

Produto Técnico e Tecnológico

Plano de Trabalho para implantação de algoritmo de floresta randômica para identificação de contribuintes do ICMS inadimplentes

Paulo Martins dos Passos

Gustavo Henrique Petean

Objetivo: Diagnóstico do panorama da dívida de ICMS no Rio Grande do Sul; identificação de padrões nos registros de inscrição da dívida ativa e identificação do perfil do inadimplente

Justificativa: A proposta de utilizar estatística descritiva para traçar um panorama da dívida ativa do ICMS e identificar padrões, e ainda um modelo de aprendizado de máquina (floresta randômica) para identificar inadimplentes do ICMS no Rio Grande do Sul é justificada pela relevância do tema em um contexto de restrições fiscais e orçamentárias. O ICMS representa a principal fonte de receita no estado, porém a dívida ativa de ICMS representa perdas de arrecadação e recuperar esses débitos é fundamental para alcançar as metas de arrecadação e evitar prejuízos financeiros. Nesse cenário, o aprendizado de máquina tem se mostrado uma ferramenta eficaz para aprimorar a fiscalização e prevenir sonegação e inadimplência tributária. A utilização da floresta randômica é uma abordagem que pode ser incorporada às rotinas para identificar a inadimplência do ICMS, permitindo uma atuação mais proativa da administração tributária. Ao classificar esses contribuintes, é possível agir preventivamente com a implementação de estratégias específicas para melhorar a conformidade tributária e aumentar a arrecadação. Portanto, a aplicação da floresta randômica visa minimizar as perdas de arrecadação, garantir a sustentabilidade financeira do estado e viabilizar os direitos sociais previstos na Constituição Federal. Adicionalmente, a implementação do aprendizado de máquina pela administração tributária contribui para o desenvolvimento de novas abordagens e soluções na área tributária, beneficiando a sociedade.

Proposta

- Implementação de rotina de confecção de relatório para compreender o estado em que se encontra a dívida ativa – Estatística Descritiva e identificação de padrões e tendências comportamentais na dívida ativa¹
- Implementação de algoritmo de floresta randômica utilizando dados abertos (fatores ambientais e sociológicos) para identificação de contribuintes que possivelmente possam se tornar inadimplentes do ICMS (o algoritmo pode ser otimizado pela autoridade tributária com a utilização de dados além dos públicos/abertos).
- A partir da identificação, estabelecer estratégias de prevenção à inadimplência do contribuinte.

Desenho do algoritmo

1. Extração de registros de inadimplentes do banco de dados da receita estadual e de registros cadastrais dos contribuintes.

¹ Mais detalhes em: https://rpubs.com/pmpassos/dat_icms

2. Cruzamento dos registros extraídos construindo um banco de dados que contenha a ocorrência da inadimplência e as características cadastrais dos contribuintes.
3. Escolha de características/variáveis de interesse de acordo com a literatura já consolidada² e também de acordo com análise estatística das principais características associadas ao perfil de inadimplentes.
4. Utilização de algoritmo de floresta randômica³:
 - a. Divisão da amostra em duas partes: parte para treinamento e parte para teste.
 - b. Avaliação dos resultados (se satisfatório, incorporar às rotinas da administração tributária)
5. A partir das principais características levantadas pelo algoritmo de floresta randômica pensar estratégias para reduzir a inadimplência.

Exemplo de conclusões que podem ser obtidas pela aplicação do algoritmo para registros de inadimplência do ICMS disponíveis para cobrança (segundo a literatura):

- As variáveis de opção pelo simples, região de atuação da delegacia da receita estadual, idade da empresa, capital social, CNAE, quantidade de CNAES secundários e quantidade de sócios obtiveram maior importância no modelo de floresta randômica.
- As informações indicam uma facilidade do modelo em identificar os adimplentes, mas dificuldade em identificar inadimplentes. Salienta-se, portanto, necessidade de utilização de outras variáveis.
- Mas as variáveis que obtiveram alta importância podem sugerir ações para diminuir o número de inadimplentes, tais como:
- A opção pelo programa Simples Nacional, que tem facilidade no modelo de pagamento de impostos, tem baixa predominância de inadimplentes.
- A região de atuação implica que em certas regiões as delegacias devem ter maior atuação para evitar um desequilíbrio no número de inadimplentes entre as regiões do estado.
- A idade da empresa pode ser utilizada como indicador para contato periódico da receita estadual com as empresas.
- O tipo de atividade (CNAE) também é um indicador de setor de atuação que pode ser utilizado para focar ações junto a associações ou organizações de setores econômicos específicos.

² Para maiores informações, consultar: Passos, P. M. Diagnóstico e impacto da Inadimplência da dívida ativa do ICMS no estado do Rio Grande do Sul. 2023. 150 p. Dissertação (Mestrado em Administração Pública – PROFIAP). Universidade Federal de Goiás – PROFIAP, Aparecida de Goiânia.