *XLNet*²⁰ e *ELECTRA*²¹ (Clark et al., 2020; You et al., 2020)

²⁰ XLNet: modelo de linguagem baseado em *transformers*, que combina aprendizado autorregressivo e permutacional para capturar dependências bidirecionais (Yang et al., 2019).

²¹ ELECTRA: modelo de pré-treinamento que substitui a tarefa de *Masked Language Modeling* por um treinamento discriminativo, onde a rede aprende a distinguir tokens reais de tokens substituídos (Clark et al., 2020).

Em síntese, reinicialização de camadas, *Mixout*, *SWA* e *LLRD* oferecem um conjunto de ferramentas que: (i) ajusta condições iniciais nas camadas superiores; (ii) ancora o ajuste nos parâmetros do pré-treino para evitar afastamentos bruscos; (iii) suaviza a solução final por meio de médias de trajetórias; e (iv) gradua a taxa de aprendizado por profundidade, preservando representações úteis. Em conjunto, essas escolhas tendem a atenuar fatores que a literatura relaciona a instabilidade e variância em *Transformers*, contribuindo para fluxo de trabalho de *fine-tuning* mais estáveis e mais fáceis de reproduzir.

3.2.3.3 Modelo Hierárquico

A classificação hierárquica organiza os rótulos em níveis estruturados, permitindo representar dependências entre classes. Diferentemente da classificação plana, em que todos os rótulos são tratados no mesmo nível e não há relação entre categorias, a abordagem hierárquica reduz ambiguidades ao considerar proximidade e especialização entre rótulos (Zangari et al., 2024).

No contexto do DL, diferentes estratégias são utilizadas para lidar com essa estrutura, incluindo decisões sequenciais ao longo dos níveis e métodos que processam a hierarquia de forma conjunta. Essas soluções têm sido aplicadas em tarefas de análise de intenções, nas quais categorias gerais são refinadas em subcategorias para aumentar a precisão em domínios especializados (Zangari et al., 2024).

A avaliação também requer métricas específicas, já que medidas tradicionais tratam todos os erros de forma equivalente. Métricas hierárquicas consideram a distância entre classes na estrutura, penalizando de modo menos severo confusões entre categorias próximas e de forma mais rígida erros em níveis distantes, resultando em avaliações mais alinhadas à organização dos rótulos (Zangari et al., 2024).

3.3 TRABALHOS RELACIONADOS

No campo da classificação de intenções em sistemas de diálogo e agentes conversacionais, diversas abordagens foram propostas ao longo dos anos, abrangendo desde métodos estatísticos tradicionais até arquiteturas avançadas de aprendizado profundo e modelos baseados em transformadores. Essa diversidade

reflete a complexidade da tarefa de interpretar corretamente a intenção dos usuários em diferentes domínios e contextos.

De modo geral, os primeiros estudos concentraram-se em modelos probabilísticos e algoritmos clássicos de aprendizado de máquina, que serviram como referência inicial para experimentos nessa área. Posteriormente, arquiteturas de redes neurais — como CNNs, RNNs e LSTMs — passaram a ocupar espaço central, explorando melhor as relações sequenciais e contextuais do texto. Mais recentemente, modelos baseados em transformadores, como o BERT e suas variantes, consolidaram-se como o estado da arte, oferecendo ganhos significativos de desempenho.

A Tabela 2 sumariza, de forma comparativa, os principais atributos de cada trabalho analisado.

Tabela 2 - Resumo dos Trabalhos Relacionados

Autor	Objetivo	Resultado
Liu, Qiu e Huang (2016)	Explorar o uso de RNNs para a classificação de textos, integrando o aprendizado de múltiplas tarefas.	A abordagem de aprendizado multitarefa melhora a precisão e robustez do modelo, com RNNs baseadas em LSTM apresentando o melhor desempenho global.
Helmi Setyawan, Awangga e Efendi (2018)	Comparar o desempenho dos algoritmos Multinomial NB e Regressão Logística na classificação de intenções.	Regressão Logística oferece maior precisão na classificação de intenções em comparação ao Multinomial NB.
Ezen-Can (2020)	Investigar a classificação de intenção em conjuntos de dados pequenos usando BERT e LSTM .	LSTM proporciona maior precisão em conjuntos de dados pequenos e requer menos tempo para ajuste, especialmente em tarefas de classificação de intenção.

Mosbach, Andriushchenko e Klakow (2021)	Investigar a estabilidade do <i>fine-tuning</i> do BERT em tarefas do benchmark GLUE, com foco em bases pequenas como MRPC e RTE.	O <i>fine-tuning</i> mostrou-se instável em <i>datasets</i> pequenos, com variação significativa entre execuções. Ajustes como menor taxa de aprendizado e mais épocas reduziram parcialmente a instabilidade, mas o problema permaneceu evidente.
Zhang et al. (2021)	Avaliar o desempenho do BERT em cenários de <i>few-shot learning</i> , utilizando subconjuntos artificiais de <i>benchmarks</i> com apenas 16 a 128 exemplos por classe.	O modelo apresentou sobreajuste e queda de desempenho em bases reduzidas. Estratégias como aumento no número de iterações, menor taxa de aprendizado e <i>fine-tuning</i> de hiperparâmetros melhoraram parcialmente os resultados, mas confirmaram a necessidade de metodologias específicas para esse contexto.
Rizou et al. (2022)	Aplicação de arquiteturas de DL para reconhecimento de entidades e classificação de intenções multilíngues.	Modelos baseados em <i>Transformers</i> , como BERT , superam LSTM e GRU , com BERT oferecendo a melhor performance geral.
Assayed, Shaalan e Alkhatib (2023)	Desenvolver o " HSchatbot " para prever intenções de consultas de alunos do ensino médio.	Random Forest supera Multinomial NB , alcançando precisão superior a 90%; Multinomial NB teve melhor desempenho com <i>CountVectorizers</i> do que com TF-IDF.

Hasan, Ibna Basher e Shawon (2023)	Apresentar o BNIntent30 e propor o modelo GAN-BnBERT para aprimorar a representação contextual em classificação de intenções .	GAN-BnBERT supera BiLSTM e BERT , alcançando 96,73% de precisão na classificação de intenções e 98% na extração de entidades nomeadas.
Wu et al. (2024)	Utilizar o modelo CBLMA-B para processamento multinível e classificação em reconhecimento de intenção .	O método proposto alcança excelente desempenho na classificação em tarefas de reconhecimento de intenção .
Lee et al. (2024)	Propor um <i>chatbot</i> hierárquico usando BERT para classificação de intenção e reconhecimento de entidades em dados chineses.	Desempenho superior com 95% de precisão na classificação e 98% na extração de entidades , com generalização para múltiplos idiomas.

Fonte: A autora (2025)

O artigo de Liu, Qiu e Huang (2016) explora o uso de RNNs para a classificação de textos, integrando o aprendizado de múltiplas tarefas. O estudo avalia como as RNNs, especialmente com mecanismos de memória como LSTMs e GRUs, podem ser adaptadas para realizar simultaneamente a classificação de intenções e a análise de sentimentos, entre outras tarefas relacionadas. Os resultados mostram que a abordagem de aprendizado multitarefa melhora a precisão e a robustez do modelo em comparação com métodos de aprendizado único. Entre os modelos testados, as RNNs com LSTM demonstraram o melhor desempenho global, superando as GRUs em termos de precisão e capacidade de capturar dependências de longo prazo no texto. A pesquisa destaca que a integração de múltiplas tarefas na RNN proporciona uma performance mais eficiente e geral na classificação de textos.

O artigo de Helmi Setyawan, Awangga e Efendi (2018) compara o desempenho dos algoritmos Multinomial NB e Regressão Logística na tarefa de classificação de intenções em *chatbots*. O estudo avalia como cada algoritmo lida com a identificação de intenções em textos de usuários, analisando métricas de

desempenho como precisão e eficiência. Os resultados indicam que, embora ambos os métodos tenham seus pontos fortes, a Regressão Logística geralmente oferece maior precisão na classificação das intenções em comparação ao Multinomial NB. A pesquisa destaca a importância de escolher o modelo adequado para melhorar a eficácia dos sistemas de *chatbot*.

O estudo de Ezen-Can (2020) investigou a classificação de intenção de um pequeno conjunto de dados usando os modelos BERT e LSTM. Este conjunto de dados possui 150 classes de intenção, cada uma com 100 exemplos de treinamento, além de 20 exemplos de validação e 30 de teste por classe e, também, consultas fora do escopo que não pertencem a nenhuma das classes. A análise mostrou que, em conjuntos de dados pequenos, o LSTM pode proporcionar maior precisão e requer menos tempo para construir e ajustar modelos, especialmente para tarefas de classificação de intenção.

Mosbach, Andriushchenko e Klakow (2021) investigaram a estabilidade do *fine-tuning* do BERT em diferentes tarefas do benchmark GLUE, com destaque para *datasets* pequenos como MRPC e RTE. Os autores observaram que, nesse cenário, os resultados variam significativamente entre execuções, mesmo com condições experimentais semelhantes. Essa instabilidade compromete a reprodutibilidade e dificulta a comparação justa entre abordagens. O estudo sugeriu que o aumento do número de épocas de treinamento, a adoção de taxas de aprendizado menores e a implementação de boas práticas de regularização podem reduzir as variações, mas ressaltou que o problema persiste de forma mais acentuada em conjuntos de dados limitados.

Zhang et al. (2021) analisaram o desempenho do BERT em situações de *few-shot learning*, criando subconjuntos artificiais de benchmarks com apenas 16 a 128 exemplos por classe. Os experimentos evidenciaram dificuldades do modelo em generalizar em cenários de dados reduzidos, apresentando sobreajuste e queda de desempenho. Para enfrentar esse desafio, os autores propuseram ajustes no processo de treinamento, incluindo maior número de iterações, taxas de aprendizado menores e calibração de hiperparâmetros. O estudo reforçou que a adaptação cuidadosa do BERT é essencial para obter resultados satisfatórios em bases pequenas, destacando a necessidade de metodologias específicas nesse contexto.

O estudo de Rizou et al. (2022) explora a aplicação de arquiteturas de aprendizado profundo para o reconhecimento de entidades nomeadas e a

classificação de intenções em múltiplos idiomas. O estudo utiliza e compara diversos modelos de DL, incluindo LSTM, GRU e Transformers, para avaliar seu desempenho em tarefas multilíngues. Os resultados indicam que os modelos baseados em Transformers, como BERT, superaram significativamente os LSTM e GRU em precisão tanto na identificação de entidades nomeadas quanto na classificação de intenções. Especificamente, BERT demonstrou a melhor performance geral, oferecendo alta precisão e eficiência, destacando-se como a arquitetura mais eficaz para essas tarefas em contextos multilíngues. A pesquisa reforça a eficácia dos Transformers em tarefas complexas de PLN em diversos idiomas.

O estudo de Assayed, Shaalan e Alkhatib (2023) desenvolveu o "HSchatbot" para prever as intenções das consultas dos alunos do ensino médio, que precisam de suporte imediato para suas escolhas de carreira. Utilizando os classificadores Multinomial NB e *Random Forest* com técnicas de extração de características como CountVectorizers e TF-IDF, os resultados mostraram que o *Random Forest* superou o Multinomial NB em várias métricas, alcançando uma precisão superior a 90%. No entanto, o Multinomial NB teve um desempenho melhor com CountVectorizers do que com TF-IDF. No futuro, será necessário investigar mais a fundo os fatores que afetam o desempenho do Multinomial NB e usar um corpus maior de consultas dos alunos.

A análise de Hasan, Ibna Basher e Shawon (2023) apresenta o BNIntent30²², um conjunto de dados bengali com 30 classes para classificação de intenções, e propõe o modelo GAN-BnBERT²³, que combina BERT com uma rede adversária generativa para aprimorar a representação contextual e a diversidade das classes. Os experimentos demonstram que o GAN-BnBERT supera o BiLSTM e o BERT tradicional, alcançando 96,73% de precisão na classificação de intenções e 98% na extração de entidades nomeadas. Apesar do desempenho superior do modelo, algumas classificações incorretas ainda ocorrem devido a confusões linguísticas.

A pesquisa conduzida por Wu et al. (2024) utiliza o modelo CBLMA-B, uma abordagem de processamento multinível, combinando CNN para capturar informações locais, BiLSTM para extrair recursos sequenciais, e mecanismos de atenção multi-cabeças e BERT para enriquecer e estruturar ainda mais as

²² BNIntent30: conjunto de dados em bengali, com 30 classes de intenção, desenvolvido para pesquisa em classificação de intenções.

²³ GAN-BnBERT: modelo que combina redes adversárias generativas (GANs) com o BERT para classificação de intenções em textos em bengali.

características das frases. Os experimentos comparativos e de análise detalhada mostram que o método proposto alcança um excelente desempenho na classificação em tarefas de reconhecimento de intenção.

O artigo de Lee et al. (2024) propõe um *chatbot* hierárquico que usa o modelo BERT pré-treinado para a classificação de intenção e reconhecimento de entidades nomeadas em um conjunto de dados chinês, demonstrando desempenho superior com 95% de precisão na classificação e 98% na extração de entidades. O sistema é projetado para lidar com dúvidas sobre medicamentos, avaliando a intenção, a entidade e a pontuação de risco das informações, e fornece respostas adaptadas. Além disso, o modelo é ajustado para melhorar a classificação de texto e a identificação de entidades, com uma arquitetura que permite a generalização para múltiplos idiomas.

A revisão dos trabalhos relacionados mostra a evolução da classificação de intenções até os modelos baseados em *Transformers*, como o BERT. Entretanto, esses modelos ainda apresentam instabilidade e queda de desempenho em bases pequenas e com linguagem técnico-financeira, como as do setor de petróleo e gás (Zangari et al., 2024).

Esta tese propõe adaptar o BERT ao domínio técnico-financeiro por meio de três ações: (i) pré-treinamento adicional com banco de dados não rotulado do setor; (ii) *fine-tuning* com técnicas de estabilidade (Mixout, SWA e LLRD); e (iii) avaliação comparativa com modelos de referência.

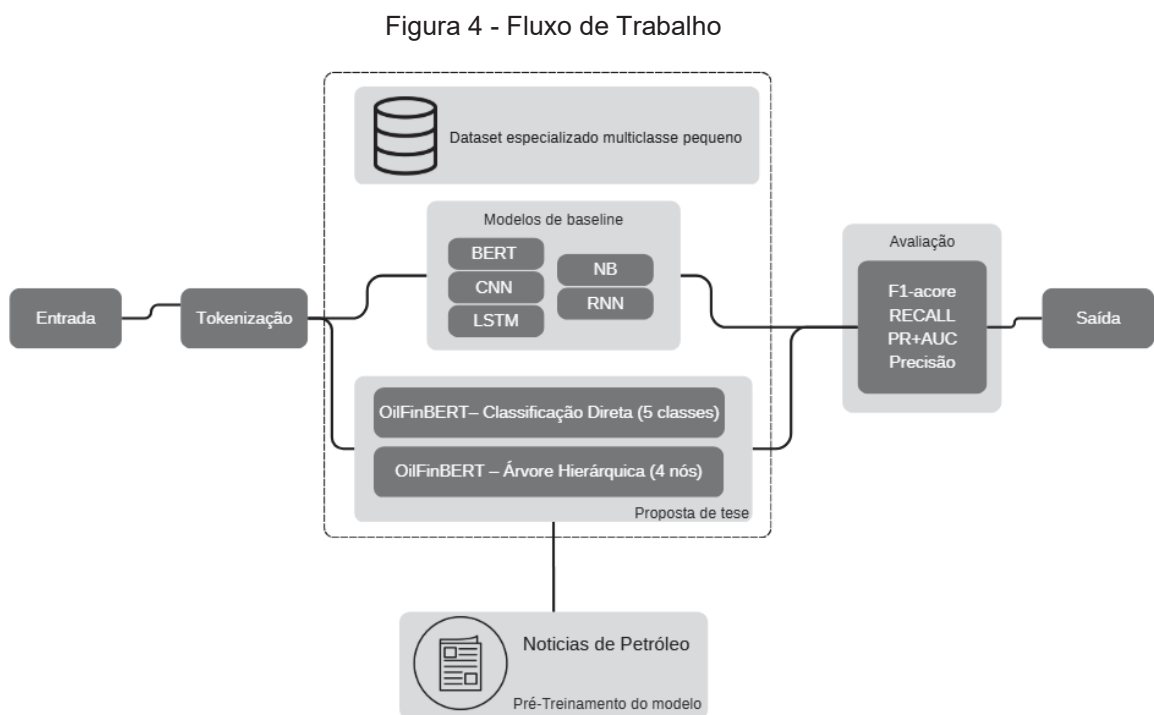
A hipótese é que essa combinação melhora a generalização e a consistência entre execuções, especialmente em classes com baixa representatividade. A eficiência será medida por F1-macro, precisão-macro, recall-macro e variabilidade entre *seeds*. Embora aplicada ao petróleo e gás, a metodologia pode ser replicada em outros domínios com escassez de dados rotulados.

4 PROPOSTA DE MODELO

O estudo segue um fluxo estruturado para o desenvolvimento do OilFinBERT, conforme ilustrado na Figura 4. O processo inicia-se com a definição do BERT-*base* como modelo inicial e a seleção de um conjunto não rotulado de notícias do setor para o pré-treinamento adicional. Em paralelo, são definidos os modelos de referência (NB, CNN, LSTM, RNN e BERT-*base*) utilizados para comparação. Na etapa seguinte, é construído um conjunto de dados rotulado e balanceado com cinco classes de intenção, validado por especialista do domínio.

O treinamento compreende duas fases: pré-treinamento no corpus especializado e *fine-tuning* supervisionado no conjunto de dados de intenções, em duas variantes de modelagem — classificação direta e abordagem hierárquica em quatro níveis. São avaliadas também estratégias de ajuste (treinamento padrão, reinicialização de camadas, LLRD e combinação) e técnicas de regularização, como *Mixout* e *SWA*.

A comparação dos modelos utiliza F1-macro, precisão-macro, recall-macro e PR-AUC, seguindo configurações experimentais uniformes descritas nas subseções deste capítulo.



Fonte: Autor (2025)

4.1 SELEÇÃO DO MODELO

O modelo adotado neste estudo é o BERT-*base*, arquitetura proposta por Devlin et al. (2019), composta por aproximadamente 110 milhões de parâmetros treináveis, 12 camadas *Transformer*, 12 mecanismos de atenção paralelos e dimensão oculta de 768 unidades. Esse modelo foi selecionado por apresentar desempenho consistente em tarefas de classificação de texto e por ser amplamente utilizado como referência em estudos de adaptação de domínio em PLN (Devlin et al., 2019). A partir dessa configuração, foi desenvolvido o OilFinBERT, que mantém a estrutura original do BERT-*base*, mas passa por pré-treinamento adicional com uma base não rotulada de notícias do setor de petróleo e gás, com o objetivo de aproximar o modelo das características linguísticas técnico-financeiras antes do *fine-tuning* supervisionado.

4.2 DEFINIÇÃO DA BASE DE DADOS NÃO ROTULADA PARA PRÉ-TREINAMENTO

A base de dados não rotulada utilizado no pré-treinamento adicional do OilFinBERT foi composto a partir de 16 fontes públicas de notícias do setor de petróleo e gás, obtidas por meio de *web scraping* em [Business Insider, Reuters, Financial Times, CNN, Bloomberg, International Business Times, Forbes, MarketWatch, WSJ, Yahoo, Fortune, Inc., The Economist, The New York Times, CNBC, MSN, Oil & Gas Journal]. A coleta considerou notícias publicadas entre janeiro de 2015 e maio de 2025. Foi considerado todas as notícias que eram relevantes ao tema de petróleo e gás, resultando inicialmente em 403.043 títulos.

Foram excluídos automaticamente itens duplicados, conteúdo fora do domínio (por exemplo, notícias sem relação com petróleo, empresas do setor ou fatores que afetam o preço da *commodity*) e registros contendo ruídos estruturais, como erros de extração, HTML ou *strings* inválidas.

Após esse processo, a base não rotulada foi submetida a uma etapa de verificação por especialista do domínio, utilizando amostragem aleatória para confirmar aderência temática e ausência de duplicidade residual. O resultado

consolidou-se em um conjunto textual consistente para adaptação do modelo ao vocabulário técnico-financeiro do setor de petróleo e gás.

4.3 DEFINIÇÃO DOS MODELOS DE REFERÊNCIA

Na fase inicial do estudo, foi estabelecido um conjunto de modelos de referência composto por modelos de classificação amplamente utilizados em tarefas de PLN. Foram considerados cinco modelos: NB, CNN, RNN, LSTM e BERT-*base*.

Esses modelos foram escolhidos por representarem diferentes paradigmas de aprendizado: o NB como método probabilístico de baixo custo computacional; as redes CNN, RNN e LSTM como arquiteturas clássicas de redes neurais aplicadas à classificação de texto; e o BERT-*base* como modelo pré-treinado de referência, capaz de capturar dependências contextuais de forma bidirecional.

A definição desse conjunto de modelos de referência permite comparar o desempenho do OilFinBERT com abordagens de diferentes níveis de complexidade, abrangendo desde métodos mais simples até arquiteturas baseadas em *Transformers*.

4.4 CONSTRUÇÃO DO CONJUNTO DE DADOS DE INTENÇÃO

Para a etapa de treinamento supervisionado, foi construído um conjunto de dados intitulado OilFinIntent, composto por mensagens curtas em inglês relacionadas ao setor de petróleo e gás. O conjunto foi rotulado em cinco classes de intenção: *previsao_preco*, *sentimento_noticia*, *preco_atual*, *conceito* e *outros*. Essas categorias foram definidas a partir das necessidades do domínio estudado, abrangendo desde consultas sobre valores de mercado e previsão de preços até análises de sentimento e questões conceituais. O conjunto de dados completo é apresentado no Apêndice A e uma amostra do conjunto de dados é apresentada Tabela 3.

Tabela 3 - Amostra do conjunto de dados OilFinIntent

Text	label
What confidence interval can we expect when forecasting Group 3 base oils?	<i>previsao_preco</i>
What is the prevailing sentiment in recent oil news?	<i>sentimento_noticia</i>
What is the current market price of Group 1?	<i>preco_atual</i>

Can univariate analysis be used on time series?	<i>conceito</i>
What is Petrobras' role in the Brazilian market?	<i>outros</i>

Fonte: Autor (2025)

O OilFinIntent foi organizado de forma balanceada, com 75 exemplos em cada classe, totalizando 375 instâncias. Essa distribuição busca preservar a proporção entre as categorias e minimizar distorções durante o treinamento. Todas as frases foram elaboradas em formato curto e objetivo, simulando consultas que poderiam ser realizadas em interações reais com *chatbots* ou sistemas de análise de dados. A construção do conjunto levou em consideração a especificidade do setor de petróleo e gás, no qual as informações combinam aspectos econômicos e técnicos. Após a elaboração das instâncias, o conjunto foi submetido à validação por um especialista do domínio, que verificou a adequação das frases e dos respectivos rótulos, assegurando a consistência do material utilizado no treinamento.

4.5 PRÉ-TREINAMENTO DO OILFINBERT

Para adaptar o modelo BERT ao domínio técnico-financeiro do setor de petróleo e gás, foi realizado um pré-treinamento adicional utilizando a base de dados não rotulada de 403.043 títulos de notícias, coletados entre janeiro de 2015 e maio de 2025. O procedimento adotou a tarefa de *Masked Language Modeling (MLM)*, com 15% das palavras mascaradas em cada sequência, permitindo que o modelo aprendesse a prever termos específicos a partir do contexto.

A base de dados foi processada com limite máximo de 128 *tokens* por sequência. O treinamento foi realizado durante 3 épocas, com *batch size* efetivo²⁴ de 32 (*per-device batch size*²⁵ 16 com *gradient accumulation*²⁶ de 2), utilizando GPU. O modelo resultante, denominado OilFinBERT, foi salvo para uso na etapa posterior de *fine-tuning* supervisionado.

4.6 FINE-TUNING SUPERVISIONADO

²⁴ *Batch size* efetivo — diferença entre batch real e efetivo

²⁵ *Per-device batch size* — que se refere ao *batch* por GPU, não total.

²⁶ *Gradient accumulation* — finalidade (permitir *batch* maior com menos memória).

Após o pré-treinamento adicional, o modelo OilFinBERT foi ajustado no conjunto de dados rotulados de intenções. O conjunto foi dividido de forma estratificada, reservando-se 80% para treinamento e 20% para validação, preservando a proporção entre as cinco classes. O processamento de entrada utilizou o tokenizador do BERT-base, com limite máximo de 256 tokens por sequência. Para uniformização, foram aplicadas as operações *padding*, que insere símbolos especiais de preenchimento de modo que todas as sequências tenham o mesmo tamanho, e truncamento, que reduz sequências mais longas do que o limite estabelecido, garantindo uniformidade no comprimento das entradas, garantindo que todas as sequências apresentassem comprimento fixo.

O treinamento supervisionado foi conduzido durante três épocas, utilizando lotes de 32 exemplos, otimizador AdamW com taxa de aprendizado inicial de 2×10^{-5} , *weight decay* de 0,01 e semente fixa para reprodutibilidade. A função de perda adotada foi a *cross-entropy*, apropriada para tarefas de classificação multiclasse. Ao final de cada época, um checkpoint foi salvo e o modelo de melhor desempenho na métrica F1-score macro na validação foi selecionado para a etapa de avaliação final.

4.7 ESTRATÉGIAS DE AJUSTE NO *FINE-TUNING*

Além do treinamento padrão, foram avaliadas três estratégias complementares de ajuste do modelo, com o objetivo de examinar seus efeitos sobre a adaptação do *Transformer* ao domínio técnico-financeiro.

4.7.1 Treinamento padrão

Essa configuração ajusta todos os pesos do modelo sem qualquer modificação estrutural adicional. Representa a forma canônica de *fine-tuning* utilizada em modelos derivados do BERT-base.

4.7.2 Reinicialização das camadas superiores

Nesta configuração, os pesos das duas últimas camadas do codificador *Transformer* e da camada final de classificação foram reiniciados antes do treinamento. Essa abordagem permite que as camadas superiores sejam ajustadas de maneira mais sensível ao domínio, enquanto as camadas inferiores preservam o conhecimento adquirido no pré-treinamento.

4.7.3 LLRD

O LLRD aplica taxas de aprendizado diferenciadas entre as camadas do modelo, utilizando valores maiores nas camadas superiores e menores nas camadas inferiores. O valor inicial global da taxa de aprendizado foi mantido, sendo aplicado um fator de decaimento para camadas progressivamente mais profundas. Essa estratégia busca limitar alterações estruturais nas primeiras camadas, preservando informações gerais aprendidas no pré-treinamento.

4.7.4 Estratégia combinada

Também foi avaliada a combinação entre reinicialização das camadas superiores e LLRD. Nessa configuração, as camadas reinicializadas recebem taxas de aprendizado mais elevadas, enquanto as demais camadas seguem o esquema de decaimento. Essa abordagem permite explorar eventuais efeitos complementares entre as duas técnicas.

4.8 MODELAGEM HIERÁRQUICA

Além da abordagem direta de classificação multiclasse, foi implementada uma estrutura hierárquica composta por quatro níveis decisórios. A árvore hierárquica é organizada por classificadores binários sucessivos, responsáveis por decidir se uma instância segue para um ramo específico do fluxo decisório. No último nível, um classificador multiclasse define o rótulo final entre as classes remanescentes. Essa estrutura permite analisar o desempenho de um arranjo de classificação incremental em comparação com a modelagem direta de cinco classes.

4.9 TÉCNICAS DE REGULARIZAÇÃO E ESTABILIDADE

Para mitigar variações decorrentes de instabilidade no processo de *fine-tuning* em conjuntos de dados reduzidos, foram aplicadas técnicas adicionais de regularização e estabilização, avaliadas separadamente das estratégias de ajuste.

4.9.1 Mixout

O Mixout substitui o *dropout* tradicional ao interpolar pesos do modelo com os valores do pré-treinamento. Essa estratégia reduz oscilações entre execuções e limita desvios significativos em relação à representação inicial do *Transformer*.

4.9.2 SWA

O SWA mantém uma média dos pesos ao longo das últimas etapas do treinamento. Essa média é utilizada para compor o modelo final, buscando estabilizar a convergência e reduzir dependência de oscilações específicas de uma única trajetória de otimização.

Essas técnicas foram avaliadas individualmente e não foram combinadas entre si, permitindo verificar seus efeitos de maneira isolada.

4.10 MÉTRICAS DE AVALIAÇÃO

A avaliação dos modelos foi realizada utilizando quatro métricas complementares: precisão, recall, F1-score e PR-AUC. O F1-score macro foi adotado como métrica principal para seleção do melhor modelo em cada configuração. Essa escolha permite avaliar o desempenho médio entre as classes, sem favorecer categorias mais frequentes.

O cálculo das métricas foi feito sobre o conjunto de validação, mantendo-se consistência com as configurações experimentais definidas para as etapas anteriores.

4.11 PROCEDIMENTOS EXPERIMENTAIS

Todas as configurações foram executadas sob as mesmas condições globais de treinamento (número de épocas, tamanho do lote, taxa de aprendizado inicial e *weight decay*), variando-se apenas a estratégia de ajuste ou regularização correspondente a cada experimento. Foi utilizada a mesma semente de inicialização em todas as execuções, garantindo reprodutibilidade. Os experimentos foram conduzidos em ambiente com GPU, e todos os logs, *checkpoints* e métricas foram registrados para posterior análise comparativa.

5 RESULTADO E DISCUSSÃO

Este capítulo apresenta os resultados obtidos nos experimentos e compara o desempenho do OilFinBERT com os modelos de referência. Também são analisados os efeitos das diferentes estratégias de *fine-tuning*, regularização e modelagem utilizadas, bem como o impacto dessas escolhas nas métricas de avaliação aplicadas ao conjunto de intenções.

5.1 PROTOTIPAGEM E FERRAMENTAS COMPUTACIONAIS UTILIZADAS

O modelo foi avaliado utilizando como referência outros modelos, entre eles NB, CNN, LSTM, RNN e BERT-Base. Para essa análise, empregou-se um conjunto de dados elaborado pela autora, direcionado ao domínio financeiro do setor de petróleo e gás.

Para o modelo OilFinBERT, foi aplicado o processo de *fine-tuning*, mediante técnicas avançadas de otimização, como a reinicialização das camadas, o emprego de diversos otimizadores, a regularização *Mixout*. Além disso, para melhorar a utilização dos recursos computacionais, o treinamento foi realizado em precisão reduzida (FP16), o que contribui para preservar a relação entre eficiência e desempenho.

A tese foi implementada em Python 3. A construção do fluxo de trabalho seguiu uma abordagem modular, permitindo reprodutibilidade dos resultados e flexibilidade para ajustes nos testes e análises.

Para a implementação do classificador de intenções utilizou-se a biblioteca *transformers* (Hugging Face), responsável pelo carregamento do modelo e do *tokenizador*, bem como pelos utilitários de treinamento, avaliação e salvamento de checkpoints. O gerenciamento do conjunto de dados em formato JSONL foi realizado com a biblioteca *datasets*, que oferece integração direta com rotinas de pré-processamento, mapeamento de rótulos e divisão dos dados. As métricas empregadas (acurácia, precisão-macro, *recall*-macro, F1-macro e PR-AUC) foram calculadas a partir de funções do *scikit-learn* integradas ao ciclo de avaliação do *Trainer*.

Os experimentos foram conduzidos no Google Colab, com aceleração por GPU NVIDIA T4. O *Google Drive* foi montado para leitura do conjunto de dados e

armazenamento dos artefatos (*checkpoints* e relatórios), garantindo organização dos caminhos de entrada e saída. Para assegurar reprodutibilidade, adotou-se execução determinística com fixação explícita das sementes pseudoaleatórias nas bibliotecas Python, NumPy e PyTorch (incluindo CUDA quando disponível).

O *dataset*, em JSONL, contém as chaves *text* e *label*. Após verificação de esquema, os rótulos foram codificados para índices inteiros por meio de *class_encode_column*, preservando o mapeamento *id2label/label2id*. A partição dos dados seguiu um esquema estratificado 80/20 (*train/val*) com *train_test_split*, a fim de manter a distribuição das classes nas duas dobras. A tokenização empregou o *AutoTokenizer* associado ao modelo de base adaptado, com truncamento em 128 *tokens*. O *DataCollatorWithPadding* foi utilizado para padronizar dinamicamente o comprimento dos lotes em tempo de treinamento.

O modelo foi carregado via *AutoModelForSequenceClassification*, configurando-se o número de classes e os dicionários *id2label/label2id*. O treinamento utilizou o Trainer da *transformers* com avaliação ao final de cada época, carregamento do melhor modelo ao término e *early stopping* (paciência 2 épocas). Os hiperparâmetros principais foram definidos empiricamente: taxa de aprendizado $2e-5$, *batch size* 32 (treino e validação), três épocas e *weight decay* 0,01. A métrica-alvo para seleção do melhor checkpoint foi o F1-macro (exposto como “f1” no *compute_metrics*). Ao final, o melhor modelo e o *tokenizador* foram persistidos, e executou-se uma rotina de inferência simples para verificação qualitativa.

O módulo *compute_metrics* consolidou as métricas agregadas: acurácia; precisão, recall e F1 na média macro (e, adicionalmente, F1 ponderado); e PR-AUC macro e ponderado. Para o cálculo do PR-AUC multiclasse, as probabilidades previstas foram comparadas à binarização *one-vs-rest* dos rótulos, utilizando *average_precision_score*. Complementarmente, gerou-se o *classification_report* por classe (scikit-learn), permitindo diagnóstico de precisão, recall e F1 por rótulo com seus respectivos suportes.

- Outras bibliotecas de apoio empregadas foram: *numpy* e *pandas*, para manipulação e inspeção de dados;
- *scikit-learn*, para métricas e a divisão estratificada;
- *torch* (*PyTorch*), como *backend* de treinamento;
- *os* e *gc*, para controle do ambiente e limpeza;

- google.colab.drive, para montagem do Google Drive e orquestração de caminhos.

Essa configuração assegura rastreabilidade dos experimentos (controle de sementes, versões de bibliotecas e *checkpoints*), reprodutibilidade dos resultados e facilidade de extensão do fluxo de trabalho para novas variações de modelo, ajustes de hiperparâmetros e experimentos adicionais.

5.2 RESULTADOS

Os experimentos realizados permitiram comparar diferentes configurações de treinamento do modelo OilFinBERT, bem como estabelecer um paralelo com abordagens clássicas de aprendizado de máquina e outras arquiteturas de redes neurais profundas. A Tabela 4 apresenta os resultados obtidos em termos de F1-macro, precisão-macro, recall-macro e PR-AUC.

Tabela 4 - Resultados dos Modelos

Modelo	F1-macro	Precisão-macro	Recall-macro	PR-AUC
OilFinBERT (AdamW)	0,9333	0,9431	0,9333	0,9770
OilFinBERT (AdamP)	0,9333	0,9431	0,9333	0,9727
OilFinBERT (SGD)	0,7159	0,7180	0,7200	0,8407
OilFinBERT (reinic. de camada)	0,5849	0,7597	0,6214	0,7863
OilFinBERT (hierárquico)	0,9474	0,9522	0,9467	0,9874
OilFinBERT (LLRD)	0,8802	0,8917	0,8800	0,9621
OilFinBERT (reinic. de camada + LLRD)	0,8330	0,8333	0,8357	0,8878
OilFinBERT (Mixout)	0,9470	0,9498	0,9466	0,9692
OilFinBERT (SWA)	0,9462	0,9541	0,9466	0,9787
NB	0,9013	0,9078	0,9032	0,9609
CNN	0,8410	0,8674	0,8503	0,9041
LSTM	0,6840	0,8044	0,7229	0,7970
RNN	0,1149	0,0849	0,2286	0,2530
BERT-base	0,6803	0,8004	0,6933	0,9871

Fonte: Autor (2025)

De forma geral, os resultados indicam que o OilFinBERT hierárquico apresentou o melhor desempenho, com F1-macro de 0,9474 e PR-AUC de 0,9874.

Essa configuração superou as demais, ainda que por margens estreitas em relação ao Mixout (F1-macro = 0,9470) e ao SWA (F1-macro = 0,9462). Considerando que o F1-macro atribui peso igual a todas as classes, independentemente de sua frequência, esse resultado indica que o modelo apresentou desempenho consistente ao longo das diferentes categorias de intenção, sem concentração de acertos em classes específicas. Esse aspecto é particularmente relevante em cenários nos quais as classes apresentam proximidade semântica, como ocorre no domínio técnico-financeiro do setor de petróleo e gás.

No caso específico do OilFinBERT hierárquico, o ganho observado em F1-macro indica que a estrutura proposta contribuiu para reduzir assimetrias entre precisão e recall ao longo das diferentes categorias de intenção. Esse comportamento sugere que a organização hierárquica das classes auxilia o modelo a lidar com ambiguidades semânticas recorrentes em textos técnico-financeiros, resultando em decisões mais equilibradas entre identificação correta e controle de erros. Em termos práticos, isso implica uma menor incidência de classificações enviesadas para classes específicas, característica desejável em aplicações nas quais diferentes tipos de intenção apresentam relevância semelhante. Essa evidência está em consonância com a literatura, que aponta que a classificação hierárquica tende a apresentar ganhos em domínios técnicos e especializados. Zangari et al. (2024), por exemplo, destacam que estruturas hierárquicas permitem capturar dependências entre classes, reduzindo erros em categorias correlacionadas e proporcionando maior consistência nas previsões.

Na análise dos otimizadores, observou-se que tanto AdamW quanto AdamP obtiveram desempenho superior em comparação ao SGD. Enquanto AdamW e AdamP atingiram F1-macro de 0,9333, com PR-AUC de 0,9770 e 0,9727, respectivamente, o SGD apresentou desempenho inferior, com F1-macro de 0,7159 e PR-AUC de 0,8407. Esses resultados indicam que os métodos adaptativos proporcionaram maior estabilidade durante o processo de *fine-tuning*, refletindo-se em métricas mais equilibradas. Esse comportamento é coerente com os resultados reportados por Wilson et al. (2018), que apontam que, embora o SGD tenda a apresentar boa generalização em bases de grande porte, ele sofre com instabilidade e baixa eficiência em conjuntos de dados pequenos. De modo semelhante, Mosbach, Andriushchenko e Klakow (2021) enfatizam que, em tarefas de *fine-tuning* de *Transformers* com conjuntos reduzidos, o AdamW apresenta estabilidade superior,

resultando em métricas mais consistentes. Além disso, a performance competitiva do AdamP confirma os achados de Heo et al. (2021), que mostraram que o algoritmo se adapta bem a arquiteturas invariantes à escala, estabilizando a convergência e mitigando problemas de crescimento excessivo da norma dos pesos.

No que se refere às técnicas de regularização, o *Mixout* e o SWA apresentaram desempenhos próximos ao do modelo hierárquico. O *Mixout* atingiu F1-macro de 0,9470, indicando que a estratégia de restringir parcialmente o afastamento dos parâmetros em relação ao modelo pré-treinado contribuiu para estabilizar o processo de ajuste em um cenário de dados limitados. Esse resultado confirma os achados de Lee, Cho e Kang (2020), que demonstraram ganhos consistentes de estabilidade em bases pequenas com o uso dessa técnica. Já o SWA destacou-se ao atingir o maior valor de PR-AUC (0,9787), ainda que com um F1-macro levemente inferior ao do *Mixout*. Esse comportamento é esperado, uma vez que o SWA busca soluções mais planas (*flat mínima*), favorecendo a robustez frente a mudanças de limiar. Os resultados corroboram estudos como os de Izmailov et al. (2019) que introduziram a técnica, e de Talman et al. (2023), que confirmaram melhorias de generalização em modelos de linguagem com SWA.

As estratégias de ajuste por camadas também apresentaram comportamentos distintos. A reinicialização isolada de camadas resultou em queda significativa de desempenho, com F1-macro de 0,5849 e PR-AUC de 0,7863, indicando perda de representações relevantes adquiridas durante o pré-treinamento. Em contrapartida, o LLRD obteve F1-macro de 0,8802 e PR-AUC de 0,9621, evidenciando sua capacidade de preservar o conhecimento geral nas camadas inferiores, ao mesmo tempo em que permite maior adaptação das camadas superiores à tarefa específica. A literatura reforça essa observação: Howard e Ruder (2018) e Zhang et al. (2021a) mostram que o ajuste discriminativo por camada melhora a performance de modelos pré-treinados em cenários de *fine-tuning*, equilibrando exploração e preservação de conhecimento.

Na comparação com os modelos de referência, o Naive Bayes apresentou F1-macro de 0,9013 e PR-AUC de 0,9609, mantendo desempenho competitivo mesmo frente a arquiteturas mais complexas. Esse resultado indica que, em bases pequenas e balanceadas, modelos probabilísticos simples podem capturar padrões relevantes quando as hipóteses de independência são razoavelmente atendidas, conforme discutido por Palanivinayagam, El-Bayeh e Damaševičius (2023). A CNN apresentou desempenho intermediário (F1-macro = 0,8410), confirmando sua capacidade de

capturar padrões locais, mas limitada diante de modelos mais sofisticados. Já a LSTM (F1-macro = 0,6840) e a RNN (F1-macro = 0,1149) tiveram resultados inferiores, o que está em linha com a literatura, segundo a qual essas arquiteturas, embora eficientes em dependências temporais longas, exigem mais dados para alcançar seu pleno potencial (Greff et al., 2017; Otter; Medina; Kalita, 2021).

Um aspecto relevante foi observado no comportamento do BERT-base, que apresentou PR-AUC elevado (0,9871), mas F1-macro reduzido (0,6803). Essa discrepância indica que o modelo foi capaz de ranquear adequadamente as instâncias em termos de probabilidade, mas apresentou dificuldades na conversão dessas estimativas probabilísticas em decisões finais consistentes no limiar padrão. Esse comportamento é indicativo de instabilidades no fine-tuning em cenários com conjuntos de dados reduzidos e está em linha com as observações de Mosbach, Andriushchenko e Klakow (2021), reforçando a importância de estratégias adicionais de adaptação ao domínio ou calibração de probabilidades.

Em síntese, os resultados obtidos confirmam a relevância de adaptar o BERT a domínios específicos por meio de pré-treinamento contínuo, regularização e ajustes de *fine-tuning*. O desempenho superior do OilFinBERT hierárquico demonstra que estruturas que exploram dependências entre classes são mais adequadas em contextos técnicos, enquanto técnicas como *Mixout* e SWA contribuem para reduzir instabilidades em bases pequenas. A comparação com *baselines* tradicionais reforça que, embora métodos probabilísticos como o NB ainda apresentem competitividade, as variantes do BERT oferecem desempenho superior quando adequadamente ajustadas. Finalmente, a análise crítica dos resultados evidencia que a tese não apenas confirma tendências já apontadas na literatura, como também contribui ao aplicá-las em um domínio de grande relevância prática, mas ainda pouco explorado, como o de petróleo e gás.

5.3 ANÁLISE DE FALHA

Embora os resultados gerais tenham demonstrado desempenho satisfatório do OilFinBERT e de suas variantes, a análise de falhas revela pontos importantes para compreender as limitações do modelo. Em diversas situações, erros de classificação ocorreram em frases cuja intenção apresentava proximidade semântica entre classes. Por exemplo, sentenças originalmente rotuladas como *previsao_preco* foram, em

alguns casos, confundidas com *sentimento_noticia*. Esse tipo de erro é compreensível, pois notícias que tratam de expectativas de mercado frequentemente carregam uma carga avaliativa, o que torna difícil separar previsão de sentimento. Situação semelhante foi observada entre *preco_atual* e *conceito*, em que expressões descritivas sobre movimentações financeiras ou definições técnicas levaram a classificações equivocadas.

Esses resultados reforçam a constatação já relatada na literatura de que classes semânticas próximas tendem a gerar fronteiras de decisão sobrepostas, mesmo em modelos de linguagem avançados (Guo et al., 2017; Niculescu-Mizil; Caruana, 2005). Além disso, confirmam as observações de Mosbach, Andriushchenko e Klakow (2021), segundo os quais o *fine-tuning* de modelos *Transformers* em bases pequenas pode resultar em instabilidades e classificações inconsistentes. Em alguns casos específicos, verificou-se que o modelo atribuía probabilidades semelhantes a duas classes distintas — por exemplo, 0,45 para *previsao_preco* e 0,43 para *sentimento_noticia*. Embora a predição final fosse incorreta, o valor próximo das probabilidades sugere que o modelo reconheceu a relação entre categorias, mas não conseguiu converter esse ranqueamento em uma decisão correta no ponto de corte.

Outro aspecto relevante é a discrepância observada entre métricas como PR-AUC e F1-macro em determinados modelos. Enquanto o BERT-*base* apresentou PR-AUC elevado (0,9871), seu F1-macro foi inferior (0,6803). Esse comportamento indica que, apesar de o modelo ordenar corretamente as instâncias em termos de probabilidade, a etapa de conversão dessas probabilidades em rótulos finais não refletiu o mesmo nível de desempenho. Tal diferença sugere que ajustes na calibração das probabilidades, ou na forma como as pontuações são transformadas em decisões de classe, podem melhorar a consistência entre as métricas (Niculescu-Mizil; Caruana, 2005). Essa constatação é relevante para futuras aplicações, uma vez que pequenas alterações no limiar de decisão poderiam reduzir substancialmente a ocorrência de erros em classes vizinhas.

Portanto, a análise de falhas evidencia que os equívocos não decorrem apenas de limitações do modelo, mas também da natureza ambígua das intenções envolvidas. Mais do que fragilidades, essas situações fornecem insumos para ajustes futuros, seja por meio da curadoria de exemplos adicionais em classes próximas, seja pela adoção de estratégias complementares, como calibradores de probabilidade ou técnicas de rejeição em casos incertos.

5.4 CONSIDERAÇÕES FINAIS

Os experimentos realizados demonstraram que o OilFinBERT hierárquico apresentou o melhor desempenho entre todas as configurações avaliadas, confirmando a importância de explorar dependências entre classes em contextos especializados. Técnicas de regularização, como *Mixout* e SWA, também se mostraram eficazes, aproximando-se dos resultados do modelo hierárquico e evidenciando sua relevância para reduzir instabilidades em bases pequenas.

Na análise dos otimizadores, os métodos adaptativos AdamW e AdamP apresentaram desempenho consistente e superior ao SGD, o que corrobora a literatura ao indicar que estratégias adaptativas são mais adequadas em tarefas de *fine-tuning* com conjuntos de dados reduzidos. Além disso, a comparação com os modelos de referência revelou que o NB manteve desempenho competitivo, superando CNN e LSTM em algumas métricas, enquanto a RNN mostrou-se pouco eficaz, confirmando sua limitação em cenários de maior complexidade sem dados abundantes.

Por fim, destaca-se o comportamento do BERT-*base*, que obteve alto PR-AUC, mas F1-macro relativamente baixo. Essa discrepância sugere que o modelo consegue ranquear adequadamente as probabilidades, mas encontra dificuldades na conversão em rótulos finais, apontando para a necessidade de calibração. De modo geral, os experimentos confirmam que a combinação de pré-treinamento contínuo, regularização apropriada e modelagem hierárquica é decisiva para alcançar resultados estáveis e robustos em contextos de dados limitados, fundamentando as discussões apresentadas no capítulo seguinte.

6 CONCLUSÃO

O objetivo geral desta tese consistiu em avaliar a eficácia de uma abordagem de análise de intenção de textos aplicada ao setor de petróleo e gás em cenários com conjuntos de dados pequenos. Para atender a esse objetivo, foram definidos como objetivos específicos o desenvolvimento de um modelo de análise de intenção baseado no BERT, adaptado ao domínio técnico-financeiro, e a construção de uma base de dados rotulada adequada à avaliação de modelos nesse contexto.

Os resultados obtidos permitem afirmar que o objetivo geral foi alcançado. A abordagem proposta, materializada no modelo OilFinBERT, demonstrou desempenho consistente em um cenário caracterizado por linguagem especializada e limitação de dados rotulados, condição recorrente no setor de petróleo e gás. Em particular, a adoção de uma estrutura de classificação hierárquica mostrou-se decisiva para melhorar o equilíbrio entre precisão e recall, refletido nos maiores valores de F1-macro observados entre todas as configurações avaliadas.

No que se refere ao primeiro objetivo específico, os experimentos confirmaram que a adaptação do BERT ao domínio técnico-financeiro, combinada com estratégias de *fine-tuning*, regularização e escolha adequada de otimizadores, influencia diretamente o desempenho do classificador de intenções. Entre as variações analisadas, o OilFinBERT hierárquico apresentou resultados superiores em comparação às abordagens multiclasse planas, indicando que a exploração explícita das relações semânticas entre classes reduz ambiguidades recorrentes em intenções conceitualmente próximas, como aquelas relacionadas a preços, previsões e interpretações de notícias.

Em relação ao segundo objetivo específico, a construção da base OilFinIntent, organizada em cinco classes de intenção relevantes ao domínio estudado, foi fundamental para viabilizar uma avaliação sistemática dos modelos. A utilização dessa base permitiu comparar diferentes estratégias de modelagem e evidenciar que arquiteturas hierárquicas e técnicas de regularização são particularmente adequadas quando o volume de dados rotulados é reduzido. Dessa forma, a base de dados constitui uma contribuição adicional da tese, tanto para a reprodutibilidade dos experimentos quanto para o desenvolvimento de estudos futuros no setor.

A comparação com modelos de referência reforçou a contribuição da abordagem proposta. Embora métodos tradicionais, como o *Naive Bayes*, tenham apresentado desempenho competitivo, nenhum deles alcançou os resultados obtidos pelo OilFinBERT hierárquico. Da mesma forma, o BERT-base, quando aplicado sem adaptação ao domínio, apresentou discrepâncias entre métricas de ranqueamento e métricas de classificação direta, evidenciando limitações em cenários especializados e com poucos dados. Nesse contexto, o OilFinBERT hierárquico mostrou-se mais consistente, ao combinar capacidade discriminativa com estabilidade nas decisões.

Assim, a principal contribuição desta tese reside na demonstração empírica de que a combinação de pré-treinamento contínuo, regularização apropriada e modelagem hierárquica constitui uma abordagem eficaz para análise de intenção em domínios técnicos com dados limitados. O modelo OilFinBERT hierárquico representa um avanço metodológico ao explorar explicitamente a estrutura semântica das intenções, oferecendo uma solução adequada ao setor de petróleo e gás e contribuindo para o avanço da pesquisa em PLN aplicado a contextos especializados.

6.1 LIMITAÇÕES

Entre as limitações do estudo, destacam-se: (i) o tamanho reduzido do *dataset* de intenções, que limita a capacidade de generalização para interações mais variadas; (ii) a aplicação restrita a textos em inglês, não explorando particularidades linguísticas do português ou de outros idiomas utilizados no setor; e (iii) a ausência de testes em cenários reais de uso contínuo do *chatbot*, nos quais fatores como ambiguidade, ruído e variação de estilo poderiam impactar o desempenho.

6.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

A partir dos resultados e limitações observados, algumas recomendações podem ser feitas para trabalhos futuros. Em primeiro lugar, sugere-se ampliar o *dataset* de intenções com dados coletados em interações reais, de modo a enriquecer a diversidade de exemplos e permitir maior robustez na classificação. Outra possibilidade é explorar a adaptação do modelo para o português, seja por meio de treinamento multilíngue ou de pré-treinamentos específicos em corpora nacionais do setor.

Adicionalmente, recomenda-se investigar técnicas de calibração de probabilidades e ajuste de limiares de decisão, a fim de reduzir discrepâncias entre métricas como PR-AUC e F1-macro. A aplicação de arquiteturas mais recentes, como DeBERTa ou LLMs especializados em finanças, também se apresenta como caminho promissor para verificar ganhos adicionais de desempenho. Por fim, destaca-se a importância de validar o modelo em cenários de produção, avaliando não apenas métricas quantitativas, mas também indicadores de usabilidade, tempo de resposta e aceitação por parte dos usuários.

Essas recomendações apontam para a continuidade natural da pesquisa, buscando consolidar e expandir os avanços apresentados neste trabalho. Ao integrar novas fontes de dados, explorar idiomas distintos e testar em ambientes reais, será possível evoluir o modelo para aplicações mais amplas, fortalecendo tanto sua relevância acadêmica quanto sua utilidade prática no setor de petróleo e gás.

6.3 PUBLICAÇÕES

Decorrentes desta pesquisa, foram obtidas três publicações:

1. PEREIRA, Rúbia Maria; SCARPIN, Cassius Tadeu; VALLE, Pablo Deivid. Classificação de notícias relacionadas ao setor de óleo e gás. In: XII SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 13, 2025. Anais [...]. João Pessoa-PB, Hotel Caiçara, 2024. ISSN: 2318-9258. Disponível em: <https://dspace.sti.ufcg.edu.br/handle/riufcg/43923>. Acesso em: 28/01/2026
2. **Artigo Análise de Intenção em Conjuntos de Dados Pequenos: Comparação do Desempenho do Naive Bayes e do BERT**, — apresentado no XLIV Congresso Nacional de Matemática Aplicada e Computacional – CNMAC. O trabalho será publicado no Proceeding Series of the Brazilian Society of Computational and Applied Mathematics
3. PEREIRA, Rúbia Maria; SCARPIN, CASSIUS TADEU; VALLE, Pablo Deivid. ABORDAGENS DE PLN PARA CLASSIFICAÇÃO DE INTENÇÕES TÉCNICAS E FINANCEIRAS: COMPARANDO CNN E BERT.. In: Anais - SIMPEP. Anais...Bauru(SP) UNESP, 2025. Disponível em: https://www.even3.com.br/anais/xxxii_simpep/1207525-ABORDAGENS-DE-PLN-PARA-CLASSIFICACAO-DE-INTENCOES-TECNICAS-E-FINANCEIRAS--COMPARANDO-CNN-E-BERT. Acesso em: 28/01/2026

REFERÊNCIAS

- ANANDARAJAN, Murugan; HILL, Chelsey; NOLAN, Thomas. Text Preprocessing. *In: Practical Text Analytics. Advances in Analytics and Data Science. [S.l.]*: Springer, 2019. v. 2 p. 45–59.
- ASLAM, Farhan. The Impact of Artificial Intelligence on Chatbot Technology: A Study on the Current Advancements and Leading Innovations. **European Journal of Technology**, v. 7, n. 3, 2023.
- ASSAYED, Suha K.; SHAALAN, Khaled; ALKHATIB, Manar. A Chatbot Intent Classifier for Supporting High School Students. **EAI Endorsed Transactions on Scalable Information Systems**, v. 10, n. 3, 2023.
- AYADI, Souha; LACHIRI, Zied. Deep neural network architectures for audio emotion recognition performed on song and speech modalities. **International Journal of Speech Technology**, v. 26, n. 4, 2023.
- BILAH, Chiva Olivia; ADJI, Teguh Bharata; SETIAWAN, Noor Akhmad. Intent Detection on Indonesian Text Using Convolutional Neural Network. *In: 2022*.
- BOTTOU, Léon. Large-Scale Machine Learning with Stochastic Gradient Descent. *In: Proceedings of COMPSTAT'2010*. Heidelberg: Physica-Verlag HD, 2010. p. 177–186.
- CAÑAS, Héctor *et al.* **Implementing Industry 4.0 principles. Computers and Industrial Engineering**, 2021.
- CHAKRABORTY, Chiranjib *et al.* From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare. **Current Research in Biotechnology**, v. 7, p. 100164, 2024.
- CHEN, Qian; ZHUO, Zhu; WANG, Wen. BERT for Joint Intent Classification and Slot Filling. **arXiv:1902.10909**, 28 fev. 2019.
- CHO, Kyunghyun *et al.* Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. 3 set. 2014.
- CLARK, Kevin *et al.* ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. 23 mar. 2020.
- DENG, Song *et al.* A Chinese power text classification algorithm based on deep active learning. **Applied Soft Computing**, v. 150, 2024.
- DEVLIN, Jacob *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. *In: 2019*.
- DIAB, Diab M.; EL HINDI, Khalil M. Using differential evolution for fine tuning naïve Bayesian classifiers and its application for text classification. **Applied Soft Computing Journal**, v. 54, 2017.

EZEN-CAN, Aysu. A Comparison of LSTM and BERT for Small Corpus. **arXiv:2009.05451**, 14 set. 2020.

FERNÁNDEZ-MARTÍNEZ, Fernando *et al.* Fine-Tuning BERT Models for Intent Recognition Using a Frequency Cut-Off Strategy for Domain-Specific Vocabulary Extension. **Applied Sciences (Switzerland)**, v. 12, n. 3, 2022.

FU, Zihao; SO, Anthony Man-Cho; COLLIER, Nigel. A Stability Analysis of Fine-Tuning a Pre-Trained Model. 7 dez. 2023.

GHOBAKHLOO, Morteza. **Industry 4.0, digitization, and opportunities for sustainability. Journal of Cleaner Production**, 2020.

GOMES, Diogo da Silva Magalhães *et al.* Portuguese word embeddings for the oil and gas industry: Development and evaluation. **Computers in Industry**, v. 124, 2021.

GREFF, Klaus *et al.* LSTM: A Search Space Odyssey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 28, n. 10, 2017.

GUO, Chuan *et al.* On Calibration of Modern Neural Networks. 3 ago. 2017.

GUO, Hao; JIN, Jiyong; LIU, Bin. Stochastic Weight Averaging Revisited. **Applied Sciences**, v. 13, n. 5, p. 2935, 24 fev. 2023.

HASAN, Mehedi; IBNA BASHER, Mohammad Jahid; SHAWON, Md Tanvir Rouf. Bengali Intent Classification with Generative Adversarial BERT. *In*: 2023.

HELM SETYAWAN, Muhammad Yusril; AWANGGA, Rolly Maulana; EFENDI, Safif Rafi. Comparison Of Multinomial Naive Bayes Algorithm And Logistic Regression For Intent Classification In Chatbot. *In*: IEEE, out. 2018.

HEO, Byeongho *et al.* AdamP: Slowing Down the Slowdown for Momentum Optimizers on Scale-invariant Weights. 18 jan. 2021.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1 nov. 1997.

HOWARD, Jeremy; RUDER, Sebastian. Universal Language Model Fine-tuning for Text Classification. *In*: Stroudsburg, PA, USA: Association for Computational Linguistics, 2018.

HUANG, Xuejian *et al.* An effective multimodal representation and fusion method for multimodal intent recognition. **Neurocomputing**, v. 548, 2023.

IZMAILOV, Pavel *et al.* Averaging Weights Leads to Wider Optima and Better Generalization. 25 fev. 2019.

JANIESCH, Christian; ZSCHECH, Patrick; HEINRICH, Kai. Machine learning and deep learning. **Electronic Markets**, v. 31, n. 3, 2021.

JIANG, Liangxiao *et al.* Deep feature weighting for naive Bayes and its application to text classification. **Engineering Applications of Artificial Intelligence**, v. 52, 2016.

KIM, Yoon. Convolutional Neural Networks for Sentence Classification. 3 set. 2014.

KINGMA, Diederik P.; BA, Jimmy. Adam: A Method for Stochastic Optimization. 30 jan. 2017.

KOKOL, Peter; KOKOL, Marko; ZAGORANSKI, Sašo. Machine learning on small size samples: A synthetic knowledge synthesis. **Science Progress**, v. 105, n. 1, 27 jan. 2022.

KOWSARI, Kamran *et al.* **Text classification algorithms: A survey. Information (Switzerland)**, 2019.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 28 maio 2015.

LEE, Cheolhyoung; CHO, Kyunghyun; KANG, Wanmo. Mixout: Effective Regularization to Finetune Large-scale Pretrained Language Models. 23 jan. 2020.

LEE, Jui Hsuan *et al.* Anti-Drugs Chatbot: Chinese BERT-Based Cognitive Intent Analysis. **IEEE Transactions on Computational Social Systems**, v. 11, n. 1, 2024.

LI, Junyi *et al.* Pre-Trained Language Models for Text Generation: A Survey. **ACM Computing Surveys**, v. 56, n. 9, p. 1–39, 31 out. 2024.

LI, Qian *et al.* **A Survey on Text Classification: From Traditional to Deep Learning. ACM Transactions on Intelligent Systems and Technology**, 2022.

LI, Xuerong; SHANG, Wei; WANG, Shouyang. Text-based crude oil price forecasting: A deep learning approach. **International Journal of Forecasting**, v. 35, n. 4, 2019.

LIN, Chien Chang; HUANG, Anna Y. Q.; YANG, Stephen J. H. **A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999–2022). Sustainability (Switzerland)**, 2023.

LIU, Cong; XU, Xiaolong. AMFF: A new attention-based multi-feature fusion method for intention recognition. **Knowledge-Based Systems**, v. 233, 5 dez. 2021.

LIU, Pengfei; QIU, Xipeng; HUANG, Xuanjing. Recurrent Neural Network for Text Classification with Multi-Task Learning. **arXiv:1605.05101**, 17 maio 2016.

LOSHCHILOV, Ilya; HUTTER, Frank. Decoupled Weight Decay Regularization. 4 jan. 2019.

LU, Hongfang *et al.* **Oil and Gas 4.0 era: A systematic review and outlook. Computers in Industry**, 2019.

LU, Peng *et al.* Improving Generalization of Pre-trained Language Models via Stochastic Weight Averaging. 16 dez. 2022.

LUO, Xiaoyu. Efficient English text classification using selected Machine Learning Techniques. **Alexandria Engineering Journal**, v. 60, n. 3, 2021.

MCCALLUM, Andrew; NIGAM, Kamal. **A Comparison of Event Models for Naive Bayes Text Classification**. [S.l.: S.n.]. Disponível em: <www.aaii.org>.

MILOSEVIC, Nikola; DEGHANTANHA, Ali; CHOO, Kim Kwang Raymond. Machine learning aided Android malware classification. **Computers and Electrical Engineering**, v. 61, 2017.

MOSBACH, Marius; ANDRIUSHCHENKO, Maksym; KLAKOW, Dietrich. On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines. 25 mar. 2021.

NICULESCU-MIZIL, Alexandru; CARUANA, Rich. Predicting good probabilities with supervised learning. *In*: New York, New York, USA: ACM Press, 2005.

ONAL, Emre *et al.* Gaussian Stochastic Weight Averaging for Bayesian Low-Rank Adaptation of Large Language Models. 20 jul. 2024.

OTTER, Daniel W.; MEDINA, Julian R.; KALITA, Jugal K. A Survey of the Usages of Deep Learning for Natural Language Processing. **IEEE Transactions on Neural Networks and Learning Systems**, v. 32, n. 2, 2021.

PALANIVINAYAGAM, Ashokkumar; EL-BAYEH, Claude Ziad; DAMAŠEVIČIUS, Robertas. Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review. **Algorithms**, v. 16, n. 5, 2023.

PÉREZ-SOLER, Sara; GUERRA, Esther; DE LARA, Juan. A model-based solution for automated (Re-)engineering of task-oriented chatbots. **Journal of Systems and Software**, v. 231, p. 112600, jan. 2026.

QIU, Xi Peng *et al.* **Pre-trained models for natural language processing: A survey. Science China Technological Sciences**, 2020.

RIZOU, S. *et al.* Multilingual Name Entity Recognition and Intent Classification employing Deep Learning architectures. **Simulation Modelling Practice and Theory**, v. 120, 2022.

SELMY, Hend A.; MOHAMED, Hoda K.; MEDHAT, Walaa. **Big data analytics deep learning techniques and applications: A survey. Information Systems**, 2024.

SINGH, Ajay; JIA, Tianxia; NALAGATLA, Varun. Generative AI Enabled Conversational Chatbot for Drilling and Production Analytics. *In*: 2023.

SIRCAR, Anirbid *et al.* **Application of machine learning and artificial intelligence in oil and gas industry.** *Petroleum Research*, 2021.

SUDHIR, Prajval; SURESH, Varun Deshakulkarni. Comparative study of various approaches, applications and classifiers for sentiment analysis. **Global Transitions Proceedings**, v. 2, n. 2, 2021.

TALMAN, Aarne *et al.* Uncertainty-Aware Natural Language Inference with Stochastic Weight Averaging. 10 abr. 2023.

THANGARAJ, M.; SIVAKAMI, M. Text classification techniques: A literature review. **Interdisciplinary Journal of Information, Knowledge, and Management**, v. 13, 2018.

VASWANI, Ashish *et al.* Attention Is All You Need. 2 ago. 2017.

WANASINGHE, Thumeera R. *et al.* **Digital Twin for the Oil and Gas Industry: Overview, Research Trends, Opportunities, and Challenges.** *IEEE Access*, 2020.

WILLIAMS, Ronald J.; ZIPSER, David. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. **Neural Computation**, v. 1, n. 2, 1989.

WILSON, Ashia C. *et al.* The Marginal Value of Adaptive Gradient Methods in Machine Learning. 22 maio 2018.

WU, Tiefeng *et al.* Intent recognition model based on sequential information and sentence features. **Neurocomputing**, v. 566, 21 jan. 2024.

XU, Jingyun *et al.* Incorporating context-relevant concepts into convolutional neural networks for short text classification. **Neurocomputing**, v. 386, 2020.

XU, Pengcheng *et al.* Small data machine learning in materials science. **npj Computational Materials**, v. 9, n. 1, p. 42, 25 mar. 2023.

YOU, Yang *et al.* Large Batch Optimization for Deep Learning: Training BERT in 76 minutes. 3 jan. 2020.

ZANGARI, Alessandro *et al.* Hierarchical Text Classification and Its Foundations: A Review of Current Research. **Electronics**, v. 13, n. 7, p. 1199, 25 mar. 2024.

ZHANG, Tianyi *et al.* Revisiting Few-sample BERT Fine-tuning. 11 mar. 2021.

ZHANG, Xiang; ZHAO, Junbo; LECUN, Yann. Character-level convolutional networks for text classification. **Advances in neural information processing systems**, v. 28, 2015.

ZHANG, Yongxiang *et al.* Business chatbots with deep learning technologies: state-of-the-art, taxonomies, and future research directions. **Artificial Intelligence Review**, v. 57, n. 5, p. 113, 11 abr. 2024.

APÊNDICE A – DATASET OINFININTENT

```
{ "text": "How important is it to compare forecasts between the three groups before investing?", "label": "previsao_preco" }
{ "text": "In which scenarios is a hybrid model preferable?", "label": "conceito" }
{ "text": "What is univariate analysis?", "label": "conceito" }
{ "text": "Does Yahoo Finance's news bring more pessimism?", "label": "sentimento_noticia" }
{ "text": "What is the difference between RMSE and MAE?", "label": "conceito" }
{ "text": "How much is base oil 30 being charged today?", "label": "preco_atual" }
{ "text": "How does the stock market work?", "label": "outros" }
{ "text": "What is EBITDA?", "label": "outros" }
{ "text": "How much is base oil 22 right now?", "label": "preco_atual" }
{ "text": "How does the hedging strategy work for airlines that buy fuel?", "label": "outros" }
{ "text": "What is the current market price of Group 2?", "label": "preco_atual" }
{ "text": "In what situations is ARIMAX used?", "label": "conceito" }
{ "text": "What is the current price of diesel today? How much does kerosene cost right now?", "label": "preco_atual" }
{ "text": "What does RMSE mean?", "label": "conceito" }
{ "text": "What is base oil 13 now?", "label": "preco_atual" }
{ "text": "How much are petroleum products costing today?", "label": "preco_atual" }
{ "text": "What is the BERT model?", "label": "conceito" }
{ "text": "How much does Group 2 cost right now?", "label": "preco_atual" }
{ "text": "Did the model identify negative sentiment after sharp price drops?", "label": "sentimento_noticia" }
{ "text": "What are government subsidies in the oil sector?", "label": "outros" }
{ "text": "What are the limitations of using only historical data?", "label": "conceito" }
{ "text": "Do hybrid methods provide more stable price forecasts?", "label": "previsao_preco" }
{ "text": "How does an RBF neural network work?", "label": "conceito" }
{ "text": "How much is Group 2 being charged today?", "label": "preco_atual" }
{ "text": "Which group of base oils has the highest price today?", "label": "preco_atual" }
{ "text": "What is marginal cost of production?", "label": "outros" }
{ "text": "Does the model calculate low sentiment values in MSN's news?", "label": "sentimento_noticia" }
{ "text": "How do statistical price forecasting models work?", "label": "previsao_preco" }
{ "text": "Can the ARIMA model predict Group 2 base oil prices with good accuracy?", "label": "previsao_preco" }
{ "text": "Does the analysis indicate a predominance of negative news currently?", "label": "sentimento_noticia" }
{ "text": "What is sentiment analysis?", "label": "conceito" }
{ "text": "What is an option contract?", "label": "outros" }
{ "text": "How to interpret the MAPE value?", "label": "conceito" }
{ "text": "Does the model show that sentiment has improved since 2015?", "label": "sentimento_noticia" }
{ "text": "Can forecasting for all 30 base oils be done with the same model, or is it better to separate them by group?", "label": "previsao_preco" }
{ "text": "What does political risk mean in investments?", "label": "outros" }
```

{\"text\": \"What are the limitations of linear regression?\", \"label\": \"conceito\"}

{\"text\": \"What is the ARIMA model for?\", \"label\": \"conceito\"}

{\"text\": \"Is it better to use univariate or multivariate analysis to forecast prices for one of the 30 base oils?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is the difference between using MLP and RBF in price forecasting?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is the current price of base oil 6?\", \"label\": \"preco_atual\"}

{\"text\": \"What is the average price of Group 3 base oils right now?\", \"label\": \"preco_atual\"}

{\"text\": \"What is a hybrid model?\", \"label\": \"conceito\"}

{\"text\": \"Does the model indicate a significant variation in the base oil 23 price over the next 6 months?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is a financial derivative?\", \"label\": \"outros\"}

{\"text\": \"What is vertical integration in the energy sector?\", \"label\": \"outros\"}

{\"text\": \"What does GDP mean?\", \"label\": \"outros\"}

{\"text\": \"How to interpret a low RMSE value?\", \"label\": \"conceito\"}

{\"text\": \"What does it mean to use historical data in a prediction?\", \"label\": \"conceito\"}

{\"text\": \"What is MAE?\", \"label\": \"conceito\"}

{\"text\": \"What is the price of Group 3 base oils today?\", \"label\": \"preco_atual\"}

{\"text\": \"What does disruptive innovation mean in the energy sector?\", \"label\": \"outros\"}

{\"text\": \"What is the current price of base oil 14?\", \"label\": \"preco_atual\"}

{\"text\": \"Is the base oil 15 price likely to rise next month?\", \"label\": \"previsao_preco\"}

{\"text\": \"Is it possible to forecast prices for all 30 base oils with a single model?\", \"label\": \"previsao_preco\"}

{\"text\": \"Does the calculated sentiment value indicate stability or variation?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What is the importance of exchange rates in international trade?\", \"label\": \"outros\"}

{\"text\": \"Does BERT require large volumes of data to function effectively?\", \"label\": \"conceito\"}

{\"text\": \"How can forecasts be used even when knowing they contain uncertainties?\", \"label\": \"previsao_preco\"}

{\"text\": \"Which countries are part of OPEC?\", \"label\": \"outros\"}

{\"text\": \"What does local content mean in the oil and gas industry?\", \"label\": \"outros\"}

{\"text\": \"What is the price of ethanol today?\", \"label\": \"preco_atual\"}

{\"text\": \"What do upstream and downstream mean in the oil industry?\", \"label\": \"outros\"}

{\"text\": \"How to interpret a low MAE value?\", \"label\": \"conceito\"}

{\"text\": \"Were oil stories in 2015 more pessimistic?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"In what contexts should MLP be used instead of RBF?\", \"label\": \"conceito\"}

{\"text\": \"How much is base oil 12 today?\", \"label\": \"preco_atual\"}

{\"text\": \"Which method is more accurate in forecasting Group 1 prices compared to Group 2?\", \"label\": \"previsao_preco\"}

{\"text\": \"Are Group 1 forecasts more stable than Group 2?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is the difference between fixed income and variable income?\", \"label\": \"outros\"}

{\"text\": \"What does the Selic rate represent?\", \"label\": \"outros\"}

{\"text\": \"How can forecasts with different errors between groups 1, 2, and 3 be interpreted?\", \"label\": \"previsao_preco\"}

{ "text": "Were the most recent sentiment analyses positive?", "label": "sentimento_noticia" }

{ "text": "Is the tone of Business Insider reports more neutral?", "label": "sentimento_noticia" }

{ "text": "Has neutral sentiment predominated in 2021?", "label": "sentimento_noticia" }

{ "text": "Can price forecasts support hedging strategies?", "label": "previsao_preco" }

{ "text": "Who are the world's largest oil producers?", "label": "outros" }

{ "text": "What is working capital?", "label": "outros" }

{ "text": "Has there been more confidence in the news since 2018?", "label": "sentimento_noticia" }

{ "text": "Can univariate analysis be used in time series?", "label": "conceito" }

{ "text": "How to treat redundant variables in multivariate analyses?", "label": "conceito" }

{ "text": "Can you tell me the price of base oil 24?", "label": "preco_atual" }

{ "text": "Why is historical data important in predictive models?", "label": "conceito" }

{ "text": "How does risk management work in oil companies?", "label": "outros" }

{ "text": "What is financial leverage?", "label": "outros" }

{ "text": "Is The Economist's average sentiment higher than CNBC's?", "label": "sentimento_noticia" }

{ "text": "Is the average sentiment value in Forbes's articles neutral?", "label": "sentimento_noticia" }

{ "text": "What is the average price of Group 1 base oils today?", "label": "preco_atual" }

{ "text": "What differences can occur in short-term and long-term forecasts for Group 1?", "label": "previsao_preco" }

{ "text": "What is exchange rate risk?", "label": "outros" }

{ "text": "Can you tell me the price of base oil 19?", "label": "preco_atual" }

{ "text": "What is linear regression?", "label": "conceito" }

{ "text": "Can the hybrid model improve price forecasting for a specific base oil?", "label": "previsao_preco" }

{ "text": "What is the price difference between Group 1 and Group 2 today?", "label": "preco_atual" }

{ "text": "What is the Group 2 price trend for the end of the year?", "label": "previsao_preco" }

{ "text": "How much does Base Oil 2 cost right now? ", "label": "preco_atual" }

{ "text": "How important is oil in Brazilian foreign trade?", "label": "outros" }

{ "text": "What indicators are used in univariate analysis?", "label": "conceito" }

{ "text": "To what extent should managers rely on price forecasts?", "label": "previsao_preco" }

{ "text": "How does ARIMA handle time series?", "label": "conceito" }

{ "text": "How can I assess forecast consistency over time?", "label": "previsao_preco" }

{ "text": "Did the model detect optimism in Business Insider's analyses?", "label": "sentimento_noticia" }

{ "text": "Was the sentiment value of this week's articles positive or negative?", "label": "sentimento_noticia" }

{ "text": "What is the base oil 30 price forecast for the end of the year?", "label": "previsao_preco" }

{ "text": "What is arbitrage in the financial market?", "label": "outros" }

{ "text": "Is the sentiment in Reuters news about oil positive or negative?", "label": "sentimento_noticia" }

{ "text": "Has there been an increase in the number of positive news stories?", "label": "sentimento_noticia" }

{ "text": "Is the overall sentiment more optimistic than pessimistic?", "label": "sentimento_noticia" }

{ "text": "Was sentiment in 2022 more neutral than in 2021?", "label": "sentimento_noticia" }

{ "text": "What is the current price of base oil 3?", "label": "preco_atual" }

{ "text": "What is the current price of fuel oil?", "label": "preco_atual" }

{ "text": "How can machine learning models predict prices?", "label": "previsao_preco" }

{ "text": "Does the model indicate positive sentiment in articles about Petrobras?", "label": "sentimento_noticia" }

{ "text": "In which cases is univariate analysis sufficient?", "label": "conceito" }

{ "text": "Do Oil & Gas Journal's articles tend to be more positive?", "label": "sentimento_noticia" }

{ "text": "What is Petrobras' role in the Brazilian market?", "label": "outros" }

{ "text": "Which external variables most affect Group 3 forecasts?", "label": "previsao_preco" }

{ "text": "What is the difference between descriptive and univariate analysis?", "label": "conceito" }

{ "text": "Is the overall sentiment on oil prices one of confidence or uncertainty?", "label": "sentimento_noticia" }

{ "text": "How do rating agencies work?", "label": "outros" }

{ "text": "How much is base oil 25 being charged today?", "label": "preco_atual" }

{ "text": "What does net income mean?", "label": "outros" }

{ "text": "Is it possible to obtain 95% confidence forecasts for Group 1 prices?", "label": "previsao_preco" }

{ "text": "How is the performance of a hybrid model evaluated?", "label": "conceito" }

{ "text": "What are the limitations of ARIMAX?", "label": "conceito" }

{ "text": "How do energy companies use price forecasts for the 30 base oils for strategic decisions?", "label": "previsao_preco" }

{ "text": "Does the Group 1 forecast indicate stability or variation in the next 30 days?", "label": "previsao_preco" }

{ "text": "How much does Group 1 cost right now?", "label": "preco_atual" }

{ "text": "How much does a base oil cost on average today?", "label": "preco_atual" }

{ "text": "What is trade balance?", "label": "outros" }

{ "text": "How can an MLP neural network predict prices?", "label": "previsao_preco" }

{ "text": "Do hybrid models improve price forecast accuracy?", "label": "previsao_preco" }

{ "text": "Is RMSE sensitive to outliers?", "label": "conceito" }

{ "text": "What is the difference between risk and uncertainty?", "label": "outros" }

{ "text": "Was the sentiment value for articles about diesel neutral?", "label": "sentimento_noticia" }

{ "text": "What is the difference between historical data and real-time data?", "label": "conceito" }

{ "text": "What is the highest price among the 30 base oils today?", "label": "preco_atual" }

{ "text": "What is the current price of base oil 5?", "label": "preco_atual" }

{\"text\": \"What is multivariate analysis?\", \"label\": \"conceito\"}

{\"text\": \"Is RMSE sufficient to evaluate base oil price forecasts?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is environmental regulation in the energy sector?\", \"label\": \"outros\"}

{\"text\": \"What is market volatility?\", \"label\": \"outros\"}

{\"text\": \"Is it better to use statistical or machine learning models to predict prices?\", \"label\": \"previsao_preco\"}

{\"text\": \"Was last week's sentiment more positive than the previous month's?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"How much is Group 3 being charged today?\", \"label\": \"preco_atual\"}

{\"text\": \"Does Financial Times analysis indicate confidence or uncertainty?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What is a low-carbon economy?\", \"label\": \"outros\"}

{\"text\": \"How much is base oil 30 currently?\", \"label\": \"preco_atual\"}

{\"text\": \"How much does Group 3 cost now?\", \"label\": \"preco_atual\"}

{\"text\": \"How does carbon credit certification work?\", \"label\": \"outros\"}

{\"text\": \"Does the International Business Times's articles indicate uncertainty?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What is the role of inflation in the economy?\", \"label\": \"outros\"}

{\"text\": \"What is the risk of using too many variables in an analysis?\", \"label\": \"conceito\"}

{\"text\": \"What is the current average price of the 30 base oils combined?\", \"label\": \"preco_atual\"}

{\"text\": \"What is technical analysis in the financial market?\", \"label\": \"outros\"}

{\"text\": \"What is a commodity swap contract?\", \"label\": \"outros\"}

{\"text\": \"What is base oil 23 now?\", \"label\": \"preco_atual\"}

{\"text\": \"What is the difference between diversification and asset allocation?\", \"label\": \"outros\"}

{\"text\": \"Did the model detect pessimism in the news about Chevron?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What does multicollinearity between variables mean?\", \"label\": \"conceito\"}

{\"text\": \"What are the limitations of ARIMA?\", \"label\": \"conceito\"}

{\"text\": \"What is the difference between ARIMA and ARIMAX?\", \"label\": \"conceito\"}

{\"text\": \"What is ESG investing?\", \"label\": \"outros\"}

{\"text\": \"How important is natural gas to the energy sector?\", \"label\": \"outros\"}

{\"text\": \"How to choose relevant variables for multivariate analysis?\", \"label\": \"conceito\"}

{\"text\": \"Is Reuters's sentiment more optimistic than Bloomberg's?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"Is the sentiment value for Shell more neutral?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"Quanto est\u00e3o custando os derivados de petr\u00f3leo hoje?\", \"label\": \"preco_atual\"}

{\"text\": \"Was sentiment during the COVID period more pessimistic?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What is divestment in oil assets?\", \"label\": \"outros\"}

{\"text\": \"How important is solar energy in the electricity sector?\", \"label\": \"outros\"}

{\"text\": \"Is it better to train separate models for each group of base oils?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is hedging in the commodities market?\", \"label\": \"outros\"}

{ "text": "Does the model detect negative sentiment in WSJ news? ", "label": "sentimento_noticia" }

{ "text": "How do different price forecasting models compare in performance?", "label": "previsao_preco" }

{ "text": "Does Inc.'s news show more pessimism?", "label": "sentimento_noticia" }

{ "text": "What is the price of base oil 28 today?", "label": "preco_atual" }

{ "text": "How much is base oil 17 today?", "label": "preco_atual" }

{ "text": "What methods can be used to predict base oil prices?", "label": "previsao_preco" }

{ "text": "Is the sentiment in The New York Times's articles more negative?", "label": "sentimento_noticia" }

{ "text": "Does the model indicate market confidence based on the most recent news?", "label": "sentimento_noticia" }

{ "text": "What are the current prices of the 30 base oils?", "label": "preco_atual" }

{ "text": "What is the energy transition?", "label": "outros" }

{ "text": "Can ARIMAX handle exogenous variables?", "label": "conceito" }

{ "text": "Does multivariate analysis help reduce bias?", "label": "conceito" }

{ "text": "Do you know the current price of base oil 4?", "label": "preco_atual" }

{ "text": "What confidence interval is recommended for price forecasts?", "label": "previsao_preco" }

{ "text": "Which metrics are more robust in volatile price forecasts?", "label": "previsao_preco" }

{ "text": "What is systemic risk?", "label": "outros" }

{ "text": "In what contexts does a neural network outperform statistical models in price forecasting?", "label": "previsao_preco" }

{ "text": "What is the current price of base oil 11?", "label": "preco_atual" }

{ "text": "Tell me the current price of Group 3.", "label": "preco_atual" }

{ "text": "What is the difference between assets and liabilities?", "label": "outros" }

{ "text": "Did the model classify the latest news as neutral?", "label": "sentimento_noticia" }

{ "text": "Did news about natural gas have a pessimistic tone?", "label": "sentimento_noticia" }

{ "text": "Did the average sentiment value fall in 2020?", "label": "sentimento_noticia" }

{ "text": "Did recent news on renewable energy have positive values?", "label": "sentimento_noticia" }

{ "text": "What confidence level is recommended for Group 3 price forecasts?", "label": "previsao_preco" }

{ "text": "What is the price of Group 1 base oils today?", "label": "preco_atual" }

{ "text": "What is the difference between capex and opex?", "label": "outros" }

{ "text": "What confidence interval can we expect when forecasting Group 3 base oils?", "label": "previsao_preco" }

{ "text": "Does the average percentage error vary significantly between groups of base oils?", "label": "previsao_preco" }

{ "text": "When should MAE be used instead of MAPE in price forecasts?", "label": "previsao_preco" }

{ "text": "Was news from 2015 to 2017 more neutral than from 2018 to 2020?", "label": "sentimento_noticia" }

{ "text": "What is BERT's purpose in NLP tasks?", "label": "conceito" }

{ "text": "How is the sentiment value distributed between neutral, positive, and negative?", "label": "sentimento_noticia" }

{\"text\": \"What is the balance sheet?\", \"label\": \"outros\"}

{\"text\": \"Which group of base oils typically has the lowest forecast error?\", \"label\": \"previsao_preco\"}

{\"text\": \"Voc\u00ea pode me dizer os pre\u00e7os atuais dos principais combust\u00edveis?\", \"label\": \"preco_atual\"}

{\"text\": \"What is the difference between sentiment analysis and semantic analysis?\", \"label\": \"conceito\"}

{\"text\": \"What does a barrel of oil equivalent mean? What is renewable energy?\", \"label\": \"outros\"}

{\"text\": \"How much is base oil 15 being charged today?\", \"label\": \"preco_atual\"}

{\"text\": \"Are articles about BP optimistic?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"Do CNBC's articles show a neutral or optimistic tone?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"How do mergers and acquisitions affect energy companies?\", \"label\": \"outros\"}

{\"text\": \"What is the difference between MLP and RBF?\", \"label\": \"conceito\"}

{\"text\": \"What is the prevailing sentiment in recent oil news?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What problems can arise in sentiment analysis?\", \"label\": \"conceito\"}

{\"text\": \"What is the difference between common and preferred shares?\", \"label\": \"outros\"}

{\"text\": \"Is the sentiment for news about oil prices negative?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What is the difference between the spot market and the futures market?\", \"label\": \"outros\"}

{\"text\": \"How should the Group 3 price behave in the next six months?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is the current market price of Group 1?\", \"label\": \"preco_atual\"}

{\"text\": \"How do exploration block auctions work in Brazil?\", \"label\": \"outros\"}

{\"text\": \"Which method is best for predicting prices in the short term?\", \"label\": \"previsao_preco\"}

{\"text\": \"Does news since 2015 indicate a trend toward pessimism or optimism?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What are the advantages of MAPE?\", \"label\": \"conceito\"}

{\"text\": \"What is the calculated sentiment value for today's news?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What variables can be used in a multivariate analysis?\", \"label\": \"conceito\"}

{\"text\": \"Was the news about ExxonMobil negative?\", \"label\": \"sentimento_noticia\"}

{\"text\": \"What is considered a commodity?\", \"label\": \"outros\"}

{\"text\": \"Is it possible to obtain price forecasts with 95% confidence?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is an MLP neural network?\", \"label\": \"conceito\"}

{\"text\": \"How can we evaluate the RMSE of forecasts for Group 3 base oils?\", \"label\": \"previsao_preco\"}

{\"text\": \"What is the difference between multivariate and univariate analysis?\", \"label\": \"conceito\"}

{\"text\": \"What does ARIMAX mean?\", \"label\": \"conceito\"}

{\"text\": \"What are the applications of sentiment analysis in the oil sector?\", \"label\": \"conceito\"}

{ "text": "What are the main methods used for price forecasting?", "label": "previsao_preco" }

{ "text": "How do import tariffs influence the energy market?", "label": "outros" }

{ "text": "Can you tell me the current prices of the main fuels?", "label": "preco_atual" }

{ "text": "How much is base oil 5 being charged today?", "label": "preco_atual" }

{ "text": "Which method is best suited for forecasting Group 1 base oil prices?", "label": "previsao_preco" }

{ "text": "Can you tell me the price of Group 1 now?", "label": "preco_atual" }

{ "text": "What is the base oil 7 price trend in the coming weeks?", "label": "previsao_preco" }

{ "text": "What is the impact of economic sanctions on the oil market?", "label": "outros" }

{ "text": "What is the current price of base oil 8?", "label": "preco_atual" }

{ "text": "Is the confidence interval the same as the margin of error?", "label": "conceito" }

{ "text": "What is the price of Base Oil 1 today?", "label": "preco_atual" }

{ "text": "What is the price of base oil 21 today?", "label": "preco_atual" }

{ "text": "What is an investment portfolio?", "label": "outros" }

{ "text": "What is base oil 16 right now?", "label": "preco_atual" }

{ "text": "What does it mean to use deep learning to predict prices?", "label": "previsao_preco" }

{ "text": "Does model performance change when predicting Group 2 prices over different periods?", "label": "previsao_preco" }

{ "text": "What is oil refining?", "label": "outros" }

{ "text": "In what context is MAPE most commonly used?", "label": "conceito" }

{ "text": "How does oil futures trading work?", "label": "outros" }

{ "text": "What is base oil 26 now?", "label": "preco_atual" }

{ "text": "How can I interpret the forecast differences between the three groups?", "label": "previsao_preco" }

{ "text": "Tell me the current price of Group 2.", "label": "preco_atual" }

{ "text": "Does the latest news indicate optimism or pessimism in the energy sector?", "label": "sentimento_noticia" }

{ "text": "What is the difference between BERT and other language models?", "label": "conceito" }

{ "text": "Do you know the current price of Group 3?", "label": "preco_atual" }

{ "text": "Does the model show more confidence on MarketWatch or Yahoo Finance?", "label": "sentimento_noticia" }

{ "text": "What is the role of the Central Bank?", "label": "outros" }

{ "text": "How much is Group 1 being charged today?", "label": "preco_atual" }

{ "text": "Was the news sentiment after the 2022 war positive or negative?", "label": "sentimento_noticia" }

{ "text": "What factors make Group 2 forecasts more uncertain?", "label": "previsao_preco" }

{ "text": "What are examples of traditional price forecasting methods?", "label": "previsao_preco" }

{ "text": "What is fundamental analysis?", "label": "outros" }

{ "text": "What is the Group 1 price forecast for the next month?", "label": "previsao_preco" }

{ "text": "Can linear regression be used on time series?", "label": "conceito" }

{ "text": "What is the difference between MAE and MAPE?", "label": "conceito" }

{ "text": "How does the combination of models in a hybrid work?", "label": "conceito" }

{ "text": "How much does the price vary between the three groups of base oils right now?", "label": "preco_atual" }

{ "text": "How can I compare forecasts made today with those made six months ago?", "label": "previsao_preco" }

{ "text": "Was Fortune's news more neutral than Forbes'?", "label": "sentimento_noticia" }

{ "text": "How does sentiment analysis work in texts?", "label": "conceito" }

{ "text": "How much is base oil 27 right now?", "label": "preco_atual" }

{ "text": "What is the current market price of Group 2 base oils today?", "label": "preco_atual" }

{ "text": "What is the market price for base oil 15 today?", "label": "preco_atual" }

{ "text": "How can we interpret a narrow confidence interval when forecasting the 30 base oils?", "label": "previsao_preco" }

{ "text": "Was the predominant tone in MarketWatch publications neutral?", "label": "sentimento_noticia" }

{ "text": "What does discounted cash flow mean?", "label": "outros" }

{ "text": "What is the difference between simple and multiple linear regression?", "label": "conceito" }

{ "text": "What is the difference between gross revenue and net revenue?", "label": "outros" }

{ "text": "What does market share mean in the oil sector?", "label": "outros" }

{ "text": "How does historical data help understand market patterns?", "label": "conceito" }

{ "text": "How to interpret the coefficients of a linear regression?", "label": "conceito" }

{ "text": "What are examples of multivariate analysis techniques?", "label": "conceito" }

{ "text": "What is the difference between forecasting short-term and long-term prices for Group 1?", "label": "previsao_preco" }

{ "text": "What does MAPE mean?", "label": "conceito" }

{ "text": "Is there a predominance of negative sentiment this semester?", "label": "sentimento_noticia" }

{ "text": "Did recent news have high or low sentiment values?", "label": "sentimento_noticia" }

{ "text": "What is the base oil 2 price forecast for the next 90 days?", "label": "previsao_preco" }

{ "text": "What is the price of base oil 18 now?", "label": "preco_atual" }

{ "text": "Is the tone of the articles published this week positive or negative?", "label": "sentimento_noticia" }

{ "text": "Tell me the price of Group 2 now.", "label": "preco_atual" }

{ "text": "How can I evaluate price forecasts for Group 2 compared to Group 3?", "label": "previsao_preco" }

{ "text": "How does the RMSE vary between forecasts for Groups 1, 2, and 3?", "label": "previsao_preco" }

{ "text": "What is the current market price of CNG?", "label": "preco_atual" }

{ "text": "What is the predominant tone of the news this quarter?", "label": "sentimento_noticia" }

{ "text": "How much is base oil 7 currently?", "label": "preco_atual" }

{ "text": "Has sentiment recovered after 2021?", "label": "sentimento_noticia" }

{ "text": "Does recent news bring more confidence or insecurity?", "label": "sentimento_noticia" }

{ "text": "Was there a change in sentiment between 2020 and 2023?", "label": "sentimento_noticia" }

{ "text": "What is the current price ratio among the 30 base oils?", "label": "preco_atual" }

{ "text": "Is it useful to update price forecasts weekly?", "label": "previsao_preco" }

{ "text": "Is uncertainty greater when predicting Group 2 prices than Group 3?", "label": "previsao_preco" }

{ "text": "What specifications does MAPE have?", "label": "conceito" }

{ "text": "What is the base oil 2 price forecast for the next 90 days?", "label": "previsao_preco" }

{ "text": "Are price forecasts always useful for decision-making?", "label": "previsao_preco" }

{ "text": "Was the calculated sentiment value for CNN news positive?", "label": "sentimento_noticia" }

{ "text": "What is the current reference price of base oil 20? How much are they paying for base oil 10 right now?", "label": "preco_atual" }

{ "text": "What is a futures contract?", "label": "outros" }

{ "text": "What is the relationship between confidence interval and variability?", "label": "conceito" }

{ "text": "How does BERT work conceptually?", "label": "conceito" }

{ "text": "What are the advantages of using MAE?", "label": "conceito" }

{ "text": "What is asset liquidity?", "label": "outros" }

{ "text": "Which method is best suited for predicting Group 1 base oil prices?", "label": "previsao_preco" }

{ "text": "What advantages are there in using hybrid models?", "label": "conceito" }

{ "text": "Which group has the lowest price right now?", "label": "preco_atual" }

{ "text": "What does the real interest rate mean?", "label": "outros" }

{ "text": "How to interpret a 95% confidence interval?", "label": "conceito" }

{ "text": "What is the average price of Group 2 base oils right now?", "label": "preco_atual" }

{ "text": "Can you tell me the current price of Base Oil 9?", "label": "preco_atual" }

{ "text": "In what situations is multivariate analysis used?", "label": "conceito" }

{ "text": "Is the calculated sentiment in Fortune articles positive?", "label": "sentimento_noticia" }

{ "text": "What is the difference between WTI and Brent?", "label": "outros" }

{ "text": "What is the current market price of gasoline?", "label": "preco_atual" }

{ "text": "How can backtesting be used to validate price forecasts for one of the 30 base oils?", "label": "previsao_preco" }

{ "text": "Has the Financial Times's average sentiment been positive this year?", "label": "sentimento_noticia" }

{ "text": "What is the general perception of the news analyzed at this time?", "label": "sentimento_noticia" }

{ "text": "What is a company's valuation?", "label": "outros" }

{ "text": "Was there an increase in negative news during the pandemic?", "label": "sentimento_noticia" }

{ "text": "Was the sentiment for articles about base oils positive?", "label": "sentimento_noticia" }

{ "text": "Does the model classify The Economist's news as positive?", "label": "sentimento_noticia" }

{ "text": "How much is base oil 10 being charged today?", "label": "preco_atual" }

{ "text": "What is the practical use of the confidence interval?", "label": "conceito" }

{ "text": "What is a confidence interval?", "label": "conceito" }

{ "text": "What is the difference between quantitative and qualitative price forecasting methods?", "label": "previsao_preco" }

{ "text": "Does the confidence interval tend to widen in long-term forecasts?", "label": "previsao_preco" }

{ "text": "Does the latest news bring a sense of stability or uncertainty?", "label": "sentimento_noticia" }

{ "text": "How to interpret a narrow confidence interval in price forecasting?", "label": "previsao_preco" }

{ "text": "Are Group 1 forecasts more stable than Group 2 forecasts?", "label": "previsao_preco" }

{ "text": "Why is RMSE used specifically for evaluation?", "label": "conceito" }

{ "text": "In what cases might MAE not be ideal?", "label": "conceito" }

{ "text": "Does the model detect more positive or negative news this month?", "label": "sentimento_noticia" }

{ "text": "What is the current market price of Group 3?", "label": "preco_atual" }

{ "text": "What is the current price of Group 1?", "label": "preco_atual" }

{ "text": "When should you choose ARIMA over another model?", "label": "conceito" }

{ "text": "Do Forbes articles show optimistic sentiment?", "label": "sentimento_noticia" }

{ "text": "Does the confidence interval for Group 2 price forecasts tend to be wider?", "label": "previsao_preco" }

{ "text": "How much is base oil 20 being charged today?", "label": "preco_atual" }

{ "text": "Is it possible to continuously improve forecasts with new data from the 30 base oils?", "label": "previsao_preco" }

{ "text": "Can you tell me the price of base oil 29?", "label": "preco_atual" }

{ "text": "Does the model show optimism in the base oils market?", "label": "sentimento_noticia" }

{ "text": "Can you tell me the current price of base oil 25?", "label": "preco_atual" }

{ "text": "Do Bloomberg articles show optimism or pessimism?", "label": "sentimento_noticia" }

{ "text": "O que é economia de baixo carbono?", "label": "outros" }

{ "text": "How does the carbon market affect companies in the energy sector?", "label": "outros" }

{ "text": "How can price forecasts be compared between the three groups of base oils?", "label": "previsao_preco" }

{ "text": "Did publications about gasoline indicate optimism?", "label": "sentimento_noticia" }

{ "text": "What does credit risk mean?", "label": "outros" }

{ "text": "Besides oil, what are the main types of commodities?", "label": "outros" }

{ "text": "Is CNN's articles more pessimistic than WSJ's?", "label": "sentimento_noticia" }

{ "text": "Is the MAPE higher in Group 3 forecasts than in Group 1?", "label": "previsao_preco" }

{ "text": "What is shale gas?", "label": "outros" }

{ "text": "Was the average sentiment of the sector's news closer to neutral?", "label": "sentimento_noticia" }

{ "text": "What does ARIMA mean?", "label": "conceito" }

{ "text": "How do confidence intervals affect decision-making in price forecasting?", "label": "previsao_preco" }

```
{"text": "What advantages can an RBF have over other models?", "label": "conceito"}  
{"text": "What is margin in trading?", "label": "outros"}  
{"text": "What is the lowest price among the 30 base oils right now?", "label":  
"preco_atual"}  
{"text": "Has the proportion of negative news stories increased since last month?",  
"label": "sentimento_noticia"}  
{"text": "What does the CVM mean?", "label": "outros"}
```