



UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

Programa de Pós-Graduação em Ciência da Computação

LEANDRO BODO

**APRENDIZAGEM DE MÁQUINA PARA
ANÁLISE DE INDICADORES EM
PROCESSOS DE SOFTWARE**

Rio Claro - SP

2015

LEANDRO BODO

APRENDIZAGEM DE MÁQUINA PARA ANÁLISE DE INDICADORES EM PROCESSOS DE SOFTWARE

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (multicâmpus), área de concentração em Computação Aplicada, linha de Pesquisa em Engenharia de Software e Banco de Dados, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, para obtenção do título de Mestre.

Orientadora: Profa. Dra. Hilda Carvalho de Oliveira

Coorientador: Prof. Dr. Fabricio Aparecido Breve

Rio Claro - SP

2015

Bodo, Leandro.

Aprendizagem de máquina para análise de indicadores em processos de software / Leandro Bodo. – São José do Rio Preto, 2015.

120 f. : il., tabs.

Orientador: Hilda Carvalho de Oliveira

Coorientador: Fabricio Aparecido Breve

Dissertação (mestrado) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas

1. Computação. 2. Engenharia de software. 3. Engenharia de requisitos. 4. Software - Controle de qualidade. 5. Desempenho - Medição. 6. Aprendizado do computador. I. Oliveira, Hilda Carvalho de. II. Breve, Fabricio Aparecido. III. Universidade Estadual Paulista “Júlio de Mesquita Filho”. Instituto de Biociências, Letras e Ciências Exatas. IV. Título.

CDU – 518.72

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Campus de São José do Rio Preto

LEANDRO BODO

APRENDIZAGEM DE MÁQUINA PARA ANÁLISE DE INDICADORES EM PROCESSOS DE SOFTWARE

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (multicâmpus), área de concentração em Computação Aplicada, linha de Pesquisa em Engenharia de Software e Banco de Dados, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, para obtenção do título de Mestre.

Banca Examinadora

Profa. Dra. Hilda Carvalho de Oliveira
UNESP – Rio Claro
Orientadora

Prof. Dr. João Paulo Papa
UNESP – Bauru

Prof. Dr. Fabiano Cutigi Ferrari
UFSCar – São Carlos

Rio Claro - SP

2015

Agradecimentos

Primeiramente a Deus, pela saúde, paciência e força para me ajudar a atingir meus objetivos.

À minha esposa, Cibelle Bodo, pelo apoio, carinho, compreensão, paciência e amor nos momentos de dificuldade durante esse processo de estudo, pesquisa e aprendizado.

Ao meu pai, mãe e irmã, que sempre me apoiaram e compreenderam nossos sacrifícios e minha ausência em muitos momentos importantes.

À Profa Dra. Hilda, orientadora, pela generosidade, paciência, empenho, ajuda, esforço e dedicação extrema aos cumprimentos de nossas metas, além do apoio fundamental à ideia do projeto.

Ao Prof. Dr. Fabrício, coorientador, por todo apoio, e, principalmente, pelas aulas de Computação Inspirada pela Natureza, que ajudaram na inspiração do projeto.

Ao Prof. Dr. Danilo Medeiros Eler, pelo auxílio e contribuição sobre Visualização da Informação, enriquecendo ainda mais este trabalho.

Ao meu grande irmão, Prof. Ms. Clayton Reginaldo Pereira, pela amizade, dicas e pelo suporte provido ao longo de todos esses anos. Compartilhamos momentos bons e superamos os momentos ruins juntos, obrigado.

Ao meu amigo Prof. Ms. Alessandro Viola Pizzoleto, pelas dicas, sugestões e apoio durante o processo de Mestrado.

Ao meu amigo e Prof. Dr. João Paulo Papa, por acreditar em mim no início de tudo como um aspirante a pesquisador, esclarecendo pontos importantes relacionados ao Mestrado e a pesquisa e não medindo esforços em ajudar em qualquer momento.

À empresa SMARAPD Informática Ltda. e, principalmente, aos senhores Antônio Puças Júnior e Emerson Ricardo Cantolini, pelo tempo dispensado e pelo grande apoio às discussões realizadas, fundamentais para a execução desse trabalho.

“Stay hungry. Stay foolish”.
(Steve Jobs)

RESUMO

As organizações produtoras de software vêm enfrentando problemas e desafios em relação à qualidade do software há muitas décadas. O gerenciamento da qualidade envolve três processos básicos: planejamento da qualidade, garantia da qualidade e controle da qualidade. O processo de controle da qualidade provê informações para avaliação do desempenho e mudanças nos projetos, processos ou produtos. Para isso, indicadores de desempenho devem ser definidos e analisados para apoiar as tomadas de decisão. Durante o monitoramento dos processos da produção de software, dados dos indicadores de desempenho são coletados e armazenados em bases históricas para serem analisados pelos administradores dos processos. Técnicas de controle estatístico da qualidade auxiliam a avaliação dos dados coletados. Contudo, alguns aspectos dificultam a análise adequada para as tomadas de decisão em tempo hábil. O controle da qualidade pode requerer análise de grupos de indicadores, compostos por indicadores de vários processos e de diferentes granularidades, tipos e frequências de coleta. Além disso, à medida que o volume de dados dos indicadores aumenta, a complexidade das análises também tende a aumentar. Nesse contexto, este trabalho apresenta uma sistemática para análise dos indicadores de desempenho, utilizando técnicas de aprendizado de máquina semi-supervisionado. Essa sistemática é composta por etapas que abrangem a seleção dos indicadores, o processo de rotulagem e a análise dos dados coletados no monitoramento. Adicionalmente, o trabalho apresenta um modelo de referência para apoiar a seleção dos indicadores, baseado nos processos dos níveis G e F do modelo de qualidade MPS-SW. Outro modelo de referência desenvolvido é baseado nas perspectivas do modelo *Balanced Scorecard*, com a finalidade de apoiar a definição de grupos de indicadores. Os processos de rotulagem e análise dos dados são feitos em um único passo. São usadas técnicas de visualização de informação para apoiar o processo de rotulagem. O trabalho apresenta um estudo de caso sobre a sistemática proposta, utilizando dados reais de uma empresa produtora de software. Alguns experimentos realizados com indicadores de desempenho também são apresentados, considerando as contribuições para os caminhos adotados neste trabalho.

Palavras-chave: Qualidade de software. Controle de qualidade. Indicadores de desempenho. Aprendizagem de máquina. Aprendizado semi-supervisionado. MPS-SW. BSC. Visualização de informação.

ABSTRACT

Software development companies have been facing problems and challenges in relation to software quality for decades. Quality management involves three basic processes: quality planning, quality assurance and quality control. The quality control process provides information to evaluate the performance and changes in projects, processes or products. For this, performance indicators should be defined and analyzed in order to help decision-making. During the monitoring of the software production processes, data of the performance indicators is collected and stored in historical bases in order to be analyzed by the managers of the processes. Statistical quality control techniques aid the evaluation of the collected data. However, some aspects complicate the appropriate analysis to timely decision-making. The quality control may require analyzing groups of indicators, composed of indicators of various processes and with different granularity, types and collect frequency. Furthermore, as the volume of indicator data increases, the complexity of analysis also tends to increase. In this context, this work presents a systematic for analysis of performance indicators, using semi-supervised machine learning techniques. This systematic consists of steps that cover the selection of indicators, the process of labeling and the analysis of data collected in monitoring. In addition, the work presents a reference model to support the selection of indicators, considering the processes of the levels G and F of the MPS model for software (MPS-SW). The other developed reference model is based on the perspectives of the Balanced Scorecard model, in order to support the definition of groups of indicators. The labeling process and the data analysis process are done in a single step. Information visualization techniques are used to support the labeling process. The work presents a case study on the systematic presented, using real data from a software development company. Some experiments were performed with performance indicators and are also presented because of their contributions to the directions adopted in this work.

Keywords: *Quality software. Quality control. Performance indicators. Machine learning. Semi-supervised learning. MPS-SW. BSC. Information visualization.*

LISTA DE FIGURAS

	Página
FIGURA 1 - VISÃO GERAL DA SISTEMÁTICA DESENVOLVIDA E SUA APLICAÇÃO.	21
FIGURA 2 - NÍVEIS DE MEDIÇÕES DE INDICADORES.	26
FIGURA 3 - PERSPECTIVAS DO MODELO BSC.	31
FIGURA 4 - COMPONENTES DO PROGRAMA MPS.BR.	33
FIGURA 5 - EQUIVALÊNCIA ENTRE MR-MPS-SW E CMMI-DEV.	35
FIGURA 6 - TRECHO DO TEXTO DO GUIA DE IMPLEMENTAÇÃO DO NÍVEL G DO MR-MPS-SW.	37
FIGURA 7 - EXEMPLO DE ONTOLOGIA NO SISTEMA PROTÉGÉ.	38
FIGURA 8 - REPRESENTAÇÃO DE UM NEURÔNIO GENÉRICO.	48
FIGURA 9 - REPRESENTAÇÃO SIMPLIFICADA DE UMA REDE NEURAL ARTIFICIAL (RNA).	49
FIGURA 10 - METÁFORA DE PARTÍCULAS BASEADA NO COMPORTAMENTO DE FORMIGAS.	50
FIGURA 11 - PSEUDO-CÓDIGO: ALGORITMO DE COMPETIÇÃO E COOPERAÇÃO ENTRE PARTÍCULAS.	51
FIGURA 12 - NÍVEIS DE DOMÍNIOS INICIAIS: (A) AMOSTRA NÃO ROTULADA; (B) AMOSTRA ROTULADA.	51
FIGURA 13 - PROBABILIDADE DE ESCOLHA DA FORMIGA - MOVIMENTO GULOSO E ALEATÓRIO.	52
FIGURA 14 - AUMENTO DO NÍVEL DE DOMÍNIO DO PRÓPRIO TIME E DIMINUIÇÃO DO NÍVEL DOS OUTROS TIMES.	52
FIGURA 15 - FORTALECIMENTO DE NÍVEIS - NÓS DO MESMO TIME.	52
FIGURA 16 - ENFRAQUECIMENTO DE NÍVEIS - NÓS DE OUTROS TIMES.	53
FIGURA 17 - PROJEÇÃO LSP USADA PARA CLASSIFICAR ARTIGOS CIENTÍFICOS DE QUATRO É DIFERENTES ÁREAS.	55
FIGURA 18 - TÉCNICA CMDS USADA PARA IDENTIFICAR AS DISTÂNCIAS DAS RODOVIAS ENTRE AS PRINCIPAIS CIDADES DOS EUA.	55
FIGURA 19 - TÉCNICA DE COORDENADAS PARALELAS USADA COMO SUPORTE AO ENSINO A DISTÂNCIA.	56
FIGURA 20 - USO DE UMA RNA MLP PARA O FLUXO DE VEÍCULOS EQUIVALENTE.	58
FIGURA 21 - VISÃO PARCIAL DO MODELO DE REFERÊNCIA DE INDICADORES DESENVOLVIDO NA ETAPA 1.	63
FIGURA 22 - VISÃO PARCIAL DA ONTOLOGIA DE PIZZOLETO (2013) NO SISTEMA PROTÉGÉ.	64
FIGURA 23 - PROVÁVEL ERRO DE ROTULAGEM NA CLASSE “REGULAR” (AMARELA), USANDO A TÉCNICA LSP.	79
FIGURA 24 - PROVÁVEL ERRO DE ROTULAGEM NA CLASSE “REGULAR” (AMARELA), USANDO A TÉCNICA PC.	80
FIGURA 25 - PROVÁVEL ERRO DE ROTULAGEM NA CLASSE “REGULAR” (AMARELA), USANDO A TÉCNICA CMDS.	81

FIGURA 26 - APLICAÇÃO DA TÉCNICA LSP APÓS REVISÃO DOS RÓTULOS DAS AMOSTRAS A130, A147 E A186.....	81
FIGURA 27 - APLICAÇÃO DA TÉCNICA CMDS APÓS REVISÃO DOS RÓTULOS DAS AMOSTRAS A130, A147 E A186.	82
FIGURA 28 - APLICAÇÃO DA TÉCNICA PC NO CONJUNTO DE INSTÂNCIAS DA CLASSE "REGULAR" (AMARELA) APÓS REVISÃO DOS RÓTULOS.	82
FIGURA 29 - TAXA DE ACERTO EM GRUPO DE INDICADORES COM A TÉCNICA PCC.....	84
FIGURA 30 - COEFICIENTE KAPPA EM GRUPO DE INDICADORES COM O ALGORITMO PCC.....	85
FIGURA 31 - TEMPO DE EXECUÇÃO: ALGORITMO PCC VS HUMANO.	85
FIGURA 32 - TAXA DE ACERTO DOS ALGORITMOS MLP, KNN E PCC NO EXPERIMENTO I.	89
FIGURA 33 - TEMPO DE EXECUÇÃO DO ALGORITMO PCC VS SER HUMANO NO EXPERIMENTO I.	90
FIGURA 34 - TEMPO DE EXECUÇÃO DO TESTE: MLP VS HUM.	94
FIGURA 35 - PROJEÇÃO MULTIDIMENSIONAL DOS DADOS UTILIZANDO A TÉCNICA LSP NO EXPERIMENTO II - PARTE A.	95
FIGURA 36 - COORDENADAS PARALELAS DO CONJUNTO DE DADOS DO EXPERIMENTO II - PARTE A.	96
FIGURA 37 - COMPORTAMENTO DOS ATRIBUTOS DE ALGUMAS INSTÂNCIAS SELECIONADAS NA PROJEÇÃO LSP.....	96
FIGURA 38 - VISÃO GERAL DO MAPA MENTAL COMPLETO DOS INDICADORES (NÍVEIS G E F).....	110
FIGURA 39 - MAPA MENTAL COM OS INDICADORES PARA O PROCESSO DE MEDIÇÃO.....	111
FIGURA 40 - MAPA MENTAL COM OS INDICADORES PARA O PROCESSO DE GERÊNCIA DE REQUISITOS.....	112
FIGURA 41 - MAPA MENTAL COM OS INDICADORES PARA O PROCESSO DE GERÊNCIA DE REQUISITOS, DESTACANDO O PROCESSO GRE1.	112
FIGURA 42 - MAPA MENTAL COM OS INDICADORES PARA O PROCESSO DE GERÊNCIA DE PROJETOS (PARCIAL).	113
FIGURA 43 - MAPA MENTAL COM OS INDICADORES PARA O PROCESSO DE GERÊNCIA DE PROJETOS (CONTINUAÇÃO).	114
FIGURA 44 - MAPA MENTAL COM OS INDICADORES PARA O PROCESSO DE GARANTIA DA QUALIDADE.	115

LISTA DE TABELAS

	Página
TABELA 1 - PRINCIPAIS ATRIBUTOS DOS INDICADORES.	25
TABELA 2 - NÍVEIS DE MATURIDADE DO MODELO MPS-SW.....	34
TABELA 3 - PROCESSOS ADICIONADOS EM CADA NÍVEL DE MATURIDADE (NM) DO MPS-SW.	34
TABELA 4 - PROCESSOS DOS NÍVEIS G E F DO MODELO MPS-SW.	36
TABELA 5 - PERCENTUAL DE ACERTO DA RNA KOHONEN EM CADA CLUSTER.	59
TABELA 6 - INDICADORES AGRUPADOS SEGUNDO PERSPECTIVAS DO BSC COM TRÊS POSSÍVEIS SAÍDAS.	66
TABELA 7 - EXEMPLO DE INDICADORES ROTULADOS.....	70
TABELA 8 - INDICADORES SELECIONADOS A PARTIR DO MODELO DE REFERÊNCIA DA ETAPA 1 DA SISTEMÁTICA.....	76
TABELA 9 - DADOS ESTATÍSTICOS SOBRE A BASE DE DADOS UTILIZADA.	78
TABELA 10 - DADOS ESTATÍSTICOS SOBRE A BASE DE DADOS HISTÓRICOS UTILIZADA NO EXPERIMENTO I.	88
TABELA 11 - TAXA DE ACERTO POR CLASSE DE SAÍDA: REAL VS PREDITA (ALGORITMO PCC).	90
TABELA 12 - DADOS ESTATÍSTICOS SOBRE A BASE DE DADOS UTILIZADA NO EXPERIMENTO II.	92
TABELA 13 - TAXA DE ACERTO CLASSE DE SAÍDA: REAL VS PREDITA (ALGORITMO MLP).	94
TABELA 14 - TAXA DE ACERTO CLASSE DE SAÍDA: REAL VS PREDITA (ALGORITMO PCC).	97
TABELA 15 - COMPARATIVO DO COMPORTAMENTO DOS ALGORITMOS: PCC (PARTÍCULAS) VS MLP (PERCEPTRON).....	98
TABELA 16 - MODELO DE REFERÊNCIA PARA SELEÇÃO DOS INDICADORES - NÍVEL G, PROCESSO GRE.....	117
TABELA 17 - MODELO DE REFERÊNCIA PARA SELEÇÃO DOS INDICADORES - NÍVEL G, PROCESSO GPR.....	118
TABELA 18 - MODELO DE REFERÊNCIA PARA SELEÇÃO DOS INDICADORES - NÍVEL F, PROCESSO GQA.	119
TABELA 19 - MODELO DE REFERÊNCIA PARA SELEÇÃO DOS INDICADORES - NÍVEL F, PROCESSO MED.	120

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
BSC	<i>Balanced Scorecard</i>
AQU	Aquisição
CEP	Controle Estatístico de Processo
CISQ	<i>Consortium for IT Software Quality</i>
CMDS	<i>Classical Multidimensional Scaling</i>
CMMI	<i>Capability Maturity Model Integration</i>
CMMI-DEV	<i>Capability Maturity Model Integration for Development</i>
CMMI-SVC	<i>Capacility Maturity Model Integration for Services</i>
GCO	Gerência de configuração
GPP	Gerência de portfólio de projetos
GPR	Gerência de projetos
GRE	Gerência de requisitos
GQA	Garantia da qualidade
IA	Inteligência Artificial
ISO	<i>International Organization for Standardization</i>
KNN	<i>K-Nearest Neighbor</i>
KPI	<i>Key Performance Indicators</i>
LI	Limite Inferior da classe
LesTIC	Laboratório de Engenharia de Software e Tecnologia da Informação e Comunicação
LS	Limite Superior da classe
LSP	<i>Least Square Projection</i>
MA-MPS	Método de Avaliação
MD	Medição de Desempenho
MED	Medição
MLP	<i>Multi-Layer Perceptron</i>
MN-MPS	Modelo de Negócio do MPS.BR
MoProSoft	Modelo de Processos para a Indústria de Software
mPME	Micros, Pequenas e Médias empresas
MPS	Melhoria de Processo de Software

MPS-SW	MPS para software
MPS.BR	Melhoria de Processo do Software Brasileiro
MR-MPS-SW	Modelo de Referência MPS para Software
MR-MPS-SV	Modelo de Referência MPS para Serviços
OMG	<i>Object Management Group</i>
RNA	Rede Neural Artificial
SCAMPI	<i>Standard CMMI Appraisal Method for Process Improvement</i>
SMD	Sistema de Medição de Desempenho
SEI	<i>Software Engineering Institute</i>
TI	Tecnologia da Informação
TSVM	<i>Transductive Support Vector Machines</i>

SUMÁRIO

	Página
1 INTRODUÇÃO	14
1.1 Objetivos do trabalho.....	18
1.2 Metodologia de trabalho	21
1.3 Organização dos capítulos	23
2 INDICADORES DE DESEMPENHO EM PROCESSOS DE SOFTWARE	25
2.1 Indicadores de desempenho	28
2.2 Balanced Scorecard (BSC)	30
2.3 Modelo de referência MPS-SW	32
2.4 Trabalhos relacionados ao capítulo.....	38
3 APRENDIZAGEM DE MÁQUINA E TÉCNICAS SELECIONADAS	41
3.1 Aprendizado de máquina supervisionado	43
3.2 Aprendizado de máquina não supervisionado	43
3.3 Aprendizado de máquina semi-supervisionado.....	44
3.4 Redes Neurais Artificiais.....	47
3.5 Algoritmo de competição e cooperação entre partículas	49
3.6 Visualização de Informação	53
3.7 Trabalhos relacionados ao capítulo.....	57
4 SISTEMÁTICA PARA CONTROLE DE INDICADORES DE PROCESSOS DE SOFTWARE	60
4.1 Etapa 1: seleção dos indicadores de desempenho	62
4.2 Etapa 2: definição dos agrupamentos de indicadores.....	65
4.3 Etapa 3: rotulagem dos dados.....	67
4.4 Etapa 4: treinamento e análise dos dados dos indicadores	72
5 ESTUDO DE CASO E EXPERIMENTOS	74
5.1 Aplicação da sistemática para controle de indicadores de processos de software.....	75
5.1.1 Etapa 1: seleção dos indicadores de desempenho	75
5.1.2 Etapa 2: definição dos agrupamentos de indicadores	76
5.1.3 Etapa 3: rotulagem dos dados	77
5.1.4 Etapa 4: treinamento e análise dos dados dos indicadores	83
5.1.5 Conclusão do estudo de caso.....	85
5.2 Experimentos de apoio	87

5.2.1	Experimento I – Avaliando diferentes algoritmos de aprendizagem de máquina	87
5.2.2	Experimento II – Avaliação em um ambiente com poucas amostras	91
5.2.2.1	Experimento II – Parte A: uso do algoritmo supervisionado MLP	93
5.2.2.2	Experimento II – Parte B: uso do algoritmo PCC	97
6	CONCLUSÃO E TRABALHOS FUTUROS	99
	APÊNDICE A - Indicadores para os processos dos níveis G e F do MR-MPS-SW.....	109
	APÊNDICE B – Modelo de referência para seleção de indicadores	116

1 INTRODUÇÃO

As organizações produtoras de software vêm enfrentando problemas e desafios com relação à qualidade do software há muitas décadas, tendo como um marco histórico a conhecida “crise do software” no final dos anos 60. Essa expressão surgiu devido a problemas recorrentes enfrentados no processo de construção do software. Tais problemas abrangem todo o processo de concepção, desenvolvimento, implantação e manutenção do software, ou seja, não são restritos apenas a programas que não funcionem (REZENDE, 2005; MAFFEO, 1992).

Segundo Pressman (1995), os maiores desafios das primeiras três décadas da era do computador consistiam em desenvolver um hardware que reduzisse o custo de processamento e armazenagem de dados. No entanto, Rezende (2005) afirma que o principal desafio para essas primeiras décadas do século XXI consiste em melhorar a qualidade das soluções implementadas com software e reduzir os custos provenientes do processo de desenvolvimento.

Para enfrentar essa situação, técnicas formais para o gerenciamento da qualidade do software começaram a ser introduzidas, muitas das quais desenvolvidas a partir de métodos usados na indústria manufatureira. Essas técnicas de gerenciamento de qualidade, somadas às novas tecnologias de desenvolvimento e testes de software, têm propiciado melhorias significativas na qualidade do software (SOMMERVILLE, 2011).

O gerenciamento da qualidade de um projeto de software envolve três processos básicos, que podem ser aplicados a todos os tipos de projetos: planejamento da qualidade, garantia da qualidade e controle da qualidade. Esses processos devem interagir entre si e com os demais processos, pelo menos uma vez durante todo o ciclo de vida de um projeto (PMI, 2008).

O planejamento da qualidade consiste na identificação e documentação dos requisitos, de modo que os processos e/ou produtos estejam em conformidade com as especificações definidas (PMI, 2008). Devem ser estabelecidos critérios para se definir o que se entende como um software de “alta qualidade”. Tais critérios se fazem necessários para evitar suposições potencialmente conflitantes, acerca de quais atributos de produto

refletem as mais importantes características de qualidade (SOMMERVILLE, 2011).

O processo de garantia da qualidade é responsável por prover a confiança de que os requisitos da qualidade serão plenamente atendidos e mantidos (ABNT, 2005). É responsável por verificar e validar os requisitos em prol da conformidade dos processos e produtos. Após o término desse ciclo, um novo ciclo surge de modo iterativo, resultando em melhorias contínuas em todos os processos de software.

O processo de controle da qualidade, por sua vez, provê informações para avaliação de desempenho e mudanças nos projetos, processos e produtos. De modo geral, esse processo é responsável pelo monitoramento e registro das atividades relacionadas à qualidade (PMI, 2008). Resultados específicos do projeto devem ser monitorados para determinar se eles estão de acordo com os critérios de qualidade definidos e considerados relevantes. Técnicas de controle estatístico da qualidade auxiliam a avaliação dos resultados do controle da qualidade. O objetivo desse processo é identificar meios para “eliminar” causas de resultados insatisfatórios nos processos e/ou produtos.

Sommerville (2011) menciona que o termo “controle da qualidade” não é muito usado na indústria de software. Segundo o autor, os termos “garantia da qualidade” e “controle da qualidade” são amplamente utilizados na indústria manufatureira. Na indústria de software, porém, as empresas preferem utilizar o termo “garantia da qualidade” de maneira diferenciada, ao invés de utilizar ambos os termos. Segundo Sommerville (2011), garantia da qualidade ora abrange a definição de processos e padrões que reforçam a qualidade, ora o gerenciamento de configuração, verificação e validação do produto entregue. Apesar disso, o autor comenta que “controle da qualidade” também pode ser aplicado em todo o processo de produção do software. Nessa direção, o termo “controle da qualidade” será usado neste trabalho voltado ao contexto do desenvolvimento de software.

Em organizações com certo nível de maturidade, o controle dos processos requer medições em múltiplas escalas e frequência. Essas medições possibilitam verificar se os resultados estão em conformidade com critérios de qualidade previamente definidos. Indicadores de desempenho ou KPIs (*Key Performance Indicators*) devem ser utilizados para quantificar o desempenho do que está sendo medido (processo, atividade, ou produto). Assim, indicadores de desempenho são recursos fundamentais para o controle da qualidade. Eles fornecem uma visao local dos processos, permitindo comparação com as metas globais da organização (BOYD; COX, 1997).

De modo geral, os resultados dos indicadores de desempenho fornecem informações relevantes para as tomadas de decisão nas organizações. Geralmente, as decisões precisam ser feitas com rapidez e eficácia, considerando um indicador ou um

grupo de indicadores. Comparando os valores históricos do indicador com a meta prevista, é possível avaliar as variações ocorridas e gerar prognósticos (projeções).

Indicadores de desempenho são importantes estimadores na produção de software sempre que há demanda de um grande número de projetos, cujo gerenciamento requer uma forma eficiente de monitoração e controle. O volume de dados produzido por esses indicadores tende a aumentar significativamente com o tempo de monitoramento. As bases históricas acabam se tornando complexas, se considerada a quantidade de dados monitorados e a diversidade intrínseca aos indicadores, diferentes quanto ao tipo, granularidade e frequência de coletas/amostragem. Os indicadores podem ser analisados individualmente ou em grupos, segundo semânticas significativas para as decisões a serem tomadas. Os indicadores de desempenho possibilitam a observação do estado e da evolução de uma ou várias dimensões de um projeto, frente a metas pré-estabelecidas.

Com os avanços na área de Tecnologia da Informação (TI), observar e analisar um indicador de desempenho individualmente não é complicado, mesmo frente a uma grande quantidade de dados para esse indicador. Diversas soluções de software estão disponíveis para automatizar as coletas e exibições dos dados dos KPIs aos usuários. Essas soluções permitem uso de filtros por período, setor, processo, tipos e datas da medição, por exemplo. Segundo gerentes de projetos de software entrevistados para este trabalho, geralmente, a análise dos indicadores, realizada pelos administradores, usa média dos dados coletados, constantes na base histórica do indicador.

Há casos, porém, em que uma organização gerencia vários projetos em paralelo, com um determinado indicador sendo utilizado nos processos de cada projeto, por exemplo. Nesse caso, a média dos dados históricos desse indicador em todos os projetos possibilita uma visão global da situação, baseada na média dos dados de todos os projetos em desenvolvimento. É discutível o uso de médias nesse caso, por exemplo, requerendo estudos mais detalhados sobre a eficácia dessa ferramenta no apoio a tomadas de decisão eficientes, sejam em nível operacional, gerencial ou estratégico.

Segundo Moura (1999), uma visão mais estratégica dos indicadores de desempenho deve produzir resultados melhores do que a soma dos resultados de suas partes. Assim, muitas empresas de TI, inclusive de produção de software, vêm adotando a estratégia do *Balanced Scorecard* (BSC) (BUENO, 2009; LUNKES et al., 2009; ROCHA, 2009; PEDROSO, 2010). Trata-se de um modelo de mensuração de desempenho empresarial, baseado em indicadores que refletem a visão e estratégia empresarial (KAPLAN; NORTON, 1992). O modelo BSC é organizado em quatro perspectivas, com KPIs associados a cada uma delas: perspectiva financeira, perspectiva dos clientes, perspectiva dos processos

internos e perspectiva de aprendizado e crescimento.

De qualquer modo, o sistema de medição normalmente utilizado compara os valores obtidos na coleta de dados, durante a execução dos processos, com os valores planejados (metas) (BARCELLOS, 2009). As metas podem ser representadas por um intervalo de valores (ex.: $80 \leq \text{meta} < 100$) ou por um valor fixo (ex.: $\text{meta} \geq 80$). Quando um determinado indicador obtém um valor que satisfaça a meta, significa que esse indicador está em conformidade com a meta previamente estabelecida; caso contrário, ele está fora da meta, e uma correção ou ação corretiva deve ser realizada no processo. Entende-se como correção, a eliminação de uma não conformidade identificada, enquanto uma ação corretiva corresponde à eliminação da causa de uma não conformidade detectada (ABNT, 2005).

Algumas organizações desenvolvedoras de software possuem indicadores com limites inferiores ou superiores de controle, semelhante ao que se usa na produção industrial, como forma de prover um diagnóstico mais eficaz na prevenção e detecção de falhas nos processos (ex.: Controle Estatístico de Processo – CEP). Esse tipo de indicador possui um valor médio de operação e dois limites de controle: o limite inferior (LI) e o limite superior (LS) da classe. Cada resultado obtido através desse indicador é “plotado” em uma carta de controle (conhecida como *Shewhart*) e analisado quanto à sua tendência em se aproximar do LI ou do LS (PARK et al., 1996; DEMING, 1990; HUMPHREY, 1988). A medição funciona semelhante a um semáforo de trânsito, com sinalização verde, amarelo e vermelho. Enquanto o valor obtido estiver sob controle, o indicador está dentro da meta (sinal verde). Conforme o valor se aproxima do LI ou LS, um alerta no sistema de gestão é disparado (sinal amarelo), requerendo atenção especial nos próximos valores, pois há forte tendência de que eles saiam de controle. Por fim, caso o indicador obtenha valores fora dos limites de controle (sinal vermelho), é necessário parar e investigar as causas que levaram a esse desvio, provendo ações corretivas.

De modo geral, à medida que o sistema de monitoramento evolui, o conjunto de dados desses indicadores pode aumentar significativamente. As análises da base histórica gerada são importantes para futuras decisões estratégicas, constituindo um desafio à gestão organizacional. Assim, mecanismos de disseminação e compartilhamento das informações providas pelos indicadores de desempenho devem ser adequados. A agilidade nas tomadas de decisão constitui um fator fundamental para obtenção de vantagens competitivas no mercado corporativo. Isso é válido para todas as organizações, inclusive para as empresas produtoras de software de qualquer porte (BLOMQUIST; ÖHGREN; SANDKUHL, 2006). Devido a muitas atividades e discussões administrativas, há muitos casos em que se acaba administrando os problemas mais por exceção do que por precaução, incorrendo no risco de

esquecer causas importantes para as estratégias de negócio (LORIGGION, 2002).

Técnicas de aprendizagem de máquina têm sido utilizadas como forma de melhorar as análises dos indicadores. Essas técnicas permitem que indicadores de diferentes tipos sejam analisados em conjunto, o que não é possível quando se faz a média dos indicadores. Alguns trabalhos, por exemplo, usam Redes Neurais Artificiais (RNAs) na análise de indicadores de desempenho (MELO et al., 2011; NETO; NAGANO; MORAES, 2005; CATTINELLI et al., 2013). Esses trabalhos são voltados às diversas áreas de aplicação, como, por exemplo: setor agropecuário, tráfego rodoviário, área de saúde, entre outras. Contudo, não foram encontrados trabalhos para análise de indicadores em processos de desenvolvimento software com técnicas de aprendizagem de máquina.

Há várias técnicas de aprendizagem de máquina encontradas nas literaturas. Uma dessas técnicas, baseada em grafos, é denominada “cooperação e competição entre partículas” ou PCC (*Particle Competition and Cooperation*). Trata-se de uma técnica inspirada na comunidade de formigas (bioinspirada), elaborada por Breve (2010). Com relação ao conjunto de dados da base histórica, essa técnica requer que poucos dados sejam rotulados pelos especialistas no domínio da aplicação. Por esse motivo, é caracterizada por usar aprendizado semi-supervisionado. Segundo Breve (2010), essa técnica tem mostrado resultados com altas taxas de acerto. Além disso, sua complexidade computacional é baixa, quando comparada com outros algoritmos de mesma estrutura.

No contexto de desenvolvimento de software, a técnica PCC é interessante pelo fato de não ser necessário rotular toda a base histórica dos indicadores, individualmente ou em grupo. Isso reduz os custos no processo de rotulagem, requerendo menos esforço humano.

Diante do contexto apresentado, a *seção 1.1* traz os objetivos que nortearam este trabalho. A metodologia utilizada para o desenvolvimento das pesquisas apresentadas encontra-se na *seção 1.2*. Os demais capítulos desta dissertação estão apresentados na *seção 1.3*.

1.1 OBJETIVOS DO TRABALHO

O principal objetivo deste trabalho é apresentar uma sistemática baseada em aprendizagem de máquina para auxiliar a análise de indicadores de desempenho (KPIs) nos processos de desenvolvimento de software. A técnica utilizada foi a de cooperação e competição entre partículas ou PCC, com aprendizado semi-supervisionado (BREVE, 2010). Essa solução diminui o custo do processo de rotulagem das amostras em relação ao

aprendizado de máquina supervisionado. Isso porque ela não requer que todas as amostras sejam rotuladas para treinamento, reduzindo o tempo de especialistas na rotulagem das amostras. Por outro lado, essa solução não despreza informações relativas aos rótulos das amostras, como no aprendizado não supervisionado. É importante ressaltar que a técnica PCC ainda não foi aplicada à Engenharia de software e nem a indicadores de desempenho, segundo o autor do algoritmo PCC, que é coorientador deste trabalho, e mediante uma revisão sistemática de literaturas realizada.

Para apoiar o processo de rotulagem, a solução proposta incluiu a utilização de técnicas de visualização da informação. O objetivo é verificar a relação existente entre as instâncias dos indicadores em um plano multidimensional e, assim, identificar visualmente possíveis falhas de rotulagem. As técnicas utilizadas foram:

- (1) Projeção por mínimos quadrados ou LSP (*Least Square Projection*);
- (2) Escalonamento multidimensional clássico ou CMDS (*Classical Multidimensional Scaling*);
- (3) Coordenadas paralelas ou PC (*Parallel Coordinates*).

Embora a sistemática proposta seja indicada para agrupamentos de indicadores preferencialmente, indicadores individuais também podem ser utilizados. A intenção é tratar grandes volumes de dados que são gerados pelos KPIs de vários projetos em desenvolvimento em uma organização produtora de software. Evidentemente, ao longo do tempo de monitoramento dos processos, essas bases históricas tendem a se tornar mais volumosas e complexas. Consequentemente, as análises também se tornam mais complexas.

A sistemática permite expressar os resultados das análises através de uma escala gradual, de acordo com as metas previstas para os indicadores. Assim, os valores dos KPIs podem ser classificados segundo a metáfora do semáforo, em três classes: (1) sinal verde, indicando para seguir com o processo, pois a situação está satisfatória; (2) sinal amarelo, que alerta para se prestar atenção, porque o grau de satisfação perante a meta previamente definida é regular; (3) sinal vermelho, indicando que o processo deve ser interrompido, por estar insatisfatório, muito fora do padrão estabelecido. A metáfora do semáforo, contudo, pode ser estendida com a criação de outras classes.

A sistemática pode ser utilizada como núcleo de um sistema de apoio ao controle e monitoramento de KPIs, com uma interface de usuário que contemple a exibição de um painel de instrumentos, como *dashboards* ou *scoreboards*. A coleta dos dados dos indicadores a serem avaliados pelo sistema pode ser definida em outro módulo do sistema.

Essas interfaces e módulos não fazem parte deste trabalho, mas podem ser projetadas e implementadas em trabalhos futuros.

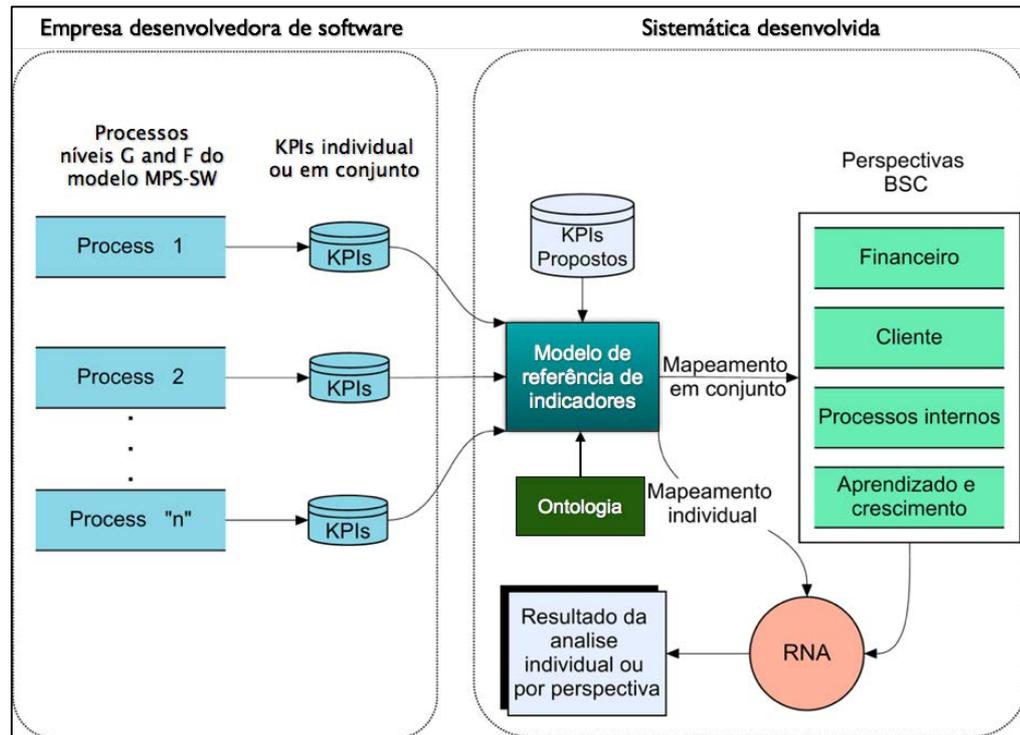
Devido à variedade de processos que podem ser utilizados na produção de software por diversas organizações, um conjunto de processos foi considerado como base para este trabalho. Esses processos pertencem aos níveis G e F do modelo de referência MPS para Software (MR-MPS-SW), que faz parte do programa para Melhoria de Processos do Software Brasileiro (MPS.BR). O MR-MPS-SW é um modelo de maturidade que compreende definições de processos no ciclo de vida de um software e conta com sete níveis de maturidade, representados por letras de G (nível mais baixo) a A (nível mais alto). O nível G possui os processos básicos necessários para o gerenciamento de projetos e de requisitos. Já o nível F possui os processos de gerenciamento de qualidade, de portfólio e de configuração, bem como os de aquisição e medição.

Com a finalidade de auxiliar as organizações na seleção dos indicadores de desempenho para uso na sistemática, dois modelos de referência são propostos. O primeiro é constituído por indicadores de desempenho sugeridos para todos os processos dos níveis G e F do MR-MPS-SW. A finalidade desse modelo é contribuir com a elaboração de um modelo de indicadores na organização, caso ela não tenha um. O segundo modelo utiliza as perspectivas do modelo BSC para apoiar as organizações a definirem grupos de indicadores de desempenho, caso elas não tenham critérios já definidos para essa finalidade.

O trabalho mostra um estudo de caso da sistemática, que utiliza dados históricos reais de indicadores de desempenho, obtidos através do monitoramento de processos de desenvolvimento de software durante quase três anos. Os dados pertencem a uma empresa produtora de software para gestão pública, que possui certificação no nível G do modelo MPS-SW. A empresa também forneceu horas de trabalho de dois gerentes de projeto, especialistas no domínio, para auxiliar o processo de rotulagem. Esse estudo de caso utilizou os modelos de referência de seleção e de agrupamento de indicadores, bem como as técnicas de visualização da informação mencionadas.

Os objetivos e a aplicação da sistemática proposta podem ser representados na **Figura 1**, considerando o tratamento dos KPIs de acordo com os interesses de cada organização. Vale salientar que os resultados das análises de grupos de indicadores provêm uma visão mais estratégica, auxiliando tomadas de decisão que não encontram apoio nas análises individuais dos dados. Esses resultados podem, assim, sinalizar o grau de conformidade com as metas definidas, bem como expressar se os processos associados precisam de ajustes.

Figura 1 - Visão geral da sistemática desenvolvida e sua aplicação.



1.2 METODOLOGIA DE TRABALHO

Esta seção apresenta a metodologia de trabalho utilizada para o desenvolvimento das atividades que conduzirão as pesquisas aos objetivos apresentados na seção 1.1.

Inicialmente foram realizados levantamentos sobre indicadores chave de desempenho e investigados os problemas decorrentes da análise em grandes volumes de dados. Foi possível observar que houve pouca mudança nos modelos de medição adotados nas organizações de modo geral, bem como na interpretação das medições, mesmo com a evolução dos processos ao longo dos anos. Observou-se também, que os indicadores podem ser classificados como quantitativos e qualitativos, sendo este último o mais difícil de ser compreendido por não ser passível de mensuração. Diversas são as áreas em que indicadores podem ser aplicados, mas geralmente seu campo de aplicação está ligado a processos que já possuem certa maturidade, com algum nível de controle.

As investigações na área de Engenharia de Software foram voltadas à qualidade dos processos de desenvolvimento de software, com foco no uso de indicadores e medição de desempenho. As literaturas encontradas ratificam a necessidade do uso de indicadores de desempenho, visando verificar se os objetivos dos processos estão sendo atingidos e se

estão em conformidade com as metas do planejamento e negócio.

Para melhor entender os indicadores de desempenho aplicados na produção de softwares, foram procuradas empresas na região de Ribeirão Preto que utilizavam indicadores de desempenho no controle de qualidade e decisões estratégicas. Foram encontradas empresas que têm ou já tiveram certificações em modelos de qualidade de processos de software, como o MR-MPS-SW e que tinham bases históricas de indicadores, As interações com gestores de qualidade dessas empresas contribuíram para direcionamentos nesta área. Ressalta-se que uma empresa de software para gestão pública, certificada no nível G do MPS-SW, propôs uma parceria entre a empresa e a Unesp, a fim de fornecer sua base histórica de dados e dar suporte às classificações dos grupos de indicadores.

Visando entender os processos utilizados nessas empresas e os indicadores que poderiam ser empregados nesses processos, as pesquisas foram direcionadas para o MR-MPS-SW. Trata-se de um modelo adequado ao cenário nacional das empresas de produção de software e que tem tido um número muito grande de certificações, sendo a grande maioria das certificações nos níveis G e F. Esses dois níveis causam grandes mudanças na organização dos processos, que impactam nas pessoas e nas atividades de modo geral da empresa. Indicadores chave de desempenho podem ser utilizados desde o nível G, sendo mandatório a partir do nível F.

Além dos guias do MR-MPS-SW, contou-se com o apoio da ontologia proposta por Pizzoleto (2013) para os níveis G e F e interações diretas com o autor, que também é membro do grupo de pesquisa do Laboratório de Engenharia de Software e Tecnologias da Informação e Comunicação (LesTIC). Essa ontologia inclui requisitos necessários para melhorar a compreensão do modelo MPS-SW, a partir de entrevistas com especialistas no modelo e empresas certificadas. Conceitos do PMBOK complementam as informações dos guias do modelo. Além disso, a ontologia traz indicadores para todos os processos (incluindo subprocessos) dos níveis G e F, acrescidos de indicadores de três perspectivas do BSC (menos a financeira) para complementar a estratégia de medições do MR-MPS-SW.

Conforme foram sendo compreendidos os processos do MPS-SW, foram sendo avaliadas possíveis formas de agrupamento de indicadores para apoio às decisões estratégicas. Observou-se a viabilidade e adequação do agrupamento de indicadores ser baseado nas perspectivas estratégicas do BSC. A ideia é utilizar o BSC para entender como os grupos de indicadores em uma empresa podem ser formados, de modo a contribuir com uma análise de desempenho mais estratégica.

As investigações sobre aprendizado de máquina foram realizadas na busca de uma

solução que fornecesse uma análise automática de grupos de indicadores, e que tivesse a dinâmica de evoluir e melhorar à medida que esses indicadores aumentavam em volume e complexidade. Foram necessários estudos em aprendizagem de máquina, de modo a contextualizar melhor o uso de algumas técnicas. Os estudos mostraram técnicas/algoritmos de aprendizagem de máquina supervisionados, semi-supervisionados e não supervisionados. Observou-se que algoritmos de aprendizagem semi-supervisionados são altamente indicados nos casos onde não é necessário rotular muitas amostras, além de não ignorarem informações valiosas de rótulos dos itens de dados.

Foram consideradas técnicas de visualização de informação no processo de rotulagem. A intenção foi ajudar os especialistas na identificação das classes dos dados, principalmente os dados mais próximos das bordas da classe. As técnicas de visualização de informação podem contribuir no processo de rotulagem dos dados para os algoritmos de aprendizado supervisionado e semi-supervisionado.

Experimentos foram realizados com indicadores de desempenho e aprendizado supervisionado e semi-supervisionado. Esses experimentos contribuíram para o desenvolvimento da sistemática para análise de indicadores de desempenho, utilizando aprendizado de máquina semi-supervisionado. Por fim, um estudo de caso foi realizado seguindo as etapas definidas na sistemática desenvolvida.

1.3 ORGANIZAÇÃO DOS CAPÍTULOS

Considerando os objetivos deste trabalho e a metodologia utilizada para atingi-los, esta dissertação está organizada em mais cinco capítulos.

O *capítulo 2* aborda tópicos relacionados à medição de desempenho, incluindo a problemática em organizações produtoras de software. Conceitos relacionados a KPIs são apresentados. O capítulo também apresenta a fundamentação para os modelos de referência definidos neste trabalho, tanto para a seleção como para agrupamentos de indicadores. Assim, o capítulo aborda o modelo BSC, bem como os aspectos considerados relevantes sobre o modelo de qualidade MPS-SW. Em complementação, o capítulo traz uma visão geral da ontologia dos níveis G e F do modelo MPS-SW, elaborada por Pizzoleto (2013), com partes no APÊNDICE A. Por fim, alguns trabalhos levantados na literatura relacionados ao tema abordado são apresentados no capítulo.

O *capítulo 3*, por sua vez, apresenta os levantamentos feitos sobre aprendizagem de máquina, abordando as categorias de aprendizado supervisionado, não supervisionado e

semi-supervisionado. O capítulo traz uma breve abordagem sobre RNAs, uma vez que alguns experimentos realizados durante o desenvolvimento deste trabalho utilizaram esse tipo de técnica. Contudo, a principal contribuição do capítulo é a apresentação do algoritmo de aprendizado semi-supervisionado utilizado neste trabalho: algoritmo de competição e cooperação entre partículas (PCC), elaborado por Breve et al. (2009). Considerando que técnicas de visualização de informação são utilizadas na sistemática desenvolvida, esse capítulo traz um breve levantamento sobre essas técnicas. Alguns trabalhos relacionados à aprendizagem de máquina, considerados de interesse ao trabalho, também são apresentados.

O *capítulo 4* apresenta a sistemática desenvolvida para apoiar o controle de qualidade na produção de software, utilizando indicadores de desempenho. A sistemática é composta por quatro etapas, que tratam da seleção de indicadores, individualmente ou em grupo, bem como do processo de rotulagem e análise dos dados. Cada etapa é apresentada com a respectiva descrição, ferramentas utilizadas e resultados esperados. Nas duas primeiras etapas também são apresentados os dois modelos de referência de indicadores desenvolvidos neste trabalho, complementado com o conteúdo do APÊNDICE B. Na terceira etapa, de rotulagem dos dados, é que são utilizadas as técnicas de visualização de informação. Já na quarta, é utilizado o algoritmo PCC.

O *capítulo 5* apresenta um estudo de caso que foi realizado com a aplicação da sistemática desenvolvida sobre dados de indicadores reais de uma empresa produtora de software. Além disso, o capítulo traz alguns experimentos realizados durante o desenvolvimento do trabalho, cujos resultados serviram para avaliação dos caminhos adotados. Esses experimentos utilizaram, além do algoritmo PCC, algoritmos de RNAs, como Perceptron de multicamadas e K-vizinhos mais próximos. São apresentadas discussões sobre cada experimento apresentado, bem como sobre o estudo de caso desenvolvido.

Por fim, o *capítulo 6* traz as considerações finais face ao apresentado nos demais capítulos, incluindo algumas propostas para trabalhos futuros.

2 INDICADORES DE DESEMPENHO EM PROCESSOS DE SOFTWARE

Segundo Goldratt (1999), indicadores são elementos que auxiliam a tomada de decisão localmente, direcionando os esforços na busca da meta global de uma organização. Os indicadores não informam o que deve ser feito para melhorar, mas oferecem relações numéricas que refletem a situação atual. Essas relações fornecem subsídios para tomadas de decisão que visam melhorias nos processos utilizados.

O valor numérico de cada indicador em determinado momento é chamado de “índice”. A partir dos índices podem ser definidos padrões e metas para os indicadores. Um “padrão” é um índice usado como referência de comparação para o indicador; esse índice pode ser definido arbitrariamente ou por convenção. Uma meta, por sua vez, é o “índice” que se almeja para um indicador. A **Tabela 1** apresenta os oito principais atributos de um indicador.

Tabela 1 - Principais atributos dos indicadores.

Fonte: extraído de Brito (2005).

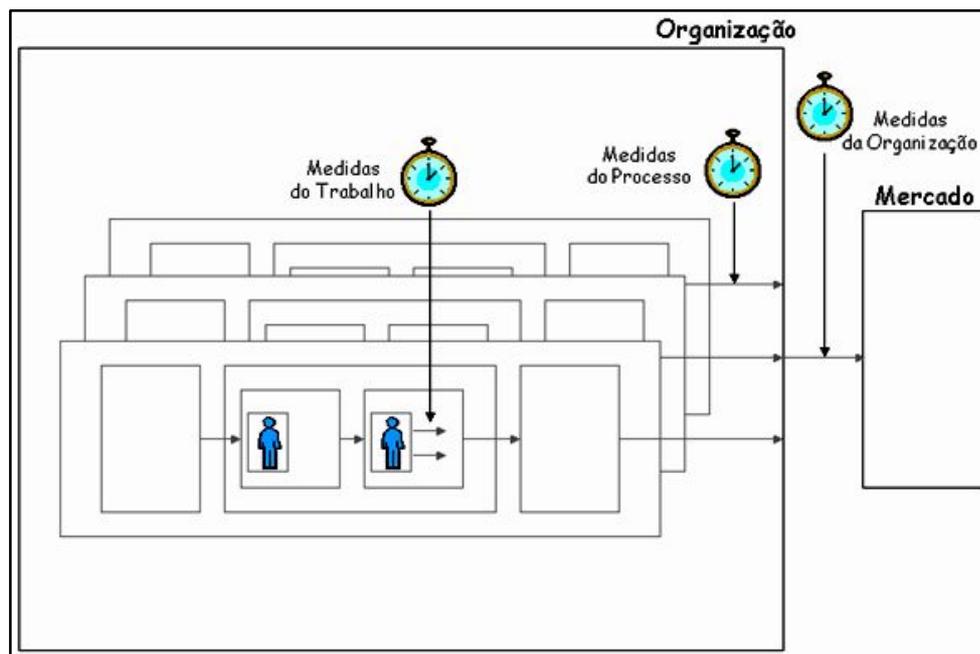
ATRIBUTO	DESCRIÇÃO
Adaptabilidade	Capacidade de resposta às mudanças
Representatividade	Cobertura das etapas mais importantes e críticas
Simplicidade	Facilidade de ser compreendido e aplicado
Rastreabilidade	Facilidade para identificação da origem dos dados
Disponibilidade	Facilidade de acesso para coleta
Economia	Baixo custo de obtenção
Praticidade	Garantia de que realmente funciona na prática
Estabilidade	Permanece ao longo do tempo

Há vários tipos de indicadores que visam controlar e melhorar o desempenho de um determinado sistema, seja no nível dos processos de uma organização ou de execução do trabalho dentro dos processos, conforme ilustra a **Figura 2**. Considerando uma organização produtora de software, os indicadores podem ser utilizados nos diferentes processos para o desenvolvimento e entrega de softwares. Esses processos podem ocorrer, inclusive, de

maneira paralela e simultânea. Em cada fase do processo de desenvolvimento de um software, diferentes níveis de medição e controle podem ser estabelecidos. Então, indicadores de desempenho podem ser utilizados para o controle da qualidade em diferentes níveis dos processos. Os principais valores desses indicadores são expressos em forma de percentual e numérico.

Figura 2 - Níveis de medições de indicadores.

Fonte: extraído de Reis (2014).



Em termos de controle da qualidade, podem ser utilizados indicadores junto ao processo, ao produto ou para avaliação da satisfação do cliente. Os indicadores voltados à eficiência do processo são chamados “indicadores de produtividade”. São usados com o objetivo de redução de custos e análise da utilização dos recursos para a geração de produtos e serviços. Esses indicadores relacionam entradas consumidas com as metas já estabelecidas. Os resultados dos indicadores de produtividade não são expressos em forma de percentual, mas sim como quantidade por homens/hora/máquina, por exemplo. Já os indicadores voltados à eficácia dos processos, ou seja, aos produtos dos processos, são chamados “indicadores de qualidade”. A ideia é relacionar as saídas dos processos com as metas pré-estabelecidas para os produtos resultantes dos processos. Por outro lado, os “indicadores de satisfação do cliente” medem a capacidade do processo em atender os requisitos dos clientes, bem como a percepção do cliente em relação ao produto.

Para o uso de um sistema de medição e avaliação de processos com indicadores, é importante que os processos sejam bem definidos, de modo a serem projetados adequadamente e estejam prontos para serem gerenciados (REIS, 2014). Isso deve ser considerado nas organizações desenvolvedoras de software, onde as informações para tomadas de decisão operacionais estão relacionadas aos processos existentes. Nesses processos, há vários indicadores, que são analisados de maneira individual ou em grupo (geralmente por médias). Cada resultado fornecido pela análise dos indicadores serve de apoio para tomadas de decisão nos processos, e, conseqüentemente, no negócio.

Nessa direção, cabe ressaltar que o modelo nacional de qualidade de software MPS-SW contempla processos bem definidos, que podem ser implementados em empresas desenvolvedoras de software. Os indicadores do modelo MPS-SW são previstos a partir dos processos do nível F, sendo mantidos até o nível A (mais alto nível de maturidade). Porém, pode ser considerada uma boa prática, a adoção de indicadores de desempenho desde o nível inicial de maturidade (nível G) do modelo MPS-SW.

Por outro lado, a maneira de conceber e gerenciar indicadores nos processos, geralmente vai de encontro aos objetivos macroorganizacionais. No contexto de estratégias de negócio, um modelo que pode ser aplicado em diversas corporações, inclusive em empresas de Tecnologia da Informação (TI), é o *Balanced Scorecard* (BSC). Esse modelo reflete o equilíbrio entre objetivos de curto e longo prazo, visando melhorar os resultados das empresas com base em quatro perspectivas: financeira, clientes, processos internos e aprendizado/crescimento. A partir dessas perspectivas, são selecionados os indicadores de controle e as metas para organizar a análise do desempenho e crescimento ao longo do tempo.

Posto isso, a seção 2.1 aborda conceitos e questões relacionadas a aplicações de indicadores de desempenho em organizações produtoras de software. É definido também o conceito de indicadores chave de desempenho (KPIs), bem como a falta de consenso em relação à definição do termo. A seção 2.2, por sua vez, apresenta o modelo BSC, proposto para o agrupamento dos KPIs. A seção 2.3 traz os aspectos considerados relevantes para o trabalho sobre o modelo de qualidade MPS para software (MPS-SW), complementado pela visão geral da ontologia dos níveis G e F do modelo, elaborada por Pizzoleto (2013). Trabalhos envolvendo indicadores e modelos de qualidade de software foram levantados na literatura, sendo que alguns estão apresentados na seção 2.4.

2.1 INDICADORES DE DESEMPENHO

Medir o desempenho de algo consiste em mensurar ações, onde medição é o processo de quantificar e as ações conduzem ao desempenho. Um “indicador de desempenho” pode ser definido como a métrica usada para quantificar a eficiência e/ou eficácia de uma ação (NEELY et al., 1995). No contexto de projetos, indicadores podem ser definidos como instrumentos que permitem avaliar a evolução do projeto sob a ótica de uma ou mais dimensões preestabelecidas (FILHO, 2010). Também é possível avaliar o projeto durante seu andamento e progressão, porém, a última análise é a mais relevante, por representar a situação atual ou mais realista do projeto (NETO, 2014).

Segundo Neely et al. (1999), o conjunto de indicadores usado para quantificar a eficiência e eficácia das ações é chamado de Sistema de Medição de Desempenho (SMD). Até a década de 90, os indicadores de um SMD eram avaliados na maneira “tradicional” de medição de desempenho. Eram usadas duas perspectivas básicas: a busca pela eficiência operacional e o gerenciamento orientado por uma visão financeira/contábil. Enquanto a primeira perspectiva buscava medir o grau de utilização dos recursos operacionais, a segunda focava no constante monitoramento e controle dos custos e lucros. Essa visão foi considerada a ideal até a década de 70, momento que simbolizou o marco da necessidade de mudança, pois algumas incongruências nessa forma de avaliar indicadores se tornaram marcantes. Assim, segundo Pereira (1999), na década de 1970 surgiu o primeiro sistema de avaliação com os chamados “indicadores chave de desempenho” ou “*Key Performance Indicators*” (KPIs), cujo foco consistia em monitorar processos.

Um indicador chave de desempenho é um conjunto de diferentes medidas para manter o controle sobre o desempenho organizacional. A ideia é que medições regulares de correção, usabilidade, qualidade e produtividade, ajustadas aos propósitos de desempenho de um projeto, ajudem a identificar o estado (*status*) do projeto em diferentes estágios do seu desenvolvimento. Isso ajuda a aumentar o desempenho organizacional, que é o fator mais importante para o sucesso atual e futuro da empresa. Assim, os KPIs indicam o que deve ser feito para aumentar significativamente o desempenho dos processos (OJHA, 2014).

Em muitas literaturas, foi observado que indicadores de desempenho e indicadores chave de desempenho são considerados como sinônimos - talvez por considerarem que todos os indicadores são importantes para organização. Para alguns autores, contudo, KPIs são indicadores de desempenho selecionados, que realmente são “chave” nos processos, imprescindíveis para tomadas de decisão. As empresas podem definir vários indicadores de

desempenho, mas a ideia é que os indicadores “chave” sejam poucos, apenas os que de fato sejam importantes. Essa forma mais restrita de considerar os KPIs também é considerada neste trabalho, inclusive na sistemática proposta, apresentada no *capítulo 4*. Cabe à organização que vai usar a sistemática definir o que considerará como KPIs.

Os KPIs podem ser classificados como qualitativos ou quantitativos. Indicadores quantitativos são mensurados de forma numérica, pois expressam as variações quantificáveis do projeto. Já os indicadores qualitativos não são expressos com valores numéricos e são difíceis de representar de forma quantificável; possuem características que podem expressar valores, atitudes e crenças extraídas das opiniões das pessoas sobre determinado assunto (FILHO, 2010).

A partir da década de 90 surgiram os “novos” SMDs, na tentativa de sanar os problemas vivenciados no modo “tradicional” de avaliação dos indicadores de desempenho. Esses sistemas buscaram minimizar e eliminar as características de incongruências geradas pelo modelo anterior, adotando diversas características específicas, como: prover um alarme antecipatório de problemas, transcender o papel de controle, realizar o tratamento da informação, entre outros. Para Bryde (2003), em gerenciamento de projetos, os SMDs costumam focar no cumprimento dos custos, prazo e especificação do produto final. Entretanto, essa visão está mudando a partir do surgimento de outros modelos de medição de desempenho, frutos de uma abordagem contingencial, que explora os resultados sob a perspectiva de todos os envolvidos no projeto (*stakeholders*).

Mesmo com a evolução dos SMDs, alguns problemas relacionados a indicadores ainda prejudicam a análise dos dados gerados por esses indicadores (MELLO, 2000):

- alto volume de dados com baixo grau de informação;
- dados inconsistentes;
- dados incorretos;
- dados tendenciosos;
- erros de digitação;
- falta de informação devido a simplificações;
- falta de padrão nos processos;
- grande quantidade de informação;
- informação desatualizada.

Atualmente, esses problemas ainda persistem e dificultam as tomadas de decisão (BERTEI; MARCHI; BUONCRISTIANI, 2015). Assim, é importante que se encontre meios de tratar o grande volume de dados gerados pelos indicadores, de modo a possibilitar que as

organizações armazenem, gerenciem e manipulem tais dados rapidamente, no tempo certo, contribuindo nas tomadas de decisão em tempo real (KUMAR; SAHNI; MARWAH, 2015).

Contudo, como este trabalho se aplica às empresas que estejam em conformidade com os processos dos níveis G e F do modelo MPS-SW, ou de outro modelo de qualidade implantado, supõe-se que as situações supracitadas tenham sido previstas e tratadas, durante a implementação e evolução dos processos.

De modo geral, a compreensão do desempenho organizacional é provida por meio da análise de seus indicadores de desempenho. Os dados providos pelos KPIs permitem que os gestores organizacionais consigam entender o nível de conformidade dos processos, podendo tomar as decisões de forma mais segura. Entretanto, a forma de agrupamento de indicadores usadas nos SMDs nem sempre proporciona uma visão estratégica. Isso tem levado as organizações a se interessarem mais pela implementação de modelos estratégicos relacionados à medição de desempenho (FERNANDES; RAJA; WHALLEY, 2006; THAKKAR et al., 2007). Esses modelos geralmente agrupam os KPIs em um painel de avaliação (*scorecard*) empresarial, de modo a fornecer um resumo histórico rápido e preciso do sucesso da empresa. Um desses modelos, que tem sido muito utilizado nos últimos anos, é o *Balanced Scorecard* (BSC) (NEELY, 2005; ASSIRI; ZAIRI; EID, 2006), apresentado na seção 2.2.

2.2 BALANCED SCORECARD (BSC)

O BSC consiste em um modelo de gestão empresarial baseado em indicadores de desempenho, que traduzem a missão e a estratégia em objetivos e medidas. Criado por Kaplan e Norton na década de 90, o *Balanced Scorecard* está centrado na necessidade de contemplar múltiplas perspectivas, ao determinar as medidas de desempenho organizacionais.

O modelo BSC propõe a criação de metas que devem ser cumpridas pela empresa dentro de um tempo previsto. Essas metas são monitoradas por meio de KPIs e estão relacionadas diretamente ao planejamento estratégico das empresas. No BSC, as medidas de desempenho estão organizadas sob quatro perspectivas (KAPLAN; NORTON, 1997):

- **Perspectiva de finanças:** tem a finalidade de abranger os resultados financeiros da empresa;
- **Perspectiva de clientes:** abrange todas as pessoas que estão interagindo com a empresa no mercado (clientes, parceiros, fornecedores) e os valores que são

entregues a eles;

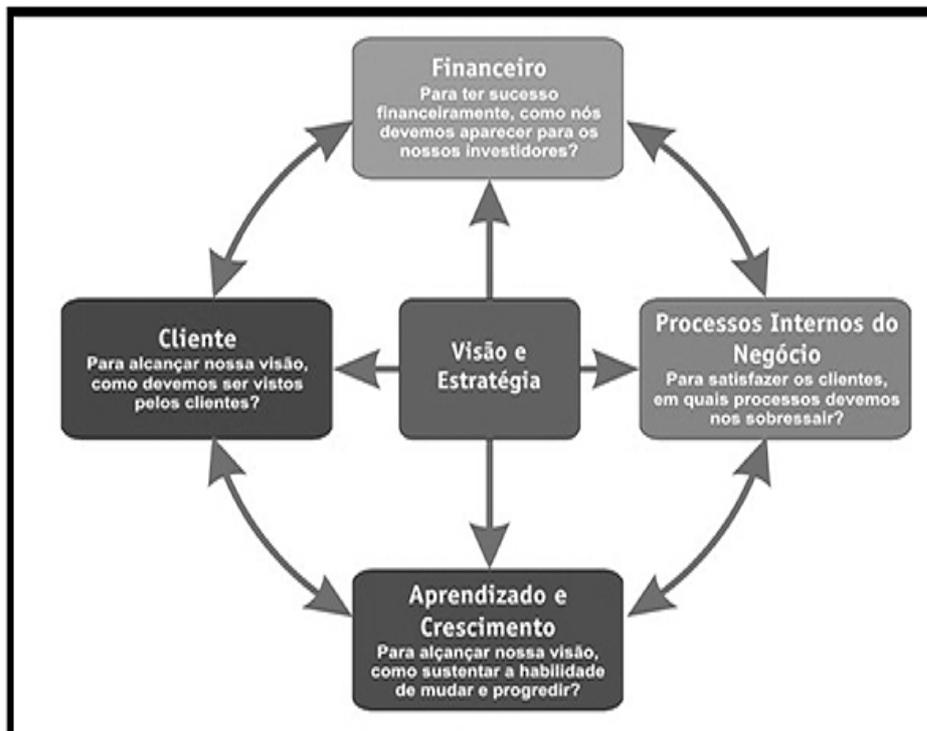
- **Perspectiva de processos internos:** abrange todos os serviços de transformação e geração de valor que ocorrem dentro da empresa (desenvolvimento de produtos, produção);
- **Perspectiva de aprendizado e crescimento:** pode ser resumida como perspectiva relacionada a pessoas, que abrange os valores, a liderança, a cultura, as competências e as ferramentas tecnológicas que a empresa possui.

O termo “*Balanced*” (balanceado) utilizado no modelo BSC diz respeito ao equilíbrio que deve haver entre os objetivos traçados, as medidas financeiras e não financeiras e indicadores de ocorrência e de tendência. Já o termo “*Scorecard*” enfatiza a forma na qual os resultados são apresentados, semelhante a um placar (*scorecard*) (HIKAGE, 2006).

As quatro perspectivas do BSC, apresentadas na **Figura 3**, estão alinhadas com uma visão sistêmica da empresa e do meio em que ela atua. Com o BSC é possível analisar desde a possibilidade de crescimento e aprendizado até como os investimentos em recursos humanos, sistemas e capacitação poderão mudar as atividades. Por ser um modelo flexível e adaptável a qualquer tipo de organização, a utilização do BSC passou a ter grande aceitação no mercado mundial, propiciando melhora na comunicação interna e externa da organização.

Figura 3 - Perspectivas do modelo BSC.

Fonte: extraído de Kaplan e Norton (1997).



Convém observar que um dos objetivos deste trabalho consiste em agrupar os KPIs de empresas desenvolvedoras de software sob as quatro perspectivas do modelo BSC. Para isso, será necessária a harmonização (mapeamento) entre os diversos indicadores e processos de desenvolvimento existentes na empresa com os processos dos níveis G e F do modelo de qualidade MPS-SW (ver seção 2.3). Essa harmonização possibilitará que os indicadores de desempenho utilizados na empresa possam ser mais facilmente agrupados de acordo com as ideias das perspectivas do BSC. Assim, o monitoramento desses indicadores facilitará as tomadas de decisão para os devidos ajustes, considerando a abordagem de cada perspectiva.

2.3 MODELO DE REFERÊNCIA MPS-SW

Para o estudo de indicadores de desempenho utilizados na produção de software, foram considerados os processos constantes do modelo de referência MPS para Software (MPS-SW). Esse modelo faz parte do programa de Melhoria de Processo do Software Brasileiro (MPS.BR), mantido pela Associação para Promoção da Excelência do Software Brasileiro (Softex).

O programa MPS.BR foi criado em 2003, com o objetivo de melhorar a qualidade do software no Brasil, bem como viabilizar o acesso das micro, pequenas e médias empresas (mPME) a modelos de qualidade de processos de software. Esse programa contempla quatro componentes, descritos através de documentos denominados “guias”:

1. Modelo de Referência MPS para Software (MR-MPS-SW);
2. Modelo de Referência MPS para Serviços (MR-MPS-SV);
3. Método de Avaliação (MA-MPS);
4. Modelo de Negócio (MN-MPS).

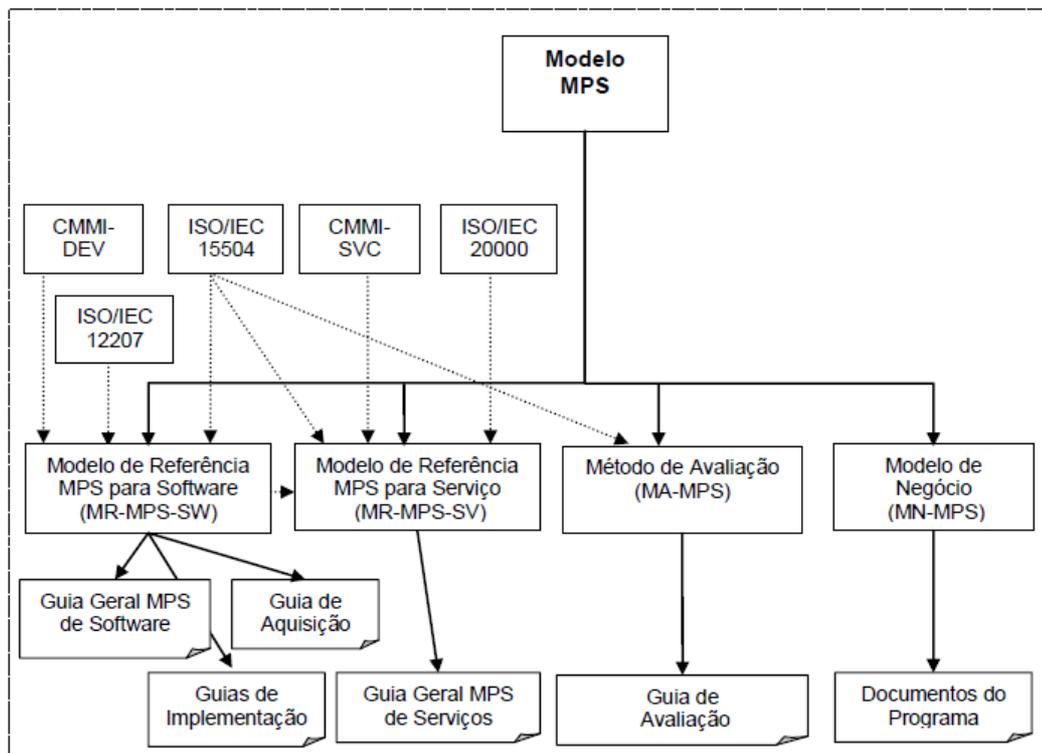
Em ambos modelos de referência, MR-MPS-SW e MR-MPS-SV, há diversos requisitos que os processos das unidades organizacionais devem atender para estarem em conformidade com o programa MPS.BR. Os modelos relacionados à avaliação (MA-MPS) descrevem o processo e o método de avaliação do modelo MPS, enquanto o modelo relacionado ao negócio (MN-MPS) descreve as regras de negócio para implementação e avaliação dos modelos, dentre outros.

A elaboração dos componentes do programa MPS.BR foi baseada em três normas internacionais relevantes para o desenvolvimento de software (ver **Figura 4**): (a) NBR

ISO/IEC 12207, referente ao processo do ciclo de vida do software; (b) NBR ISO/IEC 15504, voltada ao processo de avaliação do software; (c) NBR ISO/IEC 20000, referente ao gerenciamento de serviços de TI. A **Figura 4** mostra também que o modelo de qualidade de processos CMMI (*Capability Maturity Model Integration*) também foi usado como base para os modelos de referência do programa MPS.BR: CMMI-DEV (*CMMI for Development*) como apoio ao MR-MPS-SW e CMMI-SVC (*CMMI for Services*) como apoio ao MR-MPS-SV. A **Figura 4** destaca três tipos de guias do modelo MPS-SW: Guia de Aquisição, contendo boas práticas para aquisição de softwares correlatos, Guia Geral MPS de Software, que contém a descrição geral do modelo de referência MPS para Software, e Guias de Implementação, que se referem a um conjunto de 13 guias de orientação à implementação do MPS-SW em diversos tipos de organizações.

Figura 4 - Componentes do programa MPS.BR.

Fonte: extraído de SOFTEX (2012a).



O modelo MPS-SW compreende definições de processos no ciclo de vida de um software, descrito em termos de objetivos e resultados esperados. Ao implementar o modelo MPS-SW, cada organização deve definir as atividades e tarefas necessárias para atender os objetivos e resultados esperados dos processos, com a liberdade de adaptação às suas necessidades e políticas.

O modelo MPS-SW é incremental e possui sete níveis de maturidade, iniciando no nível G (nível mais baixo) até o nível A (nível mais alto), conforme mostrado na **Tabela 2**. Cada nível de maturidade tem processos que devem ser adicionados aos processos do nível superior. Assim, a implementação dos processos é gradual até o nível A, que preza pela continuidade da melhoria nos processos (SOFTEX, 2012a). A **Tabela 3** mostra os processos associados a cada nível do modelo MPS-SW. Convém ressaltar que para cada nível de maturidade que é implementado, há um conjunto de resultados esperados da organização, visando certo tipo de desempenho (SOFTEX, 2013).

Tabela 2 - Níveis de maturidade do modelo MPS-SW.

Fonte: extraído de SOFTEX (2012a).

Níveis	Descrição dos Níveis
A	Em otimização
B	Gerenciado Quantitativamente
C	Definido
D	Largamente Definido
E	Parcialmente Definido
F	Gerenciado
G	Parcialmente Gerenciado

Tabela 3 - Processos adicionados em cada nível de maturidade (NM) do MPS-SW.

Fonte: extraído de SOFTEX (2013).

NM	Processos
A	(nenhum novo processo é adicionado)
B	Gerência de Projetos (evolução)
C	Gerência de Riscos, Desenvolvimento de Reutilização, Gerência de Decisões
D	Verificação, Validação, Projeto de Construção do Produto, Integração do Produto e Desenvolvimento de Requisitos
E	Gerência de Projetos (evolução), Gerência de Reutilização, Gerência de Recursos Humanos, Definição de Processos Organizacionais e Avaliação e Melhoria do Processo da Organização
F	Medição, Garantia da Qualidade, Gerência de Portifólio de Projetos, Gerência de Configuração, Aquisição
G	Gerência de Requisitos e Gerência de Projetos

Na implementação do modelo MPS-SW, empresas que não possuem processos definidos e documentados adequadamente encontram mais desafios, inclusive no nível G.

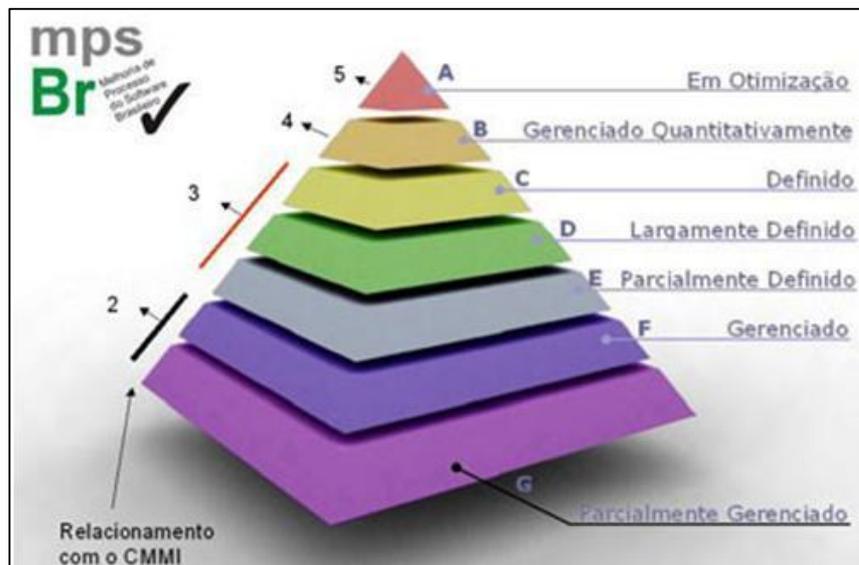
Contudo, os processos existentes podem ser ajustados para satisfazer os critérios dos processos requeridos para o nível desejado do modelo. Assim, cada processo existente na empresa produtora de software pode ser mapeado para os processos contidos em um determinado nível de maturidade do modelo MPS-SW. Cada nível do modelo representa uma evolução dos processos, de modo que o nível de maturidade em que se encontra uma organização indica qual o desempenho esperado ao executar um processo (SOFTEX, 2013).

Uma empresa pode ser certificada a partir do nível G, inclusive. Como apresentado na **Tabela 3**, esse nível de maturidade contempla dois processos de software críticos para as mPME: Gerenciamento de Requisitos e Gerenciamento de Projetos. O nível F requer o uso de indicadores de desempenho, mas considera-se uma boa prática a adoção de indicadores de desempenho desde o nível G.

A estratificação em sete níveis do modelo MPS para software foi baseada no modelo CMMI-DEV. Segundo SOFTEX (2014), há uma equivalência total do ponto de vista do MR-MPS-SW para o CMMI-DEV, entretanto, o contrário não é verdadeiro. Assim, a **Figura 5** apresenta a equivalência entre ambos modelos. O número maior de níveis do modelo MPS-SW (sete níveis) em relação ao CMMI-DEV (cinco níveis) simplifica a escala de mudanças que uma empresa necessita fazer até o caminho da certificação. Essa quantidade maior de níveis, conseqüentemente de avaliações, proporciona maior visibilidade das melhorias de processos e qualidade introduzidas.

Figura 5 - Equivalência entre MR-MPS-SW e CMMI-DEV.

Fonte: extraído de SOFTEX (2014).



Em relação à certificação, o nível mais baixo passível de certificação no modelo CMMI-DEV é o nível 2, enquanto no modelo MPS-SW é o nível G. Já o nível mais alto passível de certificação em ambos modelos consiste nos seus respectivos níveis finais (5 e A, respectivamente). Atualmente existe um acordo entre o SEI (Software Engineering Institute) e a Softex para avaliação e certificação conjunta dos modelos MPS-SW e CMMI-DEV. Esse tipo de avaliação é indicado para empresas que necessitam certificação em ambos modelos e pretendem reduzir custos e tempo. A avaliação pode ser feita de forma simultânea para ambos modelos, exigindo dos colaboradores conhecimento pleno dos dois modelos. As avaliações e certificações individuais dos modelos também são facilitadas caso se tenha uma certificação de um deles e se deseje uma do outro modelo posteriormente.

Para este trabalho, são considerados os níveis G e F do modelo MPS-SW. Essa escolha se justifica pelo grande número de empresas certificadas no MPS-SW nesses dois níveis, se comparado aos outros cinco níveis do modelo (E ao A). Como referência, até o final de 2014 as empresas certificadas nos níveis G e F representavam 87% do total acumulado de 621 certificações, sendo o nível G responsável por 58% desse montante e o nível F por 29%. Somente no ano de 2014 foram certificadas 83 empresas, das quais 42 foram no nível G, 28 no nível F, 2 no nível E, 10 no nível C e 1 no nível A. Somente em 2015 (até final de junho), 19 empresas foram certificadas no nível G, 8 no nível F e 1 no nível E (SOFTEX, 2015).

Considerando a **Tabela 3** já apresentada, a **Tabela 4** evidencia os processos dos níveis G e F, incluindo a quantidade de resultados esperados em cada processo, com a representação da identificação dos resultados esperados entre parênteses. Os processos de cada nível indicam onde a organização deve investir mais esforços para melhorias.

Tabela 4 - Processos dos níveis G e F do modelo MPS-SW.

Fonte: extraído de SOFTEX (2012b).

Processo			
Nível	Sigla	Identificação	# Resultados esperados
F	MED	Medição	7 (MED1, ..., MED7)
	GQA	Garantia da qualidade	4 (GQA1, ..., GQA4)
	GPP	Gerência de portfólio de projetos	8 (GPP1, ..., GPP8)
	GCO	Gerência de configuração	7 (GCO1, ..., GCO7)
	AQU	Aquisição	8 (AQU1, ..., AQU8)
G	GRE	Gerência de requisitos	5 (GRE1, ..., GRE5)
	GPR	Gerência de projetos	19 (GPR1, ..., GPR19)

Segundo Pizzoleto e Oliveira (2013), as mudanças causadas na empresa para a adaptação dos seus processos aos dos níveis G e F são complexas e impactantes. Há necessidade da padronização dos processos de gestão da qualidade, bem como da organização dos processos de produção do software. São envolvidos diretamente recursos humanos, tecnológicos e políticos da empresa. As mudanças requerem formalização dos processos (PIZZOLETO; OLIVEIRA, 2013).

Portanto, optar por um modelo de qualidade consiste em um grande desafio, pois requer das empresas grande reestruturação organizacional, e, conseqüentemente, altos investimentos (PIZZOLETO; OLIVEIRA, 2013). Além disso, iniciar os estudos dos guias para implementar o MR-MPS-SW exige certos desafios. A forma textual dos guias, somada à abrangência e profundidade do conteúdo, contribui para dificultar a uniformização desse entendimento (PIZZOLETO, 2013). A **Figura 6** apresenta um exemplo do texto do guia MPS-SW para implementação do nível G, onde o usuário buscou por uma palavra-chave ou termo.

Figura 6 - Trecho do texto do guia de implementação do nível G do MR-MPS-SW.

Fonte: extraído de SOFTEX (2013, pág. 9).

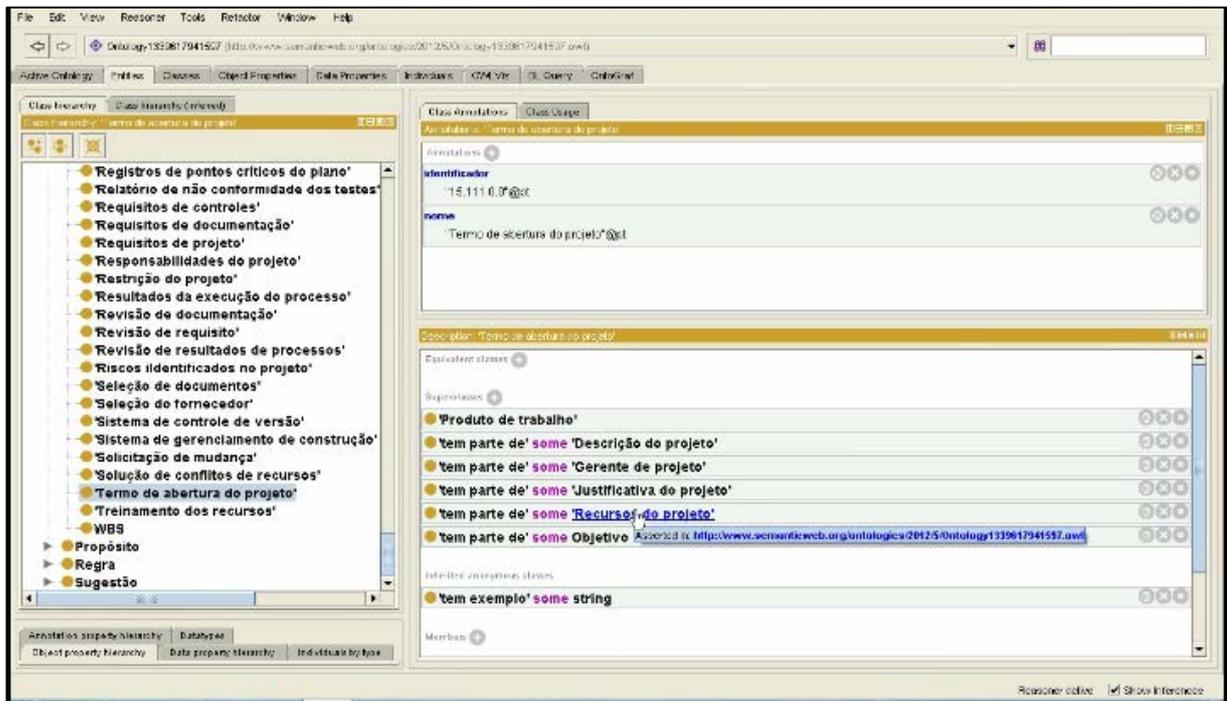
O desenvolvimento do plano do projeto inclui: identificar e estimar o escopo, os produtos de trabalho e as tarefas do projeto; estabelecer recursos necessários; identificar e analisar riscos do projeto; estabelecer compromissos; e definir cronograma de execução baseado no ciclo de vida definido para o projeto. O plano do projeto estabelece a base de execução e controle para as atividades do projeto junto aos seus interessados (especialmente o cliente). Todos os interessados devem estar comprometidos com ele.

Devido à forma da estrutura dos textos dos guias de implementação do MPS-SW, para facilitar o entendimento para fins deste trabalho, contou-se com o apoio da da ontologia proposta por Pizzoleto (2013) para os níveis G e F do modelo. De forma simplificada, uma ontologia é utilizada para descrever um determinado domínio, relacionando-o a alguns atributos e relacionamentos (GUARINO, 1998). A ontologia proposta por Pizzoleto (2013) tem como objetivo melhorar a compreensão do modelo MPS-SW e contribuir para uniformizar o entendimento do conteúdo dos guias de implementação dos níveis G e F. Dessa forma, pode-se apoiar as empresas implementadoras do modelo MPS-SW, principalmente as mPME, que possuem restrições técnicas e financeiras. Segundo Pizzoleto, a ontologia enriqueceu o conteúdo dos guias com conceitos constantes no guia de conhecimento de Gerência de Projetos, conhecido como PMBOK (*Project Management Body of Knowledge*). Além disso, a ontologia agregou conhecimento de consultores do MPS-SW e indicadores relacionados a três perspectivas do modelo BSC.

Segundo Pizzoleto (2013), as ontologias publicadas podem ser encontradas no seguinte repositório público: http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library. A ontologia pode ser visualizada através do sistema Protégé, conforme exemplificado na **Figura 7**.

Figura 7 - Exemplo de ontologia no sistema Protégé.

Fonte: extraído de Pizzoleto (2013).



Convém observar que, embora alguns guias de implementação do MPS-SW requeiram que as empresas de software possuam uma sistemática para controle da qualidade de seus processos e produtos, eles não são claros em relação aos processos mais significativos para medição. Dessa forma, o APÊNDICE A apresenta uma visão resumida desses processos em formato de mapa mental, incluindo um conjunto de indicadores para os processos. Esse conjunto inclui indicadores constantes na ontologia de Pizzoleto (2013) e outros adicionais, elaborados pelo autor deste trabalho em conjunto com Pizzoleto.

2.4 TRABALHOS RELACIONADOS AO CAPÍTULO

Na literatura foram encontrados vários trabalhos sobre indicadores de processos, porém poucos trabalhos associando indicadores a modelos de qualidade de software. A

seguir são apresentados alguns desses trabalhos.

Para Pedroso (2010), a necessidade de indicadores vem se tornando imprescindível, pois é através deles que as empresas podem verificar o que não está de acordo com os objetivos planejados. A utilização de processos de medição e de avaliação de desempenho apresenta uma série de vantagens competitivas. Esses processos estão presentes em modelos de qualidade de software, visando gerar informações e indicadores que sustentem os objetivos estratégicos do negócio. Pedroso (2010) complementa que, independentemente do modelo de maturidade usado como base para definir o processo de medição, se faz necessário que os mesmos sejam considerados estratégicos. Caso contrário, a execução poderá ser comprometida pela falta de foco nos objetivos de análise.

Moreira et al. (2009), por sua vez, apontam aspectos para viabilizar a utilização de conceitos de medição de desempenho e indicadores, de forma a contribuir com a melhoria dos processos de desenvolvimento de software de uma organização. Eles comentam a importância de se conhecer bem os processos, de modo a respeitar as características e restrições da organização quanto à sua estrutura, tamanho e domínios de aplicação em que atuam.

Santos et al. (2009) apresentam as dificuldades encontradas na implantação direta do nível E do modelo MPS-SW em uma instituição de pesquisa. Os autores salientam que a maior dificuldade foi entender os guias do modelo MPS-SW, uma vez que os mesmos não descrevem “o que fazer”. Eles comentam a importância de se conhecer os processos para se respeitar as características e restrições da organização, dependentes de fatores como: estrutura, tamanho e domínios de aplicação em que atuam. Segundo os autores, o processo de medição foi fundamental para o êxito da implantação do nível E do modelo MPS-SW em uma determinada empresa considerada no artigo. Isso porque a medição dos resultados de todos os processos chave e a análise dos indicadores possibilitaram que se observasse a evolução e melhorias gradativas dos processos. Consequentemente, as medições feitas possibilitaram acompanhar como as metas globais da organização foram alcançadas. Por fim, os autores apresentam os resultados de diversos indicadores utilizados em seus processos de desenvolvimento de software.

Para Ojha (2014), os principais fatores que devem ser levados em conta no desenvolvimento de software são: produtividade, qualidade, eficiência e desempenho dos processos. Contudo, geralmente são enfrentados três grandes desafios que dificultam a efetivação das análises desses fatores: (1) falta de dados históricos; (2) diversidade das fontes de dados; (3) medições errôneas. Para mitigar esses problemas, o autor defende a criação e manutenção de um conjunto de KPIs. Além de beneficiarem os processos de

desenvolvimento do software, esses indicadores contribuem para a entrega de um software com bom desempenho, fácil de usar, eficiente e eficaz. Para se criar uma base de dados históricos, Ojha (2014) recomenda a definição clara de KPIs, de modo que as pessoas envolvidas compreendam e colem os dados (valores) para os KPIs. Em relação à diversidade de fontes de dados dos indicadores (projetos anteriores similares ou projetos paralelos), Ojha (2014) salienta que deve ser encontrada uma forma de garantir o controle da origem dos dados. Esse controle pode ser feito manualmente ou por meio de uma ferramenta automatizada de software, seja ela adquirida ou desenvolvida para uso interno. Em relação à medição errônea, caso o processo de coleta de dados seja manual, se faz necessário que os indivíduos que realizam a medição sejam sinceros a ponto de informar seus erros, quando houver. Entretanto, o autor menciona que a medição errônea pode ser controlada por meio de monitoramento e avaliação regular dos KPIs.

Ojha (2014) também traz informações que mostram a importância de se adotar um modelo de qualidade de processos de software, como o CMMI, de modo que seja utilizado como uma referência de processos. Observa-se que no caso deste trabalho, o modelo de qualidade de processos utilizado como referência é o MPS-SW, níveis G e F, como apresentado na seção anterior. Para Ojha (2014), o uso de KPIs possibilita a identificação de problemas que interferem negativamente no alcance dos resultados esperados dos processos. Além disso, o uso de KPIs contribui para propostas de soluções para se atingir esses resultados esperados.

De modo geral, os trabalhos pesquisados evidenciaram preocupações quanto à identificação “do que medir”, bem como quanto ao custo desse processo de medição. As preocupações se estenderam ao alinhamento das medições com os objetivos estratégicos da organização, reforçando a importância dos indicadores no processo de avaliação do desempenho organizacional. Convém ressaltar que alguns trabalhos, como Santos et al. (2009) e Moreira et al. (2009), trazem comentários sobre a dificuldade que se tem para entender a forma textual dos guias do modelo MPS-SW.

3 APRENDIZAGEM DE MÁQUINA E TÉCNICAS SELECIONADAS

No contexto de Inteligência Artificial, “aprendizagem de máquina” ou “aprendizado de máquina” é uma área que envolve a construção de sistemas capazes de adquirir conhecimento automaticamente. Essa área preza pelo desenvolvimento de algoritmos que usam a “experiência” adquirida, semelhante ao comportamento humano de aprendizagem, para produzir resultados, que dependem de decisões tomadas, automaticamente, pelo próprio algoritmo. Assim, após o aprendizado adquirido, o algoritmo pode tomar decisões inteligentes com base em exemplos de decisões para dados de entrada (MITCHELL, 1997; NATARAJAN, 1991).

De modo geral, o uso de algoritmos de aprendizagem de máquina requer a análise de uma quantidade significativa de amostras, com características diferentes. Essa análise objetiva ensiná-lo a resolver problemas diferentes, dentro de um determinado contexto. Esse contexto pode possuir características potencialmente mutáveis ao longo do tempo ou ter características diferentes dependendo do tipo de aplicação e uso (ALPAYDIN, 2004).

Entre as categorias de aprendizado de máquina, destacam-se: aprendizado supervisionado, não supervisionado e semi-supervisionado. Cada categoria possui diversas técnicas de aprendizado de máquina, indicadas a determinados tipos de problemas. Entre essas técnicas, podem ser mencionadas as Redes Neurais Artificiais (RNAs) e as técnicas baseadas em Inteligência de Enxames (IE). Segundo Haykin (1994), RNAs consistem em métodos computacionais baseados em modelos matemáticos inspirados na estrutura neural de organismos inteligentes, cuja aquisição do conhecimento é realizada através da experiência. Já as técnicas baseadas em IE são inspiradas no comportamento coletivo de sistemas auto-organizados, distribuídos, flexíveis, autônomos e dinâmicos, formados por uma população de agentes computacionais simples (BONABEAU; DORIGO; THERAULAZ, 1999; SERRAPIÃO, 2009). Esses agentes se comunicam entre si e têm a capacidade de captar e modificar seu ambiente de maneira local.

De modo geral, tanto no aprendizado supervisionado como no semi-supervisionado “ensinar” consiste, basicamente, das seguintes etapas: seleção dos indicadores de

monitoramento, análise dos dados coletados e rotulagem desses dados por especialistas (humanos), para classificar os resultados em certa escala de satisfatório a insatisfatório. A quantidade de dados rotulados é que diferencia os dois tipos de aprendizado.

Nesse contexto, este trabalho utiliza uma técnica de aprendizado semi-supervisionado para análise de indicadores de desempenho na produção de software. Essa categoria de aprendizado combina dados rotulados com dados não rotulados, reduzindo custos e esforço na rotulagem por parte dos especialistas humanos, sem comprometer os benefícios no caso deste trabalho.

Devido às dificuldades encontradas pelos especialistas no processo de rotulagem dos dados, este trabalho propõe que sejam utilizadas técnicas de visualização de informação como apoio a esses especialistas. O uso dessas técnicas possibilita uma representação visual dos dados rotulados, contribuindo para que os especialistas identifiquem possíveis falhas no momento da rotulagem dos dados.

Face ao exposto, as seções 3.1 a 3.3 apresentam as principais características dos aprendizados supervisionados, não-supervisionados e semi-supervisionados, respectivamente, proporcionando uma visão geral dessas categorias.

Considerando que indicadores podem ser analisados através de RNAs, a seção 3.4 apresenta uma visão geral sobre esse tipo de técnica. De modo geral, RNAs de aprendizado supervisionado requerem treinamento da rede através de dados rotulados, para aprender como classificar os demais dados.

A seção 3.5, por sua vez, apresenta o algoritmo de aprendizado semi-supervisionado utilizado neste trabalho. Trata-se do algoritmo de competição e cooperação entre partículas, baseado em grafos, elaborado por Breve (2010). Para a escolha desse algoritmo foi levado em consideração seu tipo de aprendizado e sua baixa complexidade computacional em relação a outros algoritmos baseados em grafos. Observa-se que o mecanismo deste algoritmo não é o de uma RNA e sim de uma técnica de IE.

Considerando que no aprendizado semi-supervisionado são necessárias poucas amostras rotuladas, essas amostras precisam ser confiáveis. Portanto, este trabalho recomenda o uso de técnicas de visualização de informação para auxiliar o processo de rotulagem dos dados. As três técnicas avaliadas e adotadas neste trabalho estão apresentadas na seção 3.6: (1) Projeção por mínimos quadrados ou LSP (*Least Square Projection*); (2) Escalonamento multidimensional clássico ou CMDS (*Classical Multidimensional Scaling*); (3) Coordenadas paralelas ou PC (*Parallel Coordinates*).

Entre os trabalhos encontrados nas literaturas que aplicam técnicas de aprendizado de máquina para previsão e monitoramento de indicadores, alguns foram selecionados e

estão apresentados na *seção 3.7*.

3.1 APRENDIZADO DE MÁQUINA SUPERVISIONADO

No aprendizado supervisionado existe o papel de um especialista (entidade externa), que apresenta, ao algoritmo, alguns conjuntos de padrões de entradas e seus correspondentes padrões de saída. A saída pode ser um valor contínuo ou pode predizer um rótulo de classe para o objeto de entrada.

No caso de uma RNA, por exemplo, na fase de treinamento da rede, a cada entrada de dados, o especialista indica, de maneira explícita, se a resposta calculada é boa ou ruim (processo de rotulagem dos dados). Então, a resposta fornecida pela rede é comparada à resposta esperada. Caso o resultado seja diferente do desejado, um erro é informado à rede para que os ajustes possam ser realizados, a fim de melhorar as futuras respostas.

Nesta categoria de aprendizado, o objetivo do algoritmo consiste em obter uma função que seja capaz de predizer a saída para qualquer entrada válida, após ter visto um número suficiente de exemplos de treinamento. Para atingir esse objetivo, o algoritmo deve ter capacidade de generalização, de modo a prever, de maneira aceitável, a saída dos dados ainda não vistos. Os algoritmos desta categoria deduzem uma função a partir dos dados de treinamento.

Esses algoritmos são aplicados nas mais diversas áreas, sendo geralmente indicados à solução de problemas relacionados à mineração de dados, reconhecimento de padrão, entre outros (MITCHELL, 1997; ALPAYDIN, 2004).

3.2 APRENDIZADO DE MÁQUINA NÃO SUPERVISIONADO

Em algoritmos de aprendizado não supervisionado não existe a entidade externa para realizar o processo de rotulagem. Eles buscam determinar como os dados estão organizados, baseados somente nos padrões de entrada, sem rótulos ou valores de saída.

Algoritmos desta categoria processam as entradas disponíveis e tentam, progressivamente, estabelecer representações internas para codificar características e classificá-las automaticamente. Essa tentativa de estabelecer as representações internas é feita por meio da detecção da singularidade nas amostras de entrada, geralmente realizada através de agrupamento de dados. Desta forma, padrões dentro de um grupo são mais

similares entre si do que a padrões pertencentes a outros grupos.

Há diversos métodos para agrupamento dos dados no aprendizado não supervisionado. Os métodos baseados em otimização, por exemplo, têm a característica de maximizar as medidas de similaridade entre os elementos de um grupo e/ou maximizar as medidas de similaridade entre os diferentes grupos. Um dos mais conhecidos algoritmos dessa categoria é o algoritmo de K-médias (MACQUEEN, 1967), que particiona o conjunto de dados em K grupos com base em uma medida de similaridade.

Outro método de agrupamento de dados é o método hierárquico. Sua característica consiste na construção de uma árvore hierárquica, a qual é dividida em sucessivas junções ou separações de grupos (JAIN et al., 1999). A divisão é iniciada com um grupo contendo todos os elementos, formando a sequência de partições a partir da divisão de cada grupo a cada iteração.

Métodos de agrupamento de dados baseados em densidade, por sua vez, são apoiados na ideia de que os grupos consistem em regiões de alta densidade, separadas por regiões de baixa densidade (YIP et al., 2006).

Algoritmos de aprendizado não supervisionados vêm sendo utilizados de diversas maneiras, inclusive em aplicações para tomadas de decisão (JAIN et al., 1999).

3.3 APRENDIZADO DE MÁQUINA SEMI-SUPERVISIONADO

A categoria de aprendizado semi-supervisionado se situa entre o aprendizado supervisionado e o não supervisionado, pois faz uso tanto de dados rotulados quanto de dados não rotulados para o treinamento. Em muitos casos, o uso de alguns dados rotulados em meio aos dados não rotulados melhora consideravelmente a precisão do aprendizado (BREVE, 2010).

Atualmente, as bases de dados dão suporte a grande quantidade de dados. Rotular dados de bases grandes, para uso em algoritmos de aprendizado supervisionado, tem se tornado um processo praticamente inviável. Isso porque o processo de rotulagem é frequentemente caro e moroso, sendo imprescindível o envolvimento de especialistas humanos. Por outro lado, o fato de não haver qualquer rotulagem nos algoritmos não supervisionados faz com que se ignore informações importantes sobre os rótulos dos itens de dados.

Assim, métodos de aprendizado semi-supervisionados podem superar os problemas mencionados nas outras duas categorias: (1) oneração do tempo e custo financeiro, no caso

do aprendizado supervisionado; (2) desconsideração das informações de rótulo, no caso do aprendizado não supervisionado. Para isso, os métodos semi-supervisionados combinam poucos itens de dados rotulados com uma grande quantidade de dados não rotulados. Essa combinação visa produzir melhores classificadores, ao mesmo tempo que requer menor esforço humano (ZHU, 2005; CHAPELLE et al., 2006).

“Alguns algoritmos de aprendizado de máquina tentam eliminar a necessidade de especialistas humanos na análise de dados, enquanto outros adotam uma estratégia de colaboração entre humanos e máquinas. A intuição humana não pode ser inteiramente eliminada, pois o projetista do sistema precisa especificar como os dados serão representados e quais mecanismos serão usados para caracterizar os dados.”

Os algoritmos semi-supervisionados são indicados para diversas aplicações, como, por exemplo: reconhecimento de fala, classificação de páginas da Internet, sequenciamento de proteínas, entre outras (CHAPELLE et al., 2006). O aprendizado semi-supervisionado acontece da seguinte forma: tipicamente, existe um conjunto de dados $X = \{x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_n\} \subset \mathbb{R}^m$ e um conjunto de rótulos $L = \{1, 2, \dots, c\}$, sendo que os primeiros l pontos $x_i (i \leq l)$ tem rótulos conhecidos, ou seja, $y_i \in L$, e o restante dos pontos $x_u (l < u \leq n)$ tem os rótulos desconhecidos, ou seja, $y_u = \emptyset$. O objetivo é designar um rótulo para cada amostra não rotulada (BREVE, 2010).

A forma mais primitiva de aprendizado semi-supervisionado é o “auto-treinamento” ou auto-aprendizado. Semelhante ao aprendizado supervisionado, um classificador é treinado, porém com poucos dados rotulados, adicionando, ao conjunto de treinamento, os dados não rotulados, cujos rótulos preditos possuem maior confiança. Na sequência, o classificador é re-treinado com esse novo conjunto de dados de treinamento, sendo o processo repetido até que todos os dados sejam rotulados (ZHU, 2005; CHAPELLE et al., 2006).

Outra forma antiga de aprendizado semi-supervisionado são os modelos generativos. A partir de uma distribuição com mistura identificável, como um modelo de mistura Gaussiano (ZHU, 2005), por exemplo, contendo uma grande quantidade de dados não rotulados, é possível identificar os componentes da mistura. Dessa forma, é necessário apenas um exemplo rotulado em cada componente para determinar completamente a mistura da distribuição.

Vale observar que alguns falham ao identificar classes de formas irregulares, como é o caso do método de máquina de vetores de suporte transdutiva ou TSVM (*Transductive Support Vector Machines*) (VAPNIK, 2008). Outros métodos, baseados em grafos, possuem

ordem de complexidade computacional muito alta, o que torna seu uso limitado a pequenas bases de dados (ZHU, 2005). Essa alta complexidade caracteriza uma séria deficiência, pois as técnicas de aprendizado semi-supervisionado são aplicadas, normalmente, a base de dados com muitos dados não rotulados. Contudo, Chapelle et al. (2006) afirmam que a área de pesquisa mais ativa em aprendizado semi-supervisionado nos últimos anos tem sido a de métodos baseados em grafos.

As técnicas de agrupar e rotular consistem em utilizar técnicas de agrupamento tradicionais para encontrar a distribuição das classes de toda a base de dados (DARA et al., 2002). Em seguida, cada grupo encontrado pode ser identificado com o rótulo presente na maioria dos dados pré-rotulados. Existem, ainda, outros tipos de técnicas, como a de separação de baixa densidade e as técnicas de cotreinamento. Entretanto, vale à pena ressaltar as técnicas de agrupar e rotular dos métodos baseados em grafos.

Os métodos baseados em grafos são representados através de nós e arestas situados em um grafo. As arestas têm seu peso relacionado com a distância entre os pares de nós. Já os nós rotulados são utilizados para propagar informações de rótulos aos demais. Esses métodos utilizam um grafo $G = (V, E)$, sendo $V = \{v_1, v_2, \dots, v_n\}$ o conjunto de nós onde cada vértice v_i corresponde a uma amostra x_i , e $E \rightarrow \mathbb{R}$ o conjunto de ligações (v_i, v_j) , cujos pesos podem ser representados por uma matriz de adjacência W . Essa matriz de adjacência W_{ij} indica a similaridade entre v_i e v_j , normalmente definida através do uso de uma função Gaussiana. Vale salientar que é possível utilizar um grafo sem peso ou conectar cada nó apenas aos seus k vizinhos mais próximos, por meio de alguma medida de distância.

Diversos métodos baseados em grafos funcionam de maneira iterativa, ou seja, a estrutura do grafo é utilizada para espalhar, gradualmente, os rótulos dos nós rotulados para todo o restante do grafo. A cada iteração, cada nó espalha as informações de rótulos que possui para seus vizinhos até que haja convergência.

A propagação dos rótulos pode ocorrer de diversas formas, mas, basicamente, respeitam as seguintes regras (ZHU e GHARAMANI, 2002):

- nós rotulados iniciam com seus respectivos rótulos (1 ou -1);
- nós não rotulados iniciam com 0;
- nós propagam seus rótulos repetidamente até convergência.

Como exemplo, pode-se ter a seguinte situação: a cada passo, um nó i pode receber

a contribuição de seus vizinhos j , ponderados pelo peso normalizado da aresta (i, j) , e uma pequena contribuição adicional de seu valor inicial.

3.4 REDES NEURAIS ARTIFICIAIS

Redes Neurais Artificiais (RNAs) consistem em técnicas computacionais baseadas em modelos matemáticos inspirados na estrutura neural de organismos inteligentes, cuja aquisição do conhecimento é realizada através da experiência (HAYKIN, 1994).

Geralmente, as RNAs são utilizadas no desenvolvimento de sistemas computacionais capazes de reconhecer e classificar padrões, resolver problemas complexos em um determinado escopo, realizar processos indutivos e dedutivos, entre outros. Em linhas gerais, existem diversas vantagens na utilização de RNAs, dentre as quais se pode citar as seguintes (RAGHU; POONGODI; YEGNANARAYANA, 1995):

- os relacionamentos funcionais entre os padrões de entrada e de saída podem ser capturados por uma RNA, sem a necessidade de conhecimento ou descrição explícita desses relacionamentos;
- não há necessidade de realizar hipóteses sobre as distribuições estatísticas dos padrões de entrada; e
- mesmo que hajam alguns elementos ou conexões degradadas, seu desempenho pode ser pouco afetado.

A propriedade mais importante de uma RNA está na capacidade de aprender por meio de exemplos, podendo fazer inferências sobre o que aprendeu (RAGHU; POONGODI; YEGNANARAYANA, 1995). Essa propriedade faz com que a RNA melhore gradativamente seu desempenho.

Convém destacar que uma RNA é caracterizada, basicamente, por três elementos:

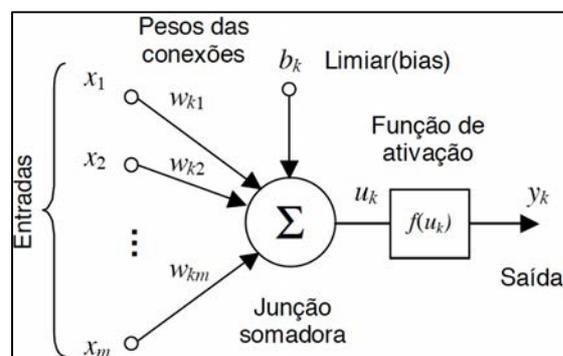
1. um padrão de conexão entre os neurônios;
2. um método de determinação dos pesos e conexões (treinamento ou aprendizagem);
3. uma função de ativação, responsável pela saída da rede.

Uma RNA possui comportamento baseado nos grupos de neurônios do cérebro

humano, que recebem e transmitem informações através dos dendritos e axônios, respectivamente (KOVACS, 1996). A partir do modelo original, Ferneda (2006) criou a imagem de um neurônio artificial genérico, ilustrado na **Figura 8**. Ao apresentar, à RNA, um conjunto específico de entradas ($b_k, x_1, x_2, \dots, x_m$) e suas respectivas saídas (y_k), ela é capaz de auto ajustar seus pesos sinápticos ($w_{k1}, w_{k2}, \dots, w_{km}$). O ajuste das conexões é obtido por meio do produto interno entre o vetor de aspecto (conjunto de entradas) e o vetor de peso, adotando uma determinada função de ativação ($f(u_k)$) como critério de treinamento (e posterior análise).

Figura 8 - Representação de um neurônio genérico.

Fonte: extraído Ferneda (2006).



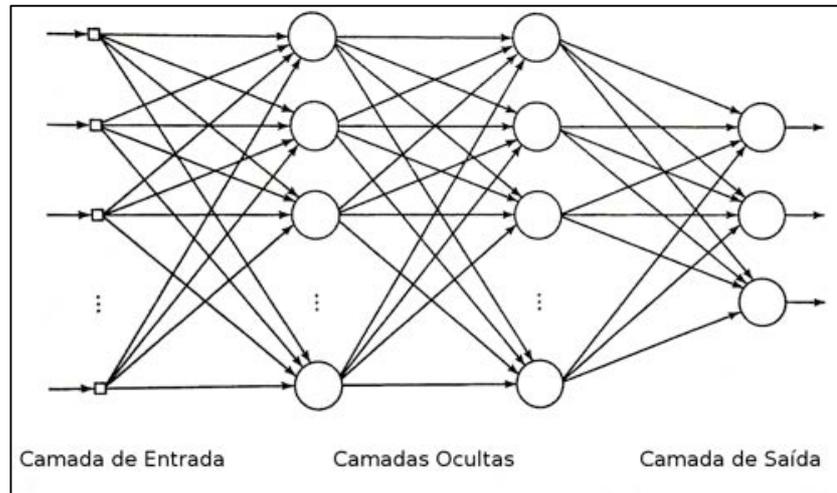
Os treinamentos das RNAs consistem no mapeamento do relacionamento funcional existente entre as entradas e as saídas. Após o treinamento, a rede deve ser capaz de tornar genérico o comportamento do processo, no momento em que outras entradas - diferentes das que foram utilizadas durante o treinamento - sejam apresentadas (BRAGA; CARVALHO; LUDEMIR, 2000).

Segundo Ferneda (2006), de maneira simplificada, uma RNA pode ser vista como um grafo, onde os nós são os neurônios e as ligações fazem a função das sinapses, como exemplificado na **Figura 9**.

Cada camada da RNA é responsável por uma especialização do aprendizado. A camada de entrada é a responsável pela entrada dos dados, enquanto a camada de saída é a que fornece os resultados obtidos na rede. Uma RNA pode ter uma ou várias camadas intermediárias (ocultas), dependendo do tipo de aplicação. Os resultados obtidos durante o treinamento serão os responsáveis por determinar a quantidade de camadas ocultas necessárias.

Figura 9 - Representação simplificada de uma Rede Neural Artificial (RNA).

Fonte: extraído Ferneda (2006).



3.5 ALGORITMO DE COMPETIÇÃO E COOPERAÇÃO ENTRE PARTÍCULAS

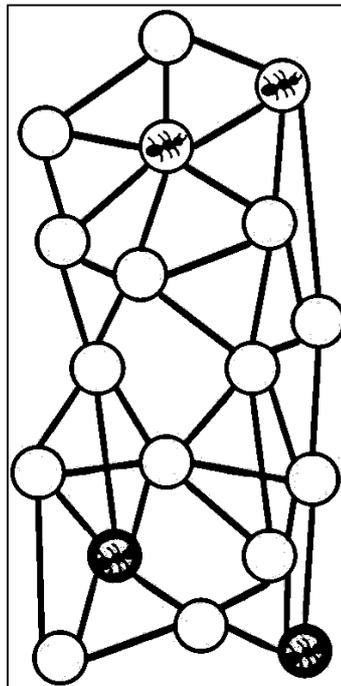
O algoritmo de competição e cooperação entre partículas foi elaborado por Breve (2010) e está inserido na categoria de métodos de aprendizado semi-supervisionado. Consiste em um grafo de redes conexas, o qual possui várias partículas caminhando pela rede. O comportamento das partículas é muito similar ao comportamento das formigas, que atuam em times e de maneira colaborativa. Assim, Breve (2010) apresenta uma metáfora de partículas baseada no comportamento das formigas, ilustrada na **Figura 10**. As partículas se organizam em forma de times, onde partículas do mesmo time caminham pela rede de maneira cooperativa, a fim de propagarem seus rótulos. Já as partículas de times diferentes competem umas com as outras para determinar as bordas da classe, rejeitando partículas intrusas.

Segundo BREVE et al. (2009), esse algoritmo difere de outros modelos de aprendizado semi-supervisionado baseado em grafos. Nesses métodos mais tradicionais, os rótulos são espalhados pela rede de maneira global. Isso significa que, a cada iteração do algoritmo, a informação é propagada de todos os nós para todos os nós, respeitando os respectivos pesos das arestas. Já o método de Breve (2010), a propagação do rótulo acontece de maneira local, de modo que, a cada passo do algoritmo, toda partícula propaga

seu rótulo para um vizinho escolhido através da regra “aleatória¹ x gulosa²”. Dessa forma, cada partícula visita apenas uma porção dos nós que potencialmente pertence ao seu time (sub-rede), evitando que operações redundantes sejam realizadas. Segundo Breve (2010), esse algoritmo acaba sendo mais eficiente que os modelos de grafos tradicionais, pois possui uma ordem de complexidade computacional bem mais atraente³, sendo apenas $O(n)$ na maioria dos casos, e não ultrapassando $O(n^2)$ no pior caso. Breve et al. (2009) definiu o algoritmo PCC com pseudo-código, conforme apresentado na **Figura 11**. No texto seguinte desta seção, será apresentada uma visão geral do funcionamento do algoritmo.

Figura 10 - Metáfora de partículas baseada no comportamento de formigas.

Fonte: extraído de Breve (2010).



¹ Caminhada aleatória: cada partícula escolhe qualquer vizinho para visitar de maneira aleatória, sem se preocupar com os níveis de domínio ou distância do nó-casa. Trata-se de um movimento para exploração e aquisição de novos nós.

² Caminhada gulosa: cada partícula visita os nós que estão mais próximos de seus nós-casa, principalmente aqueles que já são dominados por seu próprio time. Trata-se de um movimento para defesa do território de seu time.

³ Apenas como referência, a maioria dos métodos de aprendizado semi-supervisionado baseados em grafo possui ordem de complexidade cúbica, ou seja, $O(n^3)$ (ZHU, 2005).

Figura 11 - Pseudo-código: algoritmo de competição e cooperação entre partículas.

Fonte: extraído de Breve et al. (2009).

```

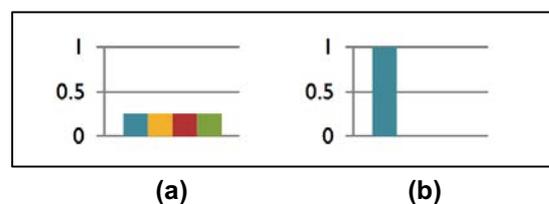
1  Construa uma matriz de adjacências W;
2  Defina os níveis de domínio dos nós;
3  Defina as posições iniciais das partículas de acordo com seus respectivos
   nós-casa;
4  Defina a força das partículas e tabelas de distância;
5  Repita
6      Para cada partícula Faça
7          Selecione entre a regra aleatória ou gulosa com as probabilidades
           previamente definidas;
8          Selecione o nó alvo para movimento aleatório ou guloso;
9          Atualize os níveis de domínio do nó alvo;
10         Atualize a força da partícula;
11         Atualize a tabela de distância das partículas;
12     Até que o critério de parada seja satisfeito;
13     Rotule cada item de dado não rotulado pelo time com maior nível de domínio;

```

Diferentemente dos algoritmos de aprendizado supervisionado, que implementam técnicas de RNA, onde, o treinamento da rede é realizado numa etapa separada e anterior à análise, o algoritmo PCC recebe todos os dados de uma única vez, rotulados e não rotulados, e utiliza estes dados para montar um grafo não direcionado e sem pesos. Todo nó dessa rede possui um vetor de elementos, responsável por representar os níveis de dominância de cada time de partículas sobre o nó. Inicialmente, cada nó correspondente a uma amostra não rotulada terá seus níveis de domínio configurados com o mesmo valor $\frac{1}{c}$, onde c = número de classes / times. A **Figura 12 (a)** exemplifica essa situação inicial com quatro classes, ou seja, nós não rotulados, com níveis de domínio 0.25: [0.25 0.25 0.25 0.25]. Em seguida, um conjunto de formigas é colocado na rede, cada uma delas representando um item de dado rotulado, o qual é configurado para o valor mais alto. A **Figura 12 (b)** ilustra essa situação para as quatro classes, com nó rotulado como classe A: [1 0 0 0]. O subconjunto de partículas que tem o mesmo rótulo recebe o nome de “time”.

Figura 12 - Níveis de domínios iniciais: (a) amostra não rotulada; (b) amostra rotulada.

Fonte: extraído de Breve (2010).



Por meio da regra “aleatória x gulosa”, toda formiga escolhe um nó vizinho para visitar a cada iteração do algoritmo, conforme ilustrado na **Figura 13**. Quando uma formiga seleciona um nó vizinho, o nível de domínio de seu time é aumentado, enquanto os níveis

de domínio de outros times são diminuídos (**Figura 14**). Caso o nó a ser visitado seja de domínio de seu próprio time, a formiga se fortalece, aumentando seus níveis de domínio (**Figura 15**). Porém, a formiga se enfraquece caso o nó a ser visitado seja de domínio de outro time (**Figura 16**).

Figura 13 - Probabilidade de escolha da formiga - movimento guloso e aleatório.

Fonte: extraído de Breve (2010).

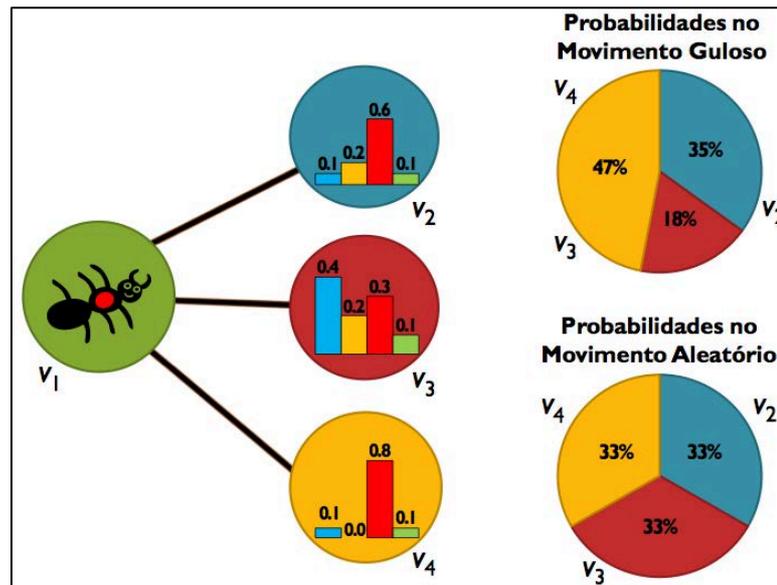


Figura 14 - Aumento do nível de domínio do próprio time e diminuição do nível dos outros times.

Fonte: extraído de Breve (2010).

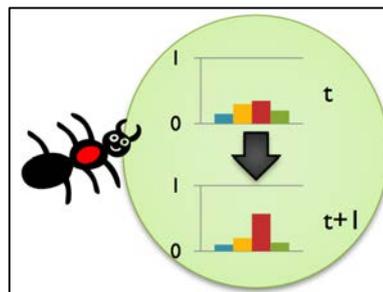


Figura 15 - Fortalecimento de níveis - nós do mesmo time.

Fonte: extraído de Breve (2010).

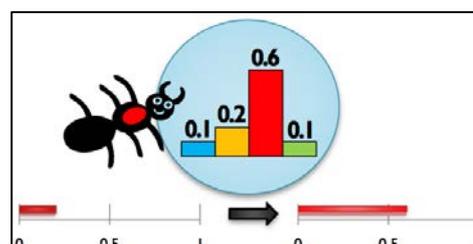
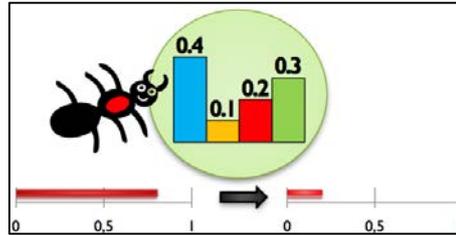


Figura 16 - Enfraquecimento de níveis - nós de outros times.

Fonte: extraído de Breve (2010).



É possível afirmar que cada formiga trabalha priorizando o domínio de suas respectivas vizinhanças (nós vizinhos). Para isso, elas têm um determinado nó como casa e cada uma mantém informações da distância entre suas respectivas casas e os outros nós da rede. Por se tratar de um método que inclui cooperação, as formigas também priorizam ajudar seus companheiros de equipe com suas vizinhanças, e, eventualmente, tentam invadir territórios de oponentes.

De maneira cooperativa, cada time tenta dominar a maior quantidade possível de nós, e, simultaneamente, tentam evitar a invasão de formigas de outros times. No final do processo iterativo, cada item de dado não rotulado será rotulado de acordo com o rótulo do time que possuir o maior nível de domínio.

3.6 VISUALIZAÇÃO DE INFORMAÇÃO

Técnicas de visualização de informação têm sido utilizadas como apoio à visualização de conjunto de dados em várias áreas de aplicação. Uma das abordagens de visualização de informação que vem sendo utilizada para este propósito são as técnicas de projeção multidimensional (TEJADA et al. 2003). Essas técnicas auxiliam a análise e a compreensão de conjuntos de dados, fornecendo uma representação gráfica desses dados.

Paiva et al. (2011) apresentam algumas vantagens para aplicação de técnicas de projeção multidimensional em mineração de dados visual, como ferramenta de apoio à classificação visual de conjuntos de dados. Entre as vantagens mencionadas, destaca-se seu uso como ferramenta para classificação manual, possibilitando ao especialista do domínio imediata exposição aos resultados de falso positivo, falso negativo, incompatibilidades e discrepâncias, permitindo nova rotulagem do conjunto de dados para corrigir e melhorar a classificação final.

Eler et al. (2009) apresentam uma ferramenta, conhecida como PEx-Image (*Projection Explorer for Images*), usada para análise da similaridade dos dados, baseada em técnicas de projeção multidimensional. Segundo os autores, a ferramenta PEx-Image auxilia as tarefas de comparação de características e de classificação orientadas para o usuário, permitindo melhor exploração para análise de dados.

As técnicas de projeção multidimensional procuram preservar as relações de similaridades do espaço original em um espaço de menor dimensão (ex.: 2D). Dessa maneira, é possível criar representações gráficas que auxiliem o usuário a explorar a similaridade entre as instâncias dos dados. Assim, pontos próximos no plano indicam instâncias com alto grau de similaridade.

Neste trabalho, foram estudadas três técnicas de projeção multidimensional: (1) Projeção por mínimos quadrados ou LSP (*Least Square Projection*); (2) Escalonamento multidimensional clássico ou CMDS (*Classical Multidimensional Scaling*); (3) Coordenadas paralelas ou PC (*Parallel Coordinates*). Para as técnicas LSP e CMDS foi utilizado o software PEx-Image, desenvolvido em Java e disponível para download em: <http://infoserver.lcad.icmc.usp.br/infovis2/PExImage>. Para a técnica PC foi utilizado um software próprio, também desenvolvido em Java, baseado no software Parvis (OLIVEIRA; SILVA; ELER, 2015).

A técnica LSP utiliza a vizinhança de pontos de controle para estabelecer as relações de similaridade entre as instâncias dos dados, evitando o cálculo de similaridade entre todo o conjunto de dados. Os resultados de sua aplicação geram grupos de pontos com grau de semelhança no espaço 2D. A **Figura 17** exemplifica uma aplicação da projeção LSP para um conjunto de documentos (artigos científicos) de quatro diferentes áreas, representadas na projeção pelas quatro diferentes cores. As características consideradas neste exemplo incluem: título, autores, resumo e referências científicas. Segundo Paulovich et al. (2008), essa técnica é mais rápida e mais precisa do que outros métodos de alta qualidade.

A técnica CMDS mostra a estrutura de dados de distância de modo semelhante a uma imagem geométrica. Essa técnica é baseada em modelos de dados lineares e, geralmente, adota a similaridade existente entre as instâncias do conjunto de dados como entrada. Apesar da restrição da característica linear, seus resultados nesse ambiente têm se mostrado relevantes dependendo da base de dados analisada. A **Figura 18** mostra um exemplo da aplicação da técnica CMDS em que se comparam as distâncias das rodovias entre as principais cidades dos Estados Unidos.

Figura 17 - Projeção LSP usada para classificar artigos científicos de quatro é diferentes áreas.

Fonte: extraído de Paulovich et al. (2008).

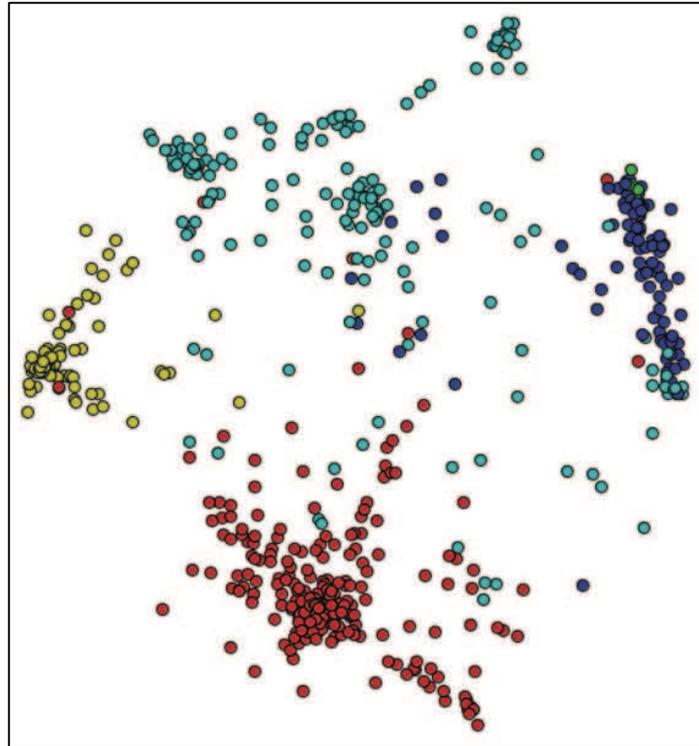
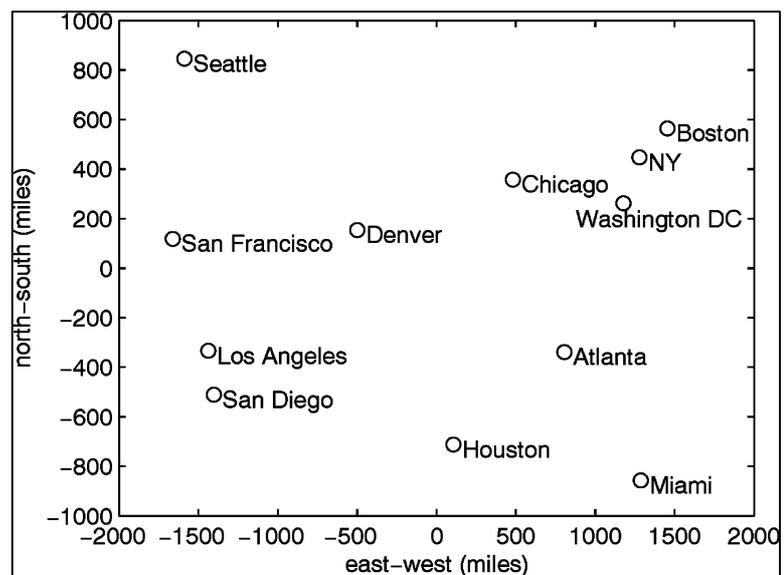


Figura 18 - Técnica CMDS usada para identificar as distâncias das rodovias entre as principais cidades dos EUA.

Fonte: extraído de Birchfield e Subramanya (2005).

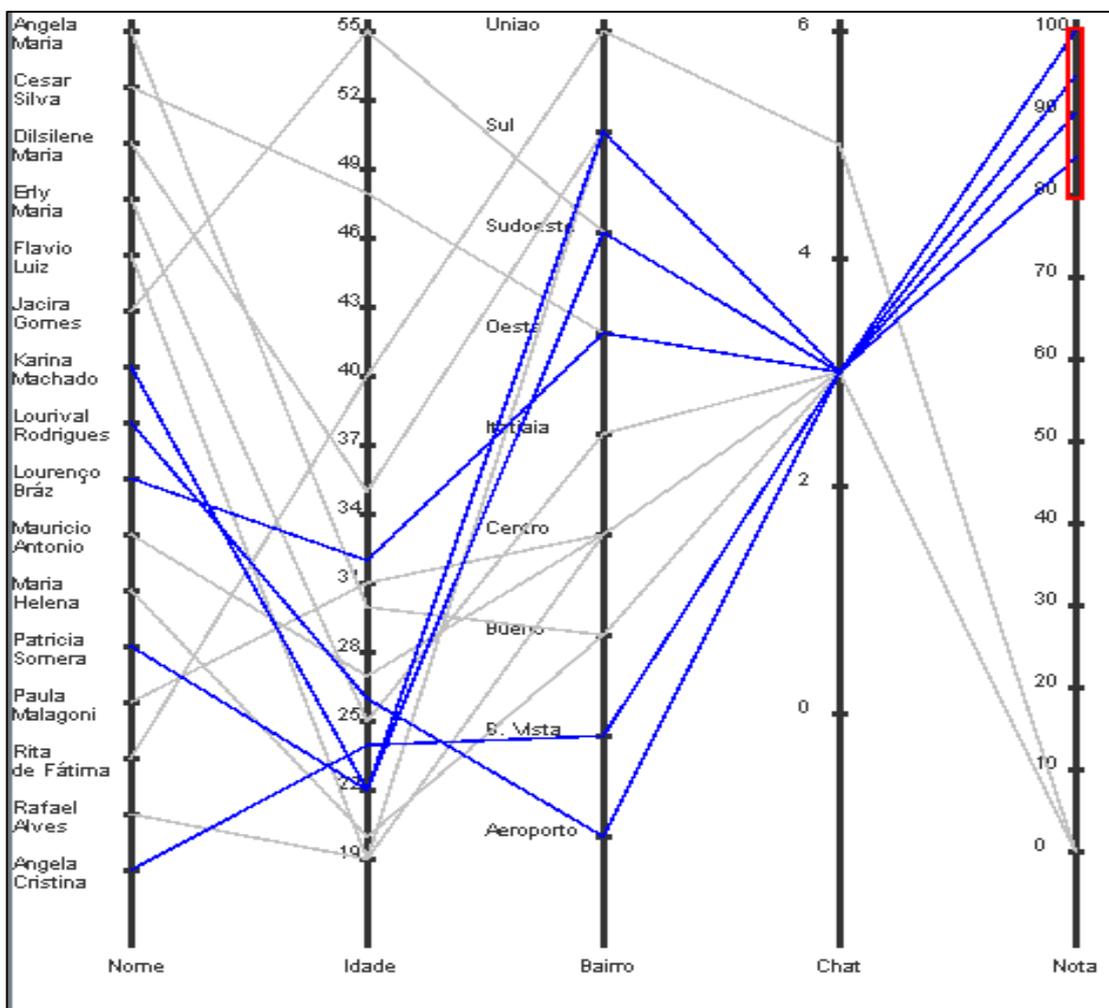


A técnica PC atribui um eixo para cada atributo (dimensão) e exibe esses eixos paralelamente em um plano. Cada instância de um conjunto de dados n-dimensional é representado por uma linha que atravessa os eixos paralelos. Assim, por meio das

“coordenadas paralelas” é possível visualizar o comportamento dos atributos (dimensões) das instâncias dos dados. Isso permite a comparação entre as instâncias das diferentes classes do conjunto. A **Figura 19** apresenta um exemplo da aplicação das coordenadas paralelas como ferramenta de suporte ao ensino à distância via Web, possibilitando a visualização de dados sobre a participação e o desempenho de alunos em um determinado curso. Nesse exemplo, os eixos representam os atributos dos alunos: nome, idade, bairro onde mora, número de acessos ao chat do curso e nota obtida; e cada instância (cada linha) é representada por dados de um determinado aluno.

Figura 19 - Técnica de coordenadas paralelas usada como suporte ao ensino a distância.

Fonte: extraído de Xavier (2004).



Experimentos utilizando as três técnicas de projeção multidimensional apresentadas foram realizados para avaliar a real contribuição na análise visual dos dados de indicadores durante o processo de rotulagem, antes da aplicação do algoritmo PCC. Esses

experimentos são apresentados no *capítulo 5, seção 5.2*. Essas técnicas contribuíram para identificar falhas no processo de rotulagem.

Desta maneira, para este trabalho, as técnicas de projeção multidimensional serão utilizadas para apoiar os especialistas no domínio durante a etapa de rotulagem dos dados. A maneira como essas técnicas serão aplicadas está descrita no *capítulo 4, seção 4.3*.

3.7 TRABALHOS RELACIONADOS AO CAPÍTULO

Nas pesquisas realizadas, foram encontrados alguns trabalhos relacionados à aplicação de RNAs na análise e monitoramento de indicadores de desempenho em diferentes áreas. Ressalta-se que não foram encontrados trabalhos que utilizam aprendizado de máquina e/ou RNAs na análise de indicadores de desempenho de processos de desenvolvimento software.

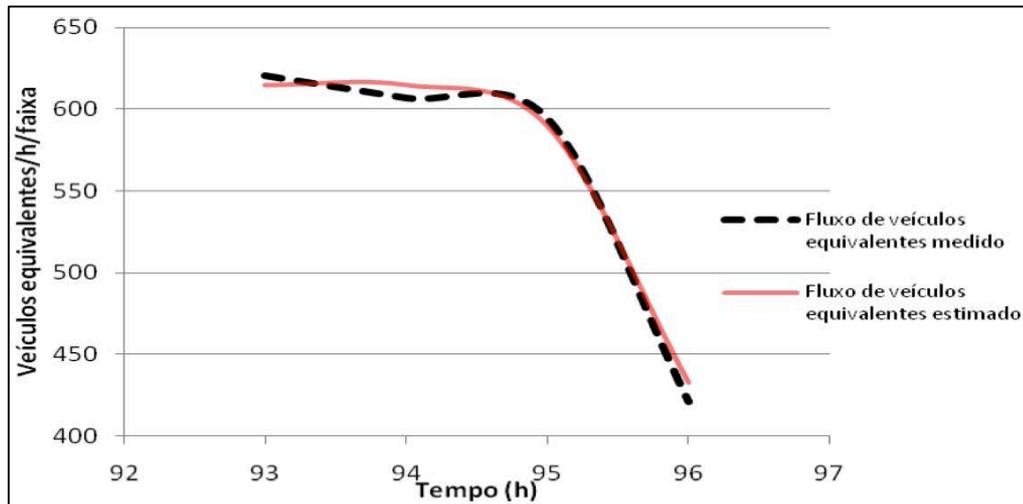
A seguir são apresentados comentários sobre três trabalhos aplicados em diferentes áreas: (1) monitoramento e prevenção do comportamento do tráfego de veículos em rodovias concessionadas; (2) classificação de cooperativas agropecuárias através de indicadores socioeconômicos; (3) classificação e análise de desempenho de clínicas de hemodiálise. O primeiro trabalho usa RNA com aprendizado supervisionado, enquanto o segundo e o terceiro trabalhos usam RNAs com aprendizado não supervisionado.

Melo et al. (2011) utilizam uma técnica de RNA supervisionada para monitoramento e prevenção do comportamento do tráfego de veículos em rodovias concessionadas. Nesse trabalho, os autores fazem uso de uma RNA Perceptron de multicamadas ou MLP (*Multilayer Perceptron*), treinada através do algoritmo "*error back-propagation*", com os seguintes parâmetros: nove neurônios de entrada, três camadas ocultas com quinze neurônios, um neurônio de saída e taxa de aprendizado constante igual a 0,1. Como valor de entrada para a RNA, foram selecionados os dados de medição de indicadores de desempenho de capacidade e nível de serviço. A **Figura 20** apresenta o resultado do fluxo de veículos equivalentes estimado pela RNA (linha vermelha) e o resultado real (medido) do fluxo de veículos equivalentes (linha preta).

Segundo Melo et al. (2011), a aplicação da RNA MLP apresentou resultado satisfatório, capaz de representar, verdadeiramente, a dinâmica dos fluxos nas rodovias (ver **Figura 20**). Adicionalmente, os autores mencionam que o uso da ferramenta pode prever oscilações no fluxo de veículos em um curto prazo tempo. Isso pode auxiliar os motoristas na escolha das melhores rotas de deslocamento.

Figura 20 - Uso de uma RNA MLP para o fluxo de veículos equivalente.

Fonte: extraído Melo et al. (2011).



Neto, Nagano e Moraes (2005), por outro lado, utilizaram uma técnica de RNA não supervisionada para classificar cooperativas agropecuárias do estado do Paraná, através de indicadores socioeconômicos. Considerando que agrupamentos de dados não têm rótulos no aprendizado não-supervisionado, os autores optaram pelo uso do algoritmo de Kohonen (*Kohonen feature maps*). Foram utilizadas treze variáveis de entrada, geradas a partir de indicadores financeiros e sociais das respectivas cooperativas. Os autores compararam a classificação da RNA, obtida por meio da formação dos *clusters*, com uma análise tradicional, identificada através do índice do Termômetro de Kanitz⁴.

Segundo Neto, Nagano e Moraes (2005), o resultado obtido após a classificação apresentou alto percentual de acerto, como pode ser observado na **Tabela 5**. Observa-se que não foram utilizadas as variáveis do índice de Kanitz, nem o próprio índice, como fator de seleção das variáveis para a RNA. Para os autores, as variáveis econômicas podem representar a dinâmica social das cooperativas analisadas. Isso porque os *clusters* são diferentes entre si e a classificação social depende do desempenho econômico. De modo geral, esse trabalho mostra que o desempenho social das cooperativas pode ser caracterizado através dos seus indicadores socioeconômicos.

Outro trabalho que usa uma RNA com aprendizado não supervisionado é o de Cattinelli et al. (2013). O trabalho mostra como foi feita a análise do desempenho de 109 clínicas de hemodiálises em três países: Itália, Portugal e Turquia. Foram selecionados 30

⁴ Refere-se apenas a índices financeiros.

KPIs relacionados às quatro perspectivas do modelo *Balanced Scorecard* (BSC): financeira, clientes, processos internos e aprendizado/crescimento. Cattinelli et al. (2013) mencionam que o algoritmo utilizado foi do tipo mapas auto-organizáveis, conhecido como SOM (*Self-Organizing Maps*), mas não especificam qual foi o algoritmo.

Tabela 5 - Percentual de acerto da RNA Kohonen em cada *cluster*.

Fonte: extraído de Neto, Nagano e Moraes (2005).

<i>Cluster</i>	Porcentagem de Acerto Cso
Grupo Fa	50
Grupo Fb	100
Grupo Fc	60
Grupo Fd	78
Grupo Fe	25
Grupo Ff*	0
Acerto Total	54

Segundo Cattinelli et al. (2013), os resultados da RNA permitiram analisar grupos específicos de clínicas dos três países. A análise identificou as clínicas com potencial para obterem melhor desempenho e as clínicas com tendência a declínios no desempenho. É interessante observar que o país com mais clínicas com tendência a aumento do desempenho foi a Turquia, seguida por Portugal e depois pela Itália. Os autores enfatizam que os resultados obtidos foram valiosos para orientar as estratégias de gestão dentro de uma política de melhoria contínua.

4

SISTEMÁTICA PARA CONTROLE DE INDICADORES DE PROCESSOS DE SOFTWARE

Este capítulo apresenta uma sistematização de ações que pode ser utilizada por organizações desenvolvedoras de software que utilizam indicadores de desempenho para controle de seus processos. Essa sistemática faz uso de um algoritmo de aprendizado semi-supervisionado para propiciar uma visão do que está acontecendo durante a execução dos processos, de modo a auxiliar tomadas de decisão operacionais, gerenciais e estratégicas.

Para uso da sistemática, a organização deve ter processos bem definidos, sendo recomendada a aderência a algum modelo de qualidade de processos, como o MPS-SW ou CMMI (*Capability Maturity Model Integration*). Através do monitoramento de indicadores chave de desempenho, os valores coletados podem ser avaliados para ver se estão dentro do esperado, requerendo ou não ajustes nos processos envolvidos.

A sistemática proposta pode ser utilizada para analisar indicadores com diferentes complexidades. Os indicadores podem ser analisados individualmente ou agrupados de um modo significativo à organização, visando tomadas de decisão dentro de certo contexto. Os agrupamentos podem ser compostos por indicadores diferentes de um mesmo processo, indicadores iguais de processos diferentes (lineares ou paralelos) ou por diferentes indicadores de diferentes processos em uma mesma cadeia de processos.

É interessante mencionar que a sistemática proposta pode ser utilizada em organizações de software com volumosas bases históricas de indicadores. Contudo, também pode ser utilizada em organizações que estejam iniciando o processo de medição e controle, com pouco volume de dados para seus indicadores.

É importante esclarecer que a sistemática pode ser utilizada como núcleo de um sistema de apoio ao controle e monitoramento de KPIs, com uma interface de usuário que contemple a exibição de um painel de instrumentos, como *dashboards* ou *scoreboards*. A coleta dos dados dos indicadores a serem avaliados pelo sistema pode ser definida em outro módulo do sistema. Essas interfaces e módulos não fazem parte deste trabalho, mas podem ser projetadas e implementadas em trabalhos futuros.

Convém ressaltar que a originalidade da sistemática proposta incide na utilização de

um algoritmo de aprendizado semi-supervisionado: o algoritmo de cooperação e competição entre partículas o PCC, apresentado na *seção 3.5*. Esse algoritmo, diferente do que se usa em Redes Neurais Artificiais (RNAs) de aprendizado supervisionado, não requer o treinamento prévio da rede (separadamente). Logo, os processos de aprendizado e classificação são realizados em um só passo.

Outro destaque é o uso de técnicas de visualização de informação com projeção multidimensional, apresentadas na *seção 3.6*, para apoiar o processo de rotulagem dos dados. Convém mencionar que as técnicas de visualização de informação utilizadas têm sido implementadas em sistemas desenvolvidos com a linguagem de programação Java - o que possibilita utilizá-los em qualquer plataforma.

A sistemática proposta está organizada em quatro etapas, que deverão ser executadas sequencialmente:

1. Seleção dos KPIs a partir de um modelo de indicadores adotado na organização;
2. Definição dos agrupamentos dos KPIs, segundo critérios significativos para as tomadas de decisão;
3. Rotulagem dos dados, com o apoio de técnicas de visualização de informação com projeção multidimensional;
4. Análise dos dados com aprendizado semi-supervisionado.

A sistemática pode ser repetida sempre que a organização deseje compor um novo quadro de indicadores para seus processos, projetos e produtos. Ressalta-se que a segunda etapa não precisa ser executada no caso de indicadores tratados individualmente. Além disso, as etapas 3 e 4 podem ser aplicadas a qualquer momento, nessa ordem, aos dados dos KPIs monitorados. Isso contribui para melhorar a acurácia dos resultados, pois quanto maior a frequência e o percentual de amostras rotuladas, melhor será o resultado da análise.

Convém informar que diferentemente de uma RNA de aprendizado supervisionado convencional, o uso do algoritmo PCC na etapa 3 requer que todos os dados, rotulados ou não, sejam usados como entrada, para então se obter o resultado da classificação. Logo, não podem ser reaproveitados em diferentes processos, produtos ou projetos, mesmo que os indicadores selecionados sejam mantidos. Isso porque cada processo, produto ou projeto tem características intrínsecas e específicas, o que faz com que os valores obtidos no monitoramento dos KPIs sejam expressivamente diferentes. Essa diferença fará com que a

rede forneça resultados incoerentes com o conjunto de dados analisado.

Observa-se que para a etapa 1, este trabalho desenvolveu um modelo de referência de indicadores de desempenho com base nos processos dos níveis G e F do modelo de MPS-SW. Esse modelo poderá ser utilizado para apoiar as organizações que ainda não têm um modelo de indicadores a definirem um modelo adequado às suas necessidades e objetivos. Assim, o modelo de indicadores de desempenho foi desenvolvido para apoiar organizações que estejam no início da implantação do modelo de referência MPS-SW, níveis G ou F. Além disso, esse modelo servirá como ferramenta de seleção de indicadores para as empresas que já possuam seus processos bem definidos.

De modo similar, para apoiar a etapa 2, foi desenvolvido um modelo de referência baseado nas quatro perspectivas do modelo *Balanced Scorecard* (BSC), as quais refletem a visão estratégica da organização. A intenção é que a organização defina seus próprios critérios de agrupamento de indicadores, podendo usar esse modelo como base, se assim o desejar.

As quatro etapas da sistemática estão descritas nas *seções 4.1 a 4.4*, respectivamente. Em cada etapa são apresentadas suas possíveis entradas e saídas, ferramentas/técnicas utilizadas e restrições quanto ao uso/aplicação. A aplicação das quatro etapas da sistemática será discutida no estudo de caso apresentado no *capítulo 5*.

4.1 ETAPA1: SELEÇÃO DOS INDICADORES DE DESEMPENHO

O objetivo desta etapa é que a organização selecione os KPIs de interesse ao monitoramento, a partir de um modelo de indicadores utilizado para os seus processos. Esse modelo deve ser composto pelos indicadores significativos para o controle dos processos, com as respectivas informações associadas, como: descrição, forma de cálculo, frequência de coleta, etc. Os KPIs selecionados deverão ser monitorados de acordo com a frequência e os valores armazenados em bases históricas. Convém ressaltar que a organização não precisa ter dados de monitoramento para poder selecionar os indicadores nesta etapa e na etapa 2.

Entretanto, nesta etapa, a seleção dos indicadores depende do modelo de indicadores que a organização mantém, onde devem estar também identificados os indicadores de desempenho dos processos. Assim, é possível escolher o conjunto de KPIs de interesse para análise com a aplicação da sistemática.

Caso esse conjunto de indicadores selecionados contemple um só KPI, então o

indicador deverá ser analisado individualmente, não requerendo que se execute a etapa 2 da sistemática, somente as etapas 3 e 4.

Como uma ferramenta de apoio a esta etapa, um modelo de referência de indicadores foi desenvolvido com base nos processos constantes no nível G e F do modelo de qualidade MPS-SW. Esse modelo pode ser utilizado mesmo por empresas não certificadas no modelo MPS-SW nos níveis G ou F.

Para cada indicador do modelo de referência são associados os seguintes atributos: descrição, objetivo, forma de cálculo, unidade de medida, frequência de coleta, frequência de apresentação de resultados e escopo de aplicação (projeto, produto ou negócio). O APÊNDICE B apresenta os indicadores disponíveis no modelo de referência, sendo que a **Figura 21** mostra uma parte do modelo através do aplicativo MS/Excel. Convém observar que o modelo de referência de indicadores apresentado contempla indicadores de modo geral, inclusive indicadores de desempenho.

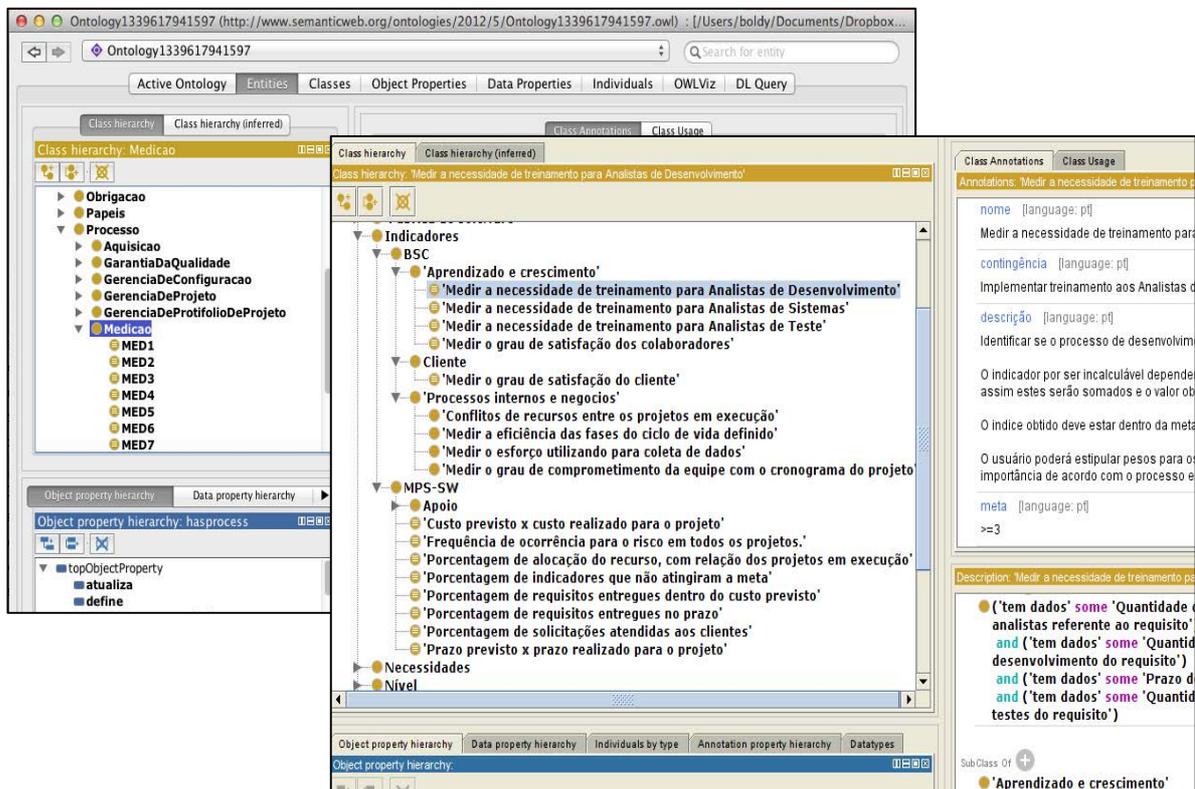
Figura 21 - Visão parcial do modelo de referência de indicadores desenvolvido na etapa 1.

Nível	Processo MPS-SW	Descrição	Objetivo	Cálculo	Escopo Aplicação	...	
F	GQA2	Aderência ao processo	Comparar a quantidade de itens auditados com a quantidade de itens em conformidade	$(\text{Total de itens em conformidade} / \text{Total de itens auditados}) * 100$	Projeto		
	GQA4	Variação do planejamento	Indicar a percentagem de variação do planejamento da sprint do projeto	$(\text{Total requisitos previsto de implementação no projeto} - \text{Total requisitos implementados no projeto}) * 100$	Projeto		
...							
G	MED5						
	MED7	GRE4	Requisitos implementados no produto	Exibir a quantidade de requisitos que foram implementados no dia	Somatório do total de requisitos implementados no dia	Produto	
		GRE5	Requisito de Impedimento	Exibir o percentual de requisitos responsável pelo impedimento do projeto	$(\text{Total requisitos de Impedimento} / \text{Total requisitos abertos no início do projeto}) * 100$	Projeto	
	...						
	G	GPR2	Eficiência vs disponibilidade	Apresentar o percentual de horas trabalhadas e apontadas pelos colaboradores com relação ao período em que os mesmos encontram-se na empresa	$(\text{Total de horas apontadas pelo colaborador} / \text{Total de horas registradas no espelho de ponto}) * 100$	Negócio	...
GPR9		Planejamento de projeto	Apresentar o percentual do tempo total em que o coordenador de projetos atuou com planejamento	$(\text{Total de horas apontadas para o planejamento do projeto} / \text{Total de horas apontadas pelo coordenador do projeto}) * 100$	Projeto		
...							

Para compor o modelo de indicadores para os níveis G e F do modelo MPS-SW, foi utilizada a ontologia proposta por Pizzoleto (2013), indicadores resultantes de estudos nas

literaturas e levantamentos junto a empresas desenvolvedoras de software. A ontologia de Pizzoleto associou, ao MPS-SW, a terminologia e conceitos do PMBOK (*Project Management Body of Knowledge*). A ontologia propõe indicadores (incluindo indicadores de desempenho) para quase todos os resultados esperados nos processos. Para complementar a estratégia de medições nos processos do MPS-SW, a ontologia inclui indicadores de três perspectivas do modelo BSC: clientes, processos internos e crescimento/aprendizado. Segundo Pizzoleto, esses indicadores contribuem para iniciativas de alinhamento com o planejamento estratégico e modelo de negócios. A **Figura 22** mostra telas do sistema Protégé com partes do processo de medição (MED) e dos indicadores baseados no modelo BSC.

Figura 22 - Visão parcial da ontologia de Pizzoleto (2013) no sistema Protégé.



Convém observar que o modelo de referência proposto pode ser utilizado por organizações que estejam iniciando o processo de implantação do modelo de referência MPS-SW e ainda não possuam um modelo de indicadores. Caso a organização já possua indicadores definidos como um modelo para uso no monitoramento e controle, o modelo apresentado no APÊNDICE B poderá ser consultado para ver se há interesse em complementar o seu modelo.

4.2 ETAPA 2: DEFINIÇÃO DOS AGRUPAMENTOS DE INDICADORES

Esta etapa deve ser executada quando se pretende agrupar indicadores para se avaliar determinadas situações que podem ocorrer em um ou mais processos/projetos. Nesses casos, as avaliações devem trazer informações relevantes para tomadas de decisão operacionais, gerenciais ou estratégicas. É interessante que a organização estabeleça critérios para o agrupamento dos indicadores, de acordo com as decisões que precisem ser tomadas, normalmente de ordem gerencial e estratégica. Recomenda-se um estudo envolvendo os interessados nos processos, de modo a se definir como grupos de indicadores podem ser compostos a fim de trazerem uma informação relevante ao processo.

Como exemplo, pode-se mencionar o monitoramento de indicadores de desempenho para ver se um conjunto de fatores para encerramento do projeto (*break point*) foi atingido. É o caso de um plano de projeto, no qual se estabeleceu que o desenvolvimento deve ser encerrado quando for ultrapassado 40% do prazo em alguma parte do processo e 30% dos custos. Assim, dois KPIs relacionados a essas características, selecionados na etapa 1, podem ser agrupados. Conforme forem sendo desenvolvidos os processos, esses valores podem ser monitorados conjuntamente. Cada resultado da dupla de KPIs pode ser classificado como “satisfatório”, “regular” ou “insatisfatório”. Como se tem apenas três classificações, no caso da sinalização ser regular ou insatisfatória podem ser empregadas soluções para mitigar ou corrigir eventuais causas, evitando-se o encerramento do projeto antes da sua conclusão.

Outro exemplo, que pode requerer análise de grupos de indicadores, pode ser direcionado a tomadas de decisão quanto à necessidade de treinamentos, para impulsionar a produtividade de programadores em várias linhas de produção de software. Um mesmo KPI, selecionado na etapa 1, pode servir para analisar o desempenho dos programadores, mas as instâncias desse indicador nos diferentes projetos podem ser agrupadas. Assim, a situação da produtividade dos programadores pode ser sinalizada se está satisfatória, regular ou insatisfatória, de modo a justificar ou não os custos de um determinado treinamento.

Takashina e Flores (1996) afirmam que é necessário que os indicadores forneçam informações que facilitem a gestão estratégica, permitindo que os administradores tomem decisões tendo percepção clara dos problemas. Esta etapa permite flexibilidade de tratamento de grupos de indicadores para orientar as tomadas de decisão desses administradores. Espera-se, assim, trazer soluções efetivas aos problemas, como, por exemplo, readequação de procedimentos ou até redefinição do processo de planejamento

ou controle, se for o caso.

A organização pode estabelecer critérios para a definição de grupos de indicadores. Entretanto, para contribuir com a elaboração desses critérios, propõe-se que os indicadores possam ser agrupados tomando como base as perspectivas do modelo *Balanced Scorecard* (BSC), apresentado na seção 2.2. Assim, a organização pode mapear seus indicadores chave, pertencentes ao seu modelo de indicadores considerado na etapa 1, para uma categoria de indicadores que corresponda a uma das quatro perspectivas do BSC: financeira, clientes, processos internos e crescimento/aprendizado. Assim, a organização pode ter um modelo de agrupamento de indicadores baseado no modelo BSC.

A **Tabela 6** exemplifica a formação de um modelo para agrupamento de indicadores com as perspectivas do BSC. Os indicadores apresentados foram selecionados do modelo de referência de indicadores, proposto neste trabalho para a etapa 1, conforme apresentado no APÊNDICE B. Como resultado de cada grupo de indicadores, a **Tabela 6** ilustra três possíveis referências de estado: verde (satisfatório), amarelo (regular) e vermelho (insatisfatório). Essa é uma estratégia inspirada em um semáforo de trânsito. A intenção é que os indicadores inseridos em uma categoria de perspectiva possam ser agrupados para posterior monitoramento e análise, refletindo se o resultado está satisfatório, regular ou insatisfatório.

Tabela 6 - Indicadores agrupados segundo perspectivas do BSC com três possíveis saídas.

BSC	Indicador	Semáforo		
Financeiro	...	Verde	Amarelo	Vermelho
	Retrabalho durante implementação do requisito			
	Eficiência do planejamento dos projetos			
	Requisitos de impedimento			
...	...	Verde	Amarelo	Vermelho
Política de atendimento aos clientes				
Número de notificações geradas pelos clientes				
Requisitos acumulados				
Processos Internos	...	Verde	Amarelo	Vermelho
	Variação do planejado			
	Planejamento de projetos			
	Aderência aos processos			
Crescimento Aprendizado	...	Verde	Amarelo	Vermelho
	Treinamento dos desenvolvedores			
	Treinamento dos analistas			
	Treinamento da equipe do projeto			
...	...	Verde	Amarelo	Vermelho

Observa-se que, na metáfora do semáforo, o sinal verde indica que é para seguir com o processo, pois a situação está satisfatória. O sinal amarelo alerta para prestar atenção, porque o grau de satisfação perante a meta previamente definida é regular. Já o sinal vermelho indica que o processo deve ser interrompido, por estar insatisfatório, muito fora do padrão estabelecido.

Embora na metáfora do semáforo só seja possível identificar três classes de estado, outras classes podem ser criadas, com valores diferentes. No entanto, usando-se apenas três classes de saída (e não quatro, por exemplo) facilita-se a decisão de rever ou não os dados dos indicadores individualmente. Quando o resultado da análise dos dados do grupo dos indicadores for “satisfatório”, significa que dados obtidos no monitoramento dos indicadores estão em conformidade com os padrões (metas) previamente estabelecidos pelos especialistas da organização. Por outro lado, se o resultado da análise for “regular” ou “insatisfatório”, significa que o grupo de indicadores tem dados que não estão em conformidade com as metas estabelecidas. Logo, será necessário analisar individualmente os dados de cada indicador, a fim de identificar qual indicador (ou indicadores) está com valores fora da meta estabelecida e, então, providenciar correções ou ações corretivas.

Esse modelo de agrupamento baseado no BSC pode ser utilizado como uma ferramenta de apoio a administradores, visando análises de possíveis grupos de indicadores. Contudo, esse modelo não precisa ser adotado, mas pode ajudar a estruturar um modelo próprio da empresa, onde as perspectivas do BSC sejam substituídas por outras categorias de interesse, mantendo-se os estados do semáforo.

4.3 ETAPA 3: ROTULAGEM DOS DADOS

Esta etapa pressupõe que a organização tenha um modelo de indicadores, dentre os quais tenha selecionado os indicadores a serem monitorados e analisados, conforme os resultados da etapa 1. Pressupõe também que grupos de indicadores possam ter sido definidos, segundo a etapa 2.

Feito isso, é necessário coletar os dados históricos referentes a esses indicadores para a rotulagem dos dados. O processo de rotulagem consiste na identificação das classes de saída (satisfatório, regular ou insatisfatório), com base nos resultados de monitoramento do indicador ou do conjunto de indicadores. Essa identificação é feita por um ou mais especialistas (humanos) no domínio.

Observa-se que a adoção de apenas três classes de saída facilita o processo de

rotulagem, reduzindo as opções de decisão do especialista na classificação das amostras dos indicadores. A ideia é que o especialista só classifique como “satisfatório” os resultados que ele tenha certeza que estão em conformidade com as metas esperadas. De modo similar, ele classifique como “insatisfatório” somente os resultados que ele tenha certeza que são indesejados. Nos resultados que ele tiver dúvida se classifica como “satisfatório” ou “insatisfatório”, ele pode classificar como “regular”. Contudo, o especialista também pode classificar como “regular” os resultados que ele sabe que requererão atenção e/ou ações corretivas.

Assim, primeiramente, os dados obtidos através do monitoramento dos indicadores devem ser recuperados das bases históricas; a quantidade desses dados pode variar conforme a disponibilidade. Apenas como *benchmarking*, observa-se que Chapelle et al. (2006) realizaram experimentos com técnicas de aprendizado semi-supervisionado, utilizando bases com 1500 amostras, variando a quantidade de amostras treinadas entre 10 e 100 amostras (aproximadamente 1% e 10%). Adicionalmente, experimentos apresentados na *subseção 5.2.2* foram bem sucedidos com a classificação dos dados a partir de poucas amostras rotuladas. Apesar disso, não é recomendável se estipular a quantidade de amostras a ser rotulada e nem o tamanho da base. Isso é feito de acordo com a taxa de acerto, sendo que o algoritmo utilizado na etapa 4 pode utilizar poucas amostras rotuladas para treinamento. Essa característica é vantajosa, pois não obriga a organização a rotular grandes quantidades de dados, economizando esforço e diminuindo custos. Porém, em bases muito pequenas, caso haja um pequeno erro de rotulagem, esse erro se torna significativo para as análises resultantes. Assim, se a organização estiver iniciando o processo de monitoramento e controle dos seus indicadores e possua poucos resultados de medição, ainda assim a sistemática pode ser aplicada. Contudo, recomenda-se acompanhamento de um especialista paralelamente. Isso, contudo, é temporário, pois à medida que o volume de medição aumenta e mais amostras são rotuladas e analisadas/treinadas, a confiança nos resultados obtidos melhora consideravelmente.

Embora tecnicamente o algoritmo utilizado na etapa 4 possa funcionar com no mínimo uma amostra rotulada por classe, recomenda-se que esses dados sejam mais significativos. Assim, em relação ao percentual de amostras que devem ser rotuladas para a aplicação da etapa 4, observa-se que foram realizados experimentos com 100% dos dados rotulados, a fim de avaliar a acurácia da análise feita pelo algoritmo utilizado na etapa 4, porém, foi apresentado apenas parte desses rótulos (1% a 10%) para o algoritmo que utilizou o restante para avaliar a acurácia de classificação do mesmo. Embora alguns experimentos exibidos no *capítulo 5* apresentaram resultados positivamente expressivos com apenas 1% de amostras treinadas, sugere-se que seja necessário rotular no mínimo

10% do total dos dados a serem analisados. Convém informar que esse percentual de treinamento pode variar, dependendo da base de dados a ser analisada. Neste trabalho, esses valores contribuíram para obtenção de resultados com taxa superior a 95% de acerto, também baseado nos resultados dos mesmos experimentos. Contudo, ressalta-se que que esses dados precisam ser representativos, ou seja, todas as classes de classificação (vermelho, amarelo, verde) precisam estar representadas no subconjunto rotulado, em proporções similares ao que se espera na base completa. Em outras palavras, se o especialista escolher aleatoriamente 10% de amostras do total de 1000 e rotular 33 amostras de cada classe (aproximadamente 3,33% por classe), algoritmo deve funcionar bem. Entretanto, se ele sair escolhendo e rotulando intencionalmente mais amostras de uma classe do que de outra, por exemplo, 2 verdes, 15 amarelas e 83 vermelhas (0,2%, 1,5% e 8,3%, respectivamente) ele pode acabar formando um subconjunto desbalanceado, que não é representativo do todo, podendo resultar em um mau funcionamento do algoritmo. Convém informar que experimentos foram realizados utilizando uma base desbalanceada quanto ao percentual de amostras treinadas e os resultados foram bastante positivos (vide *capítulo 5*).

No caso de grupos de indicadores, é importante observar que todos os indicadores tratados na n-upla de indicadores agrupados devem possuir valores, ou seja, não pode haver valor faltante na n-upla. Caso algum indicador não possua medição durante um dado momento do monitoramento, a instância do grupo de indicadores que tiver esse valor faltante deve ser excluída. Essa restrição se deve ao fato do algoritmo utilizado na etapa 4 não tratar casos com valores faltantes.

Considerando os dados selecionados, com a certeza de que não há instâncias de grupos com valores faltantes, deve ser iniciado o processo de rotulagem dos dados, por um ou mais especialistas do domínio considerado. Recomenda-se que todos os valores ou n-uplas de valores a serem rotulados sejam organizados em forma tabular, de modo que os especialistas possam fazer a análise para a rotulagem linha-a-linha da tabela. A **Tabela 7** apresenta um exemplo dessa forma tabular para uma 5-tupla de indicadores identificados como: SOL_CLI, IND_ATE, REQ_ABERT, REQ_N_IMPL, REQ_IMPED. A última coluna da **Tabela 7** mostra o rótulo atribuído a cada 5-tupla de indicadores, de acordo com o estado (*status*) dos dados em um dado momento, segundo os especialistas avaliadores. Conforme o estado avaliado, a 5-upla foi classificada com o rótulo de uma das três classes: satisfatório (classe 1), regular (classe 2) ou insatisfatório (classe 3).

Convém observar que o especialista avaliador deve seguir critérios significativos para o momento da rotulagem dos dados e não para o período de obtenção dos dados (monitoramento). Por exemplo, considere um conjunto de valores de indicadores que foi

obtido em 26 de outubro de 2012 e tenha apresentado o estado “satisfatório” em 2012. Considere que atualmente (neste ano) seu estado passou a ser “regular”. Então, no momento da rotulagem (atualmente) o especialista deve considerar o conjunto de valores como “regular”.

Tabela 7 - Exemplo de indicadores rotulados.

PRODUTO A	SOL_CLI	IND_ATE	REQ_ABERT	REQ_N_IMPL	REQ_IMPED	ROTULO
26/10/2012	1	1.00	3	29	0	1
29/10/2012	1	1.00	1	78	0	1
30/10/2012	1	1.00	9	88	0	1
31/10/2012	2	1.00	16	77	0	1
01/11/2012	28	7.00	39	178	0	2
05/11/2012	21	7.00	34	110	0	2
06/11/2012	21	7.00	29	113	0	2
07/11/2012	21	7.00	37	113	0	2
08/11/2012	41	11.80	50	261	1	3
09/11/2012	19	16.33	42	160	1	3
12/11/2012	45	16.50	35	219	0	3
13/11/2012	23	7.67	47	253	1	3

Apoio de técnicas de visualização de informação

Nesta etapa de rotulagem dos dados dos indicadores, técnicas de visualização de informação podem ser utilizadas como ferramentas apoio à rotulagem, bem como ferramentas de validação da rotulagem. Além disso, essas técnicas contribuem para “forçar” os especialistas no domínio a rotularem principalmente os dados que estão mais próximos das bordas das classes. Esse termo “próximo da borda da classe” representa as amostras que estão um pouco mais distantes da maior parte dos demais dados de sua classe, quase na fronteira de outra classe, podendo o algoritmo se equivocar quanto a sua classificação. Em outras palavras, considere as três classes (satisfatório, regular e insatisfatório) num ambiente hipotético com apenas um indicador. Dado um cenário onde esse indicador possua em sua base de dados histórica os valores de instância representados pelo conjunto { 3, 2, 1, 5, 12, 22, 22, 25, 26, 27 }, vamos supor que o especialista tenha selecionado aleatoriamente o conjunto { 3, 2 } e rotulou como satisfatório, e o conjunto { 22, 27 } como regular. Caso seja inserido ao algoritmo uma nova amostra, com o valor { 13 }, existe a possibilidade do algoritmo classifica-la em qualquer uma das duas classes (satisfatório ou regular), pois não existe uma referência mais próxima a esse valor no conjunto de dados rotulado e o mesmo está na fronteira entre as duas classes. Portanto, rotular o dado mais “próximo da borda da classe”, no exemplo apresentado, consiste em perguntar ao especialista a qual classe pertence o valor { 12 }, se satisfatório ou regular.

Assim, nesta etapa devem ser aplicadas as técnicas de projeção multidimensional

Least Square Projection (LSP), *Classical Multidimensional Scaling* (CMDS) e de *Parallel Coordinates* (PC), apresentadas na seção 3.6. É possível encontrar implementações dessas técnicas com a linguagem de programação Java, com portabilidade para várias plataformas operacionais. Recomenda-se utilizar a ferramenta de software PEx-Image para se executar as técnicas LSP e CMDS, e a ferramenta Parvis para a técnica PC.

Experimentos apresentados no capítulo 5 mostram a aplicação das técnicas de visualização LSP e CMDS através da ferramenta PEx-Image. Foi utilizada a distância Euclidiana, sem a necessidade de normalizar os dados. Observa-se que os indicadores são considerados dimensões nas técnicas de visualização de informação. Foram considerados grupos de até dez indicadores, logo as técnicas foram utilizadas com até dez dimensões. Em um ambiente com menos de dez indicadores não foi necessário reduzir o espaço dimensional na ferramenta. Já acima de dez dimensões (dez indicadores) talvez seja melhor reduzir a dimensionalidade para se obter melhores resultados na visualização. Para aplicação da técnica PC foi necessário realizar a normalização dos dados. Porém, a ferramenta utilizada disponibiliza uma opção para realizar essa normalização de maneira automática.

Quando aplicada a técnica LSP ou CMDS, deve-se procurar, visualmente, possíveis pontos (valores do conjunto de indicadores considerado) que estejam mais distantes da maioria do seu conjunto. Essa análise pode ser feita na própria tela da ferramenta ou em uma folha impressa, para facilitar a visualização.

Nessas técnicas, cada cor representa um conjunto de dados rotulados como uma das três classes: satisfatório, regular ou insatisfatório. Ainda, cada amostra representa um ponto, ou seja, um ponto corresponde aos valores de um indicador ou grupo de indicadores. Os indicadores mais similares são posicionados bem próximos no plano de projeção, enquanto os mais dissimilares (distantes) são posicionados mais afastados. Logo, pontos de uma determinada cor devem estar próximos, como na formação de um *cluster*. Caso haja algum ponto *x* próximo de um *cluster* de cor diferente, o comportamento desse ponto *x* deve ser avaliado. A técnica PC pode ajudar a avaliar esse resultado, comparando o ponto *x* com os pontos do *cluster* de onde ele está mais próximo e do *cluster* da sua própria cor. Dessa forma, é possível identificar possíveis erros de rotulagem, remetendo os dados rotulados ao especialista do domínio para uma revisão. Após a revisão, as técnicas de visualização de informação mencionadas podem ser utilizadas novamente para certificação dos resultados.

Cabe observar que, se apesar de seguidas todas as especificações das etapas, o resultado da visualização se mostrar com vários pontos espalhados ou com muitos pontos misturados ou sem nenhum conjunto agrupado, significa que pode ter ocorrido algum

equivoco na rotulagem dos dados ou que os indicadores selecionados não possuem correlação alguma no grupo de indicadores. Nesses casos, a orientação é que o processo todo de rotulagem seja rigorosamente revisto e, caso não seja constatado nenhuma falha nesse processo, novos indicadores devem ser analisados, pois os indicadores selecionados não podem ser analisados por essa sistemática.

Esta etapa 3 é finalizada quando todos os dados selecionados estiverem corretamente rotulados, respeitando todos os critérios aqui definidos. Assim, a etapa 4 pode ser iniciada.

4.4 ETAPA 4: TREINAMENTO E ANÁLISE DOS DADOS DOS INDICADORES

Nesta etapa, todos os dados obtidos com as medições dos indicadores identificados nas etapas 1 e 2 serão analisados. Esta etapa pode ser aplicada para cada grupo de indicadores selecionado na etapa 2, bem como para cada indicador selecionado individualmente na etapa 1. Para facilitar a explicação no texto, será utilizado o termo “indicadores” genericamente, para indicar uma dessas condições dos indicadores: individualmente ou em grupo.

Para possibilitar a análise dos indicadores, parte das amostras dos dados desses indicadores foi rotulada na etapa 3. As amostras rotuladas e não rotuladas serão utilizadas, nesta etapa, como entrada do algoritmo de cooperação e competição entre partículas, apresentado na seção 3.5.

A partir dessas entradas um grafo é montado e cada amostra (rotulada ou não) se transforma em um nó desse grafo. Cada amostra rotulada gera uma partícula, que é colocada no nó do grafo correspondente à amostra. Amostras não-rotuladas não geram partículas. As partículas “caminham” pelo grafo de maneira competitiva e colaborativa, espalhando seus rótulos na tentativa de dominar a maior quantidade possível de nós. Esse processo iterativo é encerrado quando o time de partículas (classe) que dominou o nó é usado para rotular esse nó. Dessa forma, os nós não rotulados são rotulados para uma classe, que no caso deste trabalho, pode ser: satisfatório (classe 1), regular (classe 2) ou insatisfatório (classe 3). Assim, os resultados da classificação podem ser obtidos para todos os valores dos indicadores, pois o treinamento e a classificação dos dados são realizados em um só passo.

Convencionalmente, foi definido que os atributos considerados no algoritmo PCC são representados pelos indicadores de desempenho em processos de software, que podem ser

considerados individualmente ou em grupo. Consequentemente, cada instância desses atributos no PCC é representada por um dado da base histórica, obtido por monitoramento dos indicadores.

Segundo Breve (2010), o algoritmo PCC foi implementado na linguagem do sistema MATLAB (*MATrix LABoratory*), apesar de poder ser implementado em outras linguagens. Para este trabalho, foi utilizada a implementação disponível para o MATLAB no seguinte endereço: <http://www.fabriciobreve.com/software/artigos/2012/tkde2012/strwalk8k.7z>. Os dados de entrada podem ser organizados em um arquivo de planilha eletrônica “.xls” ou em um arquivo “.txt”, pois a ferramenta MATLAB possibilita realizar a importação dos dados nesses formatos.

Para o treinamento não precisa ser informado o percentual de amostras que foram rotuladas; basta passar o conjunto de rótulos para o subconjunto de dados rotulados. Convém observar que o algoritmo PCC não seleciona amostras aleatoriamente. Ele precisa de pelo menos uma amostra rotulada por classe, como já foi especificado no processo de rotulagem na etapa 3.

Outro parâmetro importante é a quantidade “ k ” de vizinhos próximos que será definida para treinamento. O valor do “ k ” pode variar de acordo com o tamanho da base de dados a ser analisada, mas, em geral, seus valores ótimos são obtidos no intervalo de 1 a 9. Assim, inicialmente, pode-se definir $k = 5$ e ir ajustando seu valor para mais ou para menos. Isso é feito apresentando apenas parte das amostras rotuladas, com seus respectivos rótulos, para o algoritmo. As demais amostras são apresentadas como não rotuladas. Assim, os rótulos atribuídos pelo algoritmo podem ser comparados com os rótulos fornecidos pelo especialista, medindo-se, então, a acurácia do algoritmo para cada valor de k – o que possibilita a otimização do algoritmo. Esse processo deve ser repetido até as maiores taxas de acerto serem obtidas.

Dessa forma, com a execução desta etapa 4, todos os valores dos indicadores são classificados em uma saída que demonstra o estado dos processos, projetos ou produtos sob controle.

5 ESTUDO DE CASO E EXPERIMENTOS

Este capítulo apresenta o estudo de caso realizado com dados de uma empresa real desenvolvedora de software para gestão pública. O capítulo também contempla resultados de alguns experimentos realizados para apoiar o desenvolvimento da sistemática apresentada no *capítulo 4*.

Os dados utilizados fazem parte de uma base histórica de indicadores de desempenho da empresa considerada, que foram coletados por aproximadamente três anos, através de monitoramento dos processos de software. Convém observar que a empresa fornecedora dos dados atua no mercado de software há mais de 30 anos e tem centenas de administrações municipais como cliente. A empresa é certificada no nível G do modelo MPS-SW, mas tem processos que também atendem parcialmente o nível F do modelo MPS-SW. Dessa forma, os indicadores utilizados neste capítulo foram extraídos de um processo que é associado ao nível G, “Gerenciamento de Requisitos”, e de um processo que é associado ao nível F, “Garantia da Qualidade”.

Nesse contexto, a *seção 5.1* apresenta um estudo de caso com a aplicação de todas as etapas da sistemática apresentada no *capítulo 4*. A *seção* caracteriza o cenário considerado e traz discussões que evidenciam os benefícios da aplicação da sistemática.

A *seção 5.2* apresenta dois experimentos, que ajudaram a avaliar a sistemática de modos diferentes. Em um dos experimentos a sistemática foi aplicada, mas sem as técnicas de visualização da informação para apoiar o treinamento das amostras. No outro experimento, não foi seguida a sequência das etapas da sistemática, porque foi o primeiro experimento realizado, o qual também contribuiu para a definição da sequência das etapas da sistemática apresentada.

Tanto para o estudo de caso na *seção 5.1* como para os experimentos apresentados na *seção 5.2*, as seguintes Informações são válidas e relevantes:

- Todos os indicadores de desempenho utilizados são reais e foram rotulados com a ajuda de dois especialistas do domínio dos processos, gerentes de projetos da empresa fornecedora da base histórica;
- A seleção dos indicadores para cada cenário foi realizada a partir dos modelos

de referência de indicadores, apresentados nas *seções 4.1 e 4.2*;

- Os resultados de saída fornecidos pelos algoritmos respeitaram os critérios das três classes, conforme apresentado na *seção 4.3*, sendo: satisfatório (C1), regular (C2) e insatisfatório (C3);
- O conjunto de dados utilizado para treinamento das técnicas não fazem parte do conjunto de amostras analisadas;
- A ferramenta MATLAB foi utilizada para a implementação do algoritmo de cooperação e competição entre partículas ou PCC.

Além disso, convém salientar que o processo de coleta dos dados dos indicadores não fez parte desse trabalho, sendo fornecidos na totalidade pela empresa parceira. Contudo, ressalta-se também que o fato da empresa ter certificação MPS-SW (nível G) já propicia uma forma de garantia de uso de boas práticas que inferem alta confiabilidade dos dados.

5.1 APLICAÇÃO DA SISTEMÁTICA PARA CONTROLE DE INDICADORES DE PROCESSOS DE SOFTWARE

A partir da base histórica fornecida pela empresa desenvolvedora de software parceira deste trabalho, foi possível realizar a aplicação da sistemática desenvolvida. As *subseções 5.1.1 a 5.1.4* apresentam a aplicação das quatro etapas da sistemática, em ordem sequencial, apresentando de modo bem definido os dados de entrada e os resultados (saída). A última seção traz discussões que evidenciam os benefícios do uso da sistemática.

5.1.1 ETAPA 1: SELEÇÃO DOS INDICADORES DE DESEMPENHO

A empresa desenvolvedora de software selecionada para este estudo de caso utiliza um software proprietário para o gerenciamento de seus indicadores. Esse software permitiu exportar uma lista em Excel (arquivo “.xls”), contendo toda a relação de indicadores relacionados aos processos, projetos e produtos de software da empresa. Reuniões foram necessárias para o fomento de ideias sobre quais indicadores poderiam fazer sentido quando analisados também em conjunto. Como haviam indicadores relacionados aos processos, projetos e produtos na mesma lista, houve grande dificuldade na escolha dos indicadores adequados às análises.

Assim, o modelo de referência apresentado na etapa 1 (seção 4.1 e APÊNDICE B) da sistemática, baseado nos indicadores da ontologia dos níveis G e F de Pizzoleto (2013), foi adotado como forma de apoio à seleção dos indicadores. O modelo ajudou os especialistas a associarem os indicadores da lista aos processos e a irem substituindo os indicadores do modelo pelos indicadores próprios da empresa. Assim, a visualização dos indicadores dos processos ficou mais clara e levou a empresa a escolher cinco indicadores para o estudo de caso, cujo histórico corrente requeria maior atenção. A **Tabela 8** apresenta alguns exemplos de indicadores selecionados da empresa, associados aos processos de Gerenciamento de Requisitos” (GRE) e “Garantia da Qualidade” (GQA), após terem considerado as **Tabelas de 16 a 19**, apresentadas no APÊNDICE B.

Tabela 8 - Indicadores selecionados a partir do modelo de referência da etapa 1 da sistemática.

Nível	Processo MPS-SW	Descrição	Objetivo	Cálculo	Frequência		Aplicação	Unidade Medida
					Coleta	Apresentação		
G	GRE5	Requisito de Impedimento	Exibe o percentual de requisitos responsável pelo impedimento do projeto	$(\text{Total requisitos de Impedimento} / \text{Total requisitos abertos no início do projeto}) * 100$	Diária	Por Projeto	Projeto	%
G	GRE1	Requisitos abertos	Identifica a quantidade de requisitos que foram abertas no dia	Somatório dos requisitos abertos no dia	Diária	Mensal	Produto	Nº
F	GQA4	Requisitos não implementados	Indica a quantidade de requisitos não implementados e que encontram-se abertos no final do período	$(\text{Total Requisitos Abertos} - \text{Total Requisitos Implementados por dia}) + \text{Saldo Anterior}$	Diária	Diária	Produto	Nº
F	GQA3	Solicitação de cliente	Identifica a quantidade de solicitações e reportes realizados por cliente	Somatório do total de solicitações abertas por dia	Diária	Mensal	Produto	Nº
F	GQA3	Índice de atendimento	Indica a relação entre as solicitações abertas pelos clientes vs a quantidade de pessoas disponíveis para realizar essas solicitações de atendimentos	$\text{Total de solicitações registradas} / \text{Total de atendentes}$	Diária	Mensal	Produto	-

5.1.2 ETAPA 2: DEFINIÇÃO DOS AGRUPAMENTOS DE INDICADORES

Após a definição dos indicadores, foram realizadas algumas reuniões para definição dos grupos de indicadores, os quais deveriam ser analisados conjuntamente. O modelo de referência apresentado na etapa 2 (seção 4.2) da sistemática, baseado nas quatro perspectivas do BSC (*Balanced Scorecard*), foi adotado como forma de apoio à composição dos grupos de indicadores. O modelo completo foi apresentado aos especialistas da empresa e, a partir dele, esses especialistas puderam verificar em qual(is) das quatro

perspectivas do BSC os indicadores seriam melhor associados.

Como o foco da empresa para a tratativa desses indicadores era estratégico, com foco no cliente, foi adotada a perspectiva “Clientes”; os cinco indicadores selecionados foram agrupados, então, segundo essa perspectiva. Em conformidade com o modelo de agrupamento de indicadores da etapa 2 da sistemática, o grupo de indicadores poderá ter um dos seguintes estados como saída na análise, com base em um semáforo de trânsito: verde (satisfatório), amarelo (regular) e vermelho (insatisfatório).

5.1.3 ETAPA 3: ROTULAGEM DOS DADOS

A partir dos indicadores selecionados no modelo de referência da etapa 1 e agrupados conforme o modelo de referência da etapa 2, a aplicação da etapa 3 (seção 4.3), de rotulagem dos dados, requereu os dados históricos referentes aos indicadores selecionados. Para isso, foi utilizada a ferramenta proprietária do cliente para recuperar os dados associados aos indicadores, exportando-os para uma planilha em Excel (“.xls”). Foram recuperados dados de aproximadamente três anos de medição de cada indicador, totalizando 1977 amostras de cada indicador.

Então, os dados foram tabulados no Excel como uma 5-tupla de indicadores, seguindo o mesmo formato exemplificado na **Tabela 7** da seção 4.3. É importante observar que o próximo passo foi verificar se haviam valores faltantes. Esses valores faltantes tiveram a instância da 5-tupla excluída da tabela, possibilitando o início do processo de rotulagem dos dados.

Considerando que o custo envolvido no processo de rotulagem das amostras é bastante alto e para fins de avaliação da sistemática seria importante rotular 100% das amostras selecionadas, foi necessário extrair um conjunto aleatório de 5-tuplas de indicadores para o processo de rotulagem. Assim, foram extraídas 660 amostras para serem rotuladas, dentre as quase duas mil 5-tuplas.

A **Tabela 9** apresenta dados estatísticos relativos à base de dados utilizada, para os cinco os indicadores considerados, a saber: solicitação de clientes, índice de atendimento, requisitos abertos, requisitos não implementados e requisitos de impedimento.

Tabela 9 - Dados estatísticos sobre a base de dados utilizada.

DESCRIÇÃO	MIN	MAX	MÉD	DESV.P
Solicitações de clientes	1	59	20.29	10.11
Índice de atendimento	1.00	19.33	5.34	2.67
Requisitos abertos	1	80	27.4	12.32
Requisitos não implementados	141	375	267.81	56.54
Requisitos de impedimento	0	2	0.02	0.15

Para a rotulagem dos dados, dois gerentes de projeto da empresa (especialistas no domínio) analisaram todos os resultados de monitoramento das 660 amostras, de forma conjunta. Para cada amostra, foram identificando qual o estado de saída (satisfatório, regular ou insatisfatório) e registrando na planilha das amostras rotuladas.

Convém informar que 100% das 660 amostras foram rotuladas pelos especialistas do domínio. Contudo, na *subseção 5.1.4*, correspondente à etapa 4 da sistemática, apenas parte desses rótulos (de 1% a 10%) será utilizado no algoritmo PCC; o restante será utilizado para avaliação.

Técnicas de visualização de informação no processo de rotulagem

Segundo a etapa 3 da sistemática, para validar o processo de rotulagem deveriam ser empregadas as três técnicas de visualização de informação apresentadas na *seção 3.6*: (1) projeção por mínimos quadrados ou LSP (*Least Square Projection*); (2) escalonamento multidimensional clássico ou CMDS (*Classical Multidimensional Scaling*); (3) coordenadas paralelas ou PC (*Parallel Coordinates*). Assim, essas três técnicas foram aplicadas às 660 amostras rotuladas. Convém informar que essas técnicas utilizaram o conceito de distância euclidiana, sem redução de dimensões.

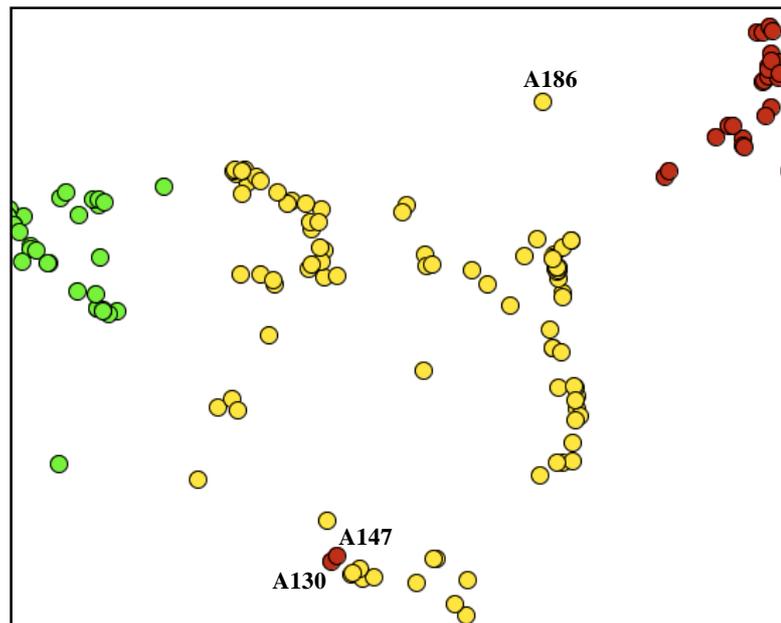
A primeira técnica aplicada ao grupo de indicadores foi a de projeção por mínimos quadrados ou LSP. Nessa técnica, as cores representam as classes e cada ponto corresponde a uma instância dos dados, isto é, cada ponto corresponde ao valor obtido nas medições do agrupamento de indicadores, identificado pela letra “A” (de “Agrupamento”). As instâncias dos dados mais similares são posicionadas bem próximas no plano de projeção, enquanto as mais dissimilares (distantes) são posicionadas mais afastadas.

A **Figura 23** mostra uma parte dos resultados obtidos com a técnica LSP, evidenciando o agrupamento da classe “regular” (amarela). Observa-se que as instâncias A130 e A147 tiveram seu rótulo atribuído à classe “insatisfatória” (vermelha). No entanto, observando seus posicionamentos no plano bidimensional, é possível identificar que elas possuem comportamento mais próximo à classe “regular” (amarela) do que de sua classe de origem (vermelha). Essa proximidade de outra classe, diferente da classe à qual a instância

pertence, indica que pode ter havido um erro no processo de rotulagem.

A **Figura 23** mostra ainda que, embora tenha sido atribuído rótulo “regular” (amarelo) para a instância A186, ela está no limite da divisa das classes “regular” e “insatisfatória”. Esse tipo de situação significa que se deve prestar mais atenção na identificação e conferência, pois a amostra A186 visivelmente poderia pertencer à classe “regular” (amarela) ou “insatisfatória” (vermelha).

Figura 23 - Provável erro de rotulagem na classe “regular” (amarela), usando a técnica LSP.

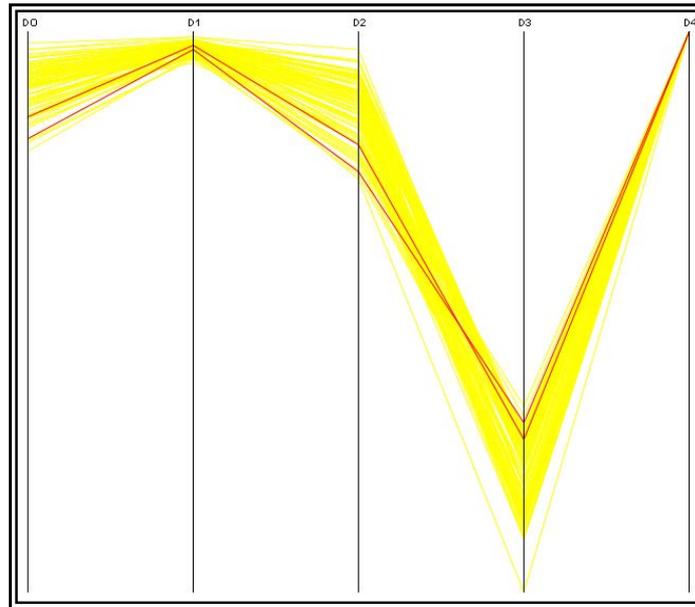


Apesar da análise visual com a técnica LSP sugerir que houve uma falha de rotulagem, não é possível afirmar, utilizando somente essa técnica, que isso de fato ocorreu com as instâncias A130, A147 e A186. Logo, é importante utilizar outras técnicas de visualização de informação para constatar se houve mesmo um problema na rotulagem dos dados.

Assim, o próximo passo foi aplicar a técnica de coordenadas paralelas ou PC para avaliar melhor o comportamento dos atributos das instâncias A130 e A147, que foram rotulados inicialmente como classe “insatisfatória” (vermelha). A técnica PC foi então aplicada ao conjunto de dados da classe “regular” (amarela), juntamente com os atributos mencionados. Nessa técnica, as cores correspondem às classes, os eixos representam os atributos e as linhas representam as instâncias dos dados. Cada linha corta o eixo paralelo no valor correspondente ao atributo. Dessa maneira, é possível observar o comportamento dos atributos dos dados, identificando padrões ou discrepâncias, considerando que as

amostras em destaque estão representadas na **Figura 24** pela cor vermelha. Dessa forma, através da técnica PC pode-se observar que as amostras A130 e A147 apresentaram comportamento muito similar aos dos atributos da classe “regular” (representados pelas linhas amarelas). Isso aumenta a probabilidade de ter ocorrido, de fato, uma falha na rotulagem das amostras A130 e A147.

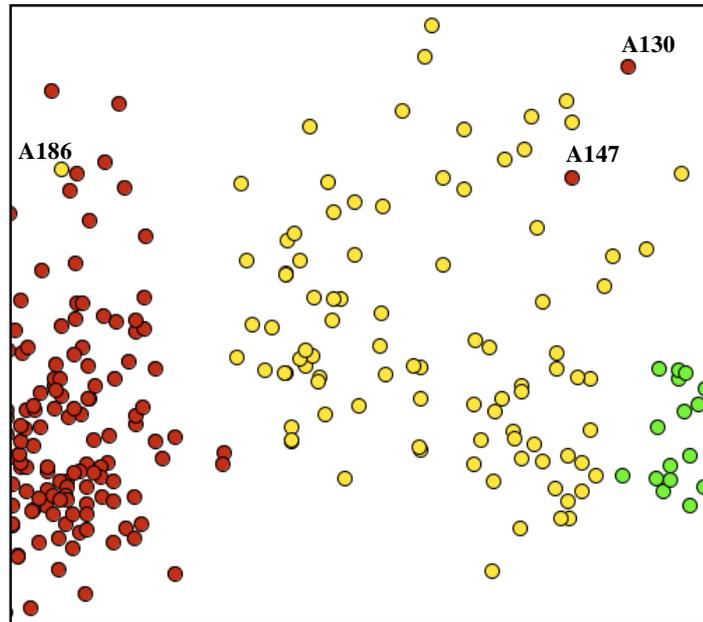
Figura 24 - Provável erro de rotulagem na classe “regular” (amarela), usando a técnica PC.



Para avaliar mais criteriosamente a relação entre as instâncias A130, A147 e A 186, evidenciadas na **Figura 23**, foi aplicada a técnica de escalonamento multidimensional clássico ou CMDS (WARREN, 1965), com resultados apresentados na **Figura 25**. Essa técnica reforçou a hipótese de falha na rotulagem dos agrupamentos A130 e A147, bem como possibilitou o questionamento sobre o rótulo atribuído à amostra A186. Então, foi possível notar que a amostra A186 possui relação bem mais próxima com a classe “insatisfatória” (vermelha) do que com a classe “regular” (amarela).

Após a aplicação das três técnicas de visualização de informação, considerando os potenciais problemas de rotulagem identificados, foram realizadas reuniões com os especialistas do domínio na empresa (gerentes de projeto), com a finalidade de revisar a rotulagem das três instâncias: A130, A147 e A186. Para isso, os resultados das técnicas foram apresentados, explicando as potenciais falhas ocorridas, bem como o impacto de se manter uma amostra rotulada erroneamente num cenário com poucas amostras rotuladas. Após algumas reuniões e discussões, os gerentes de projetos admitiram ter havido uma falha pontual na rotulagem das amostras A130, A147 e A186 e então seus rótulos foram revisados.

Figura 25 - Provável erro de rotulagem na classe “regular” (amarela), usando a técnica CMDS.



Após a revisão dos rótulos das amostras A130, A147 e A186, as três técnicas de projeção multidimensional foram aplicadas novamente, considerando o mesmo conjunto de dados usado anteriormente, porém com os rótulos das três amostras mencionadas alterados. As **Figuras 26** e **27** apresentam os resultados dessa segunda aplicação das técnicas LSP e CMDS, respectivamente. É possível observar que distribuição das classes está muito mais uniforme após os ajustes nos rotulos, sem pontos que se misturam entre classes. Já a **Figura 28** apresenta o resultado da nova aplicação da técnica PC, através da qual se pode constatar que não é possível identificar a olho nu discrepâncias entre as instâncias A130 e A147 em relação aos demais elementos do conjunto de dados.

Figura 26 - Aplicação da técnica LSP após revisão dos rótulos das amostras A130, A147 e A186.

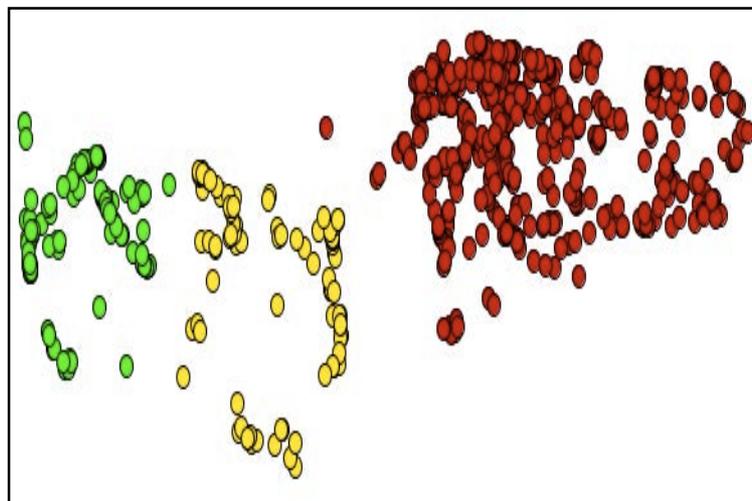


Figura 27 - Aplicação da técnica CMDS após revisão dos rótulos das amostras A130, A147 e A186.

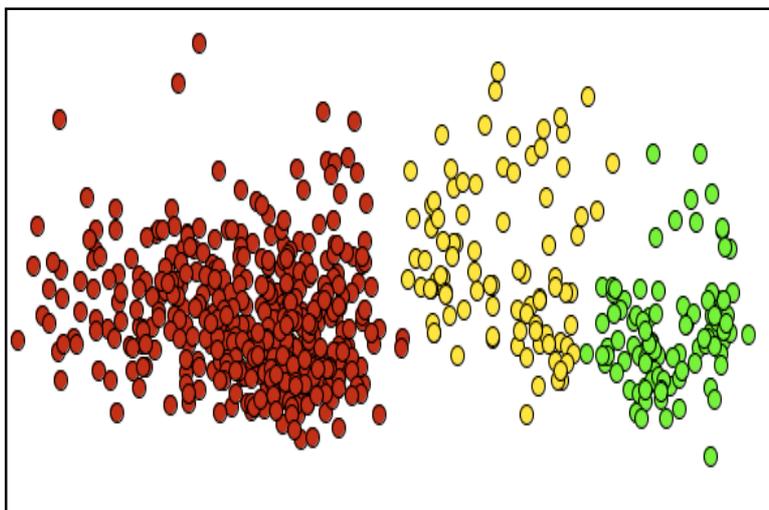
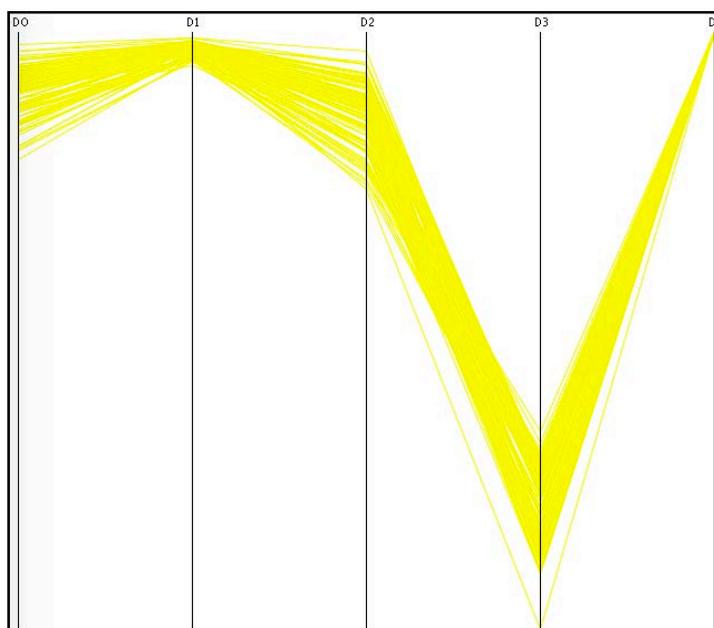


Figura 28 - Aplicação da técnica PC no conjunto de instâncias da classe “regular” (amarela) após revisão dos rótulos.



Convém informar que para fins de avaliação da sistemática proposta, foi necessário rotular 100% das 660 amostras consideradas. Como todas as amostras são rotuladas, o próprio especialista poderia verificar as amostras “suspeitas” de terem sido mal classificadas, avaliando a taxa de acerto do algoritmo. Assim, ele poderia reavaliar e reclassificar essas amostras “suspeitas”. Nesse caso, poder-se-ia até dispensar o uso das técnicas de visualização de informação. Entretanto, essa situação ideal geralmente não

ocorre, uma vez que apenas um percentual pequeno das amostras é rotulado (em torno de 10%) em cenários reais. Logo, a aplicação das técnicas de visualização de informação torna-se um recurso fundamental para a identificação de amostras que podem ter sido mal classificadas.

5.1.4 ETAPA 4: TREINAMENTO E ANÁLISE DOS DADOS DOS INDICADORES

Após a etapa de rotulagem dos dados, a etapa 4 (seção 4.4) da sistemática proposta requereu a aplicação do algoritmo de cooperação e competição entre partículas, considerando as 660 instâncias (amostras) e os cinco indicadores selecionados na etapa 1 (ver **Tabela 8**). Convém destacar que cada atributo teve seu valor normalizado antes de ser utilizado no algoritmo PCC. Para aplicação desse algoritmo, a base de dados disponível possui as seguintes características:

- Número de instâncias: 660;
- Número de atributos (indicadores): 5;
- Valores faltantes: nenhum;
- 3 Classes: C1 = Satisfatório; C2 = Regular; C3 = Insatisfatório;
- Distribuição das classes: C1 = 16%; C2 = 14%; C3 = 70%;
- Informações dos atributos (indicadores):
 - a) Solicitação de clientes;
 - b) Índice de atendimento;
 - c) Requisitos abertos;
 - d) Requisitos não implementados;
 - e) Requisitos de impedimento.

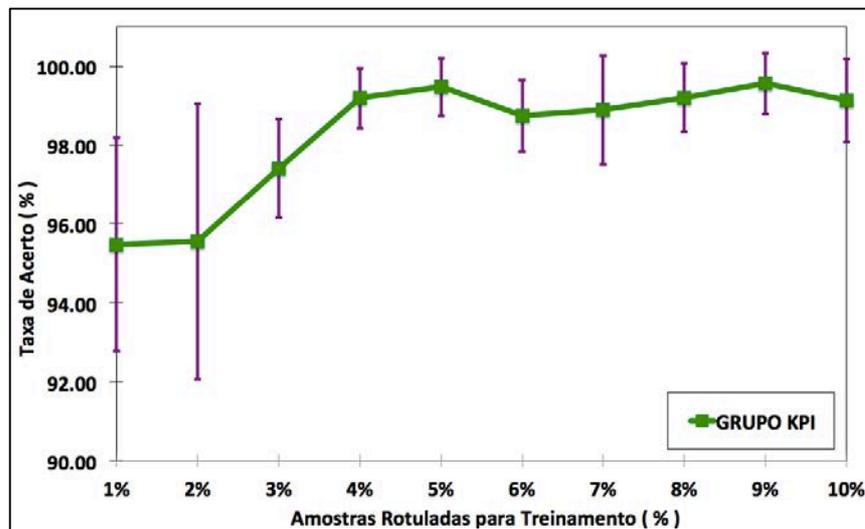
Para se usar o algoritmo semi-supervisionado PCC basta se considerar 10% das amostras rotuladas – o que é considerado um baixo custo. Porém, para demonstrar a capacidade do algoritmo em produzir bons resultados com bem poucas amostras rotuladas, os treinamentos foram realizados utilizando dez configurações diferentes, de acordo com o percentual de amostras treinadas: 1%, 2%, 3%, 4%, ..., 10%.

Considerando que 100% das amostras foram rotuladas para fins de teste e avaliação da taxa de acerto do algoritmo PCC, ele foi executado dez vezes para cada percentual treinado, sendo que o resultado apresentado é a média das dez execuções. Isso porque o algoritmo utilizado possui característica estocástica e os dados rotulados no treinamento

foram selecionados aleatoriamente pelo especialista do domínio.

Assim, o treinamento do algoritmo PCC foi realizado gradativamente, de 1% em 1%, selecionando um percentual de amostras para o treinamento e o restante para validação. A **Figura 29** mostra o resultado da taxa de acerto obtida no melhor momento da aplicação do algoritmo PCC ($k=5$), conforme o aumento do percentual de amostras treinadas. A linha verde representa a taxa de acerto do agrupamento de indicadores.

Figura 29 - Taxa de acerto em grupo de indicadores com a técnica PCC.



Considerando que a base de dados está desbalanceada quanto à distribuição das classes, a taxa de acerto pode não ser considerada uma boa medida isoladamente. Isso porque se o classificador “jogar” tudo para a classe mais provável, ele terá uma taxa de acerto considerada alta, mesmo sendo um péssimo classificador. Assim, para considerar o desbalanceamento da base de dados, foram obtidos resultados através do uso do coeficiente Kappa, apresentado na **Figura 30**. Para esse coeficiente, o resultado é melhor quanto mais próximo de 1.

Outro fator importante é o tempo de execução do algoritmo PCC para análise dos indicadores de desempenho em comparação com um ser humano (HUM), apresentado na **Figura 31**. Convém informar que a medição obtida pela análise do ser humano contou com um especialista do domínio da empresa parceira durante a etapa de rotulagem. Esse especialista analisou visualmente os resultados de cada indicador e o comparou com o respectivo resultado esperado.

O tempo gasto pelo ser humano no processo de rotulagem das amostras foi

calculado a partir do somatório de cada medição realizada. Já o tempo gasto no processo de análise feita pelo algoritmo foi medido através de função específica do MATLAB.

Figura 30 - Coeficiente Kappa em grupo de indicadores com o algoritmo PCC.

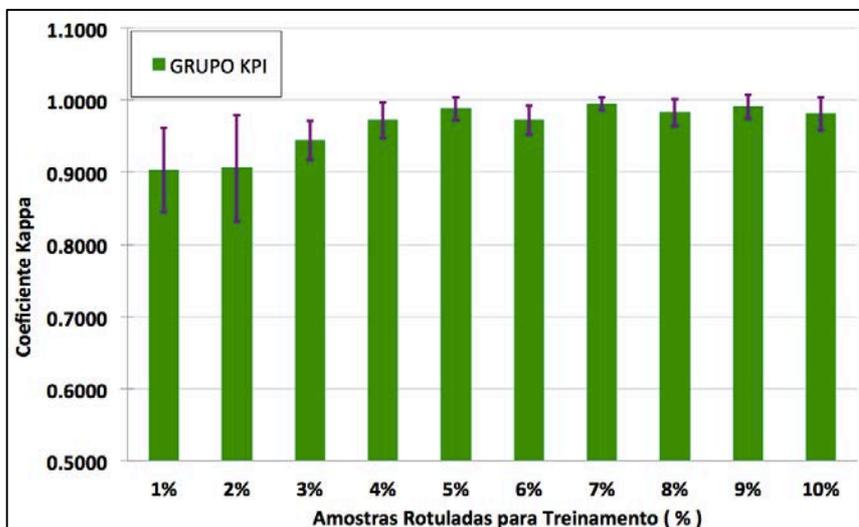
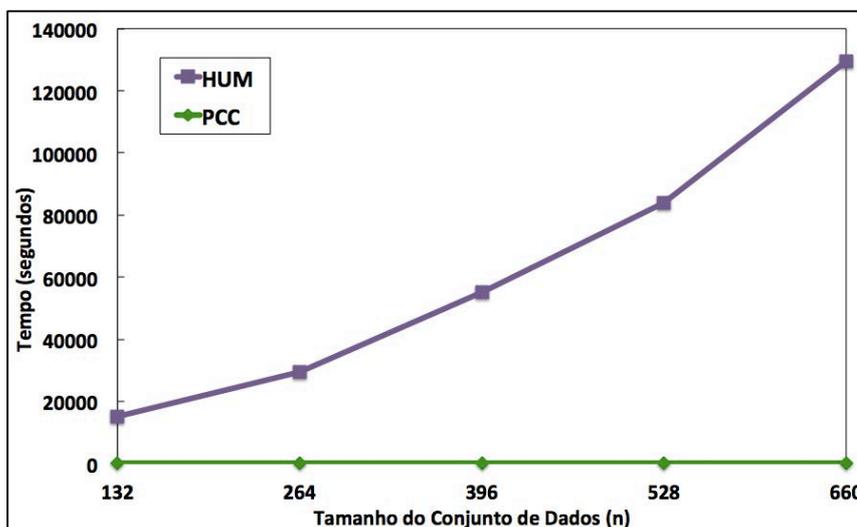


Figura 31 - Tempo de execução: algoritmo PCC vs Humano.



5.1.5 CONCLUSÃO DO ESTUDO DE CASO

Esta seção apresenta aspectos positivos e negativos que puderam ser observados no estudo de caso apresentado nesta seção 5.1, com a aplicação das quatro etapas da

sistemática apresentada no *capítulo 4*.

Os modelos de referência propostos nas etapas 1 e 2 contribuíram de fato para a seleção e o agrupamento dos cinco indicadores, respectivamente. Enquanto o primeiro modelo facilitou o entendimento e a escolha dos indicadores existentes na empresa, o segundo norteou os especialistas da empresa no tocante ao agrupamento estratégico de seus indicadores. O autor deste trabalho pode verificar essas contribuições pessoalmente, através de reuniões com os especialistas da empresa (gerentes de projetos de software).

As técnicas de visualização de informação contribuíram muito para identificar falhas ocorridas no momento da rotulagem dos dados (etapa 3), principalmente considerando a grande quantidade de amostras rotuladas pelos especialistas da empresa: 660 amostras. Apesar dos resultados das técnicas de visualização de informação serem visuais e subjetivos, eles foram decisivos para avaliação de erros no processo de rotulagem. Foi mostrado que as três técnicas utilizadas (LSP, CMDS e PC) foram utilizadas de modo complementar, sendo interessante a aplicação inicial da técnica de projeções LSP, de acordo com o que foi apresentado no estudo de caso. Com a revisão dos rótulos das amostras identificadas com possíveis erros de rotulagem, a reaplicação das três técnicas de visualização de informação assegurou, de certa forma, que o processo de rotulagem estava adequado, ou seja, com alta confiabilidade. Certamente isso contribuiu para aumentar a confiabilidade dos resultados da etapa 4. Convém ressaltar que a técnica de Coordenadas Paralelas foi a que produziu melhores resultados com as amostras normalizadas, em comparação com as outras duas técnicas de visualização multidimensional utilizadas.

Conforme mostrado na **Figura 29**, observou-se, na etapa 4, que o algoritmo PCC obteve resultados surpreendentemente satisfatórios com apenas 1% das amostras treinadas, obtendo taxa de acerto maior que 95%. Esse aumento foi gradativo até os 5% de amostras treinadas; então houve uma pequena queda e seguiu variando menos de 1%. A análise do grupo de indicadores obteve resultado com 99% de acertos quando 10% das amostras foram treinadas.

Ainda na etapa 4, analisando os dados do coeficiente Kappa apresentados na **Figura 30**, é possível notar que o algoritmo PCC obteve excelente comportamento em ambiente com vários indicadores. Analisando o grupo de indicadores, o resultado se manteve positivo e bem próximo a 1. Já em relação ao tempo de execução do algoritmo PCC, é possível observar na **Figura 31** que a dificuldade do ser humano aumenta significativamente à medida que a quantidade de dados aumenta.

5.2 EXPERIMENTOS DE APOIO

Esta seção apresenta dois experimentos realizados como forma de apoio ao desenvolvimento e/ou avaliação da sistemática proposta. A *subseção 5.2.1* apresenta um experimento que avalia o algoritmo PCC em relação a dois algoritmos de aprendizado de máquina supervisionado. Já *subseção 5.2.2* apresenta um experimento que aplicou técnicas de aprendizado de máquina e visualização de informação em um cenário com poucas amostras disponíveis.

5.2.1 EXPERIMENTO I – AVALIANDO DIFERENTES ALGORITMOS DE APRENDIZAGEM DE MÁQUINA

O primeiro experimento realizado teve como finalidade comparar o comportamento de outro tipo de técnica de aprendizado de máquina com a técnica semi-supervisionada adotada neste trabalho, visando a análise de indicadores de desempenho em controle de qualidade de software. O uso de algoritmos de aprendizado não supervisionado foi descartado, uma vez que esses algoritmos não consideram os rótulos dos dados para classificação - recurso fundamental para se resolver os problemas de controle com os indicadores.

Assim, foram selecionados dois algoritmos de aprendizado supervisionado que são muito conhecidos nas literaturas: Perceptron de múltiplas camadas ou MLP (*Multilayer Perceptron*) e o algoritmo de K-vizinhos mais próximos ou KNN (*K-Nearest Neighbor*).

A partir do modelo de referência sugerido na etapa 1 da sistemática proposta (*seção 4.1*), foi selecionado apenas um indicador, denominado “Requisitos abertos”, com frequência de coleta diária, como mostrado na **Tabela 16**, no APÊNDICE B. Os dados para esse indicador foram coletados de três diferentes projetos de desenvolvimento de software, fornecidos pela empresa parceira deste trabalho e identificados como “A”, “B” e “C”.

O modelo de agrupamento de indicadores segundo as perspectivas do modelo BSC, proposto na etapa 2 (*seção 4.2*), foi alterado para possibilitar o uso dos resultados de diferentes projetos para um mesmo indicador. Foram então agrupados indicadores similares dos três projetos considerados neste experimento. O objetivo desse agrupamento foi possibilitar uma análise global do andamento de mais de um projeto de desenvolvimento de software através dos indicadores selecionados.

Para a rotulagem dos dados, foram recuperados dados de medição de cada indicador, totalizando 300 amostras de cada um. Após essa coleta, os valores dos indicadores foram tabulados no Excel como 3-tuplas, semelhante ao exemplo da **Tabela 7** (seção 4.4). Na sequência, foi verificado se haviam valores faltantes, acarretando na exclusão das instâncias (3-tuplas) nessa situação da tabela das amostras.

Assim, o processo de rotulagem dos dados poderia ser iniciado. A **Tabela 10** apresenta dados estatísticos relativos à base de dados históricos utilizada, considerando os três projetos alvo deste experimento.

Tabela 10 - Dados estatísticos sobre a base de dados históricos utilizada no experimento I.

DESCRIÇÃO	MIN	MAX	MÉD	DESVP
Projeto A	5	80	28.55	13.61
Projeto B	1	76	29.66	11.00
Projeto C	4	70	24.86	11.00

Para a rotulagem dos dados, dois gerentes de projeto da empresa (especialistas no domínio) analisaram os resultados de monitoramento das 300 amostras em conjunto, como grupos de indicadores. Cada 3-tripla foi então rotulada com o estado satisfatório, regular ou insatisfatório, fazendo os devidos registros na planilha das amostras.

No experimento, 100% das 300 amostras foram rotuladas pelos gerentes de projeto. Contudo, na etapa 4, de treinamento e análise dos indicadores com os algoritmos de aprendizado de máquina selecionados, apenas parte desses rótulos (de 2% a 20%) foi considerada; o restante foi utilizado para a análise. Convém observar que técnicas de visualização não foram utilizadas neste experimento. Assim, para a aplicação da etapa 4, os dados utilizados são caracterizados da seguinte maneira:

- Número de instâncias: 300;
- Número de atributos (indicadores dos projetos): 3;
- Valores faltantes: nenhum;
- 3 Classes: C1 = Satisfatório; C2 = Regular; C3 = Insatisfatório;
- Distribuição das classes: 33% para cada uma das classes;
- Informações dos atributos (indicadores dos projetos):
 - a) “Requisitos abertos” por dia no Projeto A;
 - b) “Requisitos abertos” por dia no Projeto B;
 - c) “Requisitos abertos” por dia no Projeto C.

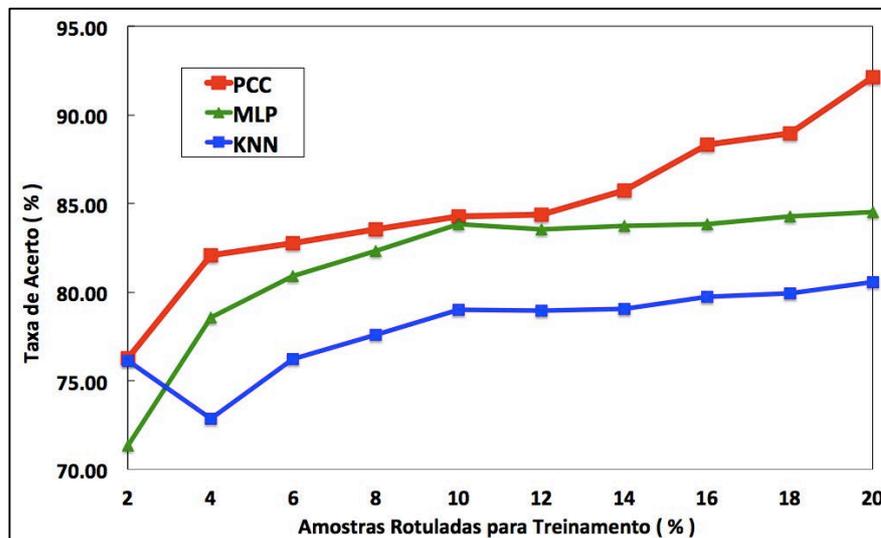
É importante ressaltar que cada atributo teve seu valor normalizado antes de ser

utilizado nos algoritmos. A ferramenta de software utilizada para a aplicação dos algoritmos de aprendizado supervisionado foi a Weka, disponibilizada gratuitamente pela Universidade de Waikato, na Nova Zelândia.

Os treinamentos foram realizados utilizando dez configurações diferentes, de acordo com o percentual de amostras treinadas: 2%, 4%, 6%, 8%, ..., 20%. Assim, cada algoritmo foi executado dez vezes para cada percentual treinado, sendo que o resultado apresentado é a média das dez execuções. Isso foi possível porque os algoritmos utilizados possuem característica estocástica e os dados rotulados no treinamento para o propósito do experimento são aleatórios.

Assim, os algoritmos supervisionados MLP e KNN, bem como o semi-supervisionado PCC foram utilizados usando as 300 amostras rotuladas oriundas das medições do indicador “Requisitos abertos” dos projetos A, B e C. O treinamento foi realizado gradativamente, de 2% em 2%, selecionando um percentual de amostras para treinamento e o restante para validação. A **Figura 32** exibe o resultado da taxa de acerto de cada algoritmo, conforme o aumento do percentual de amostras treinadas.

Figura 32 - Taxa de acerto dos algoritmos MLP, KNN e PCC no experimento I.

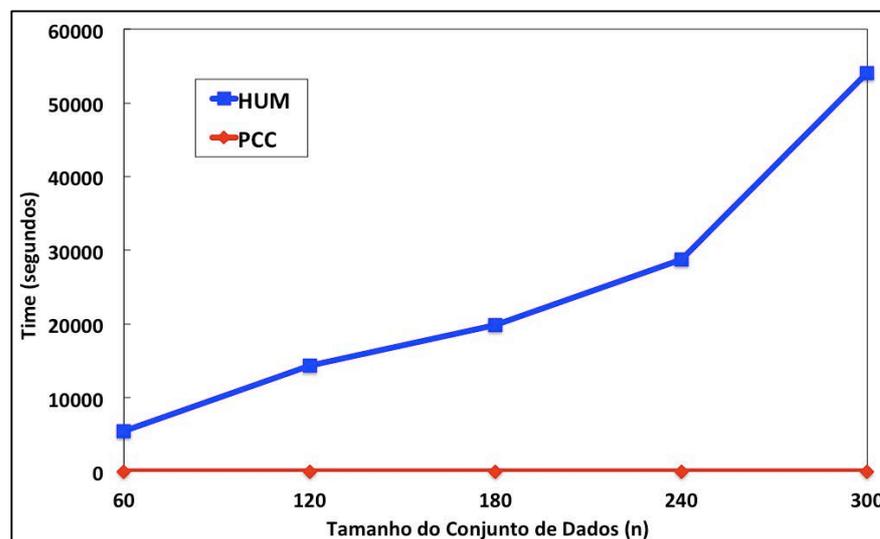


Em relação à taxa de acerto por classe de estados de saída, a **Tabela 11** apresenta o resultado do melhor momento ($k=5$) do algoritmo PCC. O algoritmo acertou 96 amostras das classes C1 (Satisfatório) e errou 4, classificando-as como C2 (Regular). Na classe C2, o algoritmo PCC classificou 86 amostras corretamente e 14 erradas, sendo 2 como classe C1 e 12 como classe C3 (Insatisfatório). Na classe C3, ele acertou 86 amostras e errou 14, classificando-a como classe C2 (Regular).

Tabela 11 - Taxa de acerto por classe de saída: real vs predita (algoritmo PCC).

		PREDITA		
		C1	C2	C3
REAL	C1	96	4	0
	C2	2	86	12
	C3	0	14	86

Outro ponto importante a destacar é o desempenho do algoritmo PCC para análise dos indicadores de desempenho se comparado ao ser humano (HUM). Para ilustrar essa diferença, a **Figura 33** apresenta o resultado de uma comparação entre o tempo de análise dos indicadores realizado pelo algoritmo PCC e o realizado por um especialista no domínio. Observa-se que a medição humana foi realizada por um especialista no domínio durante a etapa de rotulagem, o qual analisou visualmente os resultados do indicador “Requisitos abertos” em cada projeto e o comparou com o respectivo resultado esperado. O tempo gasto por um ser humano no processo de rotulagem das amostras foi calculado a partir do somatório de cada medição realizada. Já o tempo gasto no processo de análise feita pelo algoritmo foi medido através de função específica na própria ferramenta MATLAB.

Figura 33 - Tempo de execução do algoritmo PCC vs ser humano no experimento I.

Discussão sobre o experimento I

Essa discussão aponta aspectos positivos e negativos em relação à aplicação dos algoritmos MLP, KNN e PCC no cenário do experimento I, apresentado nesta seção.

Em relação à taxa de acerto dos algoritmos, foi possível notar na **Figura 32** que o algoritmo KNN apresentou um resultado interessante com 2% de amostras treinadas, mas

só obteve resultado próximo dos 81% de acerto, quando 20% das amostras foram treinadas. Já o algoritmo MLP obteve um crescimento acentuado no início e gradativo até o final, ficando abaixo dos 85% de acerto. Por fim, o algoritmo com melhor resultado durante toda variação de treinamento foi o PCC, finalizando com pouco mais de 92% de acerto.

Analisando os resultados da taxa de acerto por classe (*cluster*), obtidos no melhor momento do algoritmo PCC (**Tabela 11**), é possível notar que ele classificou apenas duas amostras erradas, que não são da classe C1 (“satisfatório”). Esse aspecto é importante para a confiabilidade da proposta deste trabalho, bem como para tomadas de decisão em um processo/projeto. Isso porque se a análise dos indicadores do processo/projeto mostra que a situação tem estado “satisfatório” (C1), significa que nenhuma ação para correção deve ser realizada. Por outro lado, se o algoritmo PCC tivesse classificado muitas amostras como “satisfatório” erroneamente, seria bastante prejudicial às tomadas de decisão, uma vez que essas amostras não seriam reavaliadas, trazendo problemas de confiabilidade ao processo.

Quanto ao tempo de execução do algoritmo PCC, é possível observar na **Figura 33** que, à medida que a quantidade de dados aumenta, a dificuldade do ser humano em realizar uma análise eficiente desses indicadores de desempenho também aumenta. Isso indica que o uso de aprendizado de máquina se mostra bastante superior em termos de tempo de análise quando comparado com um especialista humano.

5.2.2 EXPERIMENTO II – AVALIAÇÃO EM UM AMBIENTE COM POCAS AMOSTRAS

Embora o objetivo deste trabalho seja tratar grandes volumes de dados gerados por indicadores de desempenho de empresas desenvolvedoras de software, se o escopo do trabalho for ampliado pode ser possível ajustar a sistemática desenvolvida para avaliação do comportamento de indicadores em outros processos que não de software.

Dessa forma, este segundo experimento apresenta a aplicação do algoritmo semi-supervisionado PCC e de um algoritmo supervisionado em conjunto com técnicas de visualização de informação. O objetivo é analisar indicadores de desempenho de um setor de produção industrial, com poucas amostras fornecidas por uma empresa parceira, que atua há mais de 19 anos na área de automação industrial.

É importante ressaltar que os resultados deste experimento serão úteis para aplicação da sistemática em empresas de desenvolvimento de software que estão começando a padronizar seus processos segundo os níveis G e/ou F do modelo MPS-SW.

Nessa situação, o cenário também será pouco favorável, com poucos indicadores e poucas amostras para treinamento.

Convém informar que, apesar de estar sendo apresentado no final deste capítulo, este experimento foi realizado antes da sistemática ter suas etapas bem definidas e foi importante para o desenvolvimento da sistemática; logo, ele não segue a sequência das quatro etapas da sistemática proposta.

Além disso, este experimento foi desenvolvido em duas partes, onde cada uma delas requereu discussões que contribuiriam para a sistemática desenvolvida. Assim, a *subseção 5.2.2.1* apresenta a primeira parte do experimento (Parte A), que usa o algoritmo de aprendizado supervisionado Perceptron de múltiplas camadas ou MLP (*Multilayer Perceptron*). Já a *subseção 5.2.2.2* apresenta a segunda parte (Parte B), que utiliza o algoritmo de aprendizado semi-supervisionado PCC.

Para essas duas partes do experimento, algumas ações preliminares foram feitas e utilizadas em ambas as partes. Inicialmente, foram selecionados quatro indicadores: “disponibilidade de serviço”, “produtos entregues no prazo”, “nível de atendimento às requisições” e “quebra de equipamentos”. Esses indicadores não constam do modelo de referência de seleção de indicadores sugerido na etapa 1 da sistemática (*seção 4.1*). Contudo, também poderiam ser utilizados em projetos de desenvolvimento de software.

Os indicadores foram então agrupados, de modo a permitir uma análise global dos processos. O modelo de agrupamento de indicadores proposto na etapa 2 (*seção 4.2*) foi ajustado para que considerasse indicadores relacionados aos diferentes processos industriais considerados.

Para a rotulagem das amostras foram recuperados dados de medição de cada indicador, totalizando 30 amostras de cada indicador. Os indicadores foram então tabulados no Excel (.xls) como 4-tuplas de indicadores, semelhante ao formato exemplificado na **Tabela 7** (*seção 4.3*). Convém informar que nessa base de dados não havia amostras com valores faltantes. A **Tabela 12** apresenta dados estatísticos relativos à base de dados utilizada.

Tabela 12 - Dados estatísticos sobre a base de dados utilizada no experimento II.

DESCRIÇÃO	MIN	MAX	MÉD	DESV.P
Disponibilidade de serviço	80.12	97.78	89.80	4.41
Produtos entregues no prazo	82.78	97.98	90.30	4.11
Nível de atendimento às requisições	87.10	99.69	94.34	3.76
Quebra de equipamentos	0.98	19.32	9.78	5.19

Para a rotulagem dos dados, um especialista no domínio analisou os resultados de monitoramento das 30 amostras em conjunto. Cada uma das 4-tuplas foi então obtendo um valor de saída (satisfatório, regular ou insatisfatório), o qual era registrado na planilha das amostras rotuladas. Informa-se que 100% das 30 amostras foram rotuladas, sendo apresentado, posteriormente na etapa 4, apenas parte desses rótulos (10% e 50%) para os dois algoritmos utilizados; o restante foi utilizado para avaliação. Assim, os dados selecionados para uso neste experimento possuem as seguintes características:

- Número de instâncias: 30;
- Número de atributos (indicadores): 4;
- Valores faltantes: nenhum;
- 3 Classes: C1 = Satisfatório; C2 = Regular; C3 = Insatisfatório;
- Distribuição das classes: 33% para cada uma das classes;
- Informações dos atributos (indicadores):
 - a) Disponibilidade de serviço;
 - b) Produtos entregues no prazo;
 - c) Nível de atendimento às requisições;
 - d) Quebra de equipamentos.

5.2.2.1 EXPERIMENTO II – PARTE A: USO DO ALGORITMO SUPERVISIONADO MLP

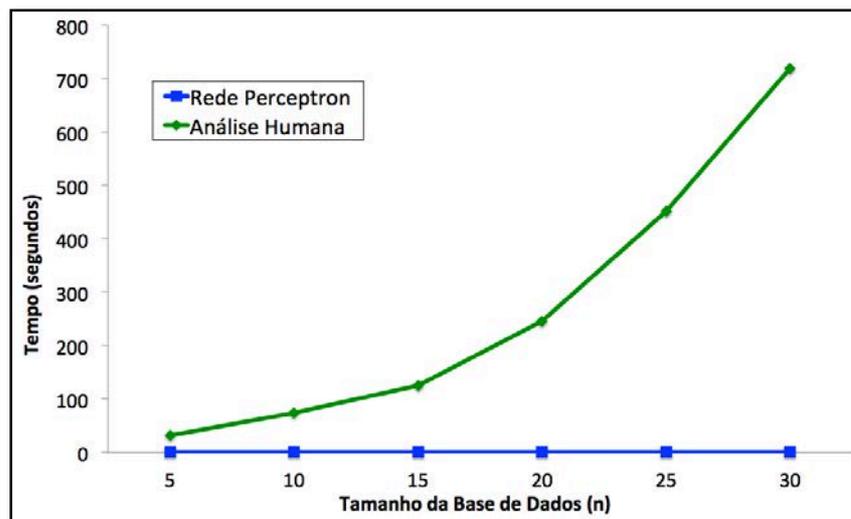
Esta parte do experimento utilizou o algoritmo MLP com as 30 instâncias do conjunto de quatro indicadores e, portanto, com quatro atributos. Primeiramente, foi realizado um experimento utilizando 50% das amostras para treinamento da rede e os outros 50% para teste. Os testes foram feitos focados em dois fatores principais: acurácia e tempo de análise.

Em relação à acurácia por classe, a **Tabela 13** apresenta o resultado no melhor caso do algoritmo MLP. O algoritmo acertou 100% das amostras das classes C1 (satisfatório) e C3 (insatisfatório). Na classe C2 (regular), houve o retorno de apenas uma amostra errada, classificando-a como classe C3 (insatisfatório).

Tabela 13 - Taxa de acerto classe de saída: real vs predita (algoritmo MLP).

		PREDITA		
		C1	C2	C3
REAL	C1	5	0	0
	C2	0	4	1
	C3	0	0	5

Quanto ao tempo de análise do algoritmo MLP, se comparado a um ser humano (HUM) o algoritmo obteve um desempenho muito superior, como é possível observar na **Figura 34**. Vale salientar que a medição humana apresentada foi realizada pelo especialista do domínio durante a etapa de rotulagem dos dados. Esse especialista analisou visualmente os resultados dos indicadores e os comparou com seus respectivos resultados esperados.

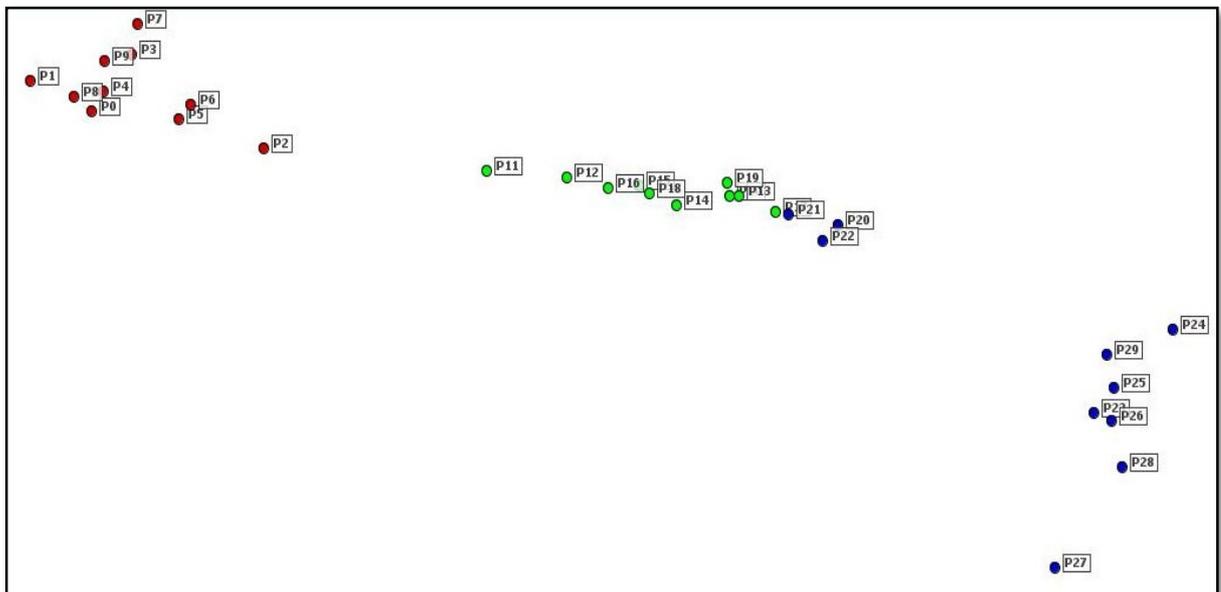
Figura 34 - Tempo de execução do teste: MLP vs HUM.

Nesse experimento, não foi possível entender o porquê do algoritmo MLP ter errado a classificação de uma amostra. Em princípio, errar a classificação de uma só amostra parece não ser significativo. Contudo, nesse cenário com poucas amostras o erro representa aproximadamente 3% das amostras, e passa a ser significativo.

A partir dos resultados deste experimento, foram realizadas pesquisas com o intuito de visualizar a relação existente entre os dados do conjunto, de modo a se verificar se poderia ter ocorrido um erro no processo de rotulagem, por exemplo. Dessa forma, foram aplicadas, no conjunto de dados classificado pelo algoritmo MLP, as três técnicas de visualização de informação já recomendadas neste trabalho.

A primeira técnica de visualização utilizada foi a de projeção por mínimos quadrados ou LSP. A **Figura 35** apresenta o resultado da projeção multidimensional dos dados, baseada nos atributos de cada amostra. Como já mencionado, nessa técnica, as cores representam as classes, cada amostra representa um ponto e cada ponto corresponde a um indicador, identificado pela letra “P”. Os indicadores mais similares são posicionados bem próximos no plano de projeção, enquanto os mais dissimilares (distantes) são posicionados mais afastados. Nesse tipo de projeção, é possível perceber que alguns indicadores de diferentes classes estão se misturando (ver o centro da projeção), mas não se pode explicar o motivo pelo qual isso ocorre.

Figura 35 - Projeção multidimensional dos dados utilizando a técnica LSP no Experimento II - Parte A.



Dessa forma, se fez necessária a busca por outra forma de visualização do conjunto de dados, cuja representação visual destacasse o relacionamento entre os atributos dos dados. Assim, a **Figura 36** apresenta o resultado da técnica de coordenadas paralelas ou PC, possibilitando analisar o comportamento das instâncias selecionadas na **Figura 35**, identificando padrões ou discrepâncias. Nessa técnica, as cores correspondem às classes, os eixos representam os atributos e as linhas representam as instâncias dos dados. Cada linha corta o eixo paralelo no valor correspondente ao atributo.

A **Figura 37** mostra alguns pontos destacados, a partir dos quais se pode observar que os atributos das instâncias possuem valores e comportamento similares, justificando o porquê delas terem sido posicionadas próximas na técnica de projeção LSP. Essa análise, embora visual e subjetiva, indica que as instâncias agrupadas na projeção exibida na **Figura**

37 poderiam ter sua classificação revisada, o que certamente contribuiria na acurácia dos testes anteriormente realizados (ver **Tabela 13**). Além disso, é possível identificar os atributos necessários para fazer a classificação das instâncias, bem como os atributos que poderiam ter um peso maior para evidenciar ou classificar uma classe específica.

Figura 36 - Coordenadas paralelas do conjunto de dados do Experimento II - Parte A.

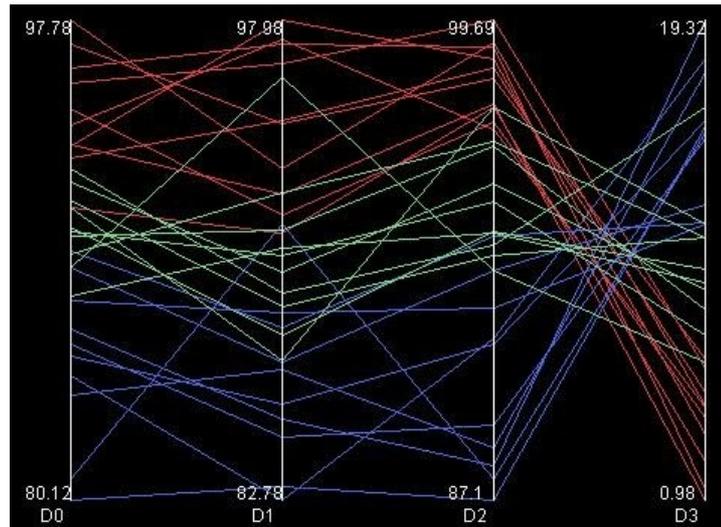
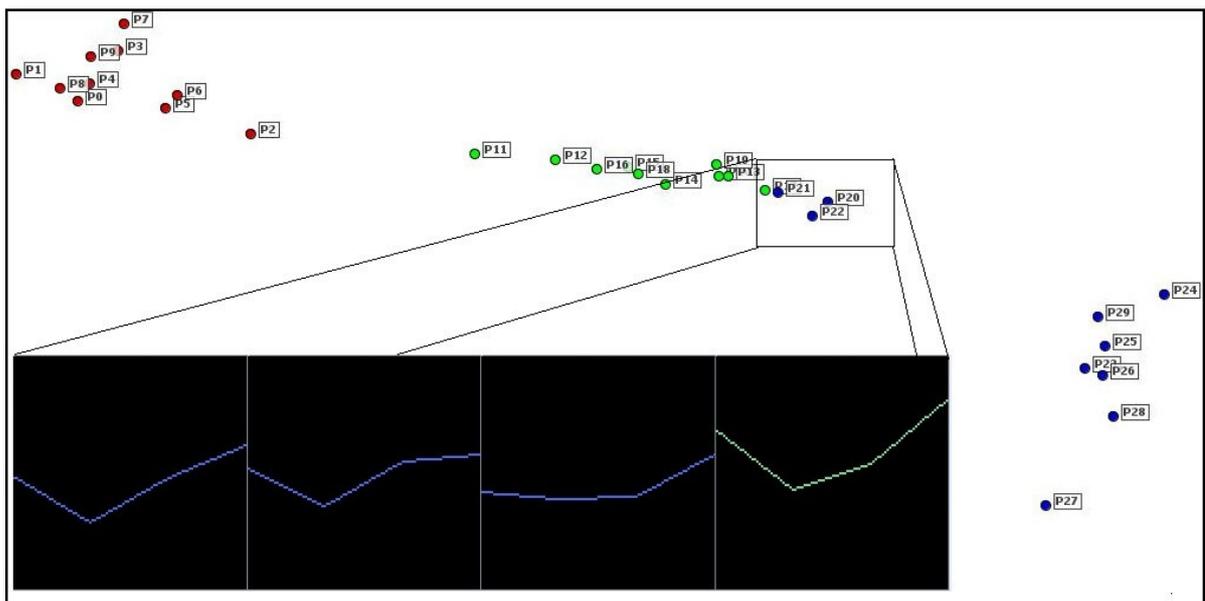


Figura 37 - Comportamento dos atributos de algumas instâncias selecionadas na projeção LSP.



Discussão sobre o Experimento II – Parte A

Como foi possível observar, as técnicas de visualização de informação possibilitaram a observação do comportamento dos atributos de tal modo que pode auxiliar o especialista no processo de rotulagem também em ambientes com poucas amostras disponíveis. Esse apoio é imprescindível, uma vez que algoritmos semi-supervisionados utilizam poucas amostras rotuladas, embora confiáveis. De modo geral, observou-se que o uso aprendido de máquina com técnicas de visualização de informação poderia, de fato, contribuir com o aumento da confiabilidade no processo de rotulagem de poucas amostras. Assim, este experimento levou a se cogitar o uso das técnicas de visualização também com o algoritmo PCC, assim como foi definido na sistemática proposta no *capítulo 4*.

5.2.2.2 EXPERIMENTO II – PARTE B: USO DO ALGORITMO PCC

Esta parte do experimento utilizou o algoritmo de aprendizado semi-supervisionado de competição e cooperação entre partículas, apresentado na *seção 3.5*. Foi utilizada a mesma base de dados da primeira parte do experimento (Parte A), com 30 amostras de quatro indicadores reais, que não se referem a processos de software. Esta seção compara o comportamento do algoritmo PCC com o algoritmo supervisionado MLP, usado na Parte A do experimento. Contudo, vale ressaltar que essa comparação não pode ser utilizada como critério de medição de desempenho entre os dois algoritmos.

O algoritmo PCC é indicado em situações onde poucos dados são rotulados, como, por exemplo, 10% ou menos. Porém, como forma de comparar o comportamento dele com o algoritmo supervisionado MLP, inicialmente foi utilizado o mesmo critério usado na Parte A do experimento (*subseção 5.2.2.1*): 50% das amostras rotuladas foram utilizadas para o treinamento do algoritmo e 50% foram selecionadas para teste. A **Tabela 14** apresenta os resultados dos testes no melhor caso, ressaltando-se que o algoritmo PCC acertou 100% das amostras em todas as classes: C1 (satisfatório), C2 (regular) e C3 (insatisfatório).

Tabela 14 - Taxa de acerto classe de saída: real vs predita (algoritmo PCC).

		PREDITA		
		C1	C2	C3
REAL	C1	5	0	0
	C2	0	5	0
	C3	0	0	5

Para analisar o comportamento do algoritmo PCC em um cenário com menos amostras rotuladas ainda, o treinamento do algoritmo foi refeito da seguinte maneira: foram usadas apenas 10% das amostras rotuladas (uma de cada classe), extraídas dos 50% das amostras utilizados para treinamento do algoritmo MLP. Os 90% restantes de amostras foram utilizadas para os testes. Os resultados obtidos foram positivos também nesse caso, conforme é possível observar na comparação apresentada na **Tabela 15**.

Tabela 15 - Comparativo do comportamento dos algoritmos: PCC (Partículas) vs MLP (Perceptron).

DESCRIÇÃO	PARTÍCULAS	PARTÍCULAS	PERCEPTRON
Amostras Rotuladas	10%	50%	50%
Média	96,22%	94,19%	79,04%
Mínimo	33,33%	40,00%	66,67%
Máximo	100,00%	100,00%	93,33%
Desvio Padrão	5,36%	6,62%	10,34%

Discussão sobre o Experimento II – Parte B

Assim como na Parte A deste experimento, os indicadores analisados foram relativos a um processo de produção de uma empresa não produtora de software. Contudo, foram observadas muitas semelhanças no tratamento de indicadores dos processos dessa empresa com os indicadores de processos de desenvolvimento de software.

Nesta parte do experimento, observou-se que, com apenas 10% das amostras rotuladas, o resultado do algoritmo PCC foi melhor que ele mesmo com 50%. Isso pode ter ocorrido devido às características da concepção do modelo semi-supervisionado. Outra possibilidade é que com um total pequeno de amostras (30), a margem de erro é grande e pode ter influenciado positivamente para esse melhor resultado.

Embora não tenha sido realizado teste com o algoritmo MLP com 10% das amostras rotuladas, pode-se supor que os resultados seriam inferiores aos obtidos com o algoritmo PCC. Essa suposição faz sentido de certa forma, considerando a característica de aprendizado supervisionado do algoritmo MLP, que necessita de uma quantidade maior de amostras rotuladas para produzir resultados mais significativos. Além disso, se o melhor resultado do algoritmo MLP (com 50% de dados rotulados) for comparado com o melhor resultado do algoritmo PCC (com apenas 10% de dados rotulados), esse último já se mostra superior ao primeiro.

6 CONCLUSÃO E TRABALHOS FUTUROS

Estudos na área de qualidade de software vêm sendo conduzidos, há algum tempo, com o objetivo de definir e controlar melhor os processos de produção de software. Modelos e normas têm colaborado com a qualidade dos processos nesse cenário, como é o caso dos modelos CMMI, MoProSoft e MPS-SW. A utilização desses modelos auxilia a definição e controle dos processos de forma evolutiva, por meio de níveis de maturidade. O controle dos processos normalmente faz uso de indicadores de desempenho, que contribuem para as tomadas de decisão, seja em nível de processo, projeto ou de negócio. Esses indicadores podem ser analisados individualmente ou em grupos, e têm, geralmente, diferentes tipos e frequências de coleta. Além disso, ao longo do tempo, a base histórica dos dados desses indicadores tende a aumentar consideravelmente, dificultando as análises para as tomadas de decisão.

Nesse contexto, o *capítulo 2* apresentou levantamentos sobre indicadores chave de desempenho, bem como os principais problemas relacionados à coleta, aplicação e análise dos dados. Esse capítulo evidenciou a importância dos indicadores para a gestão estratégica e melhorias contínuas nos processos organizacionais. Para a análise de grupos de indicadores em vários processos, a utilização de um sistema inteligente foi apresentada como uma solução, frente à complexidade e grande quantidade de dados dos indicadores geradas nas várias cadeias de projetos em desenvolvimento. Um sistema inteligente pode sinalizar quando um determinado resultado de indicadores está fora do resultado considerado adequado para o processo, como se fosse um semáforo, com diferentes posições sobre a situação (estado) dos resultados.

Nessa direção, o principal objetivo deste trabalho foi apresentar uma sistemática baseada em técnicas de aprendizado de máquina para auxiliar a análise dos indicadores de desempenho nos processos de desenvolvimento de software, tanto de forma conjunta, como isoladamente. A técnica adotada foi a de aprendizado semi-supervisionado, com o algoritmo de cooperação e competição entre partículas ou PCC (*Particle Competition and Cooperation*) de Breve (2010), que é bioinspirado em uma comunidade de formigas. A sistemática inclui a utilização de três técnicas de visualização de informação, que auxiliam, de forma complementar, o processo de rotulagem das amostras por parte dos especialistas

do domínio: (1) Projeção por mínimos quadrados ou LSP (*Least Square Projection*); (2) Escalonamento multidimensional clássico ou CMDS (*Classical Multidimensional Scaling*); (3) Coordenadas paralelas ou PC (*Parallel Coordinates*). Essas técnicas permitem a visualização da relação existente entre as amostras de indicadores em um ambiente multidimensional. Assim, especialistas do domínio podem identificar visualmente prováveis erros no processo de rotulagem e corrigirem os rótulos dos indicadores.

Ressalta-se que não foram encontrados, na literatura, trabalhos utilizando técnicas de aprendizado de máquina em análises de indicadores de desempenho de processos de desenvolvimento de software. Trabalhos encontrados nesse contexto tratam de indicadores de desempenho em processos de outras áreas de aplicação, diferentes da Engenharia de Software. Adicionalmente, o algoritmo PCC também não tinha sido utilizado ainda para tais propósitos, segundo Breve, que é coorientador deste trabalho e autor do algoritmo. Assim, a sistemática desenvolvida neste trabalho contribui de forma original com a área de desenvolvimento de software.

Os conceitos e técnicas de aprendizado de máquina avaliados para este trabalho foram apresentados no *capítulo 3*. Foi mostrado que técnicas de aprendizado supervisionado trariam alto custo no processo de rotulagem dos dados, enquanto técnicas de aprendizado não supervisionado ignorariam totalmente os rótulos das amostras. Oportunamente, as técnicas de visualização de informação adotadas neste trabalho também foram apresentadas no *capítulo 3*.

Em decorrência da sistemática desenvolvida, este trabalho também contribuiu com a proposta de dois modelos, envolvendo indicadores de processos de desenvolvimento de software, apresentados no *capítulo 4*. Esses modelos apoiam as duas primeiras etapas da sistemática, mas o uso é facultativo. A intenção é auxiliar as organizações na escolha das configurações de indicadores adequados ao controle da qualidade dos processos. Além disso, esses modelos também podem ser utilizados independentemente da sistemática proposta.

Um dos modelos contempla indicadores para os processos que compõem os sete macroprocessos dos níveis G e F: Gerência de Requisistos (GRE), Gerência de Projetos (GPR), Medição (MED), Garantia da Qualidade (GQA), Gerência de Portifólio de Projetos (GPP), Gerência de Configuração (GCO) e Aquisição (AQU). Esse modelo auxilia a organização a definir seus indicadores de desempenho ou a complementar seu conjunto de indicadores. Observa-se que aspectos do modelo MPS-SW relevantes ao trabalho foram abordados no *capítulo 2*, incluindo a ontologia de Pizzoleto (2013), que colaborou com a definição de indicadores do modelo proposto.

O segundo modelo proposto auxilia as organizações a comporem grupos de indicadores significativos para as análises, segundo as quatro perspectivas da abordagem

BSC (*Balanced Scorecard*): financeira, de clientes, dos processos internos e de aprendizado e crescimento. Esse modelo possibilita a composição de grupos de indicadores alinhados com a visão estratégica da organização.

Ressalta-se que além de contribuir com a área acadêmica, este trabalho traz contribuições para as organizações produtoras de software, em termos de gestão de seus processos de software e de gestão estratégica. O fato do trabalho ter sido direcionado aos processos dos níveis G e F do modelo MPS-SW ajuda a atingir um número significativo de empresas certificadas nesse modelo, uma vez que 87% das empresas certificadas no MPS-SW são nesses dois níveis.

Para avaliar a sistemática proposta, um estudo de caso foi realizado e apresentado no *capítulo 5*, seguindo os critérios e as etapas da sistemática. Foram utilizados dados reais de monitoramento de indicadores de processos de software de uma empresa parceira, certificada no nível G do modelo MPS-SW. Essa empresa está localizada na cidade de Ribeirão Preto e atua há mais de 30 anos no mercado de administrações municipais. O canal de interação com um cenário real de tomadas de decisão em linhas de produção de software contribuiu significativamente para o trabalho. Foram realizadas diversas reuniões com os gerentes de projetos, que apresentaram o conjunto de indicadores utilizados e as necessidades emergentes do gerenciamento de qualidade, bem como as ferramentas de software então utilizadas. A base histórica de indicadores fornecida pela empresa contemplou dados de três anos de monitoramento; os gerentes de projeto também colaboraram expressivamente com a rotulagem das amostras.

Os resultados do estudo de caso mostraram que o algoritmo PCC obteve excelente taxa de acerto, mesmo em um ambiente com, no máximo, 10% de amostras treinadas. Como a base utilizada estava desbalanceada quanto à distribuição das classes, foi analisado o coeficiente Kappa, que apresentou resultados satisfatórios na análise do grupo de indicadores considerado, obtendo valores acima de 0,9000.

O estudo de caso possibilitou constatar a viabilidade do processo de rotulagem das amostras com uma técnica de aprendizado semi-supervisionado: o algoritmo PCC. Embora tenha sido investido um tempo significativo com a rotulagem de 100% da base de dados utilizada, vale salientar que, em uma aplicação da sistemática em processos de software, esse percentual pode cair para 10%. Dessa forma, o tempo do especialista da empresa para ajudar no processo de rotulagem pode ser considerado adequado, mediante aos benefícios para várias linhas de produção de software.

Para auxiliar a definição da sistemática proposta, foram realizados dois experimentos, conforme apresentado no *capítulo 5*. O primeiro experimento comparou dois algoritmos de aprendizado supervisionado com o algoritmo semi-supervisionado PCC, usando bases de dados reais. O algoritmo PCC obteve excelente taxa de acerto, superior

aos dois algoritmos de aprendizado supervisionado. Observou-se que as três técnicas de visualização de informação adotadas na sistemática contribuíram para o aumento da confiabilidade do processo de rotulagem. Um segundo experimento contou com um ambiente com poucas amostras e com indicadores que não são de processos de software. Observa-se que a análise de poucos dados de indicadores, embora incomum em empresas com certo nível de maturidade, pode ser considerada “normal”, por um período de tempo, em empresas que estão no início da padronização de seus processos. Nesse experimento, o algoritmo PCC mostrou alta taxa de acerto, mesmo com um percentual menor ou igual a 10% de amostras treinadas.

Deve-se ressaltar que tanto nos experimentos realizados como no estudo de caso, as técnicas de visualização de informação contribuíram para o aumento da confiabilidade do processo de rotulagem dos dados. Observou-se que a técnica fundamental é a de Projeção por mínimos quadrados ou LSP. A aplicação das outras duas técnicas, CMDS e PC, talvez não seja necessária, dependendo da base de dados analisada. É importante mencionar que o algoritmo PCC obteve melhores resultados de classificação após os ajustes das amostras indicados pelas técnicas de visualização de informação – o que aumentou a confiabilidade do processo.

Propostas de trabalhos futuros

Para contribuir com a análise de indicadores em empresas desenvolvedoras de software de diferentes portes, incluindo micro, pequenas e médias empresas, trabalhos futuros poderão analisar a viabilidade de se desenvolver uma ferramenta de software, integrando as diferentes técnicas e modelos utilizados neste trabalho. Essa ferramenta poderá ter uma arquitetura em camadas ou ser implementada como um serviço, com disponibilização de acesso/consumo em nuvem. Para isso, seria interessante considerar o desenvolvimento de um sistema com interfaces específicas e integradas para as seguintes funcionalidades: definição e leitura dos indicadores, agrupamento dos indicadores, rotulagem das amostras, visualização das projeções e visualização dos resultados em painéis de controle (*dashboards*).

Por outro lado, um repositório de dados de indicadores de desempenho de processos de software poderia ser criado e disponibilizado publicamente, incluindo indicadores de outras organizações desenvolvedoras de software, sem que fossem explicitamente identificadas. Isso ajudaria a realização de experimentos e compartilhamento de dados.

Outro trabalho futuro poderá propor soluções para tratar dados de monitoramento faltantes. Na literatura existem formas disponíveis para essa tratativa, porém se faz

necessária uma pesquisa mais aprofundada sobre o melhor método a ser utilizado e como implementá-lo, alterando a sistemática proposta. Isso é importante para situações em que não se tem uma determinada medição pontual de algum indicador selecionado para análise do algoritmo PCC.

Convém ressaltar que há outras versões do algoritmo PCC que poderão ser avaliadas, as quais têm suporte a aprendizado ativo e saída que indica o grau de confiabilidade nos rótulos atribuídos (BREVE, 2014; BREVE, 2013a; BREVE, 2013b). Nesse caso, em vez de selecionar aleatoriamente todas as amostras a serem rotuladas pelo especialista, o próprio algoritmo pode começar com poucos dados rotulados e então indicar algumas amostras em que tem menos confiança no rótulo atualmente atribuído. Desse modo, os especialistas podem conferir e alterar o rótulo atribuído pelo algoritmo se for necessário, evitando que eventuais erros se propaguem por grande parte da base de dados. Consequentemente, é possível obter uma classificação com maior acurácia, sem aumentar o esforço por parte do especialista.

Normalmente, recomenda-se que a rotulagem de algumas amostras seja realizada com certa frequência para treinamento, à medida que novos dados são adicionados à base histórica dos indicadores. Porém, trabalhos futuros poderão pesquisar formas de reuso das amostras rotuladas, sem a necessidade de novos treinamentos da Rede Neural Artificial. Para isso, poder-se-ia identificar qual seria o tempo máximo para o algoritmo PCC continuar a fornecer resultados satisfatórios dentro do mesmo contexto da aplicação e sem alteração dos indicadores. Isso poderia contribuir com a redução do tempo investido pelo especialista na rotulagem de novas amostras, sem perder a credibilidade dos resultados.

REFERÊNCIAS

- ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO 9000**: sistemas de gestão da qualidade – Fundamentos e vocabulários. Rio de Janeiro, 2005.
- ALPAYDIN, E. **Introduction to machine learning**. Cambridge, Ma: The MIT Press, 2004.
- ASSIRI, A.; ZAIRI, M.; EID, R. **How to profit from the balanced scorecard: An implementation roadmap**. *Industrial Management & Data Systems*, v. 106, n. 7, p. 937-952, 2006.
- BARCELLOS, M. P. **Uma estratégia para medição de software e avaliação de bases de medidas para controle estatístico de processos de software em organizações de alta maturidade**. 2009. 419 f. Tese (Doutorado em Engenharia de Sistemas e Computação) - Curso de Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro, 2009.
- BERTEI, M.; MARCHI, L.; BUONCRISTIANI, D. Exploring Qualitative Data: the use of Big Data technology as support in strategic decision-making. **The International Journal of Digital Accounting Research**, v.15, p.99-126, 2015.
- BIRCHFIELD, S. T.; SUBRAMANYA, A. Microphone Array Position Calibration by Basis-Point Classical Multidimensional Scaling. **IEEE Transactions on Speech and Audio Processing**, v.13, n.5, p.1025-1034, 2005.
- BLOMQUIST, E.; ÖHGREN, A.; SANDKUHL, K. Ontology construction in an enterprise context: comparing and evaluating two approaches. In: INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS, 8., 2006. **Proceedings...** Paphos, 2006. p.86-93.
- BONABEAU, E.; DORIGO, M.; THERAULAZ, G. **Swarm Intelligence: From Natural to Artificial Systems**. Oxford University Press, New York, NY, 1999.
- BOYD, L. H.; COX, J. F. A cause-and-effect approach to analyzing performance measures. **Production and Inventory Management Journal**, v.38, n.3, p.25-32, 1997.
- BRAGA, A. P; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes Neurais Artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2000.
- BREVE, F. A. Query Rules Study on Active Semi-Supervised Learning using Particle Competition and Cooperation. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL (ENIAC), 6., 2014. **Anais...** São Carlos, 2014. p.134-140.
- _____. Active Semi-Supervised Learning using Particle Competition and Cooperation in Networks. In: THE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2013a, **Proceedings...** Dallas, Texas, 2013a. p.1737-1742.
- _____. Combined Active and Semi-Supervised Learning using Particle Walking Temporal Dynamics. In: COMPUTATIONAL INTELLIGENCE and BRAZILIAN CONGRESS ON COMPUTATIONAL INTELLIGENCE (BRICS-CCI and CBIC), 11., 2013b. **Proceedings...** Ipojuca, Pernambuco, 2013b. p. 15-20.
- _____. **Aprendizado de máquina utilizando dinâmica espaço-temporal em redes complexas**. 2010. 165 f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação - ICMC-USP, Universidade de São Paulo, São Carlos, 2010.

BREVE, F. A.; ZHAO, L.; QUILES, M. G.; PEDRYCZ, W.; LIU, J. Particle competition and cooperation in networks for semi-supervised learning. **IEEE transactions on knowledge and data engineering**, v.24, p. 1686-1698, 2009.

BRITO, A. **Indicadores institucionais**. 2005. Disponível em: <portal2.tcu.gov.br/portal/pls/portal/2054594.PDF>. Acesso em: agosto 2014.

BRYDE, D. J. Modeling Project Management Performance. **International Journal of Quality & Reliability Management**. v. 20, n. 2, p. 229-254, 2003.

BUENO, W. **Utilizando BSC como medição para o MPS-BR nível F**. Dissertação (Especialização em Análise, Projeto e Gerência de Sistemas) – Centro de Ciências Exatas – CCE, Universidade Estadual de Londrina, Londrina, 2009.

CATTINELLI, I; BOLZONI, E; CHERMISI, M; BELLOCCHIO, F; BARBIERI, C; MARI, F; AMATO, C; MENZER, M; STOPPER, A; GATTI, E. Computational intelligence for the Balanced Scorecard: studying performance trends of hemodialysis clinics. **Artificial intelligence in medicine**, v.58, n.3, p.165-173, 2013.

CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. **Semi-supervised learning**. MA: The MIT Press, Cambridge, 2006.

DARA, R.; KREMER, S.; STACEY, D. Clustering unlabeled data with SOMs improves classification of labeled real-world data. In: THE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE (WCCI), 1., 2002. **Proceedings...** Honolulu, Hawaii, 2002. p. 2237-2242.

DEMING, W. E. **Qualidade: a revolução da administração**. Rio de Janeiro: Marques Saraiva, 1990.

ELER, D. M.; NAKAZAKI, M. Y.; PAULOVICH, F. V.; SANTOS, D. P.; ANDERY, G. F.; OLIVEIRA, M. C. F.; NETO, J. B.; MINGHIM, R. Visual analysis of image collections. **The visual computer**. v.25, n.10, p. 923-937, 2009.

FERNANDES, K. J.; RAJA, V.; WHALLEY, A. Lessons from implementing the balanced scorecard in a small and medium size manufacturing organization. **Technovation**. v.26, n.5/6, p. 623-634, 2006.

FERNEDA, E. **Redes neurais e sua aplicação em sistemas de recuperação de informação**. 2006. Disponível em: <<http://www.scielo.br/pdf/ci/v35n1/v35n1a03.pdf>>. Acesso em: dezembro 2013.

FILHO, A. T. **Indicadores de gerenciamento de projetos**. Monitoração contínua. São Paulo: M. Books do Brasil, 2010.

GOLDRATT, E. **Teoria das restrições**. New Haven, CT, Goldratt Satellite Program/Avraham Y. Goldratt Institute do Brasil, 1999.

GUARINO, N. **Formal ontology in information systems**. Amsterdam: IOS Press, 1998.

HAYKIN, S. **Neural Networks: a comprehensive foundation**. Upper Saddle River, Nj: Prentice Hall, 1994. 768 p.

HIKAGE, O. K.; SPINOLA, M. M.; LAURINDO, F. J. B. Software de *balanced scorecard*: proposta de um roteiro de implantação. **Produção**. v.16, n.1, p.140-160, 2006.

HUMPHREY, W. S. Characterizing the software process: a maturity framework. **IEEE Software**. v.5, n.2, p. 73-79, 1988.

INSELBERG, A; DIMSDALE, B. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: THE CONFERENCE ON VISUALIZATION (VIS'90), 1., 1990. **Proceedings...** Los Alamitos, CA, USA, 1990. p. 361-378.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys**. New York, NY, p. 264-323, 1999.

KAPLAN, R.; NORTON, D. P. The Balanced Scorecard – Measures that Drive Performance. **Harvard Business Review**, p.71-79, Jan-Feb, 1992.

_____. **A estratégia em ação: balanced scorecard**. 4. ed. Rio de Janeiro: Campus, 1997.

KOVACS, Z. Lx. **Redes Neurais Artificiais: Fundamentos e Aplicações**. São Paulo: Edição Acadêmica, 1996.

KUMAR, R.; SAHNI, A.; MARWAH, D. Real Time Big Data Analytics Dependence on Network Monitoring Solutions using Tensor Networks and its Decomposition. **Network and Complex Systems**. v.5, n.2, p.30-37, 2015.

LORIGGIO, A. **De Onde Vêm os Problemas: método para um diagnóstico eficaz**. São Paulo: Negócio Editora, 2002. 211 p.

LUNKES, R. J.; SCHNORRENBARGER, D.; GASPARETTO, V.; CAPARICA, C. E. Implementação do Balanced Scorecard em uma Empresa de Software. **Pensar Contábil**, v.11, n.45, p. 24-33, 2009.

MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. In: THE BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY. 5., 1967. **Proceedings...** Los Angeles, CA, 1967. p.281-297.

MAFFEO, B. **Engenharia de software e especificação de sistemas**. Rio de Janeiro: Campus, 1992. p.484.

MELO, A. C.; SILVA, A. L.; MARTE, C. L.; FERREIRA, M. R.; SASSI, R. J.; FERREIRA, R. P. Aplicação de Técnicas da Inteligência Artificial para obtenção de Indicadores de Desempenho como Medida de Qualidade em Rodovias Cessionadas. In: CONGRESSO BRASILEIRO DE RODOVIAS E CONCESSÕES (CBR&C), 7., 2011. **Anais...** Foz do Iguaçu, 2011. p.1-16.

MELLO, R. G.; JUNIOR, D. A. P.; OLIVEIRA, J. F. G.; BREMER, C. F. **Avaliação de desempenho para o gerenciamento estratégico do chão de fábrica**. ANPAD. 2000. 14p.

MITCHELL, T. **Machine Learning**. [s.i]: McGraw Hill, 1997.

MOURA, L. R. **Gestão Estratégica da Informação**: proposição de um modelo de organização baseado no uso da informação como recurso da gestão empresarial. 1999. 117 f. Dissertação (Mestrado em Economia, Administração e Contabilidade) – Escola Politécnica, Universidade de São Paulo, São Paulo, 1999.

MOREIRA, R. T.; LIMA, G. N.; MACHADO, B. B.; MARINHO, W. T.; VASCONCELOS, A.; ROUILLER, A. C. Uma abordagem para melhoria do processo de software baseada em medição. In: SIMPÓSIO BRASILEIRO DE QUALIDADE DE SOFTWARE, 8., 2009. **Anais...** Ouro Preto, MG, 2009. p.39-53.

NATARAJAN, B. K. **Machine learning: a theoretical approach**. Morgan Kaufmann, 1991.

NEELY, A. D.; GREGORY, M. J; PLATTS, K. W. Performance measurement system design: a literature review and research agenda. **International Journal of Operations and Production Management**, v.15, n.4, p.80-116, 1995.

NEELY, A. D. The performance measurement revolution: why now and what next? **International Journal of Operations and Production Management**, v.19, n.2, p.205-228, 1999.

_____. The evolution of performance measurement research. Developments in the last decade and a research agenda for the next. **International Journal of Operations and Production Management**, v.25, n.12, p.1264-1277, 2005.

NETO, D. M. V.; PATAH, L. A.; Ampliação da visão de indicadores para projetos: foco em equipes virtuais. **GEPROS - Gestão da Produção, Operações e Sistemas**, Bauru, a.9, n.2, p.17-33, abr-jun/2014.

NETO, S. B.; NAGANO, M. S.; MORAES, M. B. C. **Utilização de redes neurais artificiais para avaliação socioeconômica: uma aplicação em cooperativas**. 2005. Disponível em: <https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC0QFjAA&url=http://www.rausp.usp.br/download.asp?file=V4101059.pdf&ei=wZnMUqvIB4qNkAeu44GQDg&usg=AFQjCNFY97jyLdni7Ub_5vxZGB-ZEuZS6Q>. Acesso em: dezembro 2013.

OJHA, T. R. **Analysis of key performance indicators in software development**. Thesis (Master's degree in Information Technology) - Faculty of Computing and Electrical Engineering, Tampere University of Technology, Finland 2014.

OLIVEIRA, R. A. P.; SILVA, L. F.; ELER, D. M. Hybrid Visualization: A New Approach to Display Instances Relationship and Attributes Behaviour in a Single View. In: INTERNATIONAL CONFERENCE ON INFORMATION VISUALIZATION, 19., 2015. **Proceedings...** Barcelona, 2015. p.277-282.

PAIVA, J. G. S.; CRUZ, L. F.; PEDRINI, H.; TELLES, G. P.; MINGHIM, R. Improved Similarity Trees and their Application to Visual Data Classification. **IEEE Transactions on visualization and computer graphics**, v.17, n.12, p.2459-2468, 2011.

PARK, R.E.; GOETHERT, W.B.; FLORAC, W.A. **Goal-Driven Software Measurement: A Guidebook**, Pittsburgh, PA, Software Engineering Institute, Carnegie Mellon University, 1996.

PAULOVICH, F. V.; NONATO, L. G.; MINGHIM, R.; LEVKOWITZ, H. Least Square Projection: a fast high precision multidimensional projection technique and its application to document mapping. **IEEE Transactions on Visualization and Computer Graphics**, v.14, n.3, p.564-575, 2008.

PEDROSO, S. L. **Processo de medição de desenvolvimento de software como suporte aos objetivos estratégicos de negócios**: estudo de caso em empresas desenvolvedoras de software. Porto Alegre: Pontifícia Universidade Católica do Rio Grande do Sul, 2010.

PEREIRA, C. A. Ambiente, empresa, gestão e eficácia. In: CATELLI, A. (Coord.). **Controladoria**: uma abordagem da gestão econômica – GECON. São Paulo: Atlas, 1999. p.35-80.

PIZZOLETO, A. V. **Ontologia Empresarial no modelo MPS.BR visando modelagem de processos de negócios, com foco nos níveis G e F**. Dissertação (Mestrado em Ciência da Computação) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto, 2013.

PIZZOLETO, A. V.; OLIVEIRA, H. C. Methodology for ontology development in support to the MPS model for software. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING RESEARCH AND PRACTICE (SERP'13), 2013. **Proceedings...** Las Vegas, Nevada, 2013. p.479-485.

PMI - PROJECT MANAGEMENT INSTITUTE INC (Pennsylvania). **Um guia do conhecimento em gerenciamento de projetos**: Guia PMBOK. 4. ed., Newtown Square, 2008.

PRESSMAN, R. S. **Engenharia de Software**. Trad. José C. Barbosa dos Santos. 3.ed. São Paulo: Makron Books do Brasil, 1995. 1056p.

RAGHU, P.P.; POONGODI, R.; YEGNANARAYANA, B. A combined neural network approach for texture classification. **Neural Networks**. Oxford, UK, v.8, p.975-987, 1995.

REZENDE, D. A. **Engenharia de software e sistemas de informação**. 3.ed. Rio de Janeiro: Brasport, 2005.

REIS, C. **Tipos de indicadores**. Disponível em: <<http://slideplayer.com.br/slide/49556/#>>. Acesso em: agosto 2014.

ROCHA, G. **Uma infra-estrutura de apoio a um processo de medição de projetos em micro e pequenas empresas de software**. Dissertação (Mestrado em Ciência da Computação) – Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

SANTOS, G.; MONTONI, M.; FILHO, R. C. S.; KATSURAYAMA, A. E.; ZANETTI, D.; BARRETO, A. O. S.; ROCHA, A. R. Indicadores da Implementação do Nível E do MR-MPS em uma Instituição de Pesquisa. In: SIMPÓSIO BRASILEIRO DE QUALIDADE DE SOFTWARE, 8., 2009. **Anais...** Ouro Preto, MG. 2009, p.382-389.

SERRAPIÃO, A. B. S. Fundamentos de otimização por inteligência de enxames: uma visão geral. **Revista Controle e automação**, v.20, n.3, 2009.

SOFTEX - ASSOCIAÇÃO PARA PROMOÇÃO DA EXCELÊNCIA DO SOFTWARE BRASILEIRO. **MPS.BR Melhoria de processo do software brasileiro: guia de Implementação – Parte 9: Implementação do MR-MPS em organizações do tipo Fábrica de Software**, 2012a. Disponível em: <http://www.softex.br/wp-content/uploads/2013/07/MPS.BR_Guia_de_Implementacao_Parte_9_20111.pdf>. Acesso em: janeiro 2013.

_____. **MPS.BR Melhoria de processo do software brasileiro: website do programa MPS**, 2012b. Disponível em: <<http://www.softex.br/mpsbr>>. Acesso em: janeiro 2013.

_____. **MPS.BR Melhoria de processo do software brasileiro: avaliações MPS-SW (Software)**, 2013. Disponível em: <http://www.softex.br/wp-content/uploads/2013/07/2Avaliacoes-MPSSW-Publicadas_11.NOV_.2013_513.pdf>. Acesso em: janeiro 2014.

_____. **MPS.BR Melhoria de processo do software brasileiro: equivalência MR-MPS-SW e CMMI-DEV**, 2014. Disponível em: <<http://www.softex.br/faqs-item/equivalencia-mr-mps-e-cmmi-2/>>. Acesso em: julho 2015.

_____. **MPS.BR Melhoria de processo do software brasileiro: avaliações MPS-SW (Software)**, 2015. Disponível em: <http://www.softex.br/wp-content/uploads/2013/07/2Avaliacoes-MPSSW-Publicadas_26.jun_.2015_649.pdf>. Acesso em: agosto 2015.

SOMMERVILLE, I. **Engenharia de Software**. 9. ed. São Paulo: Pearson, 2011. 529p.

TAKASHINA, N. T.; FLORES, M. C. X. **Indicadores da qualidade e do desempenho**. Rio de Janeiro: Qualitymark, 1996.

TEJADA, E.; MINGHIM, R.; NONATO, L. G. On improved projection techniques to support visual exploration of multidimensional data sets. **Information Visualization**, v.2, n.4, p.218–231, 2003.

THAKKAR, J.; Deshmukh, S. G.; Gupta, A. D.; Shankar, R. Development of a balanced scorecard: Na integrated approach of Integrated approach of Interpretive Structural Modeling (ISM) and Analytic Network Process (ANP). **International Journal of Productivity and Performance Management**, v.56, n. 1, p.25-59, 2007.

VAPNIK, V. N. **Statistical Learning Theory**. New York: Wiley Interscience, 2008.

WARREN, T. Multidimensional scaling of similarity. **Psychometrika**. v.30, n.4, p.379-393, 1965.

XAVIER, S. S. A. **Visualizações de informação para suporte a EAD – ferramenta VisED**. Monografia (Projeto Final de Curso) – Instituto de Informática, Universidade Federal de Goiás, Goiás, 2004.

YIP, A. M.; DING, C.; CHAN, T. F. Dynamic cluster formation using level set methods. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.28, n.6, p.877–889. 2006.

ZHU, X. **Semi-Supervised Learning Literature Survey**. Madison, WI: University Of Wisconsin-Madison, 2005. Computer Sciences Technical Report 1530.

ZHU, X.; GHAHRAMANI, Z. **Learning from labeled and unlabeled data with label propagation**. 2002. Disponível em: <<http://lvk.cs.msu.su/~bruzz/articles/classification/zhu02learning.pdf>>. Acesso em: novembro 2013.

APÊNDICE A - Indicadores para os processos dos níveis G e F do MR-MPS-SW

Devido à grande dimensão do conteúdo que compõem o mapa mental deste Apêndice, foi necessário dividi-lo em partes menores para se obter uma visualização melhor do seu conteúdo. Desta forma, a **Figura 38** apresenta a visão completa do mapa mental, que não se encontra legível devido ao tamanho da página, mas apresenta a sua estruturação. Por esse fato, segue, adicionalmente, o detalhamento dos indicadores e processos divididos em partes menores, conforme ilustram as **Figuras 39 a 44**.

Figura 38 - Visão geral do mapa mental completo dos indicadores (níveis G e F).

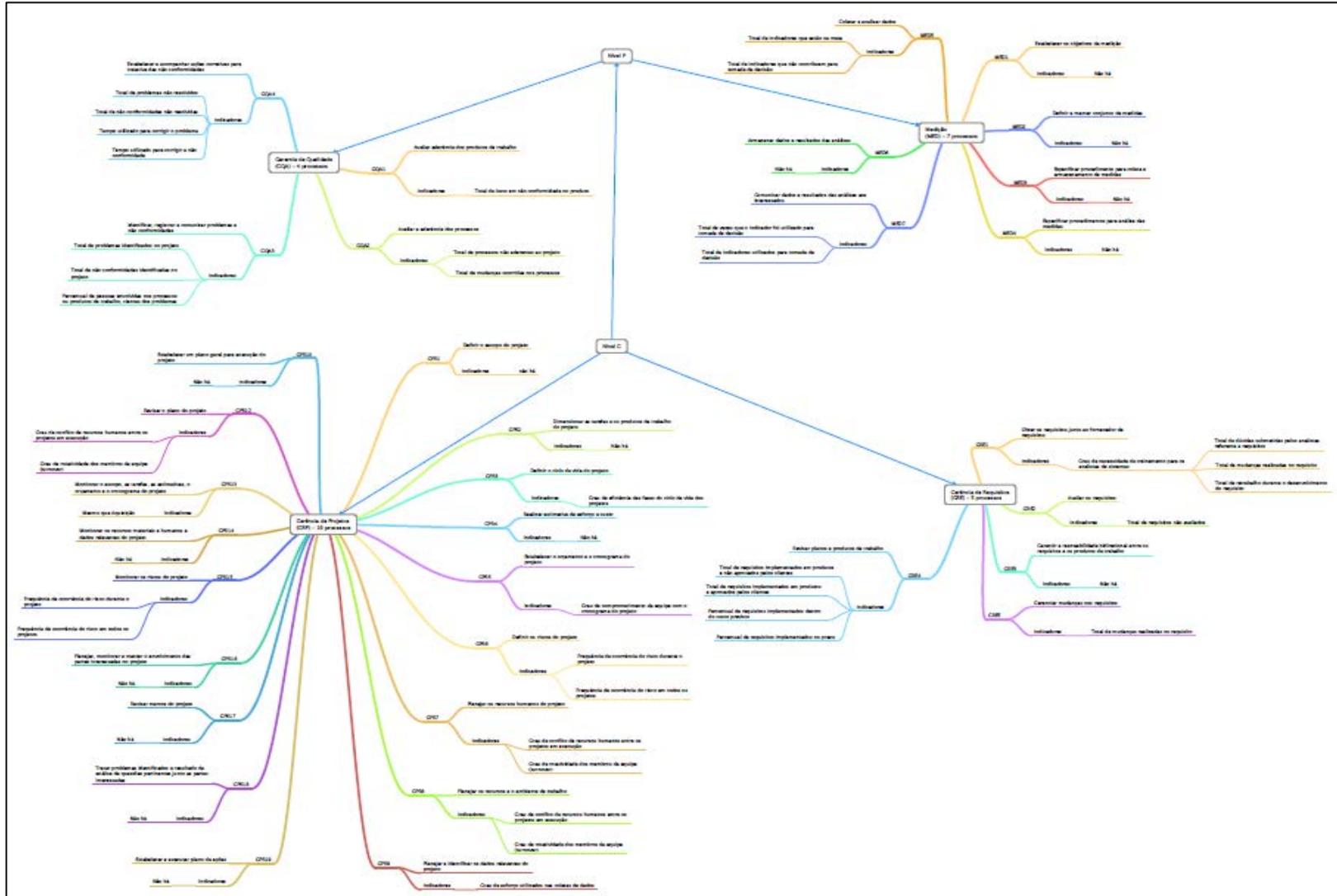


Figura 39 - Mapa mental com os indicadores para o processo de Medição.

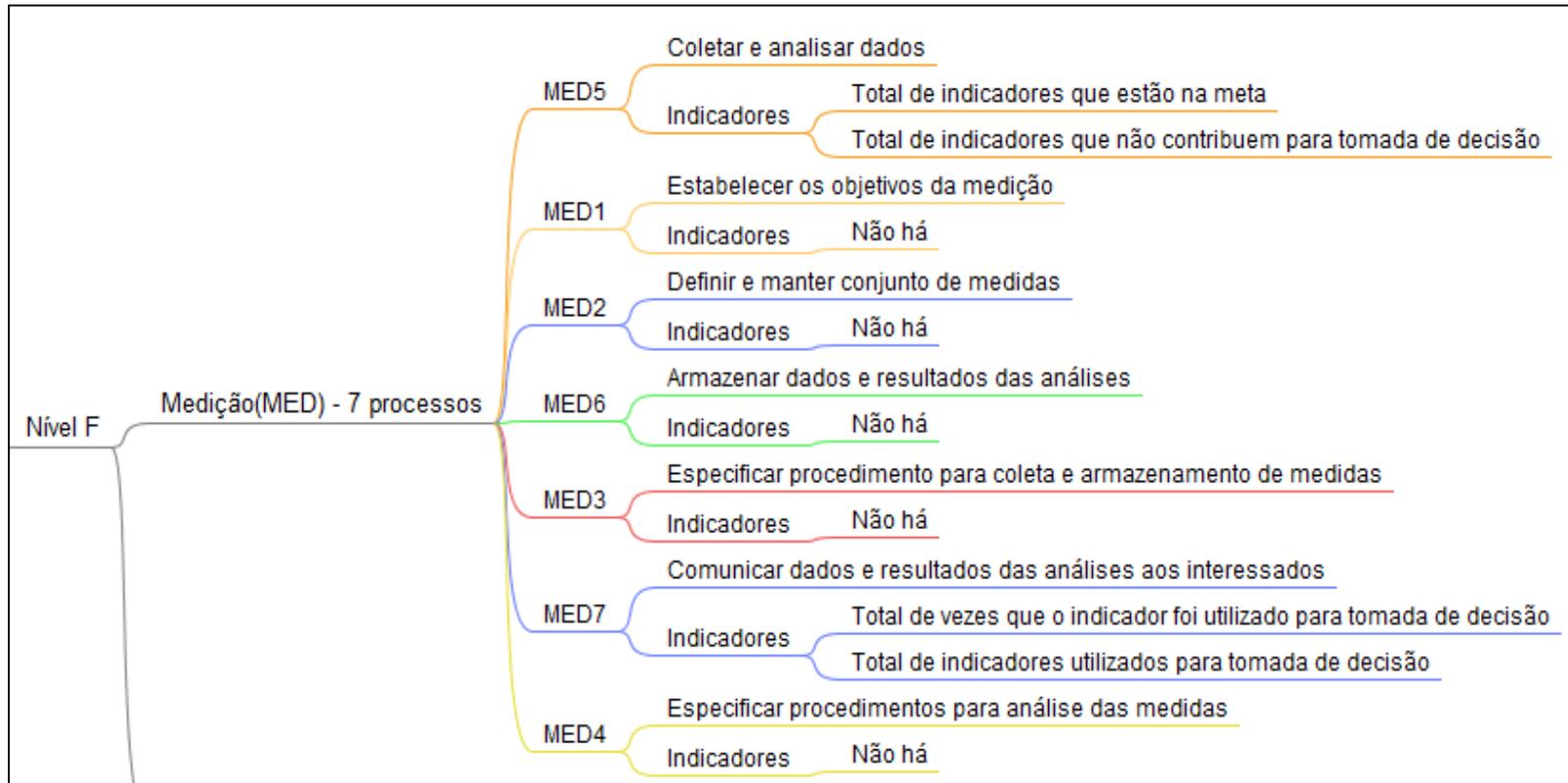


Figura 40 - Mapa mental com os indicadores para o processo de Gerência de Requisitos.



Figura 41 - Mapa Mental com os indicadores para o processo de Gerência de Requisitos, destacando o processo GRE1.

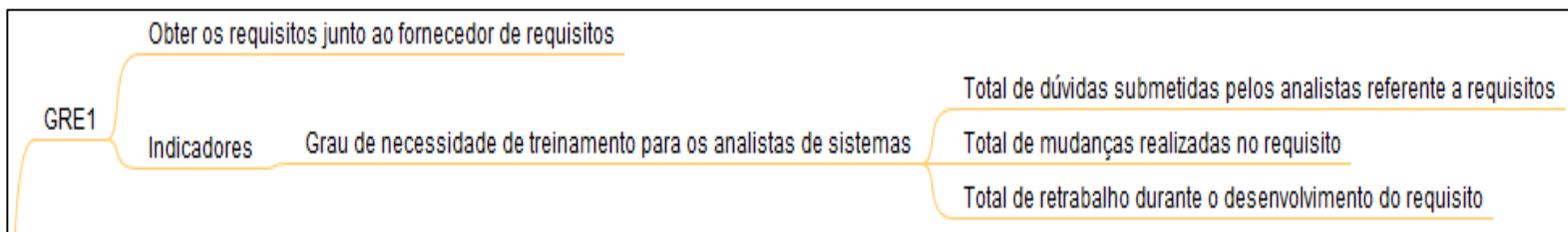


Figura 42 - Mapa mental com os indicadores para o processo de Gerência de Projetos (parcial).

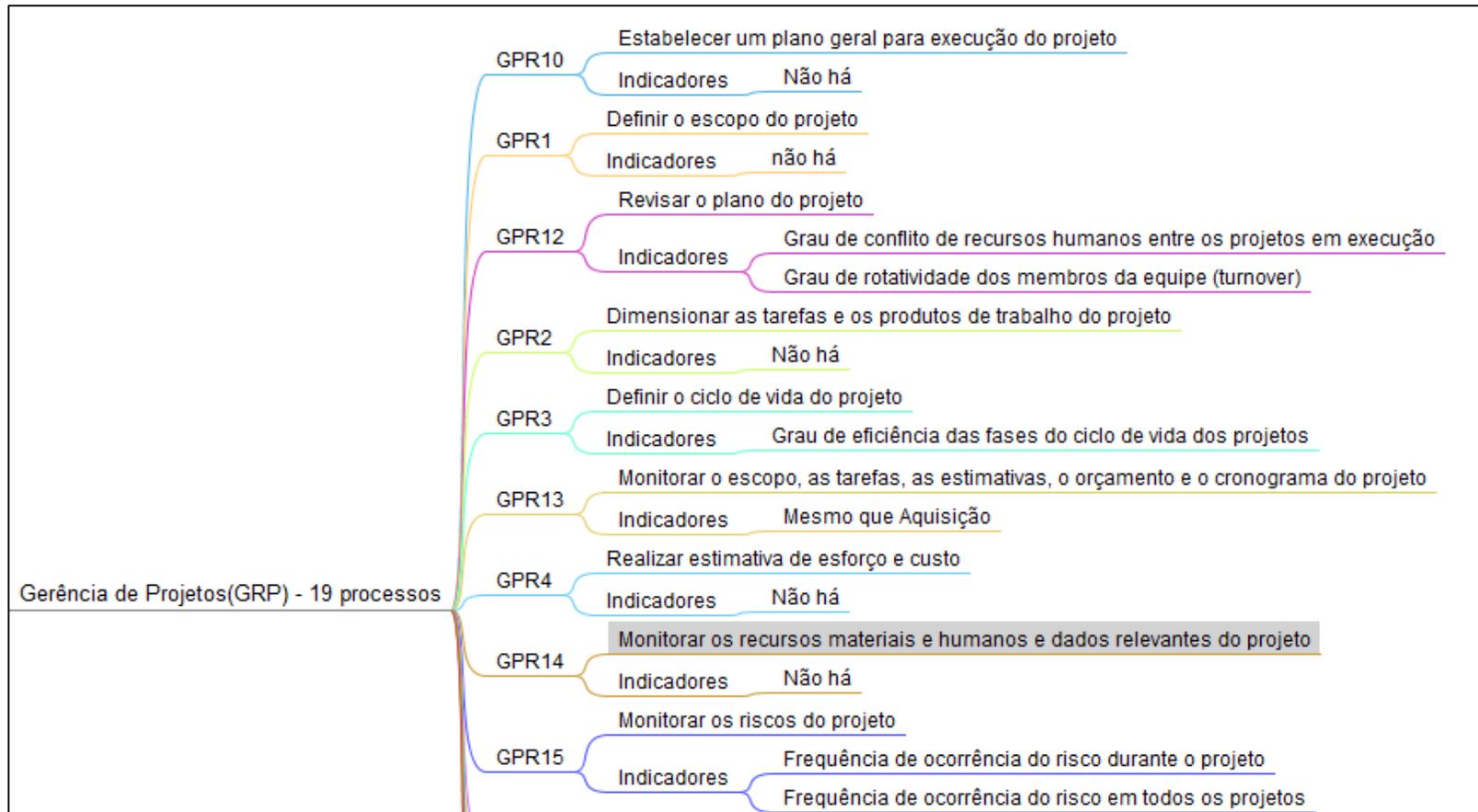


Figura 43 - Mapa mental com os indicadores para o processo de Gerência de Projetos (continuação).

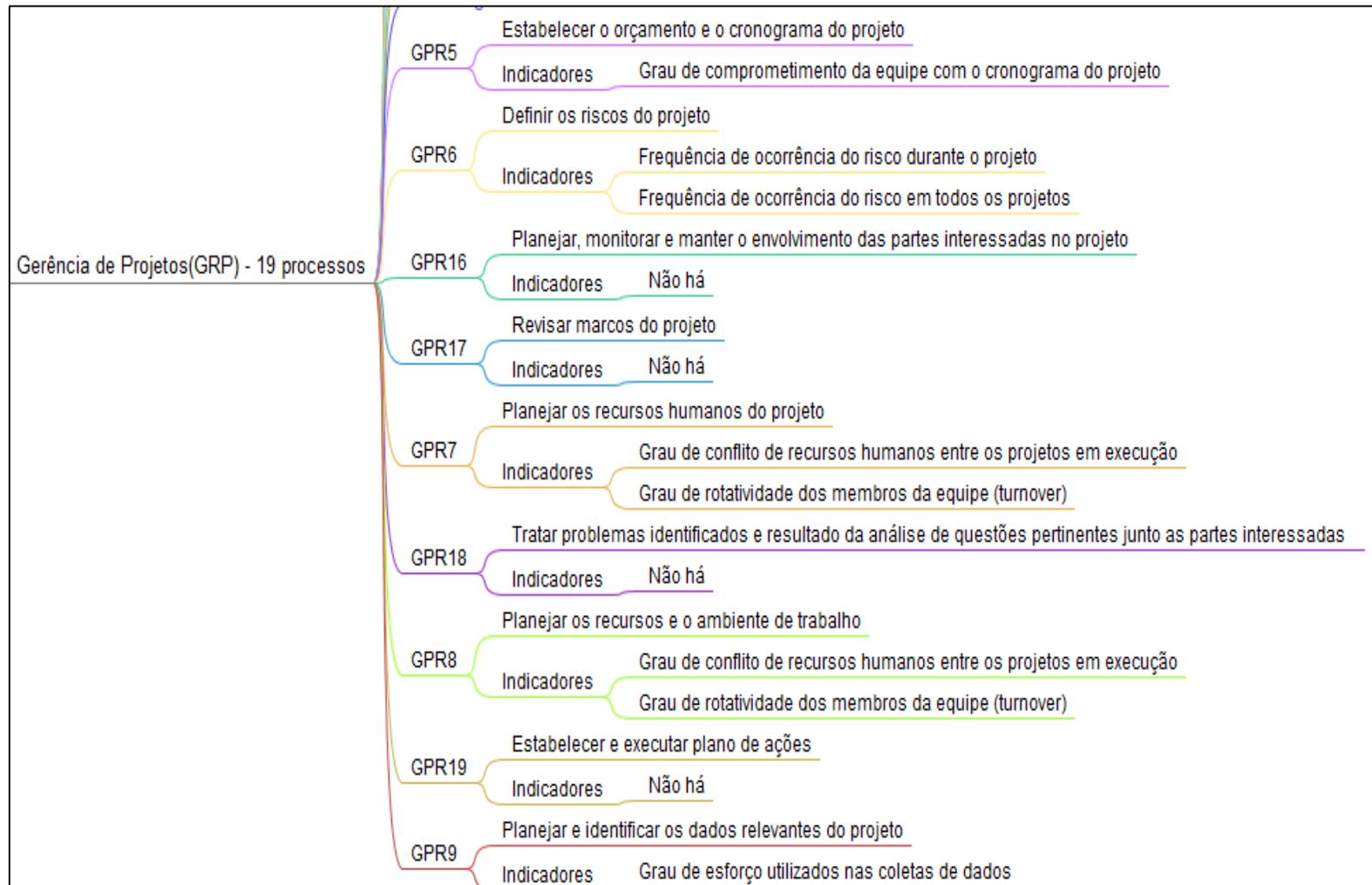
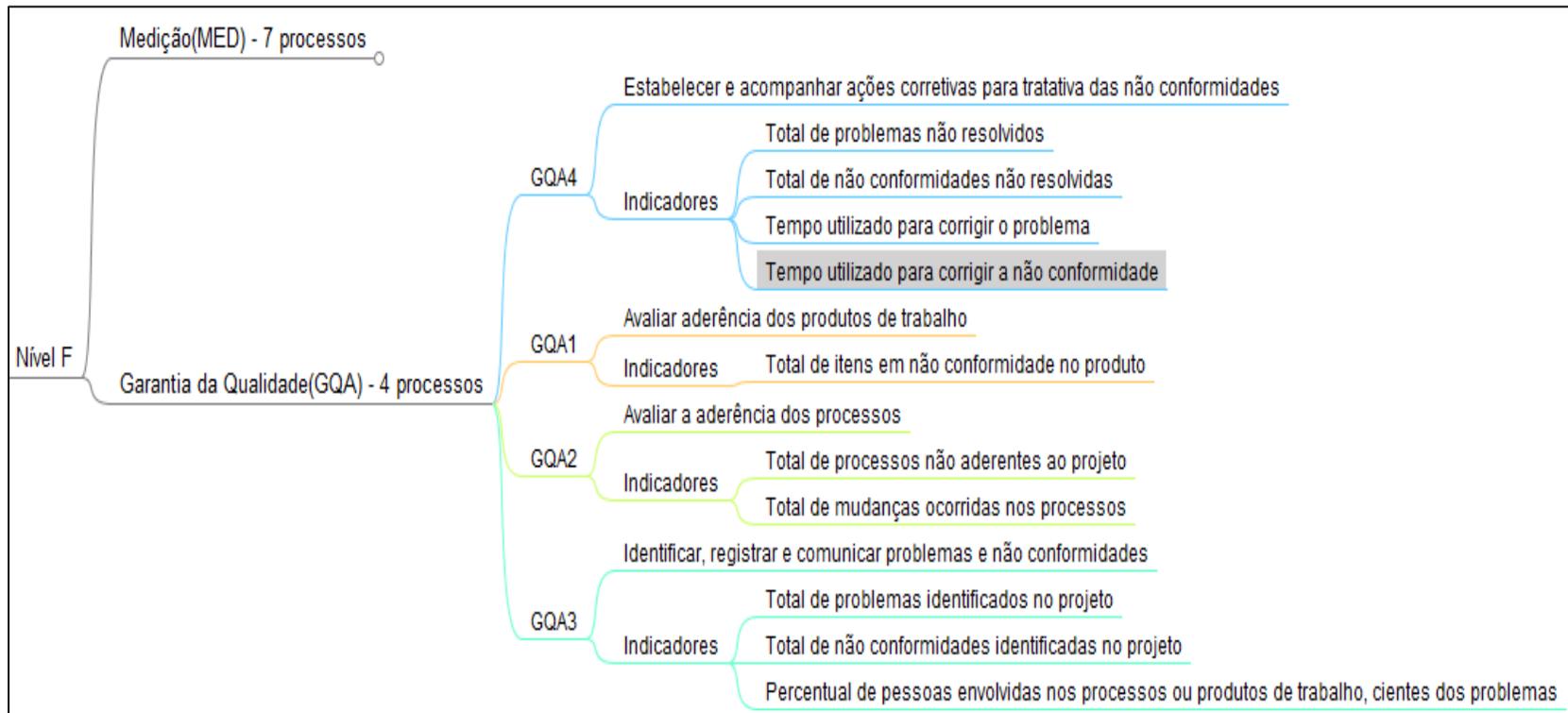


Figura 44 - Mapa mental com os indicadores para o processo de Garantia da Qualidade.



APÊNDICE B – Modelo de referência para seleção de indicadores

Este Apêndice apresenta o modelo de referência de indicadores que pode ser utilizado na etapa 1 da sistemática proposta no *capítulo 4*. Esse modelo foi desenvolvido com base nos processos dos níveis G e F do modelo MPS-SW, apresentados no APÊNDICE A. Conforme resumido na **Tabela 3**, apresentada na *seção 2.3*, são considerados 2 processos no nível G e 5 no nível F.

Para compor esse modelo foram utilizados indicadores propostos na ontologia de Pizzoleto (2013), indicadores resultantes de estudos nas literaturas e levantamentos junto a empresas desenvolvedoras de software.

O modelo de referência de indicadores está apresentado nas **Tabelas 16 a 19**, dividido por nível e processo do MPS-SW.

Nível	Processo MPS-SW	Descrição	Objetivo	Cálculo	Frequência		Aplicação	Unidade Medida
					Coleta	Apresentação		
G	GRE1	Atendimentos tramitados	Indica a quantidade de atendimentos realizados para os clientes que foram tramitados para chamados internos	Somatório do total de atendimentos tramitados	Diária	Mensal	Negócio	Nº
G	GRE1	Requisitos abertos	Identifica a quantidade de requisitos que foram abertos no dia	Somatório dos requisitos abertos no dia	Diária	Mensal	Produto	Nº
G	GRE1	Tempo de treinamento	Indica o percentual do tempo total investido pelo coordenador no treinamento da equipe por projeto	(Total de horas apontadas para treinamento da equipe do projeto / Total de horas apontadas pelo coordenador do projeto)*100	Diária	Por Projeto	Projeto	%
G	GRE1	Top 10 de solicitações abertas por cliente	Apresenta os 10 maiores clientes que fizeram solicitações no período	Somatório do total de solicitações abertas por cliente (max.10 cli)	Diária	Mensal	Negócio	Nº
G	GRE2	Requisitos não avaliados	Indica o total de requisitos não avaliados	Somatório dos requisitos não avaliados	Diária	Mensal	Projeto / Produto	Nº
G	GRE4	Requisitos implementados dentro do custo previsto	Apresenta o percentual de requisitos implementados dentro do custo previsto	(Total de requisitos implementados dentro do custo previsto / Total de requisitos implementados)*100	Diária	Mensal	Projeto / Produto	%
G	GRE4	Requisitos implementados dentro do prazo previsto	Apresenta o percentual de requisitos implementados dentro do prazo previsto	(Total de requisitos implementados dentro do prazo previsto / Total de requisitos implementados)*100	Diária	Mensal	Projeto / Produto	%
G	GRE4	Requisitos implementados no produto	Exibe a quantidade de requisitos que foram implementados no dia	Somatório do total de requisitos implementados no dia	Diária	Por Projeto	Produto	Nº
G	GRE4	Requisitos planejados não	Indica o percentual de requisitos que não estavam previstas e foram adicionados ao projeto (interferência)	(Total de requisitos que foram adicionados ao projeto / Total requisitos implementados no projeto)*100	Diária	Por Projeto	Projeto	%
G	GRE5	Mudanças realizadas no requisito	Indica o total de mudanças realizadas no requisito	Somatório de vezes que um requisito foi alterado durante o desenvolvimento do produto ou projeto	Diária	Mensal	Projeto / Produto	Nº
G	GRE5	Requisito de Impedimento	Exibe o percentual de requisitos responsável pelo impedimento do projeto	(Total requisitos de impedimento / Total requisitos abertos no início do projeto)*100	Diária	Por Projeto	Projeto	%

Tabela 16 - Modelo de referência para seleção dos indicadores - Nível G, processo GRE.

Nível	Processo MPS-SW	Descrição	Objetivo	Cálculo	Frequência		Aplicação	Unidade Medida
					Coleta	Apresentação		
G	GPR2	Eficiência do coordenador de projeto	Apresenta o percentual em que os coordenadores permaneceram em um projeto durante o sprint	$(\text{Total de horas apontadas pelo colaborador no projeto} / \text{Total de horas registradas no espelho de ponto}) * 100$	Diária	Por Projeto	Projeto	%
G	GPR2	Eficiência vs disponibilidade	Apresentar o percentual de horas trabalhadas e apontadas pelos colaboradores com relação ao período em que os mesmos encontram-se na empresa	$(\text{Total de horas apontadas pelo colaborador} / \text{Total de horas registradas no espelho de ponto}) * 100$	Diária	Por Projeto	Negócio	%
G	GPR7	Conflito de recursos humanos	Indica o grau de conflito de recursos humanos entre os projetos em execução	Somatório do total de conflitos ocorridos no projeto	Diária	Mensal	Projeto	Nº
G	GPR7	Turnover equipe	Indica o grau de rotatividade dos membros da equipe	Somatório do total de pessoas que saíram/entraram na equipe	Mensal	Mensal	Projeto	Nº
G	GPR9	Planejamento de projeto	Apresenta o percentual do tempo total em que o coordenador de projetos atuou com planejamento	$(\text{Total de horas apontadas para o planejamento do projeto} / \text{Total de horas apontadas pelo coordenador do projeto}) * 100$	Diária	Por Projeto	Projeto	%
G	GPR13	Requisitos Planejados Finalizados	Apresentar o percentual de eficácia do planejamento do projeto	$(\text{Total OS Encerrada no projeto} / \text{Total OS Aberta no início do projeto}) * 100$	Diária	Por Projeto	Projeto	%
G	GPR15	Risco durante o projeto	Apresenta a frequência de ocorrência do risco durante o projeto	Somatório de ocorrência de um risco no projeto	Diária	Mensal	Projeto	Nº

Tabela 17 - Modelo de referência para seleção dos indicadores - Nível G, processo GPR.

Nível	Processo MPS-SW	Descrição	Objetivo	Cálculo	Frequência		Aplicação	Unidade Medida
					Coleta	Apresentação		
F	GQA1	Itens em não conformidade no produto	Indica o total de itens em não conformidade no produto	Somatório de não conformidades obtidas no produto	Diária	Mensal	Produto	Nº
F	GQA2	Aderência ao processo	Compara a quantidade de itens auditados com a quantidade de itens em conformidade	(Total de itens em conformidade / Total de itens auditados)*100	Mensal	Mensal	Projeto	%
F	GQA2	Mudanças ocorridas nos processos	Apresenta o total de mudanças ocorridas nos processos	Somatório de vezes que ocorreram mudanças nos processos	Diária	Mensal	Produto	Nº
F	GQA3	Cientes que não fazem solicitação há tempos	Identifica o total de clientes que estão com mais de 3 meses sem fazer solicitação	Somatório do total de clientes que não fizeram solicitação a mais de 3 meses	Mensal	Mensal	Produto	Nº
F	GQA3	Instalações realizadas por cliente	Exibe a quantidade de sistemas (softwares) instalados por cliente no período	Somatório do total de sistemas instalados por período	Diária	Mensal	Projeto	Nº
F	GQA3	Não conformidades no projeto	Indica o total de não conformidades constatadas no projeto	Somatório de não conformidades constatadas no projeto	Diária	Mensal	Projeto	Nº
F	GQA3	Problemas no projeto	Apresenta o total de problemas identificados no projeto	Somatório de problemas identificados e registrados durante a execução do projeto	Diária	Mensal	Projeto	Nº
F	GQA4	Não conformidades não resolvidas	Indica o total de não conformidades constatadas e não resolvidas no projeto	Somatório de não conformidades constatadas e não resolvidas no projeto	Diária	Mensal	Projeto	Nº
F	GQA4	Requisitos acumulados	Indica a quantidade de requisitos não implementados e que encontram-se abertos no final do período	(Total Requisitos Abertos - Total Requisitos Implementados por DIA) + Saldo Anterior	Diária	Diária	Produto	Nº
F	GQA4	Tempo para corrigir não conformidade	Apresenta o tempo total utilizado para corrigir a não conformidade	Somatório do tempo total utilizado para corrigir uma não conformidade	Diária	Mensal	Projeto	Nº
F	GQA4	Tempo para corrigir problema	Apresenta o tempo total utilizado para corrigir um problema	Somatório do tempo total utilizado para corrigir um problema	Diária	Mensal	Projeto	Nº
F	GQA4	Variação do planejamento	Indica a percentagem de variação do planejamento da sprint do projeto	(Total requisitos previsto de implementação no projeto - Total requisitos implementados no projeto)*100	Diária	Por Projeto	Projeto	%

Tabela 18 - Modelo de referência para seleção dos indicadores - Nível F, processo GQA.

Nível	Processo MPS-SW	Descrição	Objetivo	Cálculo	Frequência		Aplicação	Unidade Medida
					Coleta	Apresentação		
F	MED5	Indicadores em conformidade	Apresenta o percentual de indicadores que estão dentro da meta	(Total de indicadores que encontram-se com valores dentro da meta / Total de indicadores)*100	Diária	Mensal	Negócio / Projeto / Processo	%
F	MED5	Indicadores não p/ tomada de decisão	Apresenta o percentual de indicadores que não contribuem para tomada de decisão	(Total de indicadores que não contribuem para tomada de decisão / Total de indicadores)*100	Diária	Mensal	Negócio / Projeto / Processo	%
F	MED7	Indicadores p/ tomada de decisão	Apresenta o percentual de indicadores que contribuem para tomada de decisão	(Total de indicadores que contribuem para tomada de decisão / Total de indicadores)*100	Diária	Mensal	Negócio / Projeto / Processo	%
F	MED7	Indicadores utilizados na tomada de decisão	Indica o total de vezes que um indicador foi utilizado para tomada de decisão	Somatório do total de vezes que um indicador foi utilizado para tomada de decisão	Diária	Mensal	Negócio / Projeto / Processo	Nº

Tabela 19 - Modelo de referência para seleção dos indicadores - Nível F, processo MED.