

Definição de Elementos Básicos do Projeto de TCC

Curso de Extensão em Data Science – ITA

Prof. Dr. Johnny Marques

Alunos

Bruno Raphaell Alves de Matos

Matheus Henrique de Oliveira Medeiros

1. Questão de Pesquisa

Em que medida variáveis macroeconômicas brasileiras, como taxa Selic, índice de atividade econômica (IBC-Br), câmbio USD/BRL, taxa de desocupação e rendimento médio real, melhoram a previsão da taxa mensal de inadimplência de pessoas físicas no Brasil, quando comparadas a modelos que utilizam apenas o histórico da própria inadimplência?

2. Objetivo Geral

Desenvolver e avaliar um pipeline reprodutível de nowcasting da inadimplência de pessoas físicas no Brasil, integrando dados públicos do Banco Central do Brasil e do IBGE, e quantificar o ganho de desempenho preditivo obtido com a incorporação de variáveis macroeconômicas defasadas em relação a modelos univariados.

3. Objetivos Específicos

Os seguintes objetivos específicos são buscados para o atingimento do objetivo geral:

- Levantar e selecionar séries temporais públicas relevantes nas bases do BCB/SGS e do IBGE/SIDRA;
- Coletar automaticamente as séries via API, armazenando os dados brutos e tratados em formatos reprodutíveis;
- Harmonizar as séries em frequência mensal, agregando séries diárias e alinhando séries divulgadas como trimestre móvel ao mês de referência;
- Construir variáveis explicativas com defasagens e transformações (lags, variações e médias móveis), garantindo prevenção de *data leakage* ao restringir o uso de informações até $t-1$;
- Implementar modelos de referência e modelos com variáveis exógenas (por exemplo, persistência, ARIMA/SARIMA, SARIMAX e um modelo de aprendizado de máquina com validação temporal);
- Validar os modelos por *backtesting walk-forward* e comparar desempenho por métricas como MAE e RMSE;

- Analisar a contribuição das variáveis e suas defasagens, identificando os fatores mais informativos para a previsão.

4. Orientador Proposto

Elton Sbruzzi