

Introdução

A antecipação do futuro é um grande desafio da ciência e economia. A previsão de séries temporais tradicionalmente usa dados quantitativos históricos, assumindo que o passado explica o amanhã. Mesmo com a utilização de modelos dinâmicos, que permitem a evolução temporal dos parâmetros, a informação empregada ainda depende exclusivamente de dados quantitativos. No entanto, em um mundo hiperconectado, os números raramente contam a história completa.

Este trabalho explora a integração de notícias como "indicador antecedente", dando o contexto semântico ausente nos modelos matemáticos, que são limitados diante de choques externos. Ao decodificar não apenas "quanto" um valor mudou, mas o "porquê", obtêm-se previsões mais resilientes e explicáveis diante da incerteza.

Metodologia

A metodologia proposta descreve o desenvolvimento e a interpretabilidade de um sistema híbrido de previsão para o mercado de commodities. A abordagem integra um modelo de previsão de séries temporais com Large Language Models (LLMs), utilizando notícias de mercado como camada explicativa das projeções quantitativas. O processo está estruturado em dois pilares fundamentais, como mostrado na Figura 1:

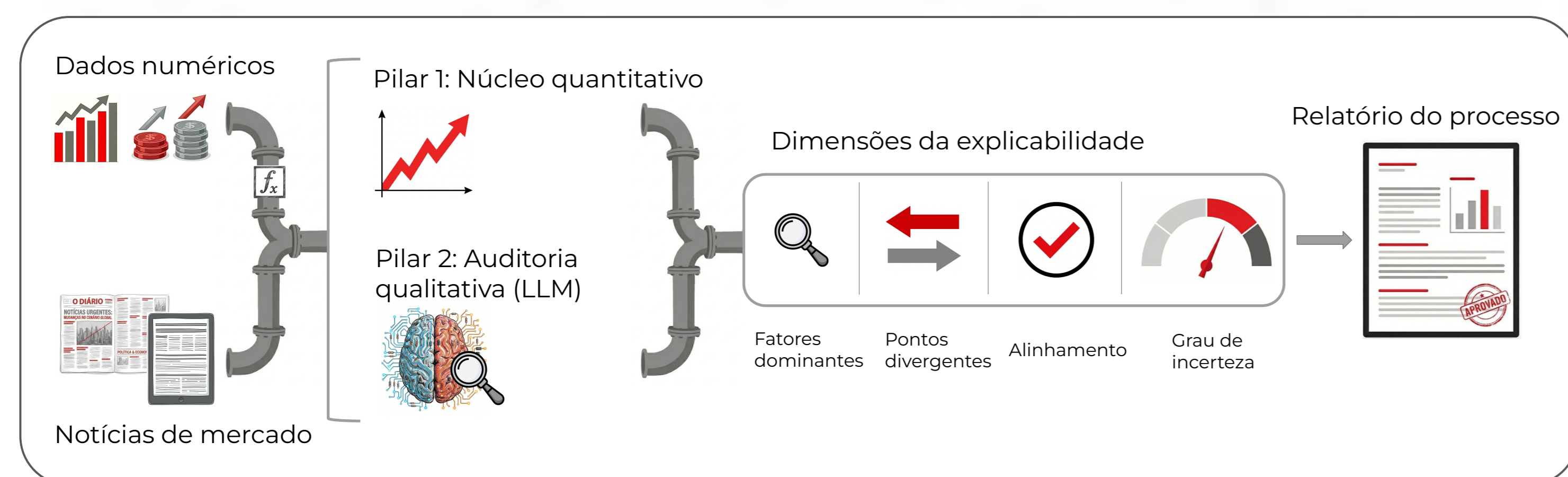


Figura 1: Ilustração do fluxo de números e notícias para previsões reais.

Descrição do fluxo:

- O modelo de série temporal analisa preços históricos e variáveis macroeconômicas para projetar a variação percentual nos próximos 3 meses.
- O LLM atua como uma camada de inteligência processando notícias do setor para validar se os números fazem sentido com a realidade atual.
- A união destas etapas gera um relatório técnico que justifica variações de preços de commodities com base em fatos reais de mercado.
- O sistema funciona como um filtro duplo, em que a matemática prevê e o LLM explica o motivo de dessa previsão.

A Figura 2 apresenta o fluxo no sistema de eixos cartesianos.

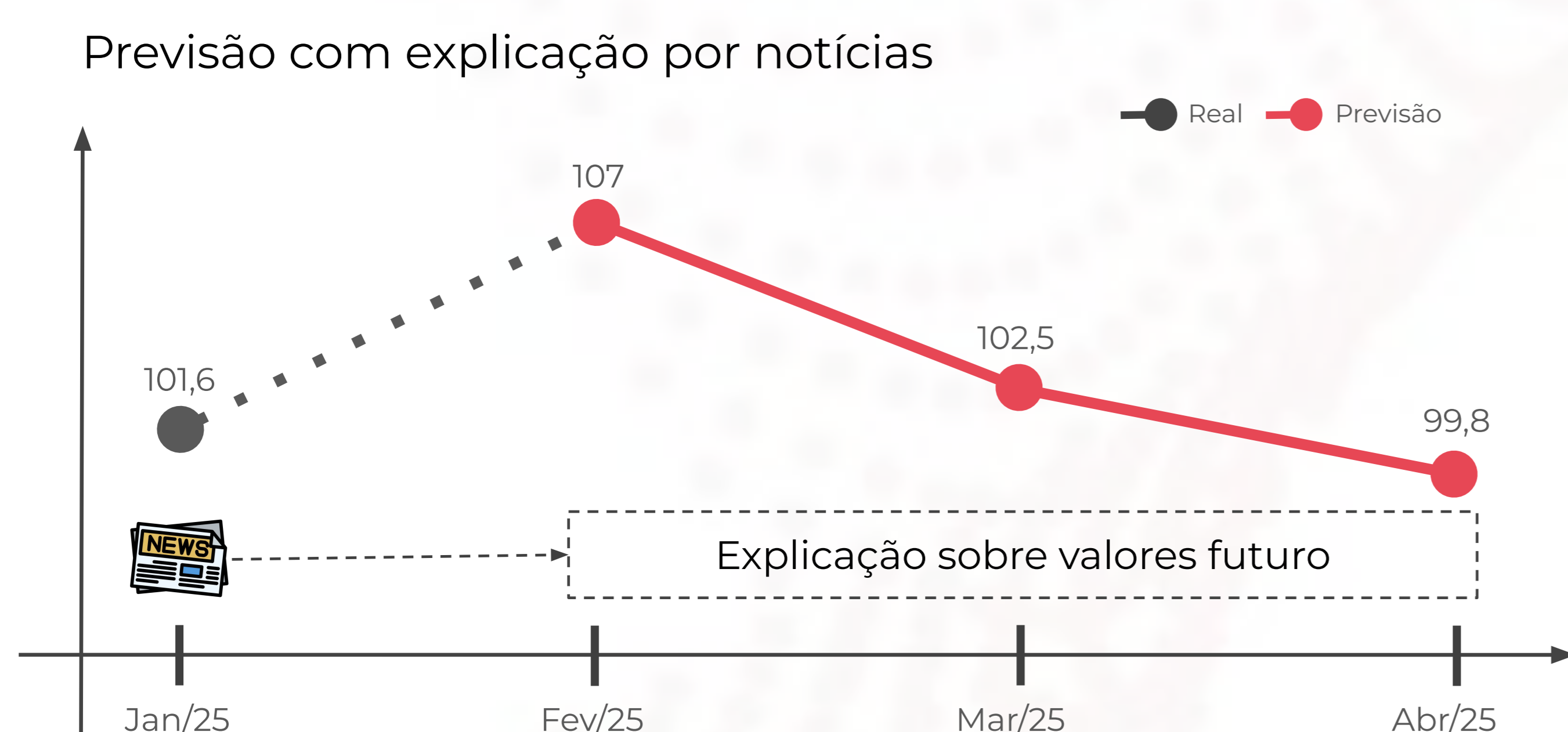


Figura 2: Ilustração do fluxo de número e notícias para previsões reais.

- Previsão dos preços a partir de dados reais.
- Explicação com base em notícias.

Resultados

Os resultados apresentados a seguir referem-se à aplicação do framework híbrido no contexto de previsão de preço do minério de ferro, com horizonte de três meses à frente. O modelo quantitativo de séries temporais gera as projeções numéricas, enquanto o LLM estrutura a interpretação qualitativa com base nas notícias de mercado do período.

As quatro tabelas consolidadas sintetizam as dimensões analíticas propostas: fatores dominantes, pontos de divergência, alinhamento e justificativa da tendência prevista, e grau de certeza. Cada coluna representa um mês projetado, permitindo analisar a evolução da narrativa explicativa ao longo do horizonte de previsão.

Essa estrutura evidencia como a abordagem integra previsão estatística e contextualização informacional, transformando projeções numéricas em análises interpretáveis e tecnicamente justificadas para o mercado de minério de ferro.

Fatores Dominantes

Fevereiro/25	Março/25	Abril/25
As notícias destacam uma demanda fraca por pellets, especialmente na China, com a expectativa de que a atividade comercial retorne após o Ano Novo Lunar. Embora haja especulações sobre um aumento na demanda pós-feriado, a maioria das fontes concorda que a atividade de compra está em um estado de espera.	As notícias sugerem que, embora a demanda possa aumentar após o feriado, a recuperação pode ser lenta e moderada. A competição com o lump ainda é um fator limitante, e os preços dos pellets podem não se estabilizar rapidamente.	As notícias indicam que a demanda por pellets pode continuar fraca, com uma possível estabilização nos preços, mas ainda sob pressão devido à preferência por lump e à oferta excessiva.

Pontos Divergentes

Fevereiro/25	Março/25	Abril/25
Algumas fontes mencionam uma expectativa de recuperação na demanda após o feriado, mas isso é contrabalançado por relatos de baixa atividade de compra e uma preferência contínua por lump em vez de pellets. Portanto, não há um consenso claro sobre a recuperação da demanda.	Enquanto algumas fontes esperam um aumento na atividade comercial, outras continuam a relatar uma demanda fraca e uma preferência por opções mais econômicas, como o lump.	Não há um consenso claro sobre a recuperação da demanda, com algumas fontes prevendo uma leve melhora e outras mantendo uma perspectiva pessimista.

Alinhamento e Justificativa

Fevereiro/25	Março/25	Abril/25
A previsão do modelo de um aumento de 5.39% no primeiro mês parece otimista, dado que as notícias indicam uma demanda fraca e uma pausa nas atividades comerciais. O sentimento cauteloso e a falta de sinais claros de recuperação sustentam a ideia de que o preço pode não subir conforme previsto.	A previsão do modelo de uma alteração de 0.91% parece razoável, dado que a recuperação da demanda pode ser mais lenta do que o esperado, mas não há evidências suficientes para uma queda acentuada.	A previsão do modelo de uma queda de 1.8% parece alinhada com a expectativa de que a demanda não se recuperou rapidamente e que os preços dos pellets continuarão sob pressão.

Grau de Certeza

Fevereiro/25	Março/25	Abril/25
Médio, pois há uma expectativa de recuperação, mas a incerteza sobre a força dessa recuperação e a continuidade da demanda fraca indicam um cenário misto.	Médio, pois as expectativas de recuperação são contraditórias e a demanda continua a ser um fator incerto.	Alto, pois as evidências de uma demanda fraca e a pressão contínua sobre os preços são apoiadas por múltiplas fontes.

Conclusões

Este trabalho apresentou um framework híbrido que integra modelos de previsão de séries temporais com análise textual via Large Language Models, com o objetivo de ampliar a interpretabilidade e a robustez das projeções. Aplicado ao contexto do minério de ferro, o método demonstrou que a incorporação estruturada de notícias como sinal explicativo adiciona contexto causal às previsões numéricas, permitindo avaliar alinhamento narrativo, identificar divergências e qualificar o grau de incerteza associado ao cenário projetado.

Embora validado neste estudo para o mercado de minério de ferro, o framework é generalizável e pode ser aplicado a qualquer série temporal sensível a fatores externos e informacionais. Exemplos incluem preços de outras commodities, ativos financeiros, demanda por energia, indicadores macroeconômicos, vendas no varejo, etc. Em todos esses casos, a integração entre dados quantitativos e contexto textual pode contribuir para previsões mais transparentes, interpretáveis e adaptáveis a mudanças estruturais do ambiente econômico.

Bibliografia

- MIGON, Helio S. et al. A review of Bayesian dynamic forecasting models: Applications in marketing. *Applied Stochastic Models in Business and Industry*, v. 39, n. 3, p. 471-493, 2023.
- BROWN, Tom et al. Language models are few-shot learners. *Advances in neural information processing systems*, v. 33, p. 1877-1901, 2020.