

MINERAÇÃO DE DADOS APLICADA À BASE DE DADOS TRIBUTÁRIA MUNICIPAL

ALCIONE DIAS DA SILVA - diasalcione@gmail.com
COMPANHIA DE TECNOLOGIA DA INFORMAÇÃO DE CACHOEIRO DE
ITAPEMIRIM - DATACI

BRUNO MISSI XAVIER - bmissix@gmail.com
UNIVERSIDADE CANDIDO MENDES - UCAM

GEÓRGIA REGINA RODRIGUES GOMES - georgia@ucam-campos.br
UNIVERSIDADE CANDIDO MENDES - UCAM

MARCELO TADEU MONTEIRO FREITAS - marcelo.tadeu@dataci.es.gov.br
DATA CI – COMPANHIA DE TECNOLOGIA DA INFORMAÇÃO DE CACHOEIRO
DE ITAPEMIRIM

Resumo: *NO SETOR GOVERNAMENTAL, A EFICÁCIA NO RECOLHIMENTO DE IMPOSTOS DEVE SER COMBINADA COM A BOA UTILIZAÇÃO DOS RECURSOS. O OBJETIVO DESTES TRABALHOS É APLICAR A MINERAÇÃO DE DADOS NA BASE DE TRIBUTOS DA CIDADE DE CACHOEIRO DE ITAPEMIRIM-ES, ESPECIFICAMENTE PARA O TRIBUTO IPTU (IMPOSTO PREDIAL TERRITORIAL URBANO), A FIM DE EXTRAIR CONHECIMENTO QUE POSSIBILITE O MELHOR ENTENDIMENTO DO QUE FOI RECOLHIDO AOS COFRES PÚBLICOS. PARA TAL FOI UTILIZADO O PROCESSO DE KDD (KNOWLEDGE DISCOVERY IN DATABASE) COMO METODOLOGIA PARA VIABILIZAR A MINERAÇÃO DE DADOS ATRAVÉS DO SOFTWARE WEKA (WAIKATO ENVIRONMENT FOR KNOWLEDGE ANALYSIS). COM O MAPEAMENTO OBTIDO ATRAVÉS DA APLICAÇÃO DOS ALGORITMOS APRIORI E J48, FORAM REVELADAS INFORMAÇÕES RELEVANTES COMO AS ÁREAS DE MAIOR INADIMPLÊNCIA E A DISTRIBUIÇÃO SÓCIO ECONÔMICA DOS IMÓVEIS LOCALIZADOS NA ZONA URBANA. AS BASES UTILIZADAS NESTE ARTIGO PODERÃO SER APLICADAS AO ESTUDO DE OUTROS TRIBUTOS BEM COMO SER IMPLEMENTADA NO CONTEXTO DE OUTROS MUNICÍPIOS BRASILEIROS.*

Palavras-chaves: *KDD; IPTU; MINERAÇÃO DE DADOS.*

Área: 8 - GESTÃO DO CONHECIMENTO ORGANIZACIONAL
Sub-Área: 8.2 - GESTÃO DA TECNOLOGIA

DATA MINING APPLIED TO MUNICIPAL DATABASE TAXES

Abstract: *IN THE GOVERNMENT SECTOR, THE EFFICIENCY IN TAX COLLECTION SHOULD BE COMBINED WITH GOOD USE OF RESOURCES. THE OBJECTIVE OF THIS WORK IS TO APPLY DATA MINING ON THE DATABASE OF TAXES IN THE CITY OF CACHOEIRO DE ITAPEMIRIMES, SPECIFICALLY TOO THE TAX IPTU (IMPOSTO PREDIAL TERRITORIAL URBANO) TO EXTRACT KNOWLEDGE THAT WILL ENABLE A BETTER UNDERSTANDING OF WHAT WAS COLLECTED TO PUBLIC COFFERS. FOR THIS WE USED THE KDD PROCESS (KNOWLEDGE DISCOVERY IN DATABASE) AS A METHODOLOGY TO ENABLE THE DATA MINING THROUGH THE SOFTWARE WEKA (WAIKATO ENVIRONMENT FOR KNOWLEDGE ANALYSIS). WITH THE MAPPING OBTAINED BY APPLYING THE ALGORITHMS APRIORI AND J48, WERE REVEALED RELEVANT INFORMATION SUCH AS AREAS OF HIGHER NON-PAYMENT AND SOCIO-ECONOMIC DISTRIBUTION OF PROPERTIES LOCATED IN THE URBAN AREA. THE BASES USED IN THIS ARTICLE MAY BE APPLIED TO THE STUDY OF OTHER TAXES AND BE IMPLEMENTED IN THE CONTEXT OF OTHER BRAZILIAN CITIES.*

Keyword: *KDD; MUNICIPAL TAXES; DATA MINING.*

1. Introdução

O setor público gerencia recursos que precisam ser utilizados de forma única, por se tratar de um patrimônio da sociedade, deve-se utilizar todo o conhecimento tecnológico para melhorar a qualidade dos serviços prestados e favorecer a construção de bens de uso comum.

A história mostra que desde o período cassita, a propriedade territorial era devidamente demarcada através das pedras designadas de Kudurru, que segundo Trindade (2007), é uma espécie de marco demarcador de fronteira composto com figuras e inscrições. De acordo com Maggi e Florissi (2002), no Brasil, o imposto IPTU (Imposto Predial Territorial Urbano) foi criado pelo Alvará de 27 de junho de 1808. A partir de agosto de 1830 passou a ser arrecadado sob a denominação de décima urbana. A Constituição de 1891 atribuiu aos estados a competência para decretá-lo (CONSTITUIÇÃO, 1891), a partir de 1934, a Constituição transferiu essa competência para os municípios (CONSTITUIÇÃO, 1934).

Para Carvalho Junior (2009), a Constituição de 1988, além de confirmar o papel do IPTU como imposto de grande potencial fiscal dos municípios, conforme as constituições anteriores passam a inovar, se preocupando com a progressividade e com os objetivos extrafiscais. Varsano (1977) comenta que, do ponto de vista de geração de receita, este imposto é talvez o mais adequado para utilização por governos municipais, pois gera receitas elevadas e estáveis, possivelmente capazes de acompanhar o crescimento das necessidades de recursos desses governos, também possibilita alto grau de autonomia financeira e não é objeto de conflitos tributários com outros municípios.

No âmbito dos governos municipais, enormes quantidades de dados têm sido sistematicamente coletados e armazenados eletronicamente. Segundo Bothorel, Serrurier e Hurter (2011), a Mineração de Dados visa extrair o máximo de conhecimento a partir de bancos de dados volumosos. Neste contexto, nota-se a necessidade de explorar a base de dados relativa ao lançamento de tributos municipais, aplicando sobre esta o processo de descoberta de conhecimento – KDD (*Knowledge Discovery in Databases*), a fim de extrair conhecimentos que possam estar “ocultos”, explicitando-os com objetivo de oferecer informações analíticas aos governos municipais para aprimorar a aplicação desta importante fonte de receita. Vale ressaltar que apesar da Mineração de Dados ser amplamente utilizada em diversas áreas de interesse, quando se trata de trabalhos relativos ao setor público, é percebido uma grande carência.

Alguns trabalhos merecem destaque no setor público, Brown, Cooper e Pidd (2005) utilizaram as técnicas de Mineração de Dados para modernizar a sistemática de funcionamento do imposto de renda de competência da Receita no Reino Unido. No que diz respeito a outro importante tributo municipal, o ISS (Imposto Sobre Serviço), Nascimento *et al.* (2009) apresentam o GIF, que é um sistema para Gestão e Inteligência Fiscal, identificando possíveis evasões fiscais por meio de técnicas de Mineração de Dados aplicadas sobre a base de dados dos serviços prestados e tomados pelos contribuintes do ISS.

Por analogia, busca-se no setor privado as relações entre pagamentos e clientes, fazendo alusão a relação contribuintes e recolhimento de impostos. Assim Vicente, Gomes e Matias (2010), utilizaram tais técnicas com objetivo de identificar o comportamento do mercado de ações brasileiro, avaliando os padrões e os relacionamentos existentes entre os mesmos. Guo e Li (2008), também tratam destas técnicas para detecção de fraudes em cartões de crédito, estes utilizam as redes neurais como algoritmo para predição das fraudes. Yeh e

Lien (2009) desenvolveram um comparativo entre seis técnicas de Mineração de Dados a fim de determinar um padrão de pagamento de clientes de cartão de crédito em Taiwan. Com maior amplitude, Ngai et al. (2011) tratam da aplicação das técnicas de Mineração de Dados para a detecção de fraudes financeiras (FFD) e realizam uma revisão da literatura sobre o assunto e das técnicas de *Mineração* que foram aplicadas. Galvão e Marin (2009) também publicaram um estudo muito útil para pesquisadores, destinado a revisão da literatura sobre as técnicas de Mineração de Dados.

Desta forma o propósito deste estudo é apresentar a técnica de Mineração de Dados para extração de conhecimento sobre a base de dados dos tributos municipais, mais especificamente para o imposto IPTU. Este trabalho identifica a metodologia e os algoritmos que melhor se adequam as características deste estudo, com objetivo de que a base de conhecimento obtida seja útil para as atividades voltadas aos anseios dos governos municipais brasileiros.

O presente artigo está estruturado em quatro seções. A Seção 1 contextualiza o imposto IPTU, fundamenta o processo de KDD e trata de uma breve revisão da literatura. Na Seção 2, é apresentada a metodologia, descrevendo as etapas do KDD e os algoritmos utilizados, na Seção 3 são demonstrados os resultados provenientes da Mineração de Dados aplicada, a Seção 4 discorre sobre as conclusões oriundas dos diversos resultados.

2. Metodologia

A necessidade de tratamento e exploração dos dados, relativo ao negócio das organizações, fez com que surgissem tecnologias para o tratamento adequado desses dados, como o KDD, OLAP (*On-Line Analytical Processing*), BI (*Business Intelligence*) e Mineração de Dados, todas com o objetivo de dar suporte à tomada de decisão.

2.1 O processo de KDD em bases de dados

Para o tratamento dos dados, objeto deste trabalho, será utilizado o conceito de descoberta de conhecimento em bases de dados, KDD. Tal processo foi definido por Fayyad, Piatetsky-shapiro e Smyth (1996), como sendo: "... o processo não trivial de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis, embutidos nos dados".

A Extração de Conhecimento é uma área multidisciplinar que incorpora técnicas utilizadas em diversas áreas como banco de dados, inteligência artificial e estatística. Desse modo, as técnicas utilizadas em KDD não devem ser vistas como substitutas de outras formas de análise, como o OLAP, mas sim, como práticas para melhorar os resultados das explorações feitas com as ferramentas atualmente usadas, Zambon e Meirelles (2001). A Mineração de Dados apresenta-se como parte do processo de KDD.

Os dados oriundos deste estudo foram extraídos da base de arrecadação tributária da Prefeitura de Cachoeiro de Itapemirim-ES. As fases de seleção e processamento contidas no KDD foram realizadas através da execução de comandos SQL (*Structured Query Language*) executadas em uma base de dados gerenciada pelo SGBD (Sistema de Gerenciamento de Banco de Dados) Oracle 10G e foram submetidas ao software de Mineração de Dados WEKA. A etapa de seleção foi responsável por identificar os dados a serem trabalhados, em seguida foram realizadas as etapas de limpeza e transformação onde foram filtrados os dados com valores inválidos ou com outros tipos de erros, sendo atribuída uma única forma de codificação para manter uma base sólida a ser transferida para o software WEKA na versão 3.7.5, que segundo Silva *et al.* (2007), começou a ser escrito em 1993, usando a linguagem Java, na Universidade de Wakato, Nova Zelândia sendo adquirido posteriormente por uma

empresa no final de 2006. O WEKA encontra-se licenciado ao abrigo da *General Public License* (GPL), portanto, sendo possível o seu estudo e alteração do respectivo código fonte.

O WEKA tem como objetivo agregar algoritmos provenientes de diferentes abordagens/paradigmas na subárea da inteligência artificial dedicada ao estudo da aprendizagem por parte de máquinas. Essa subárea pretende desenvolver algoritmos e técnicas que permitam a um computador "aprender" (no sentido de obter novo conhecimento) quer indutiva quer dedutivamente, Almeida (2010). Procede à análise computacional e estatística dos dados fornecidos, recorrendo as técnicas de Mineração de Dados, tentando indutivamente, a partir dos padrões encontrados, gerar hipóteses para soluções e nos extremos inclusive teorias sobre os dados em questão, Silva *et al.* (2007).

2.2 Pré-processamento

2.2.1 Etapa de seleção

Com base na compreensão do domínio da aplicação, obtida através de trabalhos realizados e de entrevistas com especialistas, foram explorados os dados contidos nas tabelas representadas pelo MER (Modelo Entidade Relacionamento) apresentado através da Figura 1. Então foi possível tomar a decisão sobre quais informações poderiam contribuir para o entendimento relativo aos dados do IPTU.

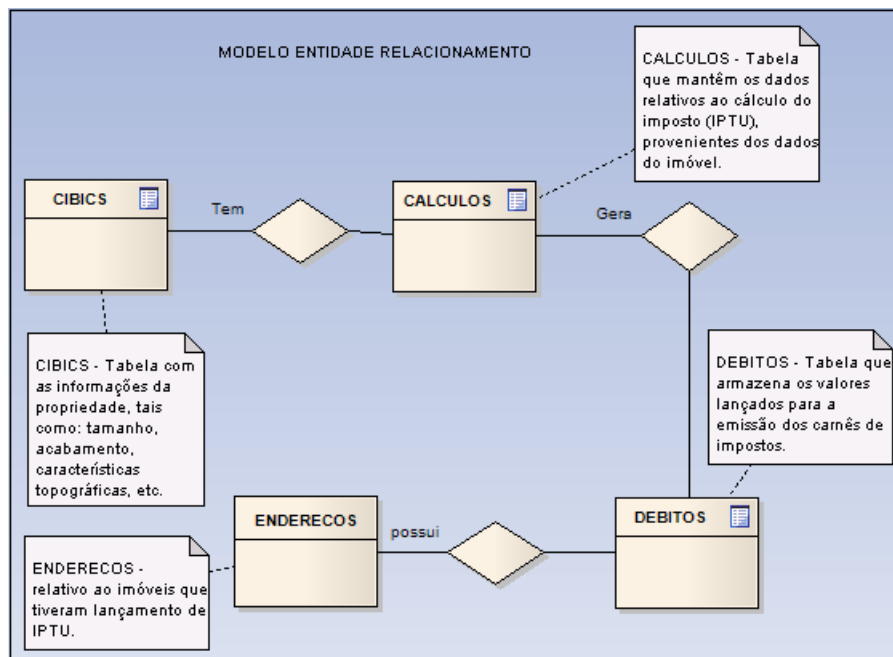


Figura 1 – Modelo Entidade Relacionamento das tabelas utilizadas na etapa de *seleção*.

O modelo entidade relacionamento, descreve as quatro tabelas de banco de dados, que foram alvo da etapa de *seleção*, onde a tabela CIBICS, armazena as características relativas aos imóveis, sendo extraída desta informações como o padrão construtivo e a natureza do lançamento. A tabela CALCULOS armazena os dados relativos a aplicação dos parâmetros de valor e alíquota para obtenção do imposto a ser cobrado. A tabela DEBITOS consolida os dados de todos os lançamentos dos diversos tributos municipais e a situação de pagamento dos mesmos pelos contribuintes. Por fim a tabela de ENDERECOS guarda a localização dos imóveis por zonas, bairros e logradouros do município.

Com base no modelo entidade relacionamento, foram obtidos três conjuntos de dados (*itemsets*) a serem utilizados na Mineração de Dados, o primeiro conjunto deu origem ao

arquivo MaioresLançamentos.arff¹ e representa uma seleção dos lançamentos de IPTU com valores maiores que 400 UPFs (Unidade Padrão Fiscal) ou seja, R\$ 9.252,00, o segundo conjunto deu origem ao arquivo Arrecadação.arff e abrange os 56374 lançamentos do imposto, o terceiro arquivo gerado, devedores.arff, representa os lançamentos do IPTU que não foram quitados. Todos os conjuntos gerados são referentes ao exercício de 2011.

2.2.2 Etapa de limpeza e integração

Nesta etapa foi realizada a limpeza de dados, remoção de ruídos, escolha de estratégias para manipular campos de dados ausentes e a formatação de dados de maneira a adequá-los à ferramenta de mineração. Informações de outros tributos municipais como o ISS foram retiradas por não fazer parte deste estudo.

Alguns atributos precisaram ser manipulados para suprir a ausência de conteúdo, como por exemplo: o atributo “AreaEdificacao”, precisou receber o conteúdo 'NAO EDIFICADO' para os lançamentos de natureza essencialmente territorial, pelo mesmo motivo o atributo “padrão” teve seu conteúdo modificado para 'TERRENO'. Os atributos “Exercicio” e “Zona”, foram alterados em seu tipo de dados, passando de numérico para alfanumérico, dada a restrição de campos numéricos para processamento do algoritmo APRIORI. Para um melhor entendimento, será apresentado na Tabela 1 o conjunto de atributos utilizados e a descrição de seu conteúdo:

Tabela 1 – Atributos constantes dos conjuntos de dados minerados.

Atributo	Descrição
Bairro	Nome do bairro relativo ao imóvel
Zona	Área geográfica relativa a um Conjunto de bairros
TipoCalculo	Referente às variações do cálculo para as naturezas predial e territorial.
Natureza	Predial para o imóvel com construção e territorial para os imóveis sem área edificada.
Exercicio	Ano base de lançamento do imposto
ParcelasDevidas	Quantidade de parcelas não pagas
ParcelasPagas	Quantidade de parcelas pagas
TotalDebito	Valor total do lançamento do imposto
Situacao	Situação relativa ao pagamento
Padrao	Padrão que designa a qualidade da construção para os impostos de natureza predial
AreaTerreno	Faixa de tamanho do terreno
AreaEdificacao	Faixa de tamanho da edificação

2.2.3 Etapa de transformação

Objetivando trazer maior granularidade ao conjunto proposto, o enriquecimento semântico das informações e a redução do número de variáveis e/ou instâncias a serem consideradas, algumas ações foram implementadas, como a criação de faixas para representar o tamanho dos terrenos (AreaTerreno) e o tamanho das edificações (AreaEdificacao), então foram transformados tais dados, conforme a Tabela 2 e a Tabela 3:

¹ As extensões de arquivo do tipo “arff” (Attribute-Relation File Format) são pertinentes aos arquivos gerados para uso do software de Mineração de Dados WEKA.

Tabela 2 – Conteúdo transformado do atributo “AreaTerreno”.

Atributo: AreaTerreno	
De:	Para:
Para todas as áreas < 100	MENOR QUE 100 m2
>= 100 e <= 220	ENTRE 100 E 220 m2
> 220 e <= 500	ENTRE 221 E 500 m2
>= 500	MAIOR QUE 500 m2

Tabela 3 – Conteúdo transformado do atributo “AreaEdificacao”.

Atributo: AreaEdificacao	
De:	Para:
Para todas as áreas < 50	MENOR QUE 50 m2
>= 50 e <=100	ENTRE 50 E 100 m2
>100 e <= 200	ENTRE 101 E 200 m2
>200 e <= 300	ENTRE 201 E 300 m2
>= 300	MAIOR QUE 300 m2

2.2.4 Etapa de Mineração de Dados

2.2.4.1 Algoritmos utilizados

O método de associação APRIORI foi adotado por encontrar padrões em um conjunto de dados que contêm itens que estão relacionados à ocorrência de outros itens. Segundo Garcia e Vieira (2008), essa aplicação gera regras que são representadas na forma de uma implicação $X \Rightarrow Y$, onde X e Y representam um conjunto de itens, e a regra encontrada representa a implicação de que onde ocorre o conjunto de itens X também ocorre o conjunto de itens Y.

Na Mineração de Dados são utilizadas medidas de interesse, sendo que, no caso das regras de associação, as mais comuns são as medidas de suporte e confiança conforme estão representadas pelas equações (1) e (2) a seguir:

$$\text{Suporte} = \frac{\text{Ocorrências de X}}{\text{Total de Ocorrências do BD}} \quad (1)$$

$$\text{Confiança} = \frac{\text{suporte}(X \cup Y)}{\text{suporte}(X)} \quad (2)$$

Os parâmetros de suporte e confiança são indispensáveis na validação da qualidade dos resultados obtidos através das regras de associação utilizadas neste estudo, onde suporte, ou seja, o percentual de vezes que o conjunto X aparece no conjunto de dados. E o segundo parâmetro, a confiança, indica o percentual de ocorrência do conjunto X e Y nas transações que contêm X.

Outra técnica aplicada diz respeito às árvores de decisões, utilizadas para gerar uma classificação dos valores semelhantes, como uma árvore, que possui seu tronco, galhos e

folhas. No trabalho proposto foi utilizado o algoritmo J48 para gerar uma árvore de decisão baseada nos atributos oriundos da Base de dados do IPTU.

De acordo com Queiroz, Gomes e Carvalho (2002), o algoritmo J48 é uma implementação em Java do algoritmo C4.5. Segundo Ingargiola (2011), a partir de uma árvore de decisão é possível derivar regras. As regras são escritas considerando o trajeto do nodo raiz até uma folha da árvore. Estes dois métodos, árvores de decisão e regras de associação, são geralmente utilizados em conjunto. Devido ao fato das árvores de decisão tender a crescer muito, de acordo com algumas aplicações, elas são muitas vezes substituídas pelas regras. Isto acontece em virtude das regras poderem ser facilmente modularizadas. Uma regra pode ser compreendida sem que haja a necessidade de se referenciar outras regras.

3 Resultados

3.1 Aplicação das regras de associação

O algoritmo APRIORI, foi utilizado para criar regras de associação, que permita extrair informações não sugeridas, referente aos dados do tributo. Nesse sentido a Tabela 4, que utiliza o conjunto Arrecadação.arff, representando a totalidade dos lançamentos do IPTU no exercício de 2011, destaca as principais regras e os respectivos fatores de confiança, onde análises diversas podem ser obtidas, principalmente por especialistas no tributo IPTU. Este tributo tem características de distribuição dos lançamentos na área urbana dos municípios.

Tabela 4 – Regras de associação utilizando os atributos “Bairro”, ”Zona”, “Situacao”, “Padrao”, “AreaTerreno” e “AreaEdificacao”.

Suporte mínimo	0.1 (5637 instances)
Confiança mínima	0.51
Número de ciclos executados	18
Regra obtida / conjunto Arrecadação.arff	Fator de confiança
1. AreaTerreno=ENTRE 100 E 220 m AreaEdificacao=ENTRE 50 E 100 m 8247 ==> Padrao=MEDIANO 6686	0.81
2. Zona=AREA-801 9511 ==> Situacao=QUITADO 7395	0.78
3. Situacao=QUITADO AreaEdificacao=ENTRE 50 E 100 m 13952 ==> Padrao=MEDIANO 10790	0.77
4. AreaEdificacao=ENTRE 50 E 100 m 20947 ==> Padrao=MEDIANO 15903	0.76
5. AreaTerreno=ENTRE 221 E 500 m AreaEdificacao=ENTRE 50 E 100 m 9191 ==> Padrao=MEDIANO 6822	0.74
6. AreaEdificacao=ENTRE 101 E 200 m 9818 ==> Situacao=QUITADO 7244	0.74
7. AreaTerreno=MAIOR QUE 500 m 8468 ==> Situacao=QUITADO 6106	0.72
8. AreaEdificacao=ENTRE 101 E 200 m 9818 ==> Padrao=MEDIANO 6950	0.71

Destacam-se então algumas regras, sendo:

- Regra número 2, com fator de confiança de 78%, a área central da cidade, compreendida principalmente pela zona 801 revela-se com boa adimplência do imposto.
- Regras número 1, 3 e 4, relativas as construções pequenas, de áreas entre 50 e 100 m², tem padrão mediano de construção e situação de pagamento quitada.

- c) Com 72% de confiança, a regra número 7, demonstra que as propriedades com áreas de terreno grandes, maiores que 500 m, também têm boa situação de quitação do IPTU.

Baseado nos resultados apresentados na Tabela 4 pode-se afirmar que a adimplência na região central da cidade pode ser fundamentada pelo fato ser um setor essencialmente comercial, portanto as empresas têm maior necessidade de manter em dia suas obrigações tributárias. Também é possível afirmar que as construções de médio porte, entre 50 e 100m², têm boa relação de pagamentos.

Foi considerado na Tabela 5, que também utiliza o conjunto de dados Arrecadação.arff, os bairros da cidade, onde os atributos preditivos “Natureza” e “Situacao” são confrontados com o atributo classe “bairro”.

Tabela 5 – Regras de associação utilizando os atributos “Bairro”, “Natureza” e “Situacao”.

Suporte mínimo	0.01 (564 instances)
Confiança mínima	0.05
Número de ciclos executados	20
Regra obtida / conjunto Arrecadação.arff	Fator de confiança
1. Bairro=GUANDU Situacao=QUITADO 948 ==> Natureza=PREDIAL 885	0.93
2. Bairro=DISTRITO DE ITAOCA 717 ==> Natureza=PREDIAL 663	0.92
3. Bairro=WALDIR FURTADO AMORIM 679 ==> Natureza=PREDIAL 620	0.91
4. Bairro=GUANDU 1221 ==> Natureza=PREDIAL 1112	0.91
5. Bairro=IBITQUARA Situacao=QUITADO 992 ==> Natureza=PREDIAL 901	0.91
6. Bairro=OTTON MARINS 685 ==> Natureza=PREDIAL 620	0.91
7. Bairro=NOVO PARQUE Situacao=QUITADO 930 ==> Natureza=PREDIAL 836	0.9
8. Bairro=SAO LUIZ GONZAGA 1019 ==> Natureza=PREDIAL 916	0.9
9. Bairro=AQUIDABAN Situacao=QUITADO 1798 ==> Natureza=PREDIAL 1611	0.9
10. Bairro=NOSSA SENHORA DE FATIMA 664 ==> Natureza=PREDIAL 594	0.89
11. Bairro=ZUMBI Situacao=QUITADO 967 ==> Natureza=PREDIAL 861	0.89
12. Bairro=NOVO PARQUE 1476 ==> Natureza=PREDIAL 1308	0.89
13. Bairro=SUMARE 872 ==> Natureza=PREDIAL 770	0.88
14. Bairro=INDEPENDENCIA Situacao=QUITADO 1315=> Natureza=PREDIAL 1161	0.88
15. Bairro=SANTO ANTONIO Situacao=QUITADO 1389 => Natureza=PREDIAL 1226	0.88
16. Bairro=IBITQUARA 1436 ==> Natureza=PREDIAL 1266	0.88
17. Bairro=ZUMBI 1793 ==> Natureza=PREDIAL 1580	0.88
18. Bairro=PARAISO Situacao=QUITADO 1044 ==> Natureza=PREDIAL 918	0.88
19. Bairro=AMARELO 848 ==> Natureza=PREDIAL 741	0.87
20. Bairro=ZUMBI Situacao=NAO PAGO 818 ==> Natureza=PREDIAL 712	0.87

As regras números 11 e 20 identificam o bairro Zumbi, região carente do município, com situação de pagamento mediana. Na primeira regra relacionada temos um fator de confiança de 0.89, situação de pagamento quitado e a segunda um fator de confiança de 0.87,

onde a situação de pagamento é desfavorável. Contudo, as demais regras indicam que os lançamentos de natureza predial têm por característica uma boa resposta do munícipe ao pagamento do tributo municipal nos bairros: Paraíso, Santo Antônio, Independência, Aquidaban, Novo Parque, Ibitiquara e Guandu. As regras apresentadas na Tabela 6 e Tabela 7, foram extraídas do conjunto devedores.arff, sendo o resultado desta exploração evidenciado a seguir.

Tabela 6 – Regras de associação utilizando os atributos “Bairro”, “AreaTerreno” e “AreaEdificacao”.

Suporte mínimo	0.05 (986 instances)
Confiança mínima	0.5
Número de ciclos executados	19
Regra obtida / conjunto devedores.arff	Fator de confiança
1. AreaEdificacao=NAO EDIFICADO 4669 => AreaTerreno=ENTRE 221 E 500 m 2753	0.59
2. AreaEdificacao=ENTRE 101 E 200 m 2574 => AreaTerreno=ENTRE 221 E 500 m 1318	0.51

Os imóveis não construídos ou com construções entre 101 e 200 m², tendo área de terreno entre 221 e 500 m², representam o perfil dos lançamentos com maior ausência de pagamento. Em seguida a técnica Apriori foi aplicada aos atributos “Bairro”, “Zona” e “Padrao”, conforme apresentado na Tabela 7.

Tabela 7 – Regras de associação utilizando os atributos “Bairro”, “Zona” e “Padrao”.

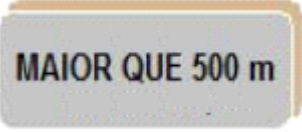
Suporte mínimo	0.05 (986 instances)
Confiança mínima	0.1
Número de ciclos executados	19
Regra obtida / conjunto devedores.arff	Fator de confiança
1. Zona=AREA-601 2800 ==> Padrao=MEDIANO 1718	0.61
2. Zona=AREA-501 3291 ==> Padrao=MEDIANO 1891	0.57
3. Zona=AREA-801 2116 ==> Padrao=MEDIANO 1130	0.53
4. Zona=AREA-401 2520 ==> Padrao=MEDIANO 1080	0.43

O conhecimento obtido evidencia as zonas da cidade onde se tem maior inadimplência com base no padrão construtivo do imóvel.

3.2 Aplicação da árvore de decisão


Em outra análise, foram criadas árvores de decisão utilizando o algoritmo J48. A seguir são demonstrados os principais resultados provenientes desta descoberta de conhecimento. Nas Tabelas 8 e 9, são apresentados os resultados extraídos do conjunto MaioresLançamentos.arff.

Tabela 8 – Área de terreno com mais de 500 m².

Métrica	Assertividade
Instâncias classificadas corretamente	98,51%
Instâncias classificadas incorretamente	1,48%
Atributo Classe : ArreaTerreno	
	

Pode-se observar que aproximadamente 99% dos lançamentos de maior valor do imposto são relativos a grandes áreas de terreno. Na Tabela 9 é apresentado o resultado da relação entre os maiores lançamentos do imposto e a situação de pagamento por natureza.

Tabela 9 – Relação ente os atributos “natureza” e “situação de pagamento”.

Métrica	Assertividade
Instâncias classificadas corretamente	67,40%
Instâncias classificadas incorretamente	32,59%
Atributo Classe : Situacao	
	

Pode-se observar que entre os maiores lançamentos do tributo a maior inadimplência está entre os imóveis de natureza territorial, por outro lado os imóveis com edificação (natureza predial) têm situação predominante de pagamento regular.

Na Tabela 10, é apresentada a assertividade quanto à classificação obtida para a árvore de decisão exposta através da Figura 2, o índice de 65,43% para as instâncias classificadas de forma correta é considerado bom, devido ao quantitativo total de 56374 registros analisados e considerando a alta granularidade do conjunto.

Tabela 10 – Classificação das instâncias relacionadas a árvore de decisão da Figura 2.

Métrica	Assertividade
Instâncias classificadas corretamente	65,43%
Instâncias classificadas incorretamente	34,56%
AtributoClasse : Padrão	

Em seguida, o algoritmo J48 nos fornece uma estratificação do zoneamento urbano, revelando alguns elementos importantes, como a relação entre os valores em débito e o padrão construtivo do imóvel. Também a relação entre a situação de pagamento, o valor do débito e o

padrão construtivo são evidenciados pela árvore de decisão obtida e apresentada na Figura 2, nesta a “árvore” foi “quebrada” em três partes para uma melhor visualização.

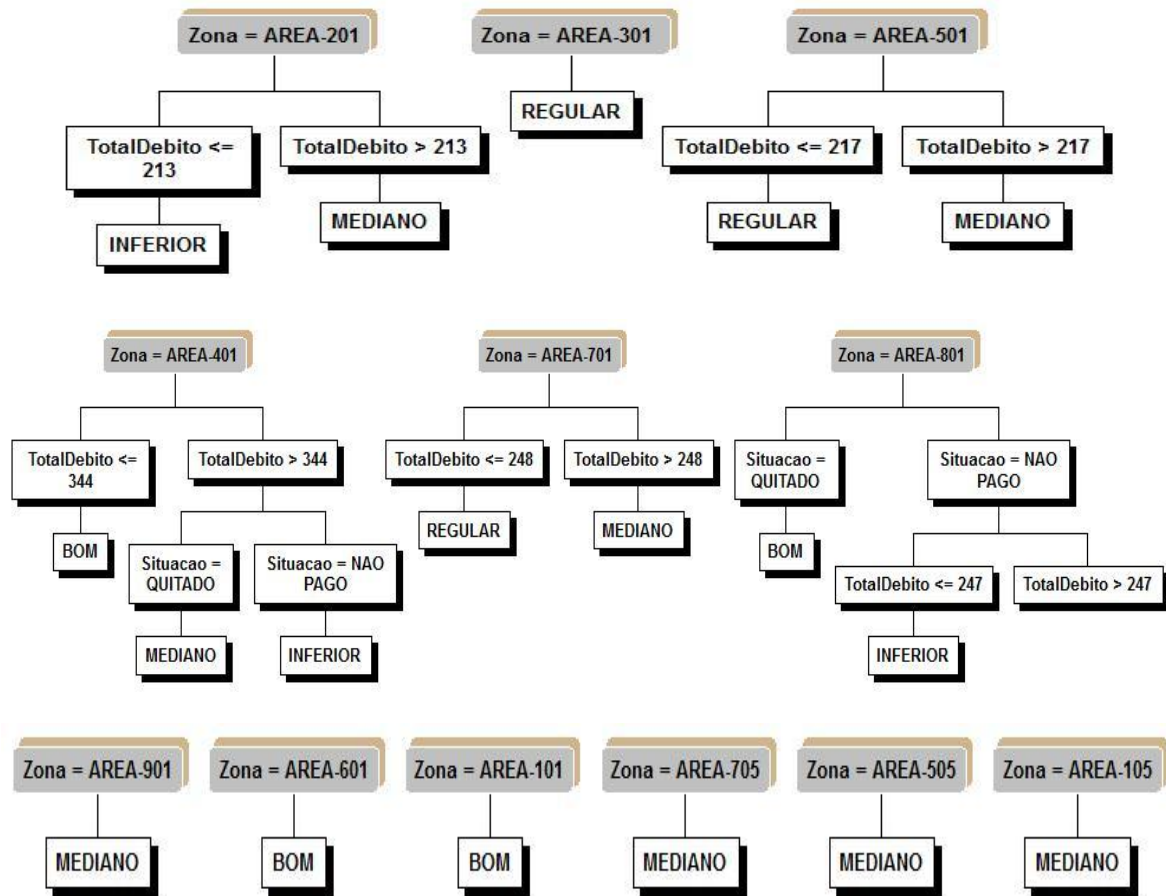


Figura 2 - Árvore de decisão do algoritmo J48, evidência o padrão construtivo por zona.

Desta forma é possível identificar em que zona da cidade estão presentes as construções relacionando-as ao padrão de construção, como é o caso da zona 101, com “BOM” nível de construções. O “galho” da zona 401 traz grande riqueza de detalhes, com destaque pela identificação do nível de construção “BOM” para os débitos lançados com valores abaixo de 344 UPF’s, para os valores acima de 344 UPF’s um novo “galho” da árvore de decisão destaca a situação de pagamento “QUITADA” para as construções de padrão “MEDIANO” e “NÃO PAGO” para as construções de padrão “INFERIOR”.

4 Conclusão

A importância da exploração das grandes bases de dados existentes nas empresas privadas, também se aplica as corporações públicas, a Mineração de Dados é uma das principais ferramentas a ser utilizada para este fim. As informações obtidas através do algoritmo J48 pruned tree, que simula uma árvore de decisão, e do algoritmo de associação APRIORI, possibilitaram a identificação de regras e análises até então não identificadas pelas ferramentas tradicionais de análise de dados, como: relatórios gerenciais e consultas por ferramentas de BI.

Um dos principais objetivos da mineração é a extração de conhecimentos inesperados e não induzidos. Desta forma, com fator de confiança de 78%, a área central da cidade, foi identificada com boa adimplência do imposto. Também com confiabilidade acima de 87% foi traçado um perfil dos bairros em face da situação de pagamento e padrão de construção dos

mesmos. O algoritmo J48 trouxe com 98.52 % de assertividade o perfil do imóvel com maior inadimplência do IPTU. Foi possível identificar também as áreas da cidade onde a inadimplência é alta, possibilitando ao poder público conhecer os motivos que levam a essa situação, o que pode culminar na decisão estratégica de remanejar obras públicas para um ou outro setor do município.

As informações apresentadas possibilitam a tomada de decisões com alto índice de acerto, o que deve fazer parte do cotidiano da gestão pública. Assim este trabalho, pretende motivar o uso das técnicas de Mineração de Dados no setor público municipal e contribuir para uma fase de maior eficácia na análise de informações.

Referências

- ALMEIDA, Ezilda Duarte. Algoritmos de Classificação Com a Opção de Rejeição. (Dissertação de Mestrado) Porto: Faculdade De Engenharia Da Universidade Do Porto, 2010. 19 p.
- BOTHOREL, Gwenaël; SERRURIER, Mathieu; HURTER, Christophe. Utilisation d'outils de Visual Data Mining pour l'exploration d'un ensemble de règles d'association. Sophia Antipolis, France: Ihm'11, 2011. 4 p.
- BRASIL. Constituição (1891). Constituição da República dos Estados Unidos do Brasil Brasília: Fundação Projeto Rondon, 1891.
- BRASIL. Constituição (1934). Constituição da República dos Estados Unidos do Brasil Brasília: Fundação Projeto Rondon, 1934.
- BROWN, Joyce; COOPER, Ceri; PIDD, Michael. A taxing problem: The complementary use of hard and soft OR in the public sector. London: Elsevier, 2005. 14 p.
- CARVALHO Jr, Pedro Humberto Bruno .ASPECTOS DISTRIBUTIVOS DO IPTU E DO PATRIMÔNIO IMOBILIÁRIO DAS FAMÍLIAS BRASILEIRAS. Rio de Janeiro: Dirur/ipea, 2009. 53 p.
- FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. Knowledge Discovery and Data Mining: Towards a Unifying Framework. Portland, Oregon: Ai Magazine, 1996.
- GALVÃO, Noemi Dreyer; MARIN, Heimar de Fátima. Técnica de mineração de dados: uma revisão da literatura. Cuiabá - Mt: Acta Paul Enferm, 2009. 5 p.
- GARCIA, Ederson; VIEIRA, Marina Teresa Pires. MINERAÇÃO DE DADOS USANDO REGRAS DE ASSOCIAÇÃO MULTI-RELACIONAL QUANTITATIVA. Piracicaba: 5º Congresso de Pós-graduação da Unimep, 2008. 96 p.
- GUO, Tao; LI, Guiyang Yang. Neural data mining for credit card fraud detection. San Diego: Proceedings Of The 7th International Conference On Machine Learning And Cybernetics, 2008. 4 p.
- INGARGIOLA, G. Building Classification Models: ID3 and C4.5. Disponível em: <http://www.cis.temple.edu/~ingargio/cis587/readings/id3-c45.html>. Acesso em: 13 jul. 2012.
- KAMBER, Micheline; HAN, Jiawei; CHIANG, Jenny Y. Data Mining - Concepts and Techniques. Nova York: Morgan Kaufmann Publishers, 2001.
- MAGGI, Rodolfo Henrique; FLORISSI, Stefano. Fundamentos da Tributação; importância e características do imposto sobre a propriedade; e um estudo de caso da equidade administrativa do iptu no município de Caxias do Sul. Porto Alegre: Faculdade de Ciências Econômicas, Universidade Federal do Rio Grande do Sul, 2002. 125 p.
- NASCIMENTO, Francisco Assis M.; LEHNEN, Fernando; MORÉ, Morecy V.; LEIZER, Salomão A. GIF: A Web-based System for Tax Management and Fiscal Intelligence in Municipal Tax Administration. Bogota, Colombia: Acm Symposium On Applied Computing, 2009. 7 p.
- NGAI, Eric; HU, Yong; WONG, Y. H.; CHEN, Yijun; SUN, Xin. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Hong Kong: Elsevier, 2011. 10 p.
- QUEIROZ, Ana Emilia de Melo; GOMES, Alex Sandro; CARVALHO, Fransisco de Assis Tenório. Mineração de Dados de IHC para Interface Educativas. Recife - Pe: Ufpe – Universidade Federal de Pernambuco, 2002. 15 p.

SILVA, Leandro Santiago; LIMA, Rafael Sá; MOREIRA, Fabiano; MONTEIRO, Odlaniger Lourenço Damaceno. **DESCOBERTA DE PADRÕES RELEVANTES NA BASE DE DADOS DA FARMÁCIA HOMEOBEL CENTER**. Santa Catarina: Cesupa, 2007.

TRINDADE, Simone. **História da Arte**. Salvador: Faculdade de Tecnologia e Ciências - Ensino A Distância, 2007. 88 p.

VARSANO, Ricardo. **O Imposto Predial e Territorial Urbano: receita, equidade e adequação aos municípios**. Rio de Janeiro: Campus, 1977.

VICENTE, Leonardo Cardoso; GOMES, Geórgia Regina Rodrigues; MATIAS, Ítalo de Oliveira. **Utilização de técnicas de mineração de dados para estudo do comportamento do mercado de ações**. Campos - Rj: Xvi - Simpósio de Engenharia de Produção (simpep), 2010. 11 p.

YEH, I-cheng; LIEN, Che-hui. **The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients**. Taiwan: Elsevier, 2009.

ZAMBON; MEIRELLES. **A Evolução do Processo Decisório e as novas Ferramentas de Apoio à Decisão: Data Warehouse, Olap e Data Mining**. São Carlos: Programa de Mestrado em Engenharia de Produção, 2001.