

# Universidade de Pernambuco

## Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

### Proposta de Dissertação de Mestrado

#### Área: Inteligência Computacional

**Título:** Sistemas Computacionais para Identificação e Seleção de Narrativas na Modelagem de Agregados Macroeconômicos

**Orientador:** Bruno José Torres Fernandes ([bjtf@ecomp.poli.br](mailto:bjtf@ecomp.poli.br))

**Coorientador:** André Luiz da Silva Xavier ([alsx@ecomp.poli.br](mailto:alsx@ecomp.poli.br))

**Descrição** – A utilização de informações textuais na previsão de variáveis macroeconômicas como: inflação, desemprego e taxa básica de juros, tem ganhado destaque diante dos avanços recentes em processamento de linguagem natural (PLN) e aprendizado de máquina [1][2]. Narrativas econômicas, compreendidas como conjuntos coerentes de ideias disseminadas por notícias, relatórios e discursos públicos, exercem influência sobre a dinâmica dos principais indicadores macroeconômicos [3][4]. Essa influência decorre do papel da informação na formação de percepções sobre riscos, incertezas e a trajetória futura da política econômica [5].

No contexto brasileiro, a previsão desses agregados é particularmente complexa devido à elevada sensibilidade a ciclos econômicos, choques de oferta, volatilidade externa e mudanças no ambiente institucional. As narrativas capturam nuances qualitativas frequentemente ausentes em séries temporais tradicionais, como percepções sobre a credibilidade da política monetária, pressões inflacionárias setoriais e expectativas relacionadas ao mercado de trabalho [6][7].

Assim, a integração de informações textuais aos modelos preditivos amplia a capacidade de representação de fenômenos dinâmicos e não lineares, especialmente em períodos de instabilidade.

Apesar desse potencial, identificar quais narrativas possuem relevância estatística e econômica constitui um desafio metodológico. A alta dimensionalidade dos textos, a redundância semântica, a variação temporal da atenção da mídia e o ruído informacional dificultam o processo de seleção de componentes realmente significativos [8].

Métodos tradicionais baseados em modelos de tópicos, técnicas de redução de dimensionalidade e medidas clássicas de importância, embora amplamente empregados, dependem de representações simplificadas do texto ou de pressupostos rígidos sobre sua estrutura latente [1][9].

Além disso, podem falhar na captura de dependências contextuais profundas, ambiguidade e relações semânticas complexas.

Nesse cenário, abordagens contemporâneas baseadas em Inteligência Artificial (IA) Generativa surgem como alternativas promissoras para identificar, sintetizar e ranquear narrativas econômicas. Esses modelos são capazes de explorar relações causais implícitas, reorganizar informações dispersas e detectar padrões que ultrapassam as capacidades dos métodos tradicionais de PLN [10].

O projeto propõe uma análise comparativa entre duas abordagens de seleção de narrativas: (i) métodos tradicionais baseados em aprendizado de máquina (AM) e modelos de tópicos; e (ii) métodos apoiados em IA Generativa. O objetivo é avaliar em que medida a IA Generativa aprimora a identificação das narrativas mais relevantes para a previsão da inflação, do desemprego e da taxa Selic, contribuindo para uma compreensão mais aprofundada do papel informacional dos textos na modelagem macroeconômica.

A pesquisa pretende oferecer: (i) avanços metodológicos no desenvolvimento de mecanismos computacionais para seleção de narrativas; (ii) evidências empíricas sobre o impacto de diferentes estratégias de extração textual na previsão de agregados

macroeconômicos; e (iii) uma compreensão mais robusta sobre como as narrativas moldam a dinâmica dos principais indicadores macroeconômicos brasileiros.

### **Referências Bibliográficas**

- [1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- [2] Atanasov, V. (2018). World output gap and global stock returns. *Journal of Empirical Finance*, 48, 181–197.
- [3] Bybee, L., Kelly, B., & Su, Y. (Forthcoming). Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text. *Review of Financial Studies*.
- [4] Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, Local, and Contagious Investor Sentiment. *Journal of Financial Economics*, 104, 272–287.
- [5] Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics*, 131, 1593–1636.
- [6] Campbell, J. Y., & Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies*, 21, 1509–1531.
- [7] Campbell, J. Y., & Vuolteenaho, T. (2004). Bad Beta, Good Beta. *American Economic Review*, 94, 1249–1275.
- [8] Chen, J., Tang, G., Yao, J., & Zhou, G. (2022). Investor Attention and Stock Returns. *Journal of Financial and Quantitative Analysis*, 57, 455–497.
- [9] Dangl, T., & Halling, M. (2012). Predictive Regressions with Time-Varying Coefficients. *Journal of Financial Economics*, 106, 157–181.
- [10] Adämmer, P., & Schüssler, R. A. (2020). Forecasting the Equity Premium: Mind the News! *Review of Finance*, 24, 1313–1355.