

DOI: 10.5748/9788599693100-11CONTECSI/PS-828
MEASURING TECHNOLOGICAL MATURITY LEVELS WITH DATA MINING

Rafael Herden Campos (Universidade do Vale do Rio dos Sinos, Rio Grande do Sul, Brasil) - rcamposrs@yahoo.com.br

Rocío Rocha (Universidad de badaantabria, Cantabria, Spain) - eliana.rocha@unican.es

Angel Cobo (Universidad de Cantabria, Cantabria, Spain) - angel.cobo@unican.es

Adolfo Alberto Vanti (Universidade do Vale do Rio dos Sinos, Rio Grande do Sul, Brasil) - avanti@unisinis.br

This paper analyzes maturity levels in sectorial technological processes using Data Mining (DM). As a methodological tool for data collection it was used COBIT 4.1 model with its 34 processes in 4 different domains. This model was used in the analysis of a sample of Brazilian companies located in southern Brazil. Clustering techniques, feature selection strategies and association rules were used. In particular, a good performance was obtained using the EM clustering algorithm and the Apriori association rule generator, both implemented in Weka software. The data mining approach provides an improvement in the risk analysis, security and governance in the companies.

Keywords: Cobit; Data Mining; Weka; EM Clustering.

MENSURAÇÃO DE NÍVEIS DE MATURIDADE TECNOLÓGICA COM O USO DE DATA MINING

Este trabalho avalia diferentes níveis maturidade em processos tecnológicos setoriais usando técnica de Data Mining (DM). Como instrumento metodológico para a coleta de dados utilizou-se o modelo Cobit 4.1 a partir da análise de 34 processos em 4 diferentes domínios junto à empresas brasileiras situadas no sul do Brasil. Como análise de dados foram utilizadas técnicas de clustering, seleção de atributos e regras de associação. Os resultados encontrados correspondem a uma melhor performance dos algoritmos EM e Apriori o que proporciona uma melhoria na análise de riscos da empresa, de segurança e de governança entre outros aspectos.

Palavras-chave: Cobit; Data Mining; Weka; EM Clustering.

1. INTRODUÇÃO

Uma das necessidades básicas de qualquer negócio é entender o estado de seus próprios sistemas de TI e decidir qual o nível de gestão e controle deve ser mantido. A obtenção de uma visão objetiva do nível de desempenho em uma empresa depende do de sua mensuração. As empresas devem medir onde estão e onde se faz necessário melhorias, implementando um conjunto de ferramentas de gestão para monitorar essas melhorias.

A avaliação da capacidade e dos processos tecnológicos baseados em modelos de maturidade pode ser realizada via o Cobit da ISACA e com isso, pode-se avaliar também de maneira setorial entre as diversas empresas como o abordado neste estudo.

Os gestores corporativos e de empresas públicas mais frequentemente são questionados o quão bem estão gerenciados os processos de TI. Portanto, se deve desenvolver um plano de negócios para melhorar e alcançar o nível adequado de gestão e controle sobre a infraestrutura de informações. Deve-se identificar o que é necessário para se alcançar um nível adequado de gestão e controle em processos de TI e fazê-lo de forma eficiente, sendo que uma das formas é o modelo internacional Cobit consolidado em sua versão 4.1 que aprimora a governança de TI das empresas.

A questão de pesquisa deste trabalho foi definida em Qual a avaliação dos níveis de maturidade de processos tecnológicos? Para responder a questão de pesquisa foi utilizada a técnica de DM aplicada em empresas localizadas no sul do Brasil. Com isso foi caracterizado o seguinte objetivo: Avaliar níveis de maturidade de processos tecnológicos com o uso da técnica de DM, o que também complementou com o objetivo específico de gerar novas regras de previsão para estudos e desenvolvimento de estratégias para melhoria dos processos de negócios.

Esse trabalho então revisa o conceito do Cobit e de Data Mining e posteriormente detalha procedimentos metodológicos de coleta de dados via Google Doc's em 78 empresas situadas no estado do Rio Grande do Sul/Brasil, bem como a forma que foi analisada os dados com a técnica de DM com suas respectivas regras e algoritmos. Finalmente apresenta os resultados encontrados e retorna à proposição e fundamentação para conclusão do trabalho.

2. COBIT

O *Control Objective for related Technology* (COBIT) é um modelo de governança de TI que auxilia na entrega de valor pela TI e busca identificar e gerir os riscos associados ao uso da tecnologia. O COBIT auxilia nas demandas do negócio, necessidade de controle e questões técnicas e busca garantir a integridade da informação e dos sistemas de informação (ITGI, 2007).

O modelo de maturidade Cobit para governança de TI é baseado em um método de avaliação em que a organização possa ser avaliada em seus processos tecnológicos entre os níveis "inexistente" (0) até "otimizado" (5). Esta abordagem é derivada do modelo de maturidade que o *Software Engineering Institute* que definiu para os níveis da capacidade de desenvolvimento de softwares.

Os níveis de maturidade são projetados como perfis de processos de TI nos quais a empresa reconheceria como possíveis estados atuais ou futuros. Eles não se aplicam como um modelo limitado onde não se pode mover para o próximo nível, sem ter cumprido todos os requisitos do nível anterior. Uma avaliação do nível de maturidade do COBIT resulta em um perfil das condições mais relevantes em diferentes níveis de maturidade.

COBIT é um framework ISACA e ITGI desenvolvido para o gerenciamento de processos de TI com um significativo enfoque de controle. Essas escalas possuem cunho prático e bem compreensível como uma forma de auditoria de sistemas. As avaliações podem ser realizadas tanto em relação às descrições no modelo de maturidade como um todo, ou mais rigorosamente, em cada uma dos processos individuais.

O *Control Objective for related Technology* (COBIT) é um modelo de governança de TI que auxilia na entrega de valor pela TI e busca identificar e gerir os riscos associados ao uso da tecnologia. O COBIT auxilia nas demandas do negócio, necessidade de controle e questões técnicas e busca garantir a integridade da informação e dos sistemas de informação.

A partir da análise do nível de maturidade, a organização pode estabelecer metas e objetivos para buscar a evolução em seus processos de tecnologia da informação. Para Rafeq (2010) a medição dos níveis de maturidade é necessária para identificar gaps em processos específicos, avaliar o estado atual do uso da TI na empresa e desenvolver planos de ação para alcançar os níveis desejados.

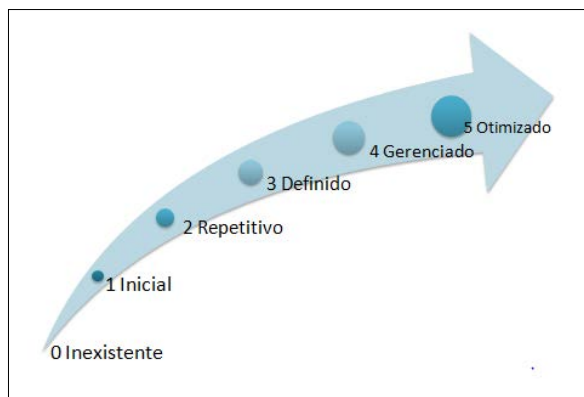
A vantagem de um modelo de maturidade é que torna relativamente fácil para a gestão de localizar-se na escala e avaliar o que deve ser feito quando se desenvolver uma melhoria. É um guia que permite que as organizações melhorarem seus processos e sua capacidade de se organizar, desenvolver, adquirir e manter produtos e serviços, em outras palavras, melhorar a governança de TI como um todo.

A escala inclui o nível inexistente (0), porque possui uma caracterização de não existirem processos formais no contexto organizacional avaliado. A escala de 0-5 é baseada em uma escala de maturidade mostrando como um processo evolui de uma capacidade inexistente até uma capacidade otimizada. No entanto, a capacidade de gestão de um processo não é o mesmo que o desempenho do mesmo.

A capacidade necessária determinada pelos objetivos de negócio e de TI pode não exigir o mesmo nível aplicado em todo o ambiente de TI, ou seja, aplicar de forma pouco consistente a apenas um número limitado de sistemas ou unidades. Nesse sentido, medir o desempenho é essencial para determinar o desempenho real da empresa em seus processos e alcançar inclusive uma mitigação de riscos.

Na Figura 1 são apresentados os níveis de maturidade do COBIT, distribuídos em uma escala de 6 diferentes níveis, tendo como limite inferior o nível Inexistente (0) e como superior os processos Otimizados (5).

Figura 1: Níveis de maturidade do Cobit.



Apesar de uma capacidade de aplicação adequada auxiliar na redução de riscos, a empresa deve analisar os controles necessários para garantir que o risco seja mitigado e que se obtenha o valor de acordo com os objetivos do negócio. Estas verificações são conduzidas pelos objetivos de controle do COBIT nos níveis de maturidade em neste trabalho foram analisados via DM junto a empresas situadas no sul do estado do RS. A versão do Cobit utilizada foi a 4.1 porque é a versão mais consolidada e bem testada internacionalmente do mesmo.

3. DATA MINING

A vasta quantidade de informações em bases de dados aliada à importância da extração de conhecimento útil como suporte à decisão têm exigido investimentos consideráveis das empresas e da comunidade científica (TAN; STEINBACH; KUMAR, 2009). Diante da geração de volumes significativos de dados, Witten e Frank (2005) afirmam que existem técnicas que tem se mostrado de muito proveitosas na obtenção de informações, as técnicas de Mineração de Dados, mais conhecidas por seu termo "Data Mining" (DM).

A mineração de dados compreende a avaliação automática ou semi-automática grandes volumes de dados possibilitando a previsão de conhecimento e comportamentos, identificando tendências e descobrindo relações entre outras vantagens para a tomada de decisão (WITTEN; FRANK, 2005). Esses grandes volumes de dados tendem ainda a serem estruturados em depósitos de dados denominados *Data Warehouse* (DW) (KIMBALL, 1998), (HARRISON, 1998) de onde são extraídos para as análises.

As técnicas de DM pertencem a um campo da ciência conhecido como Descoberta de Conhecimento em Base de Dados (DCBD) - *Knowledge Discovery in Database (KDD)* – que busca extrair com o uso de técnicas de análise inteligente e automática, conhecimentos úteis, regularidades, e informações de alto nível a partir dos conjuntos de dados relevantes em repositórios de dados, bem como sua visualização a partir de ângulos diferentes e assim possibilitando uma melhor visão do negócio (SENG; CHEN, 2010).

O Quadro 1 da continuação apresenta dois (2) diferentes e importantes técnicas de DM, sua utilização e exemplos de aplicação.

Quadro 1: Utilização e exemplos de técnicas de Data Mining

Algoritmo	Utilização	Exemplo
<i>Clustering</i>	Definição de uma métrica de similaridade para cada atributo e uma função de combinação destas métricas em uma métrica global. Os objetos são agrupados com base no princípio da maximização da similaridade intraclasse e da minimização da similaridade interclasse.	Identificar subgrupos homogêneos de clientes de uma determinada loja.
Associação	Quando a classe de uma tarefa de mineração não é determinada como no caso da classificação. O próprio algoritmo elege os atributos determinantes.	Clientes do sexo masculino, casados, com renda superior a R\$ 1.800,00 têm o seguinte hábito de consumo: roupas de grife, perfumes nacionais, relógios importados.

Fonte: Adaptado de Silva (2004).

Os algoritmos anteriormente apresentados no Quadro 1 evidenciam os tipos de algoritmos que serão utilizados neste trabalho junto a uma aplicação em 78 empresas localizadas no sul do Brasil, predominantemente dos setores de serviços, varejo e metal – mecânico, das cidades do Vale do Rio dos Sinos e de Porto Alegre. Quanto ao porte, a em sua maioria estão classificadas como de grande ou médio porte.

A mineração de dados contribui com métodos para melhorar o controle dos negócios, reduzindo o risco ao prever situações indesejáveis e fornecer recomendações com base em experiências anteriores (BAJO *et al.*, 2012). Conforme Shmueli e Koppius (2011), o poder preditivo (ou a precisão da previsão) se refere à capacidade de um modelo de gerar previsões de novas observações, em observações de um período futuro ou de observações que não foram incluídas na amostra original. No presente estudo, o algoritmo de Associação permitiu gerar regras de previsão de níveis de maturidades em processos, baseados na maturidade de outros processos, e assim proporcionar aos gestores desenvolverem estratégias para melhoria de seus processos.

A utilização de *Data Mining* nas organizações é tratada como uma inovação tecnológica devido a função de extrair conhecimento de grandes base de dados e de forma automatizada, agilizando o processo de transferência de conhecimento.

O avanço deste trabalho alcança a estruturação metodológica que enfatiza a técnica de coleta e análise de dados com captura de dados via Google Doc's recurso de Google Drive e análise com as regras e algoritmos do DM que na própria metodologia estão detalhados.

4. METODOLOGIA

Como abordagem metodológica este trabalho está contextualizada na Design Science (DS) (SIMON, 1996) e (ROMME e DAMEN, 2007) e (HEVNER e MARCH, 2004) dividindo-se em duas (2) partes: a coleta de dados e a análise de dados. Na coleta de dados utilizou-se um instrumento referenciado pelo Cobit 4.1 e estruturado em Google Doc's. Na parte de análise de dados utilizou-se então a técnica de *Data Mining* com os algoritmos de *clustering* EM, seleção de atributos (variados) e de associação Apriori .

Coleta de Dados

Para identificação dos níveis de maturidade dos processos, foi utilizado o método do COBIT 4.1, que se divide em 4 domínios, Planejar e Organizar, Adquirir e Implementar, Entregar e Suportar e por fim Monitorar e Avaliar. Os domínios estão divididos em 34 processos que foram avaliados através do uso de uma escala discreta com valores entre 0 e 5, onde 0 se refere a processo inexistentes nas organizações e 5, em que os níveis de maturidade dos processos esta otimizado, ou seja, a organização utiliza das melhores práticas com a aplicação eficiente de mensuração e controle dos processos.

Foram aplicados questionários aos colaboradores das empresas pesquisadas, sendo o mesmo dividido em duas partes, a primeira mais descritiva sobre a organização e o respondente e uma segunda parte que tratou dos processos de governança de TI. Nesta segunda parte, para os processos avaliados, os respondentes possuíam a disposição uma tabela explicativa de cada nível de maturidade segundo o modelo proveniente da *Software Engineering Institute* para níveis de maturidade.

Os questionários foram respondidos com o auxilio de ferramenta computacional de coleta de dados Google Doc's através da web, que posteriormente foi convertida em uma tabela do *software* Excel. Para tabulação dos dados utilizou-se o formulário evidenciado no Quadro 2, no qual é mostrado cada uma das variáveis coletadas.

Quadro 2: formulário de coleta de dados

Nome da Empresa: Empresa 1			
Setor: Financeiro			
Porte da empresa: Grande			
Cidade: Porto Alegre			
Cargo: Gerente			
Departamento: TI			
Data: 22/10/2013			
Processos do COBIT			
Processo	Descrição	Processo	Descrição
PO1	Define o planejamento estratégico de TI.	ES1	Define níveis e mantém os acordos de níveis de serviços.
PO2	Define a arquitetura da informação.	ES2	Gerencia os serviços de terceiros.
PO3	Determina as Diretrizes da Tecnologia.	ES3	Gerenciar desempenho e capacidade da TI.
PO4	Define a organização de TI e seus relacionamentos.	ES4	Garante a continuidade dos serviços.
PO5	Gerencia o Investimento de TI.	ES5	Garante a segurança dos sistemas.
PO6	Comunica as metas e diretrizes gerenciais.	ES6	Identifica e Aloca Custos
PO7	Gerencia os recursos humanos de TI.	ES7	Educa e treina os usuários.
PO8	Gerencia a qualidade.	ES8	Gerencia a central de serviços e incidentes.
PO9	Avalia e gerencia os riscos.	ES9	Gerencia a configuração.
PO10	Gerencia os projetos.	ES10	Gerencia os problemas.
AI1	Identifica soluções de automação.	ES11	Gerencia os dados.
AI2	Adquirir e Mantém Software Aplicativo.	ES12	Gerencia a infra-estrutura.
AI3	Adquire e mantém a arquitetura tecnológica.	ES13	Gerenciar operações.
AI4	Desenvolve e mantém procedimentos de TI.	MA1	MANitora e avalia a desempenho de TI.
AI5	Obtém recursos de TI.	MA2	MANitora e avalia o controle interno.
AI6	Gerenciar mudanças	MA3	Assegura a conformidade aos requisitos externos.
AI7	Instala e certifica soluções e mudanças.	MA4	Fornecer governança de TI.

A coleta de dados contou com oitenta e quatro (84) empresas, destas seis (6) apresentaram respostas incompletas que foram eliminadas, resultando em setenta e oito (78) empresas válidas. As respostas válidas foram organizadas através de uma tabela bidimensional, onde as respostas são as linhas (instâncias) e os atributos (classes) são as colunas e importados no *software* Weka.

O software WEKA - *Waikato Environment for Knowledge Analysis* foi utilizado para processamento dos algoritmos de *Data Mining* e sua aplicação sobre o banco de dados gerados.

Análise de Dados

A análise dos dados foi dividida em duas partes: a primeira através de técnica de mineração de dados não-supervisionada tratando de buscar os clusters (agrupamentos) entre os dados em que cada cluster é formado pela maximização da busca de dados que possuem características semelhantes intra-grupo e minimização das características entre os diferentes grupos, encontrando ao final 4 diferentes Clusters EM (Expectation Maximization). Técnicas de clustering também foram utilizadas para identificar metáforas organizacionais em empresas brasileiras (ROCHA *et al*, 2013).

O Método clusters EM é um método probabilístico com uma técnica iterativa de cálculo de máxima aceitabilidade que se utiliza para encontrar uma estimativa do conjunto de parâmetros de problemas nos que existem certos dados ocultos (Sierra 2006: 150). Uma das vantagens do método de clusters EM é que ele identifica automaticamente o número de grupos mais apropriado. O algoritmo EM fixa a cada instancia uma distribuição de probabilidade de pertinência a cada cluster, de maneira que se pode decidir quantos

clusters criar a partir de uma validação cruzada. Utiliza o modelo Gaussiano finito combinado, assumindo que todos os atributos são variáveis aleatórias independentes.

A segunda parte trata da busca de regras de associação entre as variáveis, ou seja, a influência de uma variável sobre a outra. Para o desenvolvimento desta segunda parte foram selecionados os atributos que possuíam maior poder discriminante em relação aos demais. No método de avaliação que determina a qualidade do conjunto dos atributos discriminantes, podemos distinguir duas categorias de métodos de avaliação, na primeira utiliza-se diretamente um classificador específico para medir a qualidade de cada subconjunto, gerando um alto custo computacional.

O segundo método não utiliza um classificador específico, por exemplo, o método *CfsSubsetEval*, se baseia no cálculo da correlação de cada atributo com a classe, eliminando atributos que têm uma correlação muito elevada tanto quanto atributos redundantes. Com a utilização deste classificador, se faz necessário à determinação de como será realizada a busca dos conjuntos.

Uma avaliação abrangente de todos os subconjuntos possíveis, em um alto o número de atributos, torna-se um problema combinatório intratável e por isso se seleciona algoritmos que tratam de realizar buscas de forma otimizada. Aplicaram-se métodos de buscas com diversos algoritmos com o objetivo de selecionar os atributos discriminantes.

Depois de selecionados os principais atributos foi aplicado o algoritmo de associação *Apriori*, que visa encontrar relações entre as diferentes variáveis. São algoritmos não-supervisionados, ou seja, não é conhecido a priori as relações entre as variáveis, nestas regras são avaliados o quão estatisticamente significativas são as relações entre as variáveis. Para aplicação do algoritmo, é necessário que os atributos sejam categóricos ou simbólicos e não numérico como no caso dos dados, portanto, se optou por discretizar os dados (transformar em categorias) em três diferentes níveis de maturidade: Baixo, médio e alto e aplicar o algoritmo de regras de associação.

No algoritmo *Apriori*, as regras de associação são avaliadas segundo duas métricas, confiança e suporte, também conhecido como cobertura, que indica o número de instâncias que a regra prediz corretamente, e confiança que indica a porcentagem de vezes que a regra foi cumprida.

O algoritmo pode ser configurado com diversas opções: a opção *UpperBoundSupport* indica o limite superior da cobertura necessária para aceitar um conjunto de regras. Se não é encontrado conjuntos suficientes que atendem as normas exigidas, vai sendo reduzido até o limite inferior (opção *LowerBoundSupport*). A opção *minmetric* indica confiança mínima requerida para apresentar uma regra e opção *numRules* indica o número máximo de regras a serem geradas.

Ao aplicar o algoritmo, foram consideradas as seguintes condições: (a) Suporte mínimo de 25% dos registros, ou seja, para considerar uma regra como válida é necessário que seja

observado em pelo menos um quarto das 78 empresas da amostra; (b) Confiança mínima de 0,9, ou seja, se admite um erro de 10% na verificação da regra.

5. APLICAÇÃO DO DATA MINING

A aplicação de Data Mining e de sua respectiva análise de dados considerou conforme Quadro 3 os 34 processos do Cobit, em uma escala de Inexistente (0) a otimizado (5). O algoritmo *EM* identificou automaticamente 4 grupos de empresas e para a análise de *cluster* as variáveis são do tipo nominal-numérica e representam a média para cada um dos processos.

Quadro 3: Resultado dos clusters

		Clusters			
		0 24%	1 35%	2 24%	3 17%
Código de processo	Nome do processo	Valores médios para cada processo			
PO1	Define o planejamento estratégico de TI.	3,7369	1,80960	2,68620	0,61540
PO2	Define a arquitetura da informação,	3,7895	1,33280	2,84320	0,61540
PO3	Determina as Diretrizes da Tecnologia,	4,3157	1,66520	3,25550	0,76920
PO4	Define a organização de TI e seus relacionamentos,	3,9999	1,29320	2,69100	0,76920
PO5	Gerencia o Investimento de TI,	4,3685	2,44640	3,46450	0,76920
PO6	Comunica as metas e diretrizes gerenciais,	3,7895	1,32600	2,68620	0,00000
PO7	Gerencia os recursos humanos de TI,	3,6317	0,73070	1,43100	0,07690
PO8	Gerencia a qualidade,	4,0527	1,44200	2,99380	1,07690
PO9	Avalia e gerencia os riscos,	4,2633	1,40920	3,03990	0,53850
PO10	Gerencia os projetos,	4,2632	1,44200	3,20310	0,46150
AI1	Identifica soluções de automação,	4,3684	1,70240	3,51700	1,46150
AI2	Adquiri e Mantém Software Aplicativo,	3,6843	2,29320	3,47070	0,69230
AI3	Adquire e mantém a arquitetura tecnológica,	4,0525	1,73960	3,20320	0,76920

AI4	Desenvolve e mantém procedimentos de TI,	3,9473	1,33040	2,05260	0,30770
AI5	Obtém recursos de TI,	4,2631	2,10720	3,10480	0,69230
AI6	Gerenciar mudanças	3,8420	1,92560	3,15090	0,38460
AI7	Instala e certifica soluções e mudanças,	4,2105	1,59530	2,77840	0,38460
DS1	Define níveis e mantém os acordos de níveis de serviços,	4,0000	1,95840	3,00010	0,38460
DS2	Gerencia os serviços de terceiros,	4,5264	2,40480	3,57540	0,92310
DS3	Gerenciar desempenho e capacidade da TI,	4,0002	1,73960	3,25520	0,38460
DS4	Garante a continuidade dos serviços,	4,5838	2,66960	3,56920	1,07690
DS5	Garante a segurança dos sistemas,	4,6315	2,92560	3,83700	0,84620
DS6	Identifica e Aloca Custos	4,0527	2,10720	2,89540	0,46150
DS7	Educa e treina os usuários,	3,7895	2,03720	2,00020	0,23080
DS8	Gerencia a central de serviços e incidentes,	4,4210	1,51200	2,84330	0,46150
DS9	Gerencia a configuração,	4,2106	1,55370	3,15070	0,53850
DS10	Gerencia os problemas,	4,4210	1,88840	2,68020	0,30770
DS11	Gerencia os dados,	4,2631	1,89280	3,06800	0,53850
DS12	Gerencia a infra-estrutura,	3,9475	1,80960	2,84310	0,38460
DS13	Gerenciar operações,	4,3685	2,03280	3,20930	0,53850
MO1	Monitora e avalia a desempenho de TI,	4,3686	1,58640	2,63380	0,07690
MO2	Monitora e avalia o controle interno,	4,3685	1,44200	2,88920	0,69230
MO3	Assegura a conformidade aos requisitos externos,	4,3685	2,24730	3,41850	0,53850
MO4	Fornecer governança de TI.	4,2107	1,58640	3,05220	0,46150

Fonte: Dados da pesquisa.

Se pode verificar que nos clusters encontrados, presentes no quadro 3, as empresas foram agrupados pelo nível de maturidade em seus processos. Possuindo as empresas

presentes no *cluster* 3 processos com nível de maturidade mais baixos e no *cluster* 0 processos com o nível mais alto.

O percentual de distribuição das empresas esta relativamente equilibrado, permanecendo o cluster 1 (35%) um pouco acima da média e o cluster 3 (17%) um pouco abaixo a média, mas não houve muita discrepância na alocação dos resultados bem como a distribuição das empresas.

Nota-se que entre os processos que atingiram o maior nível de maturidade entre os clusters forma: DS5 - "Garante a segurança dos sistemas". Também se destacam os processos DS2 - "Gerencia os serviços de terceiros" e o DS4 - "Garante a continuidade dos serviços".

Em contraponto, o processo PO7 - Gerencia os recursos humanos de TI possui médias muito baixas em todos os clusters, inclusive em 3 deles como o mais baixo (0, 1 e 2). Os processos PO6- "Comunica as metas e diretrizes gerenciais" e DS7 - "Educa e treina os usuários" também se apresentam entre as médias mais baixas. Se constata a preocupação em manter a segurança dos sistemas, tanto internamente quanto externamente e sua continuidade. A partir destas constatações, depreende-se que os aspectos técnicos ligados à tecnologia são priorizados em detrimento aos funcionais de recursos humanos, podendo haver vulnerabilidades e gerando riscos nos processos com níveis mais baixos de maturidade quando o enfoque é mais gerencial.

Na continuação é apresentado o Quadro 4 que contempla os Clusters, os números de empresas assinaladas em cada cluster, seu percentual, os valores mínimo e máximo encontrados nos clusters e uma descrição das principais características das empresas presentes nos referidos clusters.

Quadro 4: Descrição das principais características dos clusters.

Cluster	Número de instancias (empresas)	%	Valores médios	Valor mínimo do cluster	Valor máximo do cluster	Principais características das empresas...
0	19	24%	4,1503	3,6317	4,6315	Processos definidos, bem administrados e otimizados
1	27	35%	1,7936	0,7307	2,9256	Os processos não estão bem gerenciados, níveis iniciais, processos repetitivos e um pouco definidos
2	19	24%	2,9851	1,431	3,837	Os processos não estão bem gerenciados,

						processos repetitivos e alguns estão definidos
3	13	17%	0,5656	0	1,4615	Os processos não estão bem gerenciados, inexistentes e em níveis iniciais

Fonte: Dados da pesquisa.

O Quadro 4 está analisado e caracterizando prioritariamente através dos aspectos dos processos e seus respectivos níveis de maturidade. Nota-se que 52% das empresas possuem níveis de maturidade entre inexistentes e repetitivo (cluster 1 e 3), ou seja, possuem seus processos pouco gerenciados e em níveis iniciais. Embora o intervalo entre os níveis dos processos seja relativamente grande (*cluster 1*: diferença de 2,1949 e *cluster 3*: diferença de 1,4615), as médias demonstram que a maioria dos processos de cada *cluster* esta presentes nos níveis mais baixos dentro de seu próprio cluster. Apenas as empresas do *cluster 0*, que representa 24%, possuem seus processos bem gerenciados. Qualitativamente, se verifica que estas empresas são em sua maioria de grande porte.

A partir dos *clusters* encontrados, se destaca em vermelho o *cluster 0*, cuja diferença entre os processos com níveis mais altos e mais baixos possuem no máximo um nível, enquanto que nos demais apresentam maior variabilidade. Isso indica um maior equilíbrio de investimento e atenção nas organizações com maiores níveis de maturidade, enquanto nas de níveis menores, a maior variabilidade indica que existe uma maior preocupação de alguns processos em detrimento de outros.

Na continuação no Gráfico 1 pode ser verificada a quantidade de empresas atribuídas aos *clusters* obtidos pelo Weka, correspondentes aos 34 procesos do COBIT.

Gráfico 1: Atribuição das empresas aos clusters.

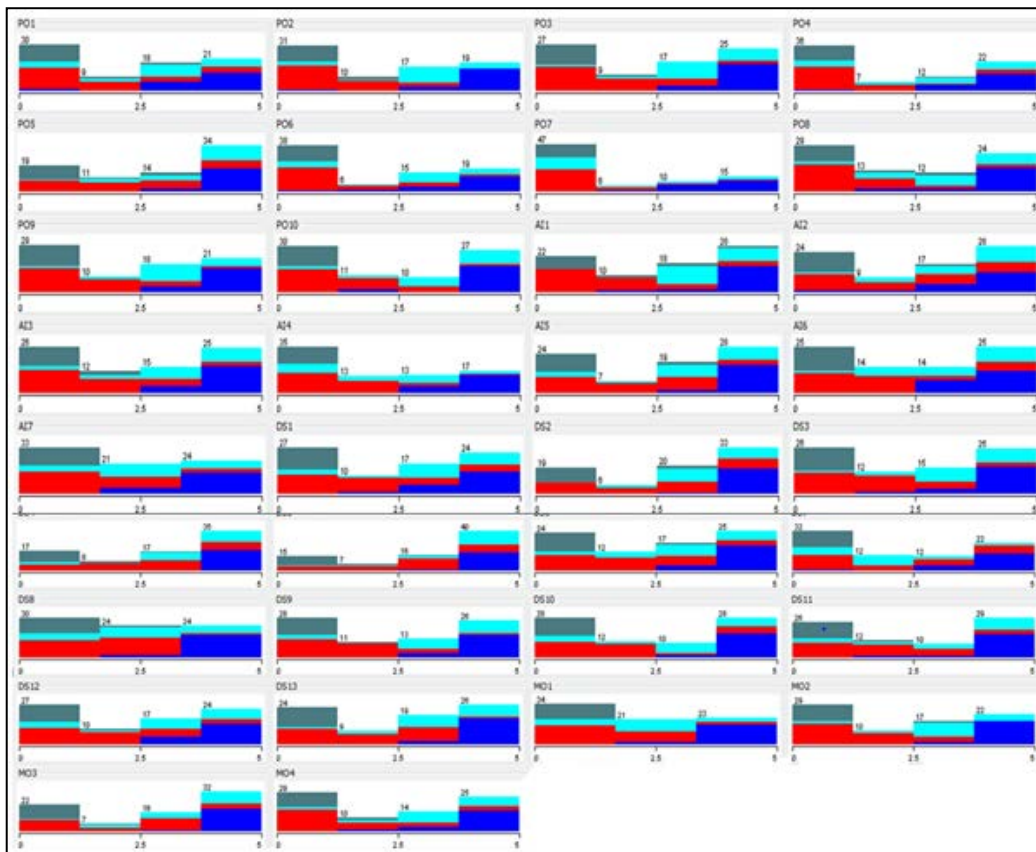


Fonte: Dados da pesquisa

No Gráfico 1 nota-se a distribuição dos clusters por número de empresas, e que os se encontram relativamente equilibrados, tendo o cluster 3 um menor número de empresas assinaladas. Este cluster é composto por empresas de menor porte ao passo que também possuem níveis mais iniciais de maturidade em seus processos.

No Gráfico 2 são apresentados os histogramas para cada processo do Cobit e a distribuição dos níveis de maturidade em relação aos clusters. As cores representam os diferentes cluster, anteriormente apresentados no Gráfico 1.

Gráfico 2: Histogramas gerados pelo Weka correspondentes aos 34 processos do COBIT.



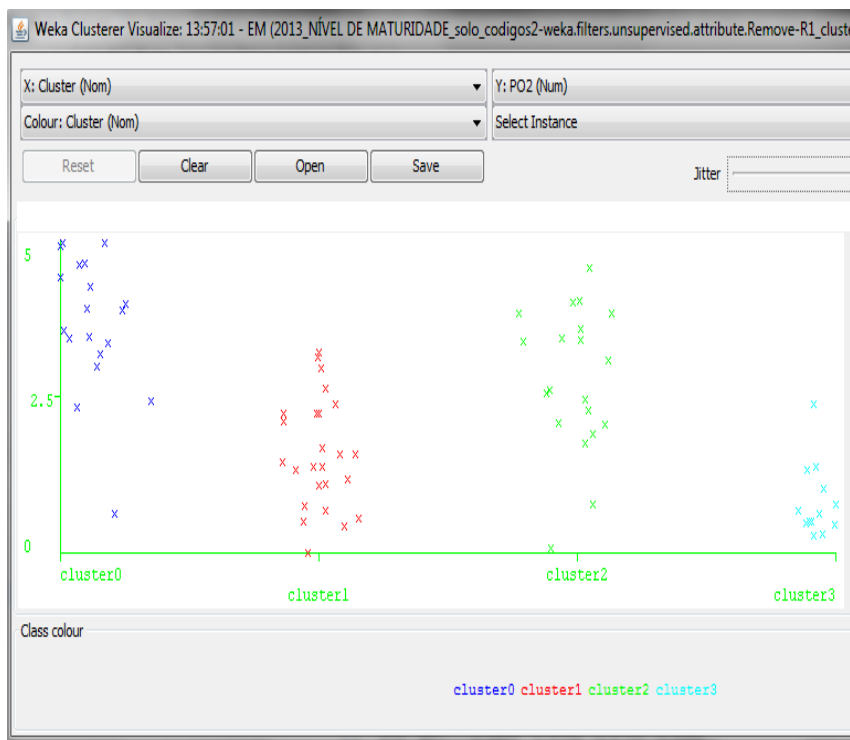
Fonte: Dados da pesquisa.

Nos resultados apresentados no Gráfico 2, ressalta-se que em sua maioria os níveis de maturidade dos processos se encontram mais presentes ou em níveis altos ou baixos, havendo uma menor distribuição nos níveis intermediários de maturidade.

Também nota-se que empresas do cluster 0 (azul) possuem processo com maior nível de maturidade e as empresas dos cluster 1 (vermelho) possuem seus processos mais iniciais, com alguns repetitivos e definidos. As empresas do cluster 2 (verde claro) possuem uma variação maior entre os processo, as vezes distribuídas mais equilibradamente (DS8, MO1), por vezes em algum ponto mais extremo (AI1). Já as presentes no cluster 3 (verde escuro) possuem seus processos fortemente indicados nos níveis iniciais.

No Gráfico 3, são apresentados os *clusters* em relação ao processo PO2 “Define a arquitetura da informação”.

Gráfico 3: Atribuição das 78 empresas a cada cluster com seus respectivos graus de maturidade.



Fonte: Dados da pesquisa.

É apresentado (Gráfico 3) somente um dos processos que é denominado PO2 “Define a arquitetura da informação”, no entanto o comportamento dos clusters (eixo x) em relação aos níveis de maturidade (eixo Y) se comportam semelhantemente nos demais processos. Nota-se uma maior maturidade no cluster 0 (azul) e a menor maturidade entre os agrupamentos no cluster 3. Verifica-se uma maior dispersão dos dados no cluster 2.

Na continuação é apresentada os resultados da Seleção de Atributos realizada para que se chegar às Regras de Associação.

Seleção de atributos

Dos 34 atributos (processos) não são todos igualmente válidos para predizer o nível de maturidade de uma organização. Neste trabalho foi utilizada a estratégia de seleção automática de atributos para identificar as questões com maior poder discriminante. O Weka incorpora ampla variedade de técnicas de seleção de atributos que tratam de explorar quais os subconjuntos de atributos melhor classificam as instâncias. Esta seleção de atributos tem dois componentes principais: um método de avaliação que determina a qualidade do conjunto de atributos que discriminam as classes e um método de pesquisa que determina a forma de busca dos conjuntos.

Na Tabela 1 pode ser observado todos os subconjuntos de atributos obtidos pelo Weka usando o método de avaliação e *CfsSubsetEval* e diferentes métodos de busca. Em todos os

casos o atributo que define a classe é o atributo *cluster* obtido a partir da análise de clusters das empresas. Como pode se verificar, os subconjuntos obtidos não são iguais, mas são muito semelhantes, e na última linha da tabela, estão incluídos todos os atributos presentes nos métodos de pesquisa.

Tabela 1. Subconjuntos de atributos selecionados com o avaliador de atributos CfsSubsetEval

Método de Busca	Nº atributos	Atributos
Best first	23	PO2,PO3,PO5,PO6,PO7,PO8,PO9,PO10, AI1,AI4,AI6,DS1,DS2,DS3,DS6,DS7,DS8,DS9,DS10,DS13,MO1,MO2,MO3
SubsetSizeForwardSelection	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
GeneticSearch	24	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI5, AI6, DS1, DS2, DS4, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
GreedyStepwise	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
LinearForwardSelection	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
RankSearch	34	PO1, PO2, PO3, PO4, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI2, AI3, AI4, AI5, AI6, AI7, DS1, DS2, DS3, DS4, DS5, DS6, DS7, DS8, DS9, DS10, DS11, DS12, DS13, MO1, MO2, MO3, MO4
ScatterSearchV1	23	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, AI6, DS1, DS2, DS3, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3

Atributos presentes em 21 todos os métodos de busca	PO2, PO3, PO5, PO6, PO7, PO8, PO9, PO10, AI1, AI4, DS1, DS2, DS6, DS7, DS8, DS9, DS10, DS13, MO1, MO2, MO3
--	---

Fonte: Dados da pesquisa.

Na tabela 1, são visualizados os algoritmos de busca, a quantidade de atributos encontrados e quais atributos. Entre os algoritmos, nota-se que o *RankSearch* selecionou todos os atributos, não encontrando os que possuem maior capacidade discriminante. Na maior parte dos resultados foram encontrados 23 atributos e um apresentou 24 atributos. A partir dos resultados obtidos, 21 atributos estavam presentes nos diferentes algoritmos de busca e estes foram selecionados para a busca de regras de associação, apresentadas na sequência.

Regras de associação: Algoritmo *Apriori*

Nesta análise, se buscou encontrar regras que relacionem os níveis de maturidade dos processos do COBIT com os 21 atributos previamente selecionados, a fim de identificar se o nível de um processo possui influência sobre os níveis de outros processos.

Para realizar a aplicação do algoritmo de regras de associação *Apriori*, os dados não podem ser numéricos e por isso se optou por discretizar os níveis de maturidade em três intervalos, de acordo com o Quadro 5:

Quadro 5: Intervalos de discretização

No existente	0	Baixo
Inicial	1	
Repetitivo	2	Médio
Definido	3	
Gerenciado	4	Alto
Otimizado	5	

Nestas condições, o algoritmo gerou 306 regras e foram eliminadas as redundantes ou que possuíam muitos atributos. As regras obtidas estão listadas na sequência

No quadro 6 é apresentada as regras encontradas, seu suporte (mínimo de 25%) e confiança (mínima de 0,90), o que significa que para a regra ser considerada válida, deve estar presente em pelo menos um quarto das observações e se admite um erro de 10% na verificação da regra.

Quadro 6: Regras encontradas

Num. regra	Se	Então	Suporte	Confiança
-------------------	-----------	--------------	----------------	------------------

1	PO10=Baixo	PO7=Baixo	30	1,00
2	AI4=Baixo	PO7=Baixo	34	0,97
3	DS13=Baixo	MO1=Baixo	23	0,96
4	MO2=Alto	MO3=Alto	21	0,95
5	PO6=Baixo	PO7=Baixo	36	0,95
6	MO1=Baixo	PO7=Baixo	32	0,94
7	PO2=Baixo	PO7=Baixo	29	0,94
8	DS8=Baixo	PO7=Baixo	28	0,93
9	PO8=Baixo	PO7=Baixo	27	0,93
10	MO2=Baixo	PO7=Baixo	27	0,93
11	PO3=Baixo	PO7=Baixo	25	0,93
12	DS6=Baixo	PO7=Baixo	22	0,92
13	DS13=Baixo	PO7=Baixo	22	0,92
14	DS13=Baixo	PO7=Baixo MO1=Baixo	22	0,92
15	MO1=Alto	PO5=Alto	21	0,91
16	MO1=Alto	DS2=Alto	21	0,91
17	AI1=Baixo	PO7=Baixo	20	0,91
18	MO3=Baixo	PO7=Baixo	20	0,91
19	MO3=Baixo	MO2=Baixo	20	0,91
20	PO10=Baixo	PO6=Baixo	27	0,90

Fonte: Dados da pesquisa.

Como pode ser visualizado no Quadro 6, a maior parte das regras, 17 entre as 20 presentes, são atribuídas à níveis baixos de maturidade. Três foram as regras atribuídas a níveis altos e nenhuma a níveis médios.

É importante ter em conta que a ordem da regra gerada deve ser repetida, pois somente se o primeiro atributo se confirmar o segundo trará seu respectivo resultado. Isto pode ser visto na regra 12, caso DS6 for baixo, PO7 também será baixo, mas o contrario não é necessariamente verdadeiro, ou seja, caso o PO7 for baixo, isso não significa que o DS6 será baixo.

Muitas das regras provem baixos níveis de maturidade no processo PO7 (Gerenciar recursos humanos de TI), quando os seguintes processos estão em estados iniciais:

- Processos do domínio PO: PO2, PO3, PO6, PO8, PO10;
- Processos do domínio DS: DS6, DS8, DS13;
- Processos do domínio AI: AI1 AI4;
- Processos do domínio MO: MO1, MO2, MO3.

A vinculação das regras do processo PO7 com outros processos do mesmo domínio são evidentes, quando ocorre o nível inadequado ou baixo de maturidade em gerenciamento de projetos (PO10) ocorre que a gestão dos recursos de TI humanos (PO7) é avaliada em um nível baixo, que significa que não esta sendo realizada adequadamente. Também se obteve

regra que é mais complexas em virtude de possuírem mais atributos envolvidos, que pode ser visto no Quadro 7.

Quadro 7: Regra com estrutura mais complexa

Se	Então	Support	Confiança
PO7=Baixo DS6=Baixo	PO6=Baixo	21	0,95

Fonte: Dados da pesquisa.

Na regra presente no Quadro 7 pode se observar que empresas que têm os processos PO7 e DS6 com valores de maturidade baixos, os processos de PO6 também são avaliados com baixos níveis de maturidade, estando estes três processos relacionados.

6. CONSIDERAÇÕES FINAIS

O DM no Cobit permitiu avaliar níveis de maturidade de processos tecnológicos de diferentes empresas e gerar regras de previsão de níveis de maturidades em novos processos, utilizando-se algoritmos de *clustering EM* e de associação *Apriori*. Essa nova previsibilidade é recurso robusto do DM e com isso permite aos gestores desenvolverem estratégias para melhoria de seus processos organizacionais e de tomada de decisão.

A análise dos dados avançou por uma técnica de mineração de dados não-supervisionada tratando de buscar os clusters (agrupamentos) entre os dados e minimização das características entre os diferentes grupos, encontrando ao final 4 diferentes clusters. Posteriormente foi realizada a busca de regras de associação entre as variáveis, permitindo assim compreender a influência de uma variável sobre a outra com a seleção de atributos que possuíam maior poder discriminante em relação aos demais.

A maior parte das regras (17 entre as 20 encontradas) foram atribuídas a níveis baixos de maturidade, sendo que três (3) foram regras atribuídas a níveis altos e nenhuma a níveis médios. Pôde-se com isso confirmar que a regra 12 com um DS6 baixo impacta num PO7 também baixo. Estas associações permitem também melhorar consequentemente a governança de TI tornando a empresa mais transparente, apresentando menores riscos e direcionando melhor seus investimentos tecnológicos.

A aplicação em empresas situadas no estado do RS pode ser expandida a outros estados brasileiros, bem como desenvolver estudos comparativos com outros países já que o Cobit possui uma caracterização internacional.

REFERÊNCIA BIBLIOGRÁFICA

BAJO, J. BORRAJO, M.L.; PAZ, J.F.D.; CORCHADO, J.M.; PELLICER, M.A. A multi-agent system for web-based risk management in small and medium business. **Expert Systems with Applications**, v. 39, n. 8, p. 6921–6931, jun. 2012.

HARRISON, T. **Intranet Data Warehouse**. São Paulo: Berkeley, 1998.

HEVNER, A.; MARCH, S. Design Science in information systems research. **MIS Quarterly**. Vol.28, no.1, pp. 75-105, march, 2004.

ITGI: Information Technology Governance Institute. CobiT 4.1 modelo, objetivos de controle, diretrizes de gerenciamento e modelos de maturidade, 2007. Disponível em: <<http://www.isaca.org/Knowledge-Center/cobit/Documents/cobit41-portuguese.pdf>>. Acesso em: 28 fev. 2012.

KIMBALL, R. – **The Data Warehouse Toolkit – Técnicas para a Construção de Data Warehouses Dimensionais**, Makron Books, São Paulo, 1998.

ROCHA, R. ; COBO, A.; JOHANN, S. ; VANTI, A. Imágenes organizacionales de las empresas brasileñas: detección y análisis con técnicas de minería de datos. **Revista de Ciências da Administração**, v. 15, p. 105-120, 2013.

RAFEQ, A. Using COBIT for Assessing IT Process Maturity: A Case Study. **COBIT Focus**, v.4, october, 2010.

ROMME, A. G. L.; DAMEN, I. C. M. Toward Science-Based Design in Organization Development: Codifying the Process. **The Journal of Applied Behavioral Science**, v. 43, n. 1, p. 108-121, march 2007.

SENG, J.L.; CHEN, T.C. An analytic approach to select data mining for business decision. **Expert Systems with Applications**, v.37, n.12, p.8042-8057, 2010.

SHMUELI, G.; KOPPIUS, O.R. Predictive analytics in information systems research. **MIS Quarterly**, v.35, n.3, 2011.

SIERRA, B. **Aprendizaje Automático: Conceptos básicos y avanzados**. Madrid: Pearson Prentice Hall. 2006.

SILVA, M.S. Mineração de dados - conceitos, aplicações e experimentos com Weka. 2004. Disponível em: <http://bibliotecadigital.sbc.org.br/?module=Public&action=PublicationObject&subject=154&publicationobjectid=9>. Acesso em 15 de setembro de 2012.

SIMON, H. A. **The Sciences of the Artificial**. USA: MIT Press, 1996.

TAN, P; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. USA: Addison-Wesley Longman Publishing, 2009.

WITTEN, I; FRANK, E. **Data mining: practical machine learning tools and techniques**. New Zealand: Morgan Kaufmann Publishers, 2005.