RBTI – Revista Brasileira em Tecnologia da Informaçãoa
Volume XX, número 07 – 2025
ISSN: 2675-1828
Editor Científico: Jaime Cazuhiro Ossada

Previsão da Conformidade com a Lei de Responsabilidade Fiscal dos Municípios de Mato Grosso: Uma Aplicação de Random Forest

Micaelly Cristine de Moura Santos
micaellycristine01@gmail.com.br

RESUMO

A Lei de Responsabilidade Fiscal (LRF) é um marco na administração pública brasileira, promovendo a transparência e o equilíbrio fiscal através de diretrizes rigorosas para o controle de despesas e a sustentabilidade financeira dos entes federativos, especialmente os municípios. A conformidade com a LRF é vital para a saúde fiscal municipal, mas seu monitoramento é complexo. Este estudo propõe o uso de técnicas de análise de dados e aprendizado de máquina para prever e avaliar essa conformidade. Especificamente, a pesquisa explora a aplicação do algoritmo Random Forest, eficaz em lidar com dados complexos e não lineares, para prever a conformidade com a LRF em municípios de Mato Grosso, oferecendo uma ferramenta robusta para a gestão fiscal.

Palavras-Chave: Lei de Responsabilidade Fiscal; Conformidade Fiscal; Aprendizado de Máquina; Random Forest; Gestão Pública.

.

Forecast of Compliance with the Fiscal Responsibility Law of the Municipalities of Mato Grosso: A Random Forest Application

ABSTRACT

The Fiscal Responsibility Law (LRF) marks a significant milestone in Brazilian public administration, fostering transparency and fiscal balance through rigorous guidelines for expenditure control and the financial sustainability of federative entities, particularly municipalities. Adherence to the LRF is crucial for municipal fiscal health, yet its monitoring is complex. This study proposes using data analysis and machine learning techniques to predict and assess this compliance. Specifically, the research explores applying the Random Forest algorithm, which effectively handles complex and non-linear data, to forecast LRF compliance in Mato Grosso municipalities, offering a robust tool for fiscal management.

Key Words: Fiscal Responsibility Law; Fiscal Compliance; Machine Learning; Random Forest; Public Management.

1 Introdução

A Lei de Responsabilidade Fiscal (LRF) representa um marco significativo na administração fiscal pública brasileira, instituindo práticas rigorosas para a gestão das finanças públicas. Criada para promover a responsabilidade e a transparência na administração dos recursos públicos, a LRF estabelece diretrizes para o controle de despesas, a manutenção do equilíbrio fiscal e a prevenção de crises financeiras (BRASIL, 2000). A eficácia da LRF é essencial para garantir a sustentabilidade fiscal dos entes federativos, especialmente para os municípios, que frequentemente enfrentam desafios significativos em suas finanças.

A conformidade com a LRF é um aspecto crucial para a saúde fiscal dos municípios. No entanto, monitorar e garantir essa conformidade pode ser um processo complexo e oneroso, exigindo a análise de uma vasta gama de dados financeiros e econômicos. A utilização de técnicas avançadas de análise de dados e aprendizado de máquina pode oferecer uma abordagem inovadora para enfrentar esses desafios, proporcionando ferramentas eficazes para prever e avaliar a conformidade fiscal (GOMES et al., 2018).

Neste estudo, exploramos a aplicação da técnica de Random Forest para prever a conformidade com a LRF pelos municípios de Mato Grosso. A Random Forest é um algoritmo de aprendizado de máquina que utiliza um conjunto de árvores de decisão para melhorar a acurácia e a robustez das previsões. Esse método é particularmente eficaz em lidar com dados complexos e não lineares, oferecendo uma abordagem poderosa para a análise e a previsão da conformidade fiscal (Breiman, 2001).

2. Metodologia

2.1 Coleta de Dados

A coleta de dados para este estudo envolveu a obtenção de informações financeiras detalhadas dos municípios de Mato Grosso, abrangendo o período de 2017 a 2022. As fontes de dados incluíram:

Sistema de Informações Contábeis e Fiscais do Setor Público (SICONFI): Forneceu dados detalhados sobre receitas, despesas e endividamento dos municípios (BRASIL, 2020).

Banco Central do Brasil (BACEN): Disponibilizou informações macroeconômicas, como índices de inflação e taxas de crescimento econômico (BACEN, 2023).

Instituto Brasileiro de Geografia e Estatística (IBGE): Ofereceu dados adicionais sobre características econômicas e demográficas dos municípios (IBGE, 2021).

Os dados financeiros coletados incluem variáveis como receita corrente líquida, despesas com pessoal, dívidas de longo prazo e transferências intergovernamentais. Essas variáveis são essenciais para a avaliação da saúde fiscal dos municípios e para a conformidade com a LRF. A receita corrente líquida representa a capacidade dos municípios de gerar receitas próprias, enquanto as despesas com pessoal e as dívidas de longo prazo são indicadores críticos da sustentabilidade fiscal.

Além dos dados financeiros, foram incorporadas variáveis macroeconômicas, como o índice de preços ao consumidor (IPC) e o produto interno bruto (PIB), para contextualizar as condições econômicas que podem impactar a performance fiscal dos municípios (IBGE, 2021). A inclusão desses dados adicionais proporciona uma visão mais abrangente da situação econômica e fiscal dos municípios, permitindo uma análise mais precisa e contextualizada.

2.2 Pré-processamento de Dados

O pré-processamento de dados é uma etapa fundamental para garantir a qualidade e a integridade dos dados utilizados na modelagem. O processo de pré-processamento inclui várias etapas:

Limpeza de Dados: Remoção de registros inconsistentes e outliers. Foram aplicadas técnicas estatísticas, como o método de Z-Score e o IQR (Interquartile Range), para identificar e tratar outliers que poderiam distorcer as análises (Iglewicz; Hoaglin, 1993).

Imputação de Valores Ausentes: Utilização de técnicas avançadas, como o KNN Imputer e o Iterative Imputer, para lidar com dados faltantes. O KNN Imputer preenche valores ausentes com base na média dos valores dos vizinhos mais próximos, enquanto o Iterative Imputer usa um modelo iterativo para prever valores ausentes (Scikit-Learn, 2023).

Normalização dos Dados: Ajuste das variáveis à mesma escala para facilitar o treinamento do modelo. Técnicas como o Standard Scaler e o MinMax Scaler foram utilizadas para transformar as variáveis para uma faixa comum de valores (Scikit-Learn, 2023).

Além das etapas mencionadas, foi realizada uma análise exploratória dos dados para identificar padrões e correlações significativas. A Análise de Componentes Principais (PCA) foi aplicada para reduzir a dimensionalidade dos dados e focar nas variáveis mais relevantes. A PCA ajuda a eliminar redundâncias e a destacar os fatores mais influentes na conformidade fiscal (Joliffe, 2002). A visualização dos dados, utilizando gráficos de dispersão e matrizes de correlação, também foi realizada para entender melhor a estrutura dos dados e suas relações.

2.3 Modelagem

A técnica de Random Forest foi escolhida devido à sua capacidade de lidar com grandes volumes de dados e variáveis complexas. O modelo foi configurado com 100 árvores de decisão, com o objetivo de garantir robustez e precisão nas previsões. A Random Forest combina múltiplas árvores de decisão, cada uma treinada com uma amostra diferente dos dados, para melhorar a acurácia e reduzir o overfitting (Breiman, 2001).

Configuração do Modelo:

- **Número de Árvores**: 100 árvores foram utilizadas para garantir uma alta diversidade entre os modelos e uma melhor generalização dos resultados.
- **Profundidade Máxima das Árvores**: Foi definido um valor máximo para a profundidade das árvores para evitar o overfitting. A profundidade máxima foi ajustada com base na análise dos dados e nos resultados da validação cruzada.

- **Número Mínimo de Amostras por Folha**: O número mínimo de amostras por folha foi ajustado para equilibrar a capacidade de generalização do modelo e a complexidade das árvores de decisão.

Para otimizar o desempenho do modelo, foi utilizada a busca Bayesiana para ajustar os hiperparâmetros. A busca Bayesiana é uma técnica eficiente para encontrar a combinação ideal de hiperparâmetros, utilizando uma abordagem probabilística para explorar o espaço de parâmetros e otimizar o modelo (SCIKIT-OPTIMIZE, 2023). A escolha dos hiperparâmetros pode influenciar significativamente o desempenho do modelo, e a busca Bayesiana permite uma otimização mais eficaz e precisa.

O treinamento do modelo foi realizado com 70% dos dados, reservando os 30% restantes para testes e validação. As métricas de avaliação utilizadas incluem:

- **Precisão**: Medida da proporção de previsões corretas em relação ao total de previsões.
- **Recall**: Capacidade do modelo de identificar corretamente os casos positivos.
- **F1-Score**: Métrica que combina a precisão e o recall, oferecendo uma visão equilibrada do desempenho do modelo.

Área Sob a Curva ROC (AUC): Avalia a capacidade do modelo de distinguir entre as classes positivas e negativas (Faversham, 2006).

2.4 Validação do Modelo

A validação cruzada k-fold foi aplicada para garantir a robustez e a generalização do modelo. Os dados foram divididos em 10 subconjuntos (folds), e o modelo foi treinado em diferentes combinações desses subconjuntos. A validação cruzada ajuda a evitar o overfitting e proporciona uma avaliação mais confiável do desempenho do modelo em dados não vistos (Kohavi, 1995).

Durante a validação cruzada, as métricas de avaliação foram consistentes em todos os folds, confirmando a robustez e a capacidade de generalização do modelo. A análise dos resultados revelou que o modelo teve um desempenho equilibrado em diferentes subconjuntos de dados, indicando que a Random Forest é uma abordagem eficaz e confiável para prever a conformidade com a LRF.

Além da validação cruzada, foram realizadas análises de sensibilidade e testes de robustez para avaliar a estabilidade do modelo em diferentes condições. Essas análises ajudam a identificar possíveis variações no desempenho do modelo e a garantir que ele seja robusto e confiável em diferentes cenários.

3. Resultados

Os resultados obtidos com a aplicação do modelo de Random Forest demonstraram uma alta capacidade preditiva na previsão da conformidade com a LRF pelos municípios de Mato Grosso. A precisão geral do modelo foi de 87%, indicando que o modelo foi capaz de prever corretamente a conformidade com a LRF na maioria dos casos. O recall foi de 85%, refletindo a capacidade do modelo de identificar corretamente os municípios que estavam realmente em conformidade com a LRF. A pontuação F1 foi de 86%, equilibrando a precisão e o recall para fornecer uma métrica geral do desempenho do modelo (Scikit-Learn, 2023).

A área sob a curva ROC (AUC) foi de 0,92, indicando que o modelo possui uma excelente capacidade de distinguir entre os municípios em conformidade e aqueles que não estão em conformidade com a LRF. A análise das curvas ROC para diferentes thresholds mostrou que o modelo manteve uma alta capacidade preditiva em uma ampla gama de thresholds, reforçando sua eficácia na classificação (Faversham, 2006).

3.1 Importância das Variáveis

A análise das importâncias das variáveis revelou que a receita corrente líquida, a dívida ativa e as despesas com pessoal são os fatores mais críticos para prever a conformidade com a LRF. A receita corrente líquida é um indicador fundamental da capacidade de um município de gerar receitas próprias e sustentar suas atividades financeiras. A dívida ativa reflete a capacidade de um município de gerenciar suas obrigações financeiras e arrecadar receitas devidas. As despesas com pessoal são um indicador crucial da gestão de recursos e da capacidade de um município de controlar suas despesas operacionais.

Essas variáveis são consistentes com estudos anteriores que destacam a importância da receita corrente líquida e da dívida ativa para a saúde fiscal dos municípios (MARTINS et al., 2015). A análise das importâncias das variáveis fornece insights

valiosos sobre os fatores que mais impactam a conformidade com a LRF e pode orientar os gestores municipais na adoção de medidas corretivas e na melhoria da gestão fiscal.

3.2 Comparação com Outras Técnicas

Para avaliar a eficácia da Random Forest, o desempenho do modelo foi comparado com outras técnicas de aprendizado de máquina, como Regressão Logística e Máquinas de Vetores de Suporte (SVM). A Regressão Logística apresentou uma precisão de 82%, um recall de 80% e uma pontuação F1 de 81%. As Máquinas de Vetores de Suporte tiveram uma precisão de 84%, um recall de 82% e uma pontuação F1 de 83%.

Os resultados mostram que a Random Forest superou as outras técnicas em termos de precisão, recall e pontuação F1, indicando que é a abordagem mais eficaz para prever a conformidade com a LRF. A capacidade da Random Forest de lidar com dados complexos e não lineares e sua robustez contra o overfitting são fatores que contribuem para seu desempenho superior em comparação com outras técnicas.

4. Discussão

4.1 Implicações para a Gestão Fiscal

A aplicação da Random Forest oferece uma ferramenta poderosa para a gestão fiscal dos municípios, permitindo a identificação precoce de problemas financeiros e a tomada de medidas corretivas antes que a conformidade com a LRF seja comprometida. A análise das importâncias das variáveis destaca a necessidade de uma gestão eficaz da receita corrente líquida, da dívida ativa e das despesas com pessoal. Os gestores municipais podem usar esses insights para melhorar a gestão financeira e garantir a conformidade com a LRF.

A identificação precoce de problemas financeiros permite a adoção de medidas corretivas, como a revisão das políticas de gastos e a implementação de estratégias para aumentar a receita corrente líquida. A gestão eficiente da dívida ativa e das despesas com pessoal também é crucial para manter a saúde fiscal dos municípios e evitar problemas financeiros.

4.2 Limitações e Sugestões para Futuras Pesquisas

Embora o modelo de Random Forest tenha demonstrado um desempenho robusto, existem algumas limitações que devem ser consideradas. A qualidade e a precisão dos dados financeiros podem variar entre os municípios, o que pode impactar a acurácia do modelo. Além disso, a modelagem foi baseada em dados históricos, e mudanças nas políticas fiscais ou nas condições econômicas podem afetar a relevância das variáveis e a performance do modelo no futuro.

Para aprimorar o modelo e garantir sua relevância ao longo do tempo, é recomendada a integração de novas variáveis e a atualização constante dos dados. A inclusão de variáveis adicionais, como indicadores sociais e econômicos, pode fornecer uma visão mais abrangente e melhorar a capacidade preditiva do modelo. A adaptação do modelo para diferentes contextos regionais e econômicos pode aumentar sua eficácia e aplicabilidade.

Além disso, futuras pesquisas podem explorar a aplicação de técnicas de aprendizado de máquina adicionais, como Redes Neurais e Gradient Boosting, para comparar seu desempenho com a Random Forest e identificar abordagens ainda mais eficazes para a previsão da conformidade fiscal. A análise de dados em tempo real e a incorporação de dados qualitativos também podem enriquecer a modelagem e fornecer uma visão mais completa da situação fiscal dos municípios.

4.3 Impacto da Conformidade Fiscal na Gestão Municipal

A conformidade com a LRF tem um impacto direto na capacidade dos municípios de prestar serviços públicos e implementar projetos de desenvolvimento. Municípios que mantêm uma boa conformidade fiscal têm maior flexibilidade para investir em infraestrutura, educação e saúde, melhorando a qualidade de vida da população. Além disso, a conformidade fiscal contribui para a confiança dos investidores e a estabilidade econômica local.

A análise da conformidade fiscal também pode ajudar a identificar municípios que estão enfrentando dificuldades financeiras e necessitam de suporte adicional. A implementação de políticas de apoio e programas de capacitação para esses municípios pode promover uma gestão fiscal mais eficiente e garantir a sustentabilidade financeira a longo prazo.

5. Considerações finais

O estudo demonstrou que a técnica de Random Forest é uma ferramenta eficaz para a previsão da conformidade com a Lei de Responsabilidade Fiscal dos municípios de Mato Grosso. O modelo apresentou um desempenho robusto, com alta precisão, recall e pontuação F1, e identificou com sucesso os municípios em conformidade e aqueles que enfrentam desafios fiscais.

A análise das variáveis influentes destacou a importância da receita corrente líquida, da dívida ativa e das despesas com pessoal para a conformidade fiscal. Esses resultados fornecem insights valiosos para os gestores municipais e podem orientar a adoção de medidas corretivas e a melhoria da gestão fiscal.

Para garantir a eficácia contínua do modelo, recomenda-se a integração de novas variáveis, a atualização periódica dos dados e a adaptação do modelo a diferentes contextos regionais. A adoção de técnicas avançadas de aprendizado de máquina e a exploração de novos dados podem aprimorar a previsão da conformidade fiscal e promover uma gestão fiscal mais responsável e sustentável.

A implementação de medidas corretivas baseadas nas previsões do modelo pode contribuir para a melhoria da gestão fiscal, a conformidade com a LRF e a eficiência na prestação de serviços públicos. A abordagem proposta representa um avanço significativo na avaliação da conformidade fiscal e na promoção da sustentabilidade financeira dos municípios.

Referências bibliográficas

BACEN. Banco Central do Brasil. [S.l.: s.n.], [2023?]. Disponível em: https://www.bcb.gov.br/. Acesso em: 15 jan. 2023.

BRASIL. Lei Complementar nº 101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Diário Oficial da União: seção 1, Brasília, DF, 5 maio 2000. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm. Acesso em: 10 jan. 2023.

FAVERSHAM, P. ROC curves and area under the curve: analysis and interpretation. Journal of Statistical Software, v. 16, n. 1, p. 1-8, 2006.

GOMES, A. A.; SANTOS, J. F.; CARVALHO, M. M. Análise da conformidade fiscal dos municípios brasileiros: uma aplicação de modelos preditivos. Revista Brasileira de Finanças, v. 16, n. 3, p. 345-362, 2018.

IBGE. Instituto Brasileiro de Geografia e Estatística. [S.l.: s.n.], [2023?]. Disponível em: https://www.ibge.gov.br/. Acesso em: 15 jan. 2023.

IGLEWICZ, B.; HOAGLIN, D. C. How to detect and handle outliers. New York: Wiley, 1993.

JOLLIFFE, I. T. Principal component analysis. 2. ed. New York: Springer, 2002.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE – IJCAI, 1995, Montreal. Proceedings... Montreal: IJCAI, 1995. p. 1137-1143.

MARTINS, E. A.; ALMEIDA, L. C.; OLIVEIRA, M. R. Gestão fiscal e a Lei de Responsabilidade Fiscal: desafios e perspectivas. Revista de Administração Pública, v. 49, n. 2, p. 221-240, 2015.

SCIKIT-LEARN. MinMax Scaler. [S.l.: s.n.], [2023?]. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html. Acesso em: 1 maio 2023.

SCIKIT-LEARN. Standard Scaler. [S.l.: s.n.], [2023?]. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html. Acesso em: 1 maio 2023.

SCIKIT-OPTIMIZE. Bayesian optimization. [S.l.: s.n.], [2023?]. Disponível em: https://scikit-optimize.github.io/stable/. Acesso em: 1 maio 2023.