

## MINERAÇÃO DE DADOS APLICADA A RELAÇÃO CLIENTES E PAGAMENTOS – ESTUDO BIBLIOMÉTRICO

### ***Alcione Dias Silva***

Mestre em Pesquisa Operacional e Inteligência Computacional / UCAM  
diasalcione@gmail.com

### ***Bruno Missi Xavier***

Mestre em Pesquisa Operacional e Inteligência Computacional / UCAM  
bruno.missi@dataci.es.gov.br

### ***Helder Gomes Costa***

Doutor em Engenharia Mecânica / PUC-Rio  
helder.uff@gmail.com

### ***Georgia Regina Rodrigues Gomes***

Doutor em Informática / PUC-Rio  
georgia@ucam-campos.br

Recebido: 15 de junho de 2012. Revisado: 04 de julho de 2013. Aceito: 28 de outubro de 2012.

Publicado online: 23 de março de 2013.

### **RESUMO**

O KDD (*Knowledge Discovery in Database*) e a Mineração de Dados incluem tarefas e métodos para extração de conhecimento útil, interessante e indispensável na tomada de decisões rápidas nas mais diversas áreas de conhecimento. Esta pesquisa tem como objetivo nortear o leitor na vasta relação de trabalhos que se utilizam desta importante ferramenta e apontar as principais publicações de Mineração de Dados utilizadas na relação clientes e pagamentos, no âmbito público e privado, além de propor chaves de pesquisa que possam contribuir para a obtenção de trabalhos relacionados com objetivo de facilitar as buscas pertinentes a este importante assunto. Quanto à metodologia, foram pesquisadas as publicações relacionadas às técnicas de Mineração de Dados aplicadas a base de dados tributária pertencente a entidades governamentais estendendo-se analogamente aos trabalhos similares na área privada. A base de dados utilizada para este estudo foi a SCOPUS no período de janeiro de 1999 a março de 2012. Dentre os 927 resultados obtidos, tem merecido destaque o setor de pagamentos eletrônicos via cartões de crédito e a rede neural artificial como técnica aplicada com maior sucesso. O idioma mais utilizado para a confecção das publicações foi o inglês. Os resultados apresentados contribuem no sentido de direcionar os autores quanto às áreas mais carentes de trabalhos científicos aplicados a Mineração de Dados, os países e idiomas com maior número de publicações, contribuindo de forma direcionada para o enriquecimento de trabalhos futuros.

**Palavras-Chave:** Mineração de Dados; Clientes; Pagamentos; Estudo bibliométrico.

## ABSTRACT

The KDD (Knowledge Discovery in Database) and the Data Mining include tasks and methods for extracting useful knowledge, interesting and indispensable in making quick decisions in several areas of knowledge. This research aims to guide the reader in the vast list of jobs that use this important tool and show the major publications data mining used in customer relationship and payments, in the public and private sector, besides propose search keys that may contribute to obtain work related with goal of facilitate the searches relevant to this important subject. Regarding the methodology, were surveyed publications related to data mining techniques applied to tax database belonging to government entities extending analogously to similar jobs in the private sector. The database used for this study was the SCOPUS from January 1999 to March 2012. Of the 927 results, the sector that has been highlighted was the electronic payments via credit cards and the artificial neural network the technique as applied with greater success. The language most often used for making the publications was the english. The results presented contribute towards directing the authors as to the neediest areas of scientific research applied to Data Mining, countries and languages with the greatest number of publications, contributing directly to the enrichment of future works.

**KeyWords: Data Mining; Customers; Payments; bibliometric study.**

## 1 – INTRODUÇÃO

Desde meados do século passado com o uso crescente de computadores, a quantidade de informações produzidas cresce substancialmente. De acordo com Duque (2001), em 1946 John Mauchly e John Eckert desenvolveram o primeiro computador eletrônico. Naquele momento o mundo contemporâneo iniciava o experimento do uso da informação automatizada. Contudo, em 1975 (TUFFI, 2011) com o surgimento do computador pessoal iniciou-se um período que pode ser chamado de revolução tecnológica, onde pessoas não necessariamente vinculadas a grandes corporações podiam sistematizar dados transformando-os em informação.

Com a popularização da internet em 1990 a revolução computacional ganhou seu maior aliado, possibilitando então que as informações entre pessoas e empresas fossem facilmente trocadas e conseqüentemente o volume das mesmas cresceu exponencialmente, gerando enormes repositórios de informações estruturadas em sistemas de bancos de dados ou mesmo não estruturadas, distribuídas em arquivos de textos, imagens, áudios, etc. O tratamento destes grandes repositórios de informação tornou-se indispensável e muito rentável. Desta forma tecnologias para exploração de informações surgiram como grandes aliadas das corporações públicas e privadas na descoberta de conhecimento.

O uso de tecnologias que viabilizem a extração de informações e conhecimentos para serem analisados na esfera decisória das organizações públicas e privadas passaram a representar um diferencial competitivo das empresas. Uma das técnicas utilizadas é a Mineração de Dados, que é parte integrante do processo de KDD (*Knowledge Discovery in Database*). Segundo Quoniam, *et al.* (2001), o KDD pode ser visto como o processo da descoberta de novas correlações, padrões e tendências significativas por meio da análise minuciosa de grandes conjuntos de dados estocados. De acordo com Fayyad, Piatetsky-shapiro e Smyth (1996), a descoberta de conhecimento em base de dados, KDD, é vista como uma disciplina mais ampla, e o termo Mineração de Dados como o componente que trata dos métodos do descobrimento do conhecimento.

As ferramentas de análise de dados tradicionais utilizadas não são capazes de traçar perfis e relacionamentos não triviais. Para este estudo, a aplicação da Mineração de Dados é motivada pela implementação de “inteligência computacional” às bases de dados e pela melhoria do conhecimento das mesmas, principalmente no poder público.

O principal objetivo deste trabalho é apresentar um estudo bibliométrico seguido de uma revisão bibliográfica das obras científicas que estejam relacionadas a Mineração de Dados aplicada à relação de clientes e pagamentos, com abrangência no setor público e privado, identificando os trabalhos de maior relevância, os autores mais importantes, além dos artigos mais alinhados com o tema proposto. Os resultados encontrados podem contribuir para que pesquisadores das técnicas de Mineração de Dados possam identificar os países, idiomas e autores que mais produzem conhecimento relacionado, além de demonstrar o que está sendo publicado na comunidade internacional. Desta forma, indicar as áreas mais carentes de pesquisas, contribuindo para o enriquecimento das próximas publicações.

O presente artigo está estruturado em quatro seções, como segue: A Seção 1 contextualiza o processo de KDD e a técnica de Mineração de Dados. A Seção 2 apresenta a metodologia utilizada para a pesquisa. A Seção 3 expõe os resultados obtidos de acordo com as chaves de busca empregadas. A Seção 4 apresenta a revisão bibliográfica e comenta os artigos mais relevantes. A seção 5 discorre sobre as conclusões oriundas dos diversos resultados.

## 2 – METODOLOGIA

A base de dados utilizada para busca dos artigos foi a SCOPUS, por tratar-se da mais volumosa biblioteca de publicações disponível para este trabalho, uma vez que foram encontrados 927 resultados, contra 8 resultados da base SCIELO e 316 da base *Web of Knowledge*.

Foi realizada uma pesquisa bibliográfica onde não foram utilizados filtros de exclusão a fim de evitar a redução dos resultados, optando-se por um retorno com a maior amplitude possível, uma vez que são poucos os trabalhos produzidos que tratam de “clientes e pagamentos” no setor público.

As palavras-chave utilizadas foram: “*data mining AND government*”, “*data mining AND public tax*” e “*payments and data mining*”, o termo em inglês “*data mining*” foi combinado com as áreas de interesse como: “*government*”, “*public tax*” e “*payments*”. O intervalo temporal da pesquisa compreendeu o período de janeiro de 1999 a março de 2012. O estudo bibliométrico, apontou os artigos de maior relevância para o tema proposto, sendo estes abordados na etapa de revisão bibliográfica.

## 3 – RESULTADOS DO ESTUDO BIBLIOMÉTRICO

### 3.1 – Pesquisas das palavras-chave “*data mining and government*”

As buscas executadas na base de conhecimento do SCOPUS foram realizadas com intuito de encontrar trabalhos relacionados à política de governo na aplicação e recolhimento de impostos. Desta forma, utilizando a chave de busca “*data mining AND government*” foram encontrados 813 resultados, com destaque para os anos de 2010 e 2011, pelo fato da pesquisa ter sido realizada em março de 2012, este ano ainda apresenta resultado simbólico.

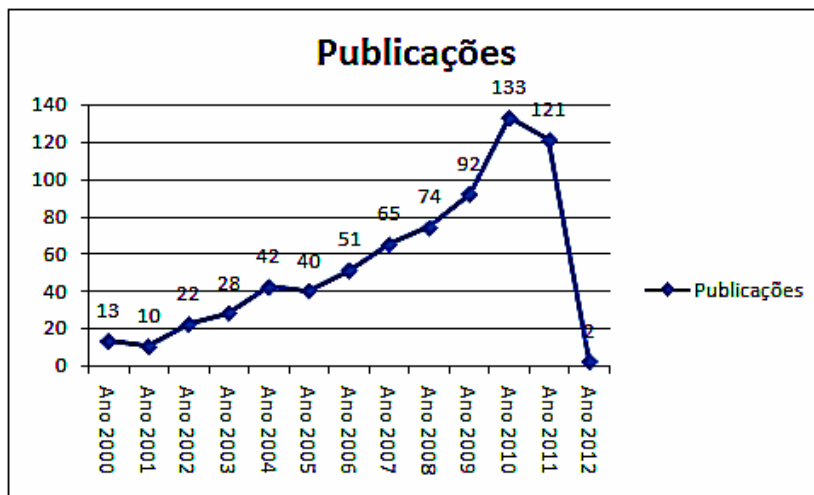


Figura 1: Publicações realizadas a partir do ano 2000.

Identifica-se através dos dados expostos na Figura 1, que a aplicação das técnicas de Mineração de Dados no setor governamental cresceu significativamente na última década, o que demonstra um interesse crescente do setor na aplicação desta ferramenta.

Os autores que mais produziram estão dispostos no Quadro 1, onde são apresentados aqueles com mais de três publicações relacionadas a chave de pesquisa.

Quadro 1: Autores com mais de 3 publicações relacionadas a chave de pesquisa.

Autor	Publicações	Autor	Publicações
Saygin, Y.	4	Lee, J.H.	3
Hong, G.H.	4	Matsui, T.	3
Memon, N.	3	Law, K.H.	3
Wiederhold, G.	3	Lau, G.T.	3
Manarvi, I.	3	Izumi, K.	3
Lu, J.	3	Anon,	3
Kim, W.	3	Cao, L.	3

A figura 2 apresenta o gráfico das áreas onde as técnicas de Mineração de Dados estão sendo mais aplicadas, com especial destaque para o setor de ciências sociais, engenharia e ciência da computação com 48% do total de publicações.

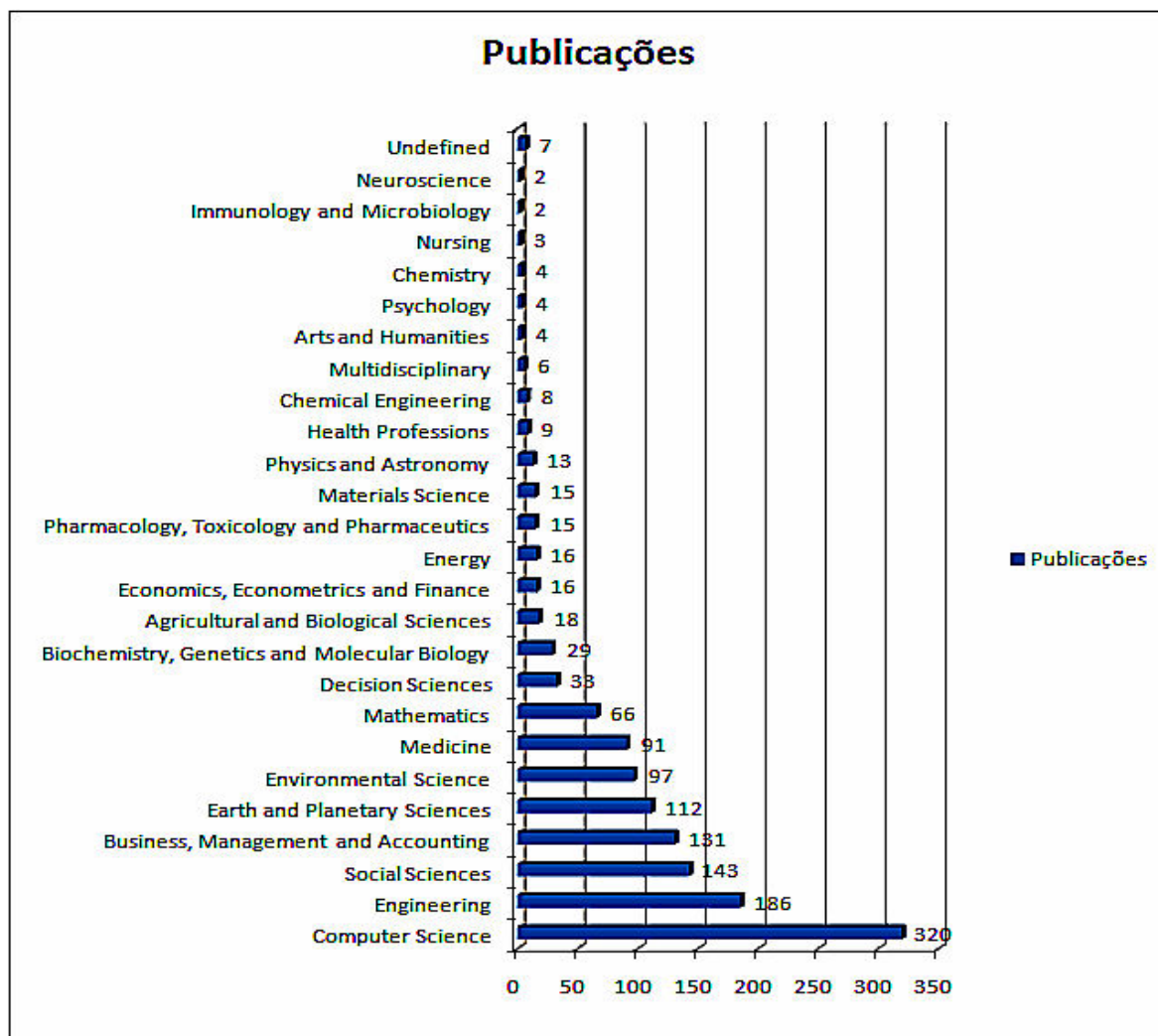


Figura 2: Publicações por áreas.

As palavras-chave relacionadas ao estudo em questão foram encontradas conforme o Quadro 2 e representam 13% do total.

Quadro 2: Palavras-chave relacionadas à área de governo.

Palavras-chave	publicações
<i>Government data processing</i>	155
<i>Government</i>	64
<i>E-Government</i>	23
<i>Public policy</i>	22

Os países que mais produziram pesquisas relacionadas são China e Estados Unidos, sendo que 98% das publicações foram escritas no idioma Inglês, de acordo com o Quadro 3.

Quadro 3: Países e idiomas com maior número de publicações.

País	Publicações	Idioma	Publicações
China	195	inglês	745
Estados Unidos	174	chinês	36
Austrália	39	francês	4
Canadá	37	alemão	4
Reino Unido	33	japonês	3
Taiwan	23	espanhol	3
Japão	15	tcheco	1
Coréia do Sul	13	italiano	1
Brasil	13	polonês	1
Índia	12	português	1
Alemanha	11		
Itália	10		

É interessante destacar que apesar de 13 publicações serem realizadas no Brasil, apenas um dos trabalhos foi escrito em português, o que demonstra o interesse dos autores em publicar seus trabalhos em âmbito internacional, tendo como preferência o idioma inglês com 745 publicações.

### 3.2 – Pesquisas das palavras-chave “data mining and public tax”

Na busca relacionada diretamente com o assunto tributos e taxas, foi utilizada a chave de busca “data mining AND public tax” sendo encontrados 7 resultados, distribuídos entre os anos de 2006 a 2011, conforme Figura 3.

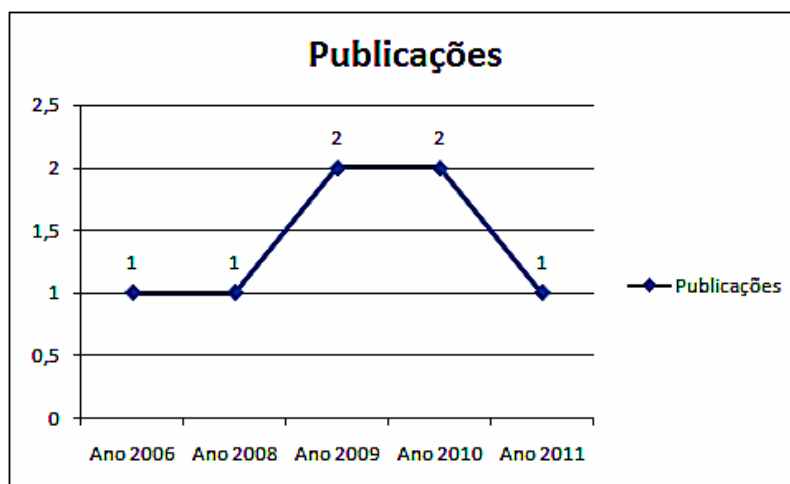


Figura 3: Publicações por ano, utilizando a chave de busca “data mining AND public tax”

Os autores de maior destaque, por terem suas obras citadas são elencados no Quadro 4.

Quadro 4: Autores que tiveram suas obras citadas, na busca por “data mining AND public tax”.

Autor	Obra	Número de citações
Fedorowicz, J., Gogan, J.L., Culnan, M.J.	<i>Barriers to interorganizational information sharing in e-government: A stakeholder analysis</i>	1
Brown, J., Cooper, C., Pidd, M.	<i>A taxing problem: The complementary use of hard and soft or in the public sector</i>	7

Na Figura 4, estão relacionadas às áreas onde as técnicas de Mineração de Dados estão sendo mais aplicadas, de acordo com a chave de busca “data mining AND public tax”, nota-se que a área de Ciências da Computação também foi uma das mais estudadas.

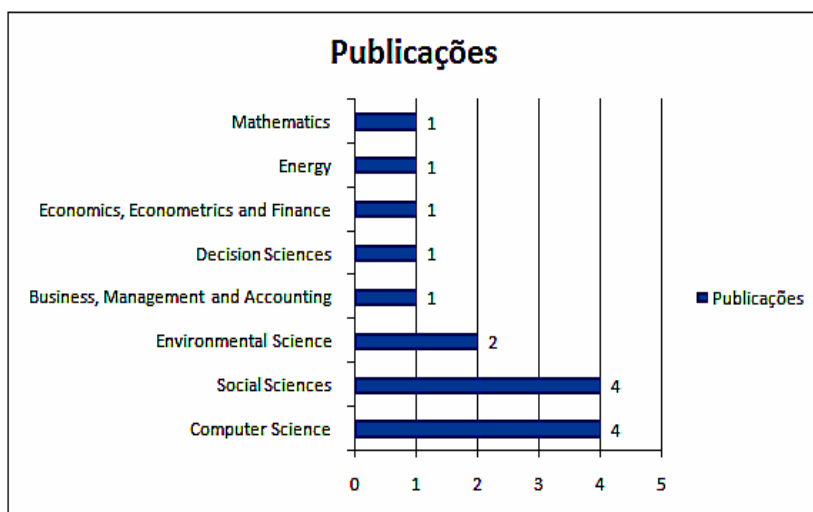


Figura 4: Publicações por áreas.

Os países relacionados a chave de busca “data mining AND public tax” estão distribuídos de acordo com o Quadro 5, nesta todas as obras foram escritas em inglês.

Quadro 5: Países e idiomas com maior número de publicações.

País	Publicações	Idioma	Publicações
Brasil	1	inglês	7
Canadá	1		
China	1		
Holanda	1		
Romênia	1		
Reino Unido	1		
Estados Unidos	1		

3.3 – Pesquisas das palavras-chave “payments and data mining”

A chave de busca “payments and data mining” foi aplicada na base do SCOPUS de forma a encontrar 107 trabalhos que por analogia possam ser utilizados, uma vez que a imposição de tributos aos contribuintes pelos governos gera para estes créditos fiscais, culminando, portanto na obrigação de pagamento dos tributos. A seguir, no Quadro 6, estão relacionadas as publicações que se destacaram pelo grande número de citações.

Quadro 6: Autores com mais de 10 citações.

Autor	Obra	Número de citações
Chan, P.K., Fan, W., Prodromidis, A.L., Stolfo, S.J.	<i>Distributed data mining in credit card fraud detection</i>	91
Desmond, N., Allen, C.F., Clift, S., Justine, B., Mzugu, J., Plummer, M.L., Watson-Jones, D., Ross, D.A	<i>A typology of groups at risk of HIV/STI in a gold mining town in north-western Tanzania</i>	26
Foster, D.P., Stine, R.A.	<i>Variable selection in data mining: Building a predictive model for bankruptcy</i>	22
Ferris, M.	<i>Insights on mobile advertising, promotion, and research</i>	11

Na Figura 5, são apresentadas as publicações relacionadas por ano, sendo elencadas apenas as que têm data de publicação maior que o ano 2000, desconsiderando os anos de 1974 a 1999, por serem poucas e muito antigas. Destacam-se os anos de 2007 e 2011 com um total de 32 publicações, representando 30% do total.



Figura 5: Publicações realizadas a partir do ano 2000.



Assim como nas buscas por “*data mining AND public tax*” e “*data mining AND government*”, as áreas de engenharia e ciência da computação recebem destaque na produção científica, sendo destacadas as principais na Figura 6.

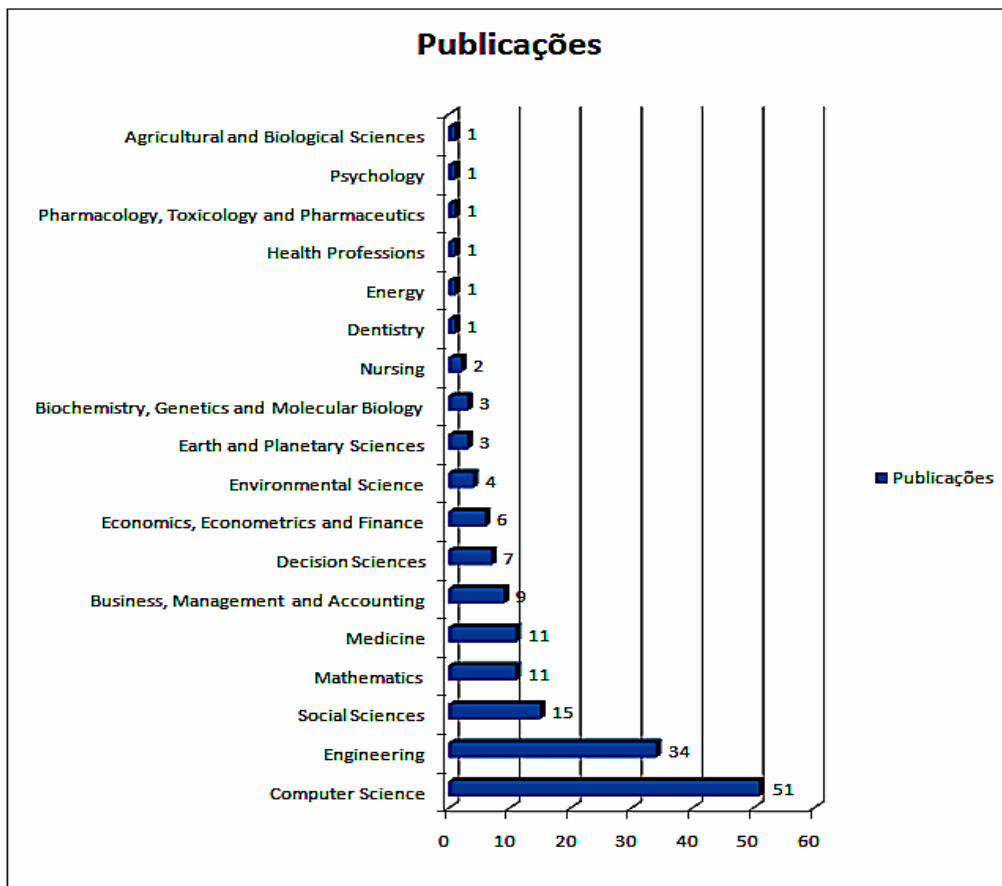


Figura 6: Publicações por áreas.

Dentre as palavras-chave contidas na chave de busca “*payments and data mining*”, foram selecionadas as constantes no Quadro 7, por terem relação direta com o estudo proposto, como por exemplo: *Payment collection*, *Financial data* e *Public utilities*.

Quadro 7: Palavras-chave relacionadas à Mineração de Dados aplicada ao estudo proposto.

Palavras-chave	publicações	Palavras-chave	publicações
<i>Data mining techniques</i>	12	<i>Associative processing</i>	3
<i>Information management</i>	11	<i>Classification</i>	3
<i>Decision trees</i>	10	<i>Computer simulation</i>	3
<i>Risk assessment</i>	7	<i>Credit evaluations</i>	2
<i>Decision support systems</i>	6	<i>Payment collection</i>	3
<i>Data mining methods</i>	5	<i>Population statistics</i>	3
<i>Fraud detection</i>	5	<i>Data Mining</i>	2
<i>Learning systems</i>	5	<i>Data analysis</i>	2
<i>Data mining technology</i>	4	<i>Demographic characteristics</i>	2
<i>Database systems</i>	4	<i>Financial data</i>	2
<i>Knowledge discovery</i>	4	<i>Financial management</i>	2
<i>Predictive modeling</i>	4	<i>Financial obligations</i>	2
<i>Profitability</i>	4	<i>Public transit systems</i>	2
<i>Artificial intelligence</i>	3	<i>Public utilities</i>	2
<i>Association rules</i>	3	<i>Relevant information</i>	2

No contexto da chave de busca “*payments and data mining*”, o Quadro 8 revela os países com maior número de produções científicas.

Quadro 8: Países com mais de duas publicações e idiomas mais utilizados.

País	Publicações	Idioma	Publicações
China	20	inglês	100
Estados Unidos	14	chinês	2
Taiwan	8	tcheco	1
Coréia do Sul	4	holandês	1
Austrália	4	alemão	1
Reino Unido	3	japonês	1
Japão	3	russo	1
Índia	3		
Canadá	3		
Alemanha	3		

#### 4 – REVISÃO BIBLIOGRÁFICA

Nesta etapa são destacados os trabalhos de maior relevância para o tema proposto em face aos resultados apresentados anteriormente. Foram oito os artigos selecionados para a revisão bibliográfica, conforme o Quadro 9, por serem considerados os mais aderentes a aplicação das técnicas de Mineração de Dados e sua aplicação na relação clientes e pagamentos.

Quadro 9: Relação das publicações mais relevantes.

Item	Publicação	Autor	Data
1	Discovering cardholders' payment-patterns based on clustering analysis.	Shih, Chien-chou; CHIANG, Ding-an; HU, Yi-jen; Chen, Chun-Chi.	2011
2	The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature	Eric Ngai Yong Hu Y. H. Wong Yi-Jun Chen Xin Sun	2010
3	GIF: A web-based system for tax management and fiscal intelligence applications to the municipal tax administration	Nascimento, F.A.M., Lehnen, F., Moré, M.V., Leizer, S.A.	2009
4	The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients	I-Cheng Yeh, Che-hui Lien	2009
5	Neural data mining for credit card fraud detection	Tao Guo Guiyang Yang Li	2008
6	A taxing problem: The complementary use of hard and soft or in the public sector	Brown, J., Cooper, C., Pidd, M.	2006
7	Variable selection in data mining: Building a predictive model for bankruptcy.	Foster, D. P.; Stine, R. A.	2004
8	Distributed Data Mining in Credit Card Fraud Detection.	Chan, Philip K.; Fan, Wei; Prodromidis, Andreas.	1999

Brown, Coper, e Pidd (2005) utilizaram as técnicas de Mineração de Dados para melhorar a sistemática de funcionamento do imposto de renda de competência da Receita no Reino Unido. Neste caso, a intenção foi a modernização da aplicação do imposto de pessoas físicas e das empresas, adotando uma missão mais ampla, incluindo uma série de outras responsabilidades, tais como o pagamento de créditos tributários para as famílias trabalhadoras e a gestão do sistema de coleta de empréstimo e reembolso dos estudantes.

No que diz respeito a outro importante tributo municipal, o ISS (Imposto Sobre Serviço), Nascimento, *et al.*(2009) apresenta o GIF, que é um sistema para Gestão e Inteligência Fiscal, um dos módulos deste sistema é o NFSe (Nota Fiscal de Serviços Eletrônica) que possibilita a emissão das notas fiscais de serviços no âmbito municipal. Neste, as empresas informam a movimentação realizada em decorrência da prestação dos serviços nas diversas atividades comerciais. Devido a volumosa geração de dados decorrente da quantidade de transações eletrônicas realizadas, foi também desenvolvido o módulo Atre (e-TA:Análise fiscal eletrônica), que realiza o tratamento das informações coletadas facilitando o trabalho dos fiscais e demais autoridades, 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.

Também no setor privado se encontram necessidades similares ao setor público no tratamento da relação clientes e pagamentos. Em ambos os setores as técnicas de Mineração de Dados são aplicadas para melhor entender o comportamento e o perfil dos clientes ou contribuintes, possibilitando aos gestores um tratamento diferenciado a fim de aperfeiçoar a condução dos negócios e minimizar os prejuízos advindos da inadimplência.

Chan *et al.* (1999) tratam da relação dos usuários de cartão de crédito e os riscos relacionados ao grande crescimento das compras feitas pelos clientes através da internet. Segundo os autores: “Em larga escala, técnicas de mineração de dados pode melhorar o estado da arte em práticas comerciais. Técnicas escaláveis para analisar uma maciça quantidade de dados e transações de forma eficiente servem como detectores de fraudes em tempo hábil, especialmente para o comércio eletrônico”. As fraudes em cartões implicam no aumento do custo do serviço ofertado aos clientes, em forma de aumento das taxas de juros cobradas. O desafio encontrado neste estudo foi a grande diversidade de dados, o que implica em uma heterogeneidade de conjuntos de dados para que os algoritmos sejam treinados. Foram utilizados os algoritmos o *AdaCost* (variante do *AdaBoost2*, 3.) e o algoritmo *AdaBoost*, onde os melhores resultados foram obtidos com o *AdaCost*, que obteve menor perda acumulada em um número esmagador de casos, além de ser mais rápido para os casos experimentados. Os bons resultados obtidos pelo trabalho de Chan, foram também utilizados no problema da invasão em redes de computadores.

Guo e Li (2008) também tratam das técnicas de Mineração de Dados para detecção de fraudes em cartões de crédito, estes utilizam as redes neurais como algoritmo para predição das fraudes, tem como forte motivação o grande número de transações de compras utilizando cartões, cerca de 400.000 por dia. Neste cenário uma redução nas fraudes de 2,5% pode economizar \$1.000.000 por ano. A principal tarefa de predição é identificar se o cartão está sendo utilizado de forma fraudulenta, devido às mudanças constantes de comportamento e novos tipos de fraudes fazem com que seja pouco eficaz a realização desta tarefa apenas por especialistas, este trabalho foi desenvolvido através de uma abordagem probabilística e neuro-adaptativa para uma base de dados de transações de cartões de crédito. Uma das características encontradas foi à associação de uma base com dados fraudulentos, “Começamos com a base de dados de transações fraudulentas comparando cada operação com todas as outras, a fim de encontrar os pares semelhantes”. O resultado obtido pelo estudo culminou em um percentual de 80% de confiança, que segundo os autores é bastante elevado para este tipo de problema.

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. As técnicas abordadas foram: *K-nearest neighbor classifiers (KNN)*, *Logistic regression (LR)*, *Discriminant analysis (DA)*, *Naive Bayesian classifier (NB)*, *Artificial neural networks (ANNs)* e *Classification trees (CTs)*. Os dados minerados foram extraídos da base de dados de um banco, sendo utilizadas 25000 observações relativas a pagamentos com cartões. Neste trabalho os autores chegaram a conclusão que a rede neural artificial é o melhor modelo entre os seis métodos comparados.

Foster e Stine (2001) apresentam um modelo de regressão estatístico que pode ser adaptado para lidar com problemas da Mineração de Dados. Para isto os autores sugerem três adaptações básicas no modelo, onde a omissão de um deles torna pobres os resultados: (i) Adição de funções para capturar não linearidades dos dados e ausência de valores; (ii) A exploração de modernas teorias de critérios de seleção de variáveis; e (iii) A definição de valores de erro padrão seguindo uma abordagem conservadora específica para dados cuja frequência de distribuição segue um padrão irregular. Para validar este modelo, foi utilizada uma base de dados relativa a transações com cartões de crédito onde o objetivo é predizer o endividamento pessoal dos usuários. O domínio abordado apresenta muitos destaques, especialmente o número de eventos de endividamentos muito baixo diante do volume de dados movimentado. Os resultados apresentados

atestam a eficiência do modelo, identificando inicialmente 8500 casos relevantes, e após um maior refinamento 1800 casos de endividamentos comprovados.

Shih et al. (2011) aplicaram técnicas de Mineração de Dados com o objetivo de avaliar o comportamento dos clientes de cartão de crédito de um banco em Taiwan. Através de algoritmos de agrupamento, os usuários são divididos em nove grupos e classificados com diferentes níveis de risco, de acordo com o padrão de consumo. Para isto, foram analisados 9,3 milhões de transações referentes a dois anos de movimentações de trezentos mil clientes. Os resultados sugerem a redução de limites de créditos a grupos identificados como risco altíssimo e alto, além do recrutamento de novos clientes identificados com baixo risco ou risco controlado.

De forma mais ampla, Ngai, Hu *et al.* (2010) tratam da aplicação das técnicas de Mineração de Dados para a detecção de fraudes financeiras (FFD) e apresentaram uma revisão da literatura sobre o assunto e das técnicas de mineração de dados que foram aplicadas. As fraudes foram categorizadas em quatro grupos: fraude bancária, fraude de seguros, valores mobiliários e commodities e seis classes de técnicas de mineração de dados: *classification, regression, clustering, prediction, outlier detection, e visualization*.

Os resultados desta análise, indicam que as tarefas de Mineração de Dados têm sido amplamente aplicadas à detecção de fraude de seguros. As fraudes de cartão de crédito também têm recebido forte atenção nos últimos anos. As principais técnicas utilizadas para mineração em FFD são: *logistic regression models, neural networks, the Bayesian belief network, e decision trees*. Segundo Ngai, Hu *et al.* (2010) este trabalho tem três objetivos: “O primeiro é o de desenvolver um quadro para classificar as aplicações de mineração de dados para FFD. A segunda é fornecer uma análise sistemática e abrangente das pesquisas existentes em artigos sobre as aplicações de mineração de dados para FFD. A terceira é a de gerar um roteiro para que pesquisadores e profissionais possam compreender melhor este campo.”

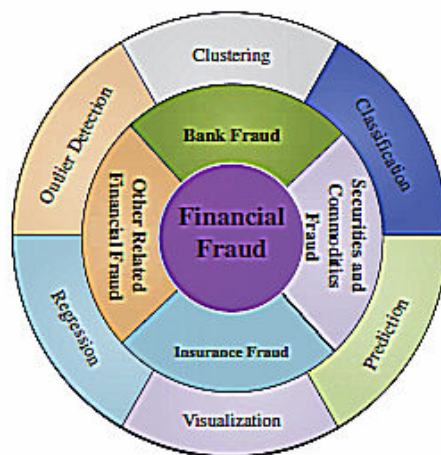


Figura 7: Estrutura conceitual para classificar as aplicações de mineração de dados para FFD, sugerido por Ngai, Hu *et al.* (2010).

O resultado deste trabalho é apresentado na forma de uma classificação entre as áreas relativas ao FFD e as técnicas de Mineração de Dados que melhor se aplicam a cada caso. Os autores destacam que o número limitado de artigos de periódicos relevantes, publicados entre 1997 e 2008 é devido à dificuldade de obtenção de dados experimentais suficientes.

## 5 - CONCLUSÃO

Dentre as chaves de busca utilizadas a string “*data mining AND government*”, foi a que retornou um maior número de resultados, 813 publicações, com destaque para os trabalhos publicados no ano de 2010, a área da ciência da computação apresenta a maioria das aplicações de trabalhos relacionados. A pesquisa por “*data mining AND public tax*” foi a que disponibilizou o menor número de resultados, 7 publicações, contudo os trabalhos encontrados foram os mais significativos quando aplicados a área de governo e impostos, destacando-se Brown, Coper, e Pidd (2005) com o trabalho: “*A taxing problem: The complementary use of hard and soft or in the public sector*”. É importante destacar que o período de 2012 compreende apenas os meses de janeiro a março, este período foi incluso nas pesquisas com objetivo de ampliar o resultado geral na busca de um número maior de publicações relevantes, contudo não devem ser comparados quantitativamente com os anos anteriores.

O idioma inglês é o de maior preferência para a publicação de trabalhos entre os diversos países relacionados na pesquisa. Tal fato pode ser atribuído a grande importância das revistas e congressos científicos com origem nos Estados Unidos.

Dentre o grupo de palavras-chave vinculadas à pesquisa “*data mining and government*” percebe-se a pouca aplicação de trabalhos relacionados às questões pertinentes à “políticas públicas”, perfazendo menos de um por cento das palavras correlacionadas a área de governo, o que destaca a necessidade de que pesquisas sejam realizadas na área de Mineração de Dados aplicadas as atividades relacionadas às diversas esferas do governo e sua atividade quanto a prestação de serviços públicos. Outra carência percebida no período investigado se dá quanto a publicações abordando estudos bibliométricos.

Devido aos poucos trabalhos diretamente relacionados ao dueto governo e impostos, a chave de busca “*payments and data mining*” foi aplicada com intuito de buscar analogamente assuntos relacionados a clientes e pagamentos. Nesta linha de pesquisa o destaque foi o setor de pagamentos eletrônicos via cartões de crédito. No capítulo de revisão bibliográfica foram relacionadas as publicações mais relevantes, tendo como principal critério de seleção a relação direta com o assunto Mineração de Dados aplicada a relação clientes e pagamentos, bem como as publicações com maior número de citações. Após esta etapa ficou claro que a técnica aplicada com maior sucesso é a da rede neural artificial.

## 6 - REFERÊNCIAS

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.

CHAN, Philip K.; FAN, Wei; PRODRUMIDIS, Andreas. **Distributed Data Mining in Credit Card Fraud Detection**. Ieee Intelligent Systems, 1999.

DESMOND, N. *et al.* **A typology of groups at risk of HIV/STI in a gold mining town in north-western Tanzania**. Social Science and Medicine, v. 60, n. 8, p. 1739-1749, 2005. ISSN 02779536 (ISSN).

DUQUE, M. D. F. S. **Evolução da tecnologia da informação e perspectivas do comércio eletrônico**. UNIT: 45 p. 2001.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **Knowledge Discovery and Data Mining: Towards a Unifying Framework**. Portland, Oregon: Ai Magazine, 1996.

FEDOROWICZ, J.; GOGAN, J. L.; CULNAN, M. J. **Barriers to interorganizational information sharing in e-government: A stakeholder analysis**. Information Society, v. 26, n. 5, p. 315-329, 2010. ISSN 01972243 (ISSN).

- FERRIS, M. **Insights on mobile advertising, promotion, and research.** Journal of Advertising Research, v. 47, n. 1, p. 28-37, 2007. ISSN 00218499 (ISSN).
- FOSTER, D. P.; STINE, R. A. **Variable selection in data mining: Building a predictive model for bankruptcy.** Journal of the American Statistical Association, v. 99, n. 466, p. 303-313, 2004. ISSN 01621459 (ISSN).
- 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.
- NASCIMENTO, Francisco Assis M.; LEHNEN, Fernando; MORÉ, Morecy V.. **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.. **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.
- QUONIAM, Luc. *et al.* **Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil.** Ciência da Informação, Brasília, v. 30, n. 2, p. 20-28, maio/ago. 2001.
- TUFFI, Eloy. **DO BASIC AO TI: A trajetória de 30 anos da Microcamp, a empresa brasileira que popularizou o ensino de informática no país.** Campinas: Mgdesign, 2011. 104 p.
- SHIH, Chien-chou; CHIANG, Ding-an; HU, Yi-jen; Chen, Chun-Chi. **Discovering cardholders' payment-patterns based on clustering analysis.** Expert Systems With Applications, n. 38, p.13284-13290, 15 set. 2011.
- 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.