

MBA
USP
ESALQ

Aplicação de rede neural artificial para auxiliar a fiscalização tributária na identificação de empresas fantasmas

Ricardo Costa Pinto

Orientadora: Patrícia Belfiore Fávero

INTRODUÇÃO

- A Administração Tributária é responsável por acompanhar, controlar e fiscalizar o cumprimento das normas que tratam das receitas tributárias, e o principal desafio diz respeito ao combate da sonegação fiscal.
- Instituição do Sistema Público de Escrituração Digital – SPED, a forma tradicional de sonegação, venda sem documento fiscal, foi dificultada em razão da obrigatoriedade de emissão de nota fiscal eletrônica em toda cadeia produtiva (indústria, atacado, varejo e serviço).
- A fiscalização tributária tem observado nos últimos anos um aumento significativo de empresas que estão sendo constituídas somente no papel (empresas fantasmas).



INTRODUÇÃO

- Necessidade de aprimoramento das técnicas de fiscalização tributária, utilizando ferramentas e tecnologias capazes de identificar padrões de empresas que são constituídas para praticar o crime de sonegação.
- O presente estudo possui como objetivo aplicar predição baseada em redes neurais artificiais capaz de identificar padrões típicos de contribuintes desaparecidos.
- O modelo classifica as empresas, a partir do conjunto de 25 (vinte e cinco) variáveis explicativas, como suspensas (desaparecidas) ou ativas.
- Dessa forma é possível determinar com mais assertividade as pessoas jurídicas a serem fiscalizadas in loco, para averiguação/confirmação da fraude praticada.



MATERIAL E METÓDOS

- Pesquisa de natureza aplicada, de cunho exploratório, que objetiva auxiliar a fiscalização tributária na detecção de fraudes no processo de constituição de empresas utilizando redes neurais artificiais.
- Selecionados dados de 27.709 empresas do cadastro de contribuintes do Estado de Goiás, pessoas jurídicas de diversos ramos de atividades econômicas localizadas na região do entorno sul do Distrito Federal.
- O trabalho foi realizado em consonância com a Lei Geral de Proteção de Dados Pessoais (LGPD) e à Instrução Normativa nº 1.455/2020-GSE da Secretaria de Economia do Estado de Goiás.



MATERIAL E METÓDOS

- Etapas do processo metodológico:



01. Análise do problema: estamos diante de um problema bastante complexo, identificar se uma empresa de fato existe ou não, baseado nas informações cadastrais e econômicas relacionadas a ela.

02. Definição de variáveis explicativas: após uma análise pormenorizada do problema, foram selecionadas 25 (vinte e cinco) variáveis que demonstram características relevantes de cada contribuinte.



MATERIAL E METÓDOS

03. Coleta de Dados: criação de 06 (seis) consultas específicas para extração dos dados do SAP BusinessObjects - software de Business Intelligence que contém os dados de diversos sistemas corporativos da Secretaria de Estado da Economia de Goiás.

04. Tratamento e Pré-processamento dos Dados: utilização das técnicas de Data Wrangling para estruturação, padronização, organização, limpeza, normalização e validação dos dados.

05. Construção do Modelo de Rede Neural: inicialmente foi definida uma rede neural com uma única camada oculta e um único neurônio, posteriormente foi aumentando o número de neurônios e/ou camadas, a depender dos resultados obtidos. As variáveis de fatores foram convertidas em variáveis dummy para otimizar a execução do modelo.



RESULTADOS E DISCUSSÃO

- As análises foram realizadas a partir de métricas estatísticas para adoção do modelo que mais se adequa à capacidade de identificar padrões de empresas fantasmas, que são constituídas para praticar o crime de sonegação.
- Avaliação de 20 (vinte) topologias de redes neurais artificiais, que sofreram variações somente em relação à quantidade de neurônios e camadas escondidas.
- Os indicadores estatísticos de cada topologia foram registrados ao término da fase de teste das redes neurais.



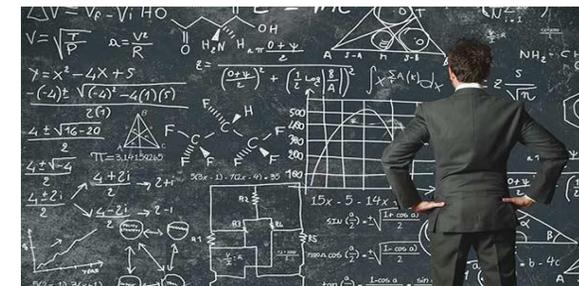
RESULTADOS E DISCUSSÃO

- Indicadores acurácia (“accuracy”), sensibilidade (“sensitivity”), especificidade (“specificity”), kappa, MSE e RMSE.

Tabela 1. Resultado das métricas para as configurações avaliadas

Id	Configuração	Acurácia	Sensibilidade	Especificidade	Kappa	MSE	RMSE
1	hidden=1	0,838	0,927	0,773	0,679	0,118	0,343
2	hidden=2	0,835	0,910	0,772	0,778	0,113	0,336
3	hidden=3	0,850	0,853	0,848	0,695	0,108	0,329
4	hidden=5	0,857	0,862	0,854	0,710	0,104	0,322
5	hidden=c(1,1)	0,840	0,920	0,781	0,681	0,117	0,342
6	hidden=c(2,1)	0,844	0,923	0,786	0,689	0,115	0,339
7	hidden=c(2,2,1)	0,841	0,902	0,682	0,678	0,119	0,345
8	hidden=c(3,1)	0,859	0,907	0,824	0,717	0,103	0,322
9	hidden=c(4,1)	0,844	0,888	0,811	0,686	0,108	0,328
10	hidden=c(4,1,1)	0,841	0,875	0,815	0,680	0,120	0,346
11	hidden=c(4,1,1,1)	0,841	0,905	0,793	0,683	0,125	0,353
12	hidden=c(5,1)	0,862	0,885	0,846	0,721	0,101	0,318
13	hidden=c(5,2)	0,863	0,901	0,835	0,725	0,103	0,321
14	hidden=c(6,1)	0,857	0,864	0,853	0,710	0,105	0,325
15	hidden=c(6,2)	0,854	0,865	0,846	0,704	0,110	0,331
16	hidden=c(7,1)	0,859	0,877	0,845	0,714	0,107	0,327
17	hidden=c(8,1)	0,857	0,867	0,849	0,709	0,110	0,331
18	hidden=c(9,1)	0,861	0,871	0,853	0,717	0,104	0,322
19	hidden=c(10,1)	0,851	0,860	0,845	0,699	0,111	0,333
20	hidden=c(15,1)	0,849	0,857	0,843	0,694	0,121	0,347
	Mínimo	0,835	0,853	0,682	0,678	0,101	0,318
	Máximo	0,863	0,923	0,854	0,778	0,125	0,353
	Mediana	0,851	0,881	0,839	0,701	0,110	0,331
	Variância	6,4E-05	4,7E-04	1,7E-03	2,3E-04	4,9E-05	1,1E-04
	Desvio Padrão	8,9E-03	2,3E-02	4,2E-02	2,3E-02	6,8E-03	1,0E-02

Fonte: Resultados originais da pesquisa



RESULTADOS E DISCUSSÃO

- Aplicando análise pela estatística Kappa podemos classificar o desempenho do modelo como ***muito bom***, desta forma, as amostras coletadas foram concordantes com as informações classificadas.
- O modelo se mostrou bastante eficiente em virtude da complexidade do problema, tendo a acurácia variando de 83,5% a 86,3%.
- O aumento de neurônios e camadas não geraram ganhos de previsão, tendo como consequência direta o aumento do custo computacional de processamento da rede neural na fase de aprendizado.

Tabela 2. Agrupamento qualitativo do coeficiente kappa

Índice Kappa	Desempenho
<0	Péssimo
$0 < k \leq 0,2$	Ruim
$0,2 < k \leq 0,4$	Razoável
$0,4 < k \leq 0,6$	Bom
$0,6 < k \leq 0,8$	Muito Bom
$0,8 < k \leq 1,0$	Excelente

Fonte: Fonseca (2000)



RESULTADOS E DISCUSSÃO

- É perceptível a contribuição para o desenvolvimento de trabalhos futuros com a aplicação do mesmo método, bem como a necessidade de se trabalhar este tipo de modelo com um número maior de variáveis explicativas relevantes, tais como:
 - ✓ dados pessoais dos sócios exclusivo da Receita Federal do Brasil (rendimentos, bens móveis e imóveis, entre outros).
 - ✓ dados das notas fiscais de venda – não foram utilizados em virtude do custo computacional e restrições tecnológicas.



CONSIDERAÇÕES FINAIS

- A rede neural artificial implementada neste trabalho demonstrou que essa tecnologia auxilia a fiscalização tributária na identificação de empresas que não existem de fato, dessa forma coibindo a sonegação fiscal.
- É notória a carência da aplicação de ferramentas desse tipo nas administrações tributárias, sendo necessário aprimorar a eficiência da ação fiscal através de técnicas de inteligência artificial e computacional.
- O ganho deste trabalho é mais vasto que demonstrado, desse modo é factível ampliar o estudo científico, baseado em técnicas de Inteligência Artificial, voltado para a fiscalização tributária.



FIM

"Existe apenas um canto do universo que você pode ter certeza de aperfeiçoar, que é você mesmo." (Aldous Huxley)

