

**Gelson Guedes Rodrigues**

***Minerando Modelos Estocásticos do RCPS com  
Vistas a um Planejamento mais Efetivo de Projetos  
de Tecnologia da Informação***

Rio de Janeiro

2007

**Gelson Guedes Rodrigues**

***Minerando Modelos Estocásticos do RCPS com  
Vistas a um Planejamento mais Efetivo de Projetos  
de Tecnologia da Informação***

Orientador:

Prof. Antonio Juarez Alencar, D.Phil.

Co-orientador:

Prof. Eber Assis Schmitz, Ph.D.

MESTRADO EM INFORMÁTICA  
INSTITUTO DE MATEMÁTICA / NÚCLEO DE COMPUTAÇÃO ELETRÔNICA  
CENTRO DE CIÊNCIAS DA MATEMÁTICA E DA NATUREZA  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Rio de Janeiro

2007

RODRIGUES, G. G.

Minerando Modelos Estocásticos do RCPS com Vistas a um Planejamento mais Efetivo de Projetos de Tecnologia da Informação ,[Rio de Janeiro], 2007. x, 92p., A8p., B11p., 29,7 cm

(IM/NCE/UFRJ, MSc., Informática, 2007)

Dissertação (Mestrado) - Universidade Federal do Rio de Janeiro,  
IM/NCE

1. Árvore de Classificação. 2. Simulação. 3. RCPS. I. Título. II. Série

Dissertação submetida ao corpo docente do Instituto de Matemática (IM) / Núcleo de Computação Eletrônica (NCE) da Universidade Federal do Rio de Janeiro (UFRJ), como parte dos requisitos necessários à obtenção do grau de Mestre em Informática.

Aprovada por:

---

Prof. Antonio Juarez Alencar, D.Phil.  
IM - UFRJ  
Co-orientador

---

Prof. Eber Assis Shmitz, Ph.D.  
IM/NCE - UFRJ  
Co-orientador

---

Prof. Gabriel Pereira da Silva, DSc.  
IM - UFRJ

---

Prof<sup>a</sup>. Renata Mendes de Araujo, DSc.  
UNIRIO

# *Dedicatória*

*Á minha família, minha namorada  
e amigos de fé.*

# *Agradecimentos*

Ao orientador Antonio Juarez Alencar, pela disponibilidade, atenção e paciência dedicadas, sendo uma constante fonte de equilíbrio e um exemplo de inteligência e bom senso. Obrigado por acreditar neste trabalho, às vezes até mais do que eu mesmo, e por saber exigir de mim nos momentos certos.

Ao coorientador Eber Assis Schmitz pelos comentários pertinentes sobre o trabalho visando sempre melhorar o meu entendimento do problema e posterior solução.

À minha família por toda compreensão em relação as minhas constantes ausências e pelo carinho que me dedicaram, me guiando ao longo da vida e construindo a base para esta realização.

Á Ana Paula, minha namorada, pela companhia em todas as vezes que estive longe da minha família, além de sempre se mostrar interessada em discutir e em saber como andava o trabalho.

Aos meus amigos Bernardo e Vinícius por terem me ajudado num momento importante deste trabalho com um suporte que somente os verdadeiros amigos podem disponibilizar. Aos outros amigos, não poderia deixar de agradecer pela simples companhia e apoio sinceros

*”Prometo, que no exercício profissional  
como bacharel em Ciência da Computação,  
serei fiel aos preceitos da honra e da ciência,  
promovendo o uso e o desenvolvimento da Informática  
em benefício do cidadão e da sociedade.”*

**SBC**

# *Resumo*

Nas últimas décadas o mundo onde as organizações fazem negócios mudou consideravelmente com grandes conseqüências para a gerencia de projetos. Mudanças nos paradigmas de negócio, a disponibilização cada vez mais rápida de novas tecnologias, a crescente competição por espaço no mercado de bens e serviços e o aumento da demanda por produtos novos e mais complexos obrigaram as organizações a aumentar a quantidade e a complexidade dos projetos que elas executam, especialmente aqueles ligados a tecnologia da informação. Além disso, estes projetos tendem a ter prazos menores e orçamentos mais apertados. Tudo isso coloca os gerentes de projetos de TI sob uma pressão enorme para entregar os projetos pelos quais são responsáveis de acordo com as especificações e com pouca margem para erros. Técnicas de mineração de dados podem ser combinadas com modelagem estocástica e simulação do problema de escalonamento de projetos com recursos limitados (RCPSP), para prover aos gerentes de projeto de TI uma visão mais abrangente dos projetos que eles executam. Esta visão torna mais fácil a antecipação de mudanças no planejamento que favorecem que projetos terminem no prazo, dentro do orçamento, em conformidade com o fluxo de caixa disponível e de acordo com as exigências que deve satisfazer.

# *Abstract*

Over the last few decades the environment in which organizations do business has changed considerably with enormous consequences for project management. Changes in business paradigms, the faster pace new technologies reach the market, the increasing competition for market space and the growing demand for new and more complex products and services have forced organizations worldwide to augment the number and complexity of the projects they run, specially those related to information technology. Moreover, these projects tend to have narrower time frames and tighter budgets. All of this has placed IT project managers under enormous pressure to deliver the projects they are responsible for according to specification and with little margin for error. Data mining techniques can be combined with RCPSP (Resource Constrained Project Scheduling Problem) stochastic modeling and simulation to provide IT project managers with better insights into the project they run. Such insights make it easier for managers to anticipate changes in planning that favor projects to be delivered on time, within budget, and in compliance with available cash flow and the requirements they were set to satisfy.

# *Lista de Figuras*

|     |   |       |
|-----|---|-------|
| 1.1 | Custo das alterações no projeto ao longo do tempo . . . . .                                     | p. 16 |
| 2.1 | Um simples sistema de gerenciamento de projetos . . . . .                                       | p. 19 |
| 2.2 | Exemplo de um processo de gerência de risco. . . . .  | p. 21 |
| 2.3 | Exemplo de rede CPM . . . . .   | p. 24 |
| 2.4 | Exemplo de rede PERT . . . . .  | p. 25 |
| 2.5 | Exemplo de uma distribuição dos tempos de um projeto. . . . .                                   | p. 30 |
| 2.6 | Exemplo da distribuição cumulativa dos tempos do projeto. . . . .                               | p. 30 |
| 3.1 | Árvore gerada para o caso Royal Real Estate House . . . . .                                     | p. 36 |
| 3.2 | Node 0 da Árvore de Classificação isolado . . . . .   | p. 37 |
| 3.3 | Critério de categorização da árvore em relação a variável Acompanhado . . .                     | p. 37 |
| 3.4 | Perfil do cliente com alta probabilidade de fechar negócio apresentado pela<br>árvore . . . . . | p. 39 |
| 3.5 | <i>Process Simulator</i> - Simulador de cenários de teste de um projeto . . . . .               | p. 41 |
| 4.1 | Rede de atividades do projeto. . . . .  | p. 47 |
| 4.2 | O <i>makespan</i> da Árvore de Classificação. . . . .   | p. 51 |
| 4.3 | Árvore de Classificação Decorada . . . . .  | p. 54 |

## *Lista de Tabelas*

|     |  |       |
|-----|--|-------|
| 3.1 | Perguntas do questionário . . . . .  | p. 35 |
| 3.2 | Respostas do questionário compiladas . . . . .   | p. 35 |
| 3.3 | Cenários de teste gerados pelo <i>Process Simulator</i> . . . . .                                  | p. 44 |
| 3.4 | Dados coletados durante o processo de simulações acrescidos da variável <i>Resultado</i> . . . . . | p. 44 |
| 4.1 | Recursos necessários ao desenvolvimento de cada módulo. . . . .                                    | p. 46 |
| 4.2 | Durações das atividades . . . . .  | p. 47 |
| 4.3 | Resultados da simulação. . . . .   | p. 48 |
| 4.4 | Variáveis de simulação. . . . .  | p. 49 |
| 4.5 | Regras que mostram as chances do projeto terminar no prazo. . . . .                                | p. 52 |
| 4.6 | Custos de Ana e Leo . . . . .  | p. 53 |
| 4.7 | Ações possíveis e seus respectivos custos com 80% de confiança. . . . .                            | p. 53 |

## *Lista de Siglas*

|       |  |
|-------|--|
| AON   | Activity On Node                                 |
| AOA   | Activity On Arc                                  |
| CART  | Classification and Regression Trees              |
| CHAID | Chi-squared Automatic Interaction Detector       |
| CPM   | Critical Path Method                             |
| PERT  | Program Evaluation and Review Technique          |
| PMBOK | Project Management Body of Knowledge             |
| QUEST | Quick Unbiased Efficient Statistical Trees       |
| RCPSp | Resource Constrained Project Scheduling Problems |
| TI    | Tecnologia da Informação                         |
| VPL   | Valor Presente Líquido                           |

# *Sumário*

|          |   |       |
|----------|---|-------|
| <b>1</b> | <b>Introdução</b>   | p. 14 |
| 1.1      | O mundo no qual as organizações fazem negócio mudou . . . . . | p. 14 |
| 1.2      | Conseqüências para as Organizações . . . . .                  | p. 15 |
| 1.3      | Objetivo do Trabalho . . . . .                                | p. 16 |
| 1.4      | Organização da Dissertação . . . . .                          | p. 17 |
| <b>2</b> | <b>Revisão de Literatura</b>                                  | p. 18 |
| 2.1      | Gerência de Projetos . . . . .                                | p. 18 |
| 2.2      | Gerência de Risco . . . . .                                   | p. 20 |
| 2.3      | Técnicas de Escalonamento de Projetos . . . . .               | p. 21 |
| 2.3.1    | Critical Path Method (CPM) . . . . .                          | p. 21 |
| 2.3.2    | Program Evaluation and Review Technique (PERT) . . . . .      | p. 24 |
| 2.3.3    | Resource Constrained Project Scheduling (RCPS) . . . . .      | p. 25 |
| 2.3.4    | RCPS Estocástico . . . . .                                    | p. 29 |
| 2.4      | Valor Presente Líquido - VPL . . . . .                        | p. 31 |
| <b>3</b> | <b>Métodos e Ferramentas</b>                                  | p. 33 |
| 3.1      | Árvores de Classificação . . . . .                            | p. 33 |
| 3.1.1    | Um Exemplo de Árvore de Classificação . . . . .               | p. 34 |
| 3.1.2    | Formalizando . . . . .  | p. 38 |
| 3.2      | O Simulador de cenários . . . . .                             | p. 40 |
| 3.2.1    | Heurísticas de Escalonamento . . . . .                        | p. 42 |

|          |  |              |
|----------|--|--------------|
| 3.2.2    | Os Cenários de Teste . . . . .   | p. 42        |
| 3.2.3    | Coletando Dados da Simulação . . . . .   | p. 43        |
| <b>4</b> | <b>Estudo de Caso - Sistema para elaboração de catálogos</b>                                       | <b>p. 45</b> |
| 4.1      | Modelo Estocástico do RCPS com Árvores Decoradas . . . . .   | p. 45        |
| 4.1.1    | Definindo a Duração das Atividades . . . . .   | p. 46        |
| 4.1.2    | Simulação dos cenários . . . . .   | p. 47        |
| 4.1.3    | Análise das regras de classificação . . . . .  | p. 50        |
| 4.1.4    | Análise financeira dos resultados . . . . .  | p. 53        |
| <b>5</b> | <b>Conclusões</b>  | <b>p. 55</b> |
| 5.1      | Discussão . . . . .  | p. 55        |
| 5.1.1    | Que passos permitem a construção de árvores de classificação decoradas? . . . . .                  | p. 55        |
| 5.1.2    | Como os gerentes de projeto se beneficiam de árvores de classificação decoradas? . . . . .         | p. 56        |
| 5.1.3    | As árvores de classificação decoradas podem ser usada como uma ferramenta de negociação? . . . . . | p. 57        |
| 5.2      | Conclusão . . . . .  | p. 58        |
| 5.3      | Trabalhos Futuros . . . . .  | p. 59        |
|          | <b>Referências Bibliográficas</b>  | <b>p. 60</b> |

# ***1 Introdução***

## **1.1 O mundo no qual as organizações fazem negócio mudou**

Nas últimas décadas o mundo onde as organizações fazem negócios mudou consideravelmente com grandes conseqüências para a gerência de projetos, especialmente os relacionados à Tecnologia da Informação (TI). Os paradigmas de negócio que prevaleceram durante a revolução industrial estão dando lugar aos novos, ditados pela era da informação, pela era do conhecimento e pela revolução tecnológica que nós estamos experimentando atualmente (JONES; KOVACICH; LUZWICK, 2002).

Por exemplo, a existência de linhas de produção rígidas de produtos tangíveis e não personalizados que caracterizaram a revolução industrial estão sendo desafiadas com sucesso pelo aumento da demanda por produtos e serviços altamente personalizados, força de trabalho descentralizada e produtos intangíveis, abrindo novos caminhos para o desenvolvimento industrial e de negócios (HOSNI; KHALIL, 2004). A comunicação via telefone celular, as redes de computadores, a TV sob demanda e a indústria de aprendizagem a distância são apenas alguns exemplos destes novos paradigmas (WAGNER, 2005). Algumas organizações, por exemplo, obtiveram melhorias substanciais em processamento de dados apenas trocando os mainframes, centralizadores de informação, por computadores pessoais conectados por uma rede local (GAO, 2000).

O aumento da competição por espaço em diversos setores da economia em nível mundial tornou-se uma força social importante nos tempos de hoje, colocando as organizações sob pressão para disponibilizar resultados complexos e personalizados o mais rapidamente possível. Hoje em dia, a resposta deve vir mais rápido, as decisões devem ser tomadas mais cedo e os resultados devem ser obtidos mais prontamente. Em decorrência, *time-to-market*<sup>1</sup> tornou-se um fator crítico de sucesso em diversas áreas de negócio (MEREDITH; MANTEL, 2002). Hoje em dia, para permanecerem competitivas, as organizações têm que reduzir custos e focar em satis-

---

<sup>1</sup>Tempo que um produto leva para sair da prancheta de projetos e chegar ao mercado.

fazer os clientes (FRAME, 2002). Se você não consegue entregar os produtos e os serviços que seu cliente valoriza, pelo preço que estão dispostos pagar, alguma outra organização, certamente o fará.

A pressão do mercado e a crescente demanda por produtos e serviços cada vez mais complexos requerem das organizações a prática da “destruição criativa”, destruindo formas antigas de fazer negócio a fim de criar novas (HAMILL, 2000; ARORA; FOSFURI; GAMBARDELLA, 2001). Entretanto, este processo de destruição e construção requer toda uma gama de conhecimento especializado, tais como: marketing, tecnologias emergentes, finanças, logística, estratégia de negócio, etc. Conseqüentemente, embora invisível, o conhecimento emerge como um recurso dos mais estratégicos para as organizações (KAKABADSE; KOUZMIN; KAKABADSE, 2001).

## 1.2 Conseqüências para as Organizações

Tudo isso tem obrigado as organizações em todo o mundo a aumentar a quantidade e a complexidade dos projetos que elas executam. Para permanecerem competitivas as organizações tem que inovar constantemente (GU; GERA, 2004). A introdução de cada inovação nos produtos e serviços requer a execução de um ou, mais freqüentemente, diversos projetos. Enquanto o ritmo da inovação aumenta, aumenta também o número de projetos. Por outro lado, a demanda por produtos novos e mais complexos pede a execução de projetos mais complexos; requerendo a coordenação dos esforços em equipes multidisciplinares com habilidades técnicas e de negócio avançadas, o estabelecimento de alianças estratégicas com parceiros externos, o *outsourcing* de atividades de projeto, e o uso de tecnologia recentemente desenvolvida.

Não obstante, o *time-to-market* reduzido e a competição acirrada por espaço no mercado requerem que os projetos sejam executados mais rapidamente, com orçamentos mais apertados e menos margem para erros. Indubitavelmente tais restrições tendem a pôr gerentes de projetos sob uma enorme pressão para produzir resultados. Quando a pressão pela entrega se torna muito grande, o senso comum freqüentemente vai pela janela, etapas cruciais no desenvolvimento do projeto são ignoradas, o resultado final é de má qualidade e o retrabalho para reparar os danos é muito mais caro do que se detectado em tempo de planejamento (GROUP, 2003).

O erro mais comum, e freqüentemente mais caro, cometido em projetos de software é o “trying to cut corners”<sup>2</sup> durante as fases iniciais do desenvolvimento. Conforme a Figura 1.1 mostra, se a qualidade não for construída desde o início, os custos para produzir um software

---

<sup>2</sup>Pegar atalhos.

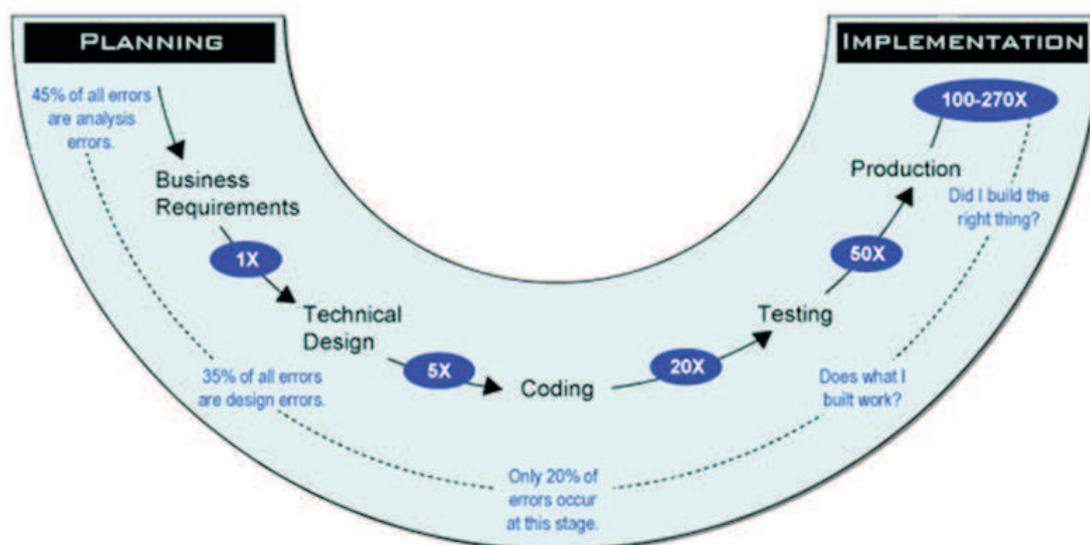


Figura 1.1: Custo das alterações no projeto ao longo do tempo

que atenda as especificações aumenta muito. Isso ocorre porque quanto mais tarde um problema for detectado maior será o custo para repará-lo. Por exemplo, se um problema custa R\$1.000,00 para ser reparado na fase de planejamento, o mesmo custará por volta de vinte vezes mais durante a codificação e até cinquenta vezes mais quando o software estiver em produção (BOEHM, 1981).

A gerência de projetos eficaz em um ambiente com poucos recursos requer bom planejamento e informação oportuna, permitindo que os problemas sejam antecipados e tratados antes que o pior aconteça (WRIGHT et al., 1999). Com a visão na antecipação de problemas que frequentemente impedem que os projetos terminem com sucesso, com o passar dos anos, os gerentes recorreram a diversos métodos de planejamento, tais como CPM (Critical Path Method), PERT (Performance Evaluation and Review Technique) e, mais recente, a modelagem estocástica e simulação do problema RCPS (Resource Constrained Project Scheduling) (LEWIS, 2000; KERZNER, 2003).

### 1.3 Objetivo do Trabalho

Embora a modelagem estocástica e a simulação do problema do RCPS tenham se mostrado ferramentas poderosas para o planejamento de projetos em ambientes restritos, a construção e análise da rede de atividades do RCPS requerem experiência considerável em construção de modelos quantitativos e conhecimentos matemáticos e estatísticos avançados.

Esta dissertação mostra como estabelecer um processo baseado em:

- Técnicas de mineração de dados;
- Modelagem estocástica do RCPS;
- Simulação de Monte Carlo, e
- Análise de risco e avaliação financeira;

Para prover informações valiosas no planejamento de projetos de TI em ambientes com restrições de precedência e recurso, diante de incertezas. Neste sentido tornamos mais fácil para os gerentes vislumbrar mudanças no planejamento que favoreçam que projetos:

- sejam entregues no prazo,
- dentro do orçamento, e
- em conformidade com os requisitos que devem satisfazer.

Estendemos o processo estocástico do RCPS para considerar heurísticas múltiplas e recursos variáveis, e decoramos a árvore com informações financeiras e ações gerenciais corretivas. Por fim discutimos os benefícios do processo, para gerência de projetos.

## 1.4 Organização da Dissertação

O restante desta dissertação esta organizada da seguinte maneira:

- Capítulo 2: Apresenta a revisão bibliográfica dos conceitos de gerência e planejamento de projetos e também das técnicas utilizadas na abordagem da proposta apresentada nesta dissertação.
- Capítulo 3: Apresenta a proposta de solução com utilização de algumas das técnicas descritas no capítulo anterior.
- Capítulo 4: Apresenta um estudo de caso com a aplicação da proposta sugerida.
- Capítulo 5: Apresenta as conclusões obtidas com este trabalho, enumera as suas contribuições e oferece perspectivas de pesquisas futuras.

## 2 *Revisão de Literatura*

### 2.1 Gerência de Projetos

Segundo o (PMBOK, 2004) um projeto é um esforço temporário empreendido para criar um produto, serviço ou resultado exclusivo. Os projetos são realizados em todos os níveis da organização, podem envolver uma única pessoa ou muitos milhares de pessoas. Sua duração varia de poucas semanas a vários anos. Os projetos são um meio de organizar atividades que não podem ser abordadas dentro dos limites operacionais normais da organização, sendo, portanto, freqüentemente utilizados como um meio de implementar o plano estratégico de uma organização, seja a equipe do projeto formada por funcionários da organização ou por prestadores de serviços contratados.

De acordo com (COMPTIA, 2006) a gerência de projetos é o processo de iniciar, planejar, executar, monitorar, controlar e encerrar um projeto aplicando habilidades específicas, ferramentas e técnicas para cumprir com sucesso um determinado conjunto de requisitos. É também satisfazer os desejos e expectativas dos *stakeholders*, isto é, das pessoas ou grupos envolvidos ativamente no projeto e que podem ganhar ou perder com seu resultado, como por exemplo, o gerente do projeto, o cliente, o executor e o patrono.

Já pelo (MIT, 2002) a gerência de projetos é a arte de combinar objetivos de um projeto, atividades e recursos para cumprir estes objetivos conforme o possível, pois estas variáveis são limitadas por tempo, dinheiro e recursos (humanos e/ou maquinários) . A Figura 2.1 mostra, em um diagrama de caixas simples, o gerenciamento de projetos como um processo que transforma insumos (*inputs*) em *deliverables*<sup>1</sup> (*outputs*).

Pos outro lado, (PMBOK, 2004) nos ensina que o gerenciamento de projetos é acompanhado de processos tais como:

- Identificação das necessidades,

---

<sup>1</sup>Entregáveis.

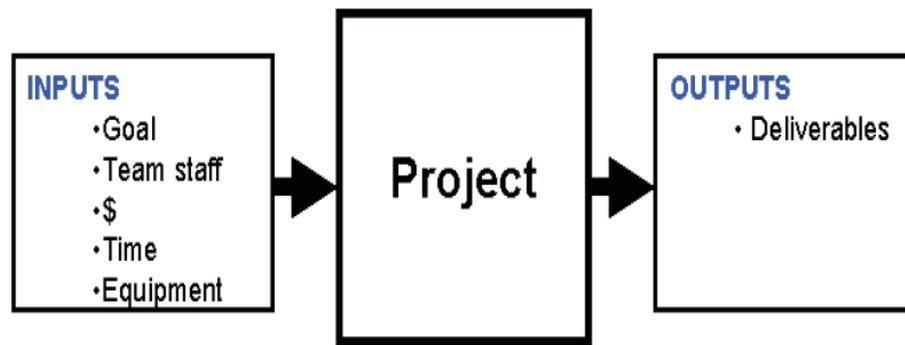


Figura 2.1: Um simples sistema de gerenciamento de projetos

- Estabelecimento de objetivos claros e alcançáveis,
- Balanceamento das demandas conflitantes de qualidade,
- Definição do escopo, e
- Estimativas de tempo e custo,
- Gestão de expectativas (adaptação das especificações, dos planos e da abordagem às diferentes preocupações e desejos das diversas partes interessadas).

São responsabilidades da gerência de projetos o planejamento de execução do projeto, a análise de risco, a manutenção do cronograma, o controle dos recursos disponíveis, a alocação destes recursos no momento apropriado, a cooperação com as pessoas envolvidas no projeto, para que o projeto obtenha sucesso, e a satisfação do cliente.

Na literatura tradicional sobre gerenciamento de projetos, o sucesso de um projeto deve ser estabelecido com a ajuda de métricas específicas, tais como tempo, custo e qualidade. Usando esta definição padrão, o sucesso de um projeto poderia ser visualizado como uma combinação de um *grid* de fatores de sucesso (KERZNER, 2005). Outras características, para um projeto alcançar o sucesso, em geral são: planejamento, uso de métricas, gerência de alterações, metodologia de desenvolvimento, gerentes de projetos experientes, pessoal técnico qualificado, especializado e motivado.

Um projeto falha quando este ultrapassa o prazo ou orçamento, é cancelado ou apresenta um baixo nível de qualidade, trazendo, conseqüentemente, insatisfação ao cliente. O (PMBOK, 2004) cita também alguns exemplos de fatores que podem fazer com que o projeto fracasse: falta de planejamento, plano de projeto que não representa a realidade, inabilidade para acompanhar o progresso, inabilidade para detectar problemas antecipadamente, mudanças de prioridades, falta de comprometimento da equipe, falta de motivação da equipe e pessoal não qualificado.

Entretanto, um dos problemas mais frequentes que a gerência de projetos enfrenta é o de minimizar o makespan<sup>2</sup> (outras possibilidades comuns incluem minimizar o custo, maximizar os resultados financeiros, maximizar as medidas de qualidade, etc.) (ZHU; BARD; YU, 2005). Um bom planejamento aliado à antecipação de problemas pode fazer com que este objetivo seja atingido.

## 2.2 Gerência de Risco

Risco é a incerteza que existe em relação a ocorrência de um evento que terá um efeito adverso em um projeto, em uma organização ou em um sistema de informação. Risco envolve tanto a probabilidade de falha quanto as possíveis conseqüências da falha (MIT, 2002). Projetos sempre possuem riscos associados, na medida em que muitas decisões precisam ser tomadas sem informações completas. Por exemplo, o gerenciamento do projeto depende de uma série de estimativas de prazos e custos, e por definição uma estimativa não é um valor sobre o qual se tenha certeza absoluta. Em qualquer estimativa existe sempre um certo grau de incerteza (GALVÃO, 2005; DEMEULEMEESTER; HERROELEN, 2002).

De acordo com (BISTARELLI; FIORAVANTI; PERETTI, 2006), a fim de focar nos riscos reais e concretos que podem afetar os projetos da organização, um processo de gerência de risco é necessário. Este processo possui atividades relacionadas com:

- Identificação dos riscos,
- Análise da probabilidade de ocorrência destes risco e do impacto sobre o projeto,
- Monitoramento da evolução dos risco no decorrer do tempo, e
- Planejamento de respostas aos riscos.

O objetivo deste processo é deixar os gerentes de projeto cientes dos possíveis riscos associados e guiá-los através de um conjunto de medidas que podem trazer os riscos a patamares aceitáveis.

Os riscos que podem afetar o projeto de forma negativa devem ser então identificados e documentados, e em seguida analisados, de forma qualitativa e/ou quantitativa. A forma qualitativa é baseada numa avaliação relativa do risco, enquanto que a quantitativa tenta gerar medidas precisas e objetivas da probabilidade de ocorrência dos riscos e do impacto esperado sobre o projeto. Note que a quantificação do risco é afetada não somente pelas incertezas do mundo

---

<sup>2</sup>Duração total do projeto.

real mas também pelo modelo da estratégia de negócio adotado e pelo processo de tomada de decisão.

A Gerência de Risco é o processo que permite equilibrar os custos operacionais de medidas preventivas com os ganhos destas medidas ao projeto (STONEBURNER; GOGUEN; FERINGA, 2002). Cada projeto tem um ambiente de tomada de decisões baseado nos objetivos, orçamento, tempo e prioridades, que é único, por isso, cada projeto deve adaptar seu processo de gerência de risco de modo a contemplar as circunstâncias únicas do ambiente em que é executado (FUSSEL; FIELD, 2005).

A Figura 2.2 contém os passos básicos do processo de gerência de risco. A forma como estes passos são executados difere, ligeiramente, de organização para organização. Entretanto todas as organizações comprometidas com o processo de gerência de risco realizam os passos fundamentais: Identificar (*Identify*), Analisar (*Analyze*), Tratar (*Handle*) e Monitorar (*Track*) os riscos a que os projetos estão sujeitos.

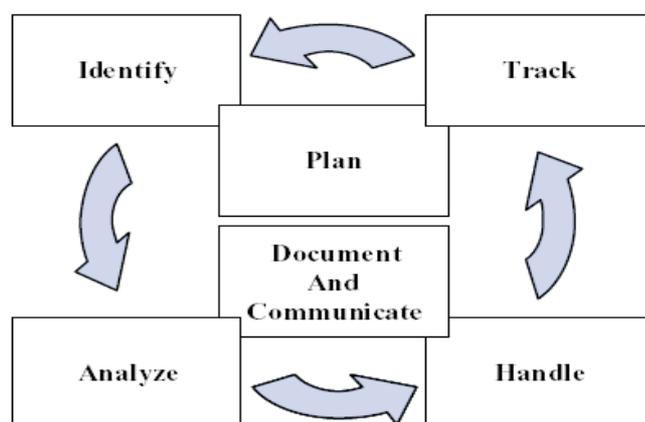


Figura 2.2: Exemplo de um processo de gerência de risco.

## 2.3 Técnicas de Escalonamento de Projetos

### 2.3.1 Critical Path Method (CPM)

Segundo (MOORE; CLAYTON, 1976) o CPM, sigla para “Critical Path Method”(Método do Caminho Crítico), foi desenvolvido no final dos anos 50 pela DuPont Company em conjunto com a divisão Univac da Remington Rand Corporation. O propósito inicial do projeto CPM era fornecer uma técnica para o controle da manutenção das indústrias químicas da DuPont. No entanto, devido à sua generalidade logo ele se tornou um modelo universal aplicado às mais diferentes áreas. Isso ocorreu principalmente porque a técnica do CPM fornecia os seguintes

benefícios:

1. Uma visão gráfica do projeto,
2. A previsão do tempo necessário para o término das atividades, e
3. Uma identificação clara das atividades que são críticas para manter o cronograma.

Conforme o nome do método diz, um dos aspectos mais importantes do CPM é a determinação do caminho crítico. O caminho crítico da rede é a seqüência de atividades que constitui o caminho de maior duração da rede, determinando o menor tempo em que um projeto pode ser realizado. O significado do caminho crítico é que as atividades que estão nele não podem ser atrasadas sem que todo o projeto seja atrasado (ORR, 2007). Por causa do seu impacto no projeto, a análise do caminho crítico é um importante aspecto no planejamento do projeto. As atividades fora do caminho crítico têm alguma folga de tempo, mas se excederem essa folga o projeto também se atrasa.

Em termos formais, dado uma rede de atividades  $A = \{a_1, a_2, \dots, a_n\}$  e a duração  $D(a_i)$  de cada atividade, para calcular o caminho crítico e as folgas, de uma determinada rede de atividades, precisamos:

- Calcular o *Earliest Start Time* (tempo mais cedo de início) e o *Earliest Finish Time* (tempo mais cedo de término) de cada atividade, considerando as relações de precedência entre as atividades:

Para  $a_1$  até  $a_n$

- Se  $a_i$  não tem predecessores, Então  $EST(a_i) = 0$  e  $EFT(a_i) = D(a_i)$
- Senão  $EST(a_i) = \text{máximo EFT dos seus predecessores e}$

$$EFT(a_i) = EST(a_i) + D(a_i)$$

- Calcular o *Latest Start Time* (tempo mais tarde de início) e o *Latest Finish Time* (tempo mais tarde de término) de cada atividade a da rede, seguindo a ordenação inversa das atividades:

Para  $a_n$  até  $a_1$

- Se  $a(i)$  não tem sucessores, Então  $LFT(a_i) = \text{máximo EFT de todas as atividades e}$

$$LST(a_i) = LFT(a_i) - D(a_i)$$

– Senão  $LFT(a_i) = \text{mínimo LST dos seus sucessores e}$

$$LST(a_i) = LFT(a_i) - D(a_i)$$

- Calcular a folga  $F$  de cada atividade da seguinte maneira:

– Para  $a_1$  até  $a_n$ ,  $F(a_i) = LST(a_i) - EST(a_i)$

As tarefas com  $F = 0$ , estão no caminho crítico e o tempo total do projeto é a soma das  $D(a_i)$  das atividades que estão no caminho crítico

Classificada como AON, sigla para “Activity On Node”(Atividade no Nó), uma rede CPM modela as atividades e os eventos de um projeto em uma rede onde as atividades são representadas como nós e os eventos, que representam o início e fim das atividades, são representados como arcos ou linhas (LUCKO, 2005). A Figura 2.3 mostra um exemplo de uma rede CPM simples, onde as caixas representam as atividades a serem executadas e as setas, conectando estas caixas, representam a ordem que as atividades devem ser executadas.

Por exemplo, a atividade B não pode ser executada antes que a atividade A seja completada. Duas setas chegam a atividade D, uma partindo da atividade B e a outra da C. Portanto, a atividade D só pode ser executada após as atividades B e C terminarem.

Os números entre parênteses abaixo de cada caixa representam o número de dias que as atividades precisam para serem completadas. Para esta rede, aplicado o método, o caminho crítico é dado pela seqüência {ABDE} que tem a maior duração, ou seja, se alguma destas atividades atrasar, o projeto inteiro irá atrasar, afetando a duração total do que é de 9 dias. A atividade C, entretanto, não está no caminho crítico.

O CPM foi desenvolvido para projetos rotineiros, com incerteza mínima dos tempos de término dos projetos (WIEST; LEVY, 1969). O método assume que o tempo necessário para executar cada atividade é dado por uma estimativa de um único valor. O método não reconhece, portanto, a incerteza existente na obtenção desta estimativa, afinal, o tempo real para execução da atividade não é conhecido *a priori*. As redes CPM não representam as restrições de recursos necessários para executar cada atividade. Para projetos mais complexos e, portanto, menos rotineiros, existem maiores incertezas quanto aos tempos de execução, e esta incerteza limita a utilidade do modelo determinista do CPM.

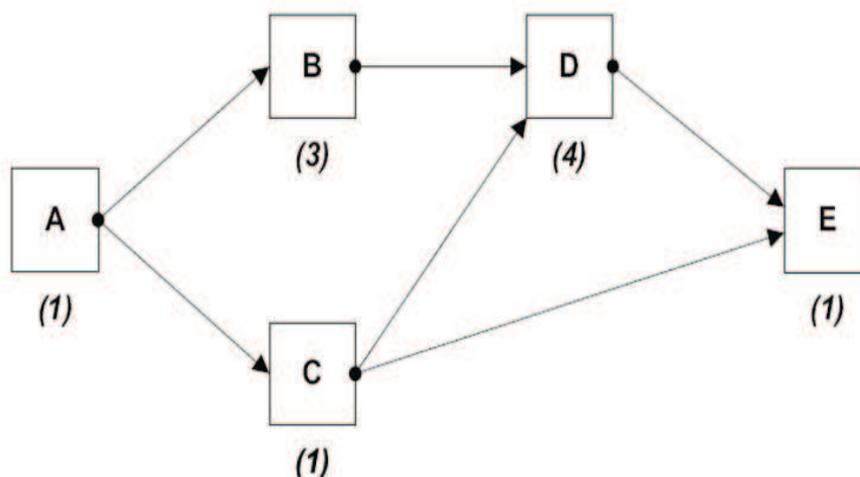


Figura 2.3: Exemplo de rede CPM

### 2.3.2 Program Evaluation and Review Technique (PERT)

O PERT, sigla para “Program Evaluation and Review Technique” (Técnica de Revisão e Avaliação de Programas), foi uma alternativa ao CPM desenvolvida em 1958 (simultaneamente ao desenvolvimento do CPM) para ajudar no planejamento do projeto do míssil Polaris da Marinha dos Estados Unidos. O sucesso magnífico do projeto Polaris é altamente responsável pela ampla aceitação obtida pelo PERT junto ao governo e à área de negócios como uma ferramenta para planejamento e controle de projetos (MOORE; CLAYTON, 1976).

Este método é similar ao CPM pois, as redes de atividades são representadas através de grafos com atividades nos arcos, e os nós são os eventos ou pontos de tempo e possuem as mesmas regras de precedência. Por esta razão uma rede PERT deste tipo é classificada como AOA “Activity On Arc” (Atividade no Arco), mas as redes PERT também podem ser representadas na forma AON (NICHOLAS, 2004). A diferença fundamental entre as duas técnicas é quanto ao modo como a estimativa do tempo para execução da atividade é fornecida. A técnica do PERT inclui um explícito reconhecimento da incerteza das estimativas de tempo permitindo que três estimativas sejam feitas para cada atividade.

As três estimativas de tempo são da seguinte forma: mínimo (menor tempo para realizá-la), mais provável, e a máximo (maior tempo). O PERT assume que o tempo para a realização da atividade toma a forma de uma distribuição Beta Pert, com estes parâmetros (KLING, 2005). Para calcular o tempo total para realização de um projeto, o PERT se baseia no cálculo do seu caminho crítico. O cálculo do caminho crítico é feito da mesma forma que no CPM, usando como tempo de realização das tarefas a média do tempo de realização de cada tarefa, que é dada

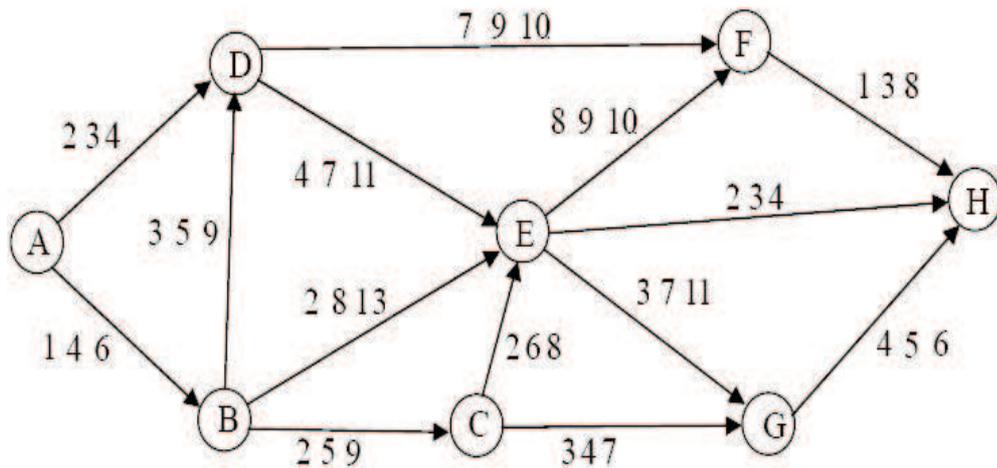


Figura 2.4: Exemplo de rede PERT

pela fórmula:

$$\frac{(\text{mínimo} + 4 \times \text{mais provável} + \text{máximo})}{6}$$

A Figura 2.4 apresenta um exemplo de uma rede PERT do tipo AOA, onde os números próximos aos arcos são as estimativas de duração das atividades (mínimo, mais provável e máximo). Assim como no CPM, o PERT também apresenta o caminho crítico, porém de uma forma um pouco diferente. Todas as atividades, em princípio, fazem parte do caminho crítico mas com diferentes probabilidades. Estabelecendo-se então um grau de confiança, pode-se dizer que um caminho crítico neste exemplo por ser dado pela seqüência  $A \rightarrow B \rightarrow D \rightarrow E \rightarrow F \rightarrow H$ .

As atividades que estão no caminho crítico são identificadas, e é possível dizer que o tempo para execução do projeto é a soma dos tempos destas atividades. Como os tempos das atividades são dados por distribuições estatísticas, não é possível porém identificar um valor exato para o tempo total. O modelo PERT pode fornecer informações sobre o tempo esperado para executar o projeto, a probabilidade de se completar o projeto antes, ou depois de determinada data, o caminho crítico que impacta a execução do projeto, as atividades que possuem folga e não estão no caminho crítico, assim como o tempo de início e fim das atividades.

### 2.3.3 Resource Constrained Project Scheduling (RCPS)

Projetos que estão sujeitos a restrições de recurso e precedência são chamados na literatura de “Resource-Constrained Project Scheduling” (Escalonamento de Projetos com Recursos Limitados), ou na sigla em inglês RCPS (VALLS; BALLEST; QUINTANILLA, 2005). Eles envolvem, portanto, a atribuição de tarefas a um recurso, ou conjunto de recursos, com capacidade

limitada de modo a atingir um determinado objetivo. Não surpreendentemente, a vasta maioria dos projetos no mundo real enfrenta problemas de RCPS (HERROELEN; LEUS, 2005). Em decorrência, tanto o mundo acadêmico como as empresas têm dado grande atenção aos esforços de tratar mais eficientemente os problemas de RCPS.

Ao longo dos anos numerosos métodos foram propostos para resolver problemas de RCPS, tais como: implicit enumeration, branch-and-bound procedure, schedule generation schemes (SGS), X-pass, etc. Uma revisão detalhada destes métodos pode ser encontrada em (DEMEULEMEESTER; HERROELEN, 2002; KOLISCH; HARTMANN, 2006). Apesar das diferenças que estes métodos podem ter, eles podem ser classificados em duas categorias: Métodos Exatos (ZAMANI, 2001) e Métodos Heurísticos (CASTILLO; MUÑOZ, 2004).

Os métodos exatos são usados para encontrar o máximo ou o mínimo exato de uma variável de interesse. Por exemplo, o tempo mínimo de um projeto ou o retorno financeiro máximo. Entretanto, estes métodos têm limitações consideráveis que se tornam evidentes quando os projetos têm um grande número de atividades ou restrições de recursos complexas. Geralmente, nestas circunstâncias, uma solução não pode ser encontrada dentro de um tempo computacional razoável.

Os métodos heurísticos, por outro lado, usam critérios de priorização de escalonamento para encontrar uma solução aproximada para os problemas de RCPS. São particularmente úteis quando o uso de métodos exatos não é viável.

Independente do método escolhido (exato ou heurístico), pode-se formalizar o problema de RCPS da seguinte forma:

- uma classe  $K$  de tipos de recursos;
- um conjunto não vazio de atividades  $A = \{a_1, a_2, \dots, a_n\}$ , onde  $a_1$  e  $a_n$  são atividades *dummy*, que têm duração zero e consomem zero recursos, que sinalizam respectivamente o início e fim do projeto ;
- para cada  $a \in A$ 
  - o tempo de início  $t_a$  da atividade  $a$ ;
  - a duração  $d_a$  da atividade  $a$  e
  - a quantidade de recursos  $k \in K$  requeridos por  $a$ , isto é  $r_{a,k}$ ;
- a quantidade total de recursos  $k \in K$  disponíveis para o projeto, isto é  $b_k$ ;

- o conjunto  $H = \{(a_i, a_j) | a_i, a_j \in A\}$  de restrições de precedência das atividades em  $A$  (observe que se  $(a_i, a_j) \in H$  então existe uma relação de precedência entre  $a_i$  e  $a_j$ );
- o relógio que controla a execução do projeto  $t$ ;

deseja-se minimizar  $t_n$ , o tempo de início da atividade  $a_n$ , tal que:

- $\forall (a_i, a_j) \in H \bullet t_{a_j} - t_{a_i} \geq d_i$  e
- $\sum r_{a,k} \leq b_k$ , onde  $a \in S_t$ , o conjunto de atividades que está sendo executada em um dado instante de tempo  $t$ , e  $k \in K$ .

Nos métodos heurísticos cada atividade é, inicialmente, uma candidata em potencial para ser escalonada. Entretanto, as atividades só podem ser escalonadas se todas as atividades precedentes tiverem terminado e os recursos requeridos para sua execução estiverem disponíveis. Infelizmente, sempre que atividades compartilham recursos, podemos nos deparar com situações na qual existirão atividades que não podem ser executadas simultaneamente.

O uso de uma heurística é então necessário para se decidir quando cada atividade deve receber os recursos necessários para sua execução. Quando nenhuma atividade candidata está disponível, o tempo de execução avança até que uma das atividades em andamento termine. Nesse ponto os recursos são liberados e o processo é repetido, isto é, atividades candidatas são identificadas, verifica-se a disponibilidade de recursos e atividades candidatas são escalonadas. A duração total de um projeto é o tempo requerido para a conclusão de todas as atividades.

De acordo com (GOERLICH; OLAGUIBEL, 1989) o algoritmo a seguir formaliza os passos gerais utilizados pelos métodos heurísticos para resolver os problemas de RCPS. Na notação utilizada para descrever o algoritmo, a relação  $\alpha \leftarrow \beta$  indica que  $\alpha$  recebe o valor de  $\beta$ . Além disso,

- $h : A \rightarrow \mathbb{R}$ , é uma função heurística que, para cada atividade  $a \in A$ , indica a prioridade que deve ser dada para escalonar  $a$  para execução;
- $PE$  é o escalonamento parcial de atividades;
- $ELG$  é o conjunto de atividades elegíveis para escalonamento, isto é, atividades cujo predecessores já foram finalizados;
- $AVL$  é o conjunto de atividades disponíveis para escalonamento;
- $ACT$  é o conjunto de atividades que estão em execução;

- $RU_k$ , para  $k \in K$ , é o conjunto de recursos em uso;

### 1. Inicialização

$$PE \leftarrow \emptyset$$

$$ELG \leftarrow \emptyset$$

$$AVL \leftarrow \emptyset$$

$$ACT \leftarrow \emptyset$$

$$t \leftarrow 0$$

$$\forall k \in K \bullet RU_k \leftarrow 0$$

### 2. Construção da $ELG$

$$ELG \leftarrow \{a_j \mid \forall (a_i, a_j) \in H \bullet t_i + d_i \leq t\}$$

Se  $(ELG = \emptyset)$  então vá para o Passo 6

### 3. Construção da $AVL$

$$AVL \leftarrow \{a_j \in ELG \mid \forall k \in K \bullet r_{a_j,k} \leq R_k - RU_k\}$$

Se  $AVL = \emptyset$  então vá para o Passo 6.

### 4. Construção do escalonamento

4.1 Seja  $a \in AVL$  a atividade que  $\forall a' \in AVL \bullet h(a') \leq h(a)$

Se  $\forall k \in K \bullet r_{a,k} \leq RU_k$  então vá para o Passo 4.2 senão vá para o Passo 5.

4.2  $ACT \leftarrow ACT \cup \{a\}$

$$t_a \leftarrow t$$

$$PS \leftarrow PS \cup \{t_a\}$$

$$\forall k \in K \bullet RU_k \leftarrow RU_k + r_{a,k}$$

vá para o Passo 4.1

5. Determinando um novo tempo  $t$ 

Seja  $a \in A$  a atividade que  $t_a + d_a = \min\{t_{a_i} + d_{a_i} \mid a_i \in ACT\}$

$t \leftarrow t_a + d_a;$

$ACT \leftarrow ACT - \{a_j \in ACT \mid t_{a_j} + d_{a_j} = t\};$

$\forall k \in K \bullet RU_k \leftarrow RU_k - r_{a,k}$

vá para o Passo 2

## 6. Pare

Todas as atividades foram escalonadas

São exemplos de heurísticas comumente utilizadas para lidar com problemas de RCPS: menor tempo de processamento (*shortest processing time*), sucessores imediatos (*most immediate successors*), tempo de execução mais tardio (*late start time*) e o primeiro a solicitar é servido (*first come first served*). Um estudo detalhando destas heurísticas pode ser encontrado em (GOERLICH; OLAGUIBEL, 1989).

Sendo um problema matemático o RCPS pertence a uma categoria onde, em geral, uma solução ótima não pode ser encontrada em um tempo linear ou polinomial, isto é, a categoria NP-difícil, ou *NP-hard* (BLAZEWICZ; LENSTRA; KAN, 1983).

### 2.3.4 RCPS Estocástico

O RCPS Estocástico é uma especialização dos problemas de RCPS criada com o objetivo de adicionar ao problema básico a noção de incerteza quanto à duração das atividades (YANG; GEUNES; O'BRIEN, 2001). No RCPS Estocástico o tempo de processamento de cada atividade é uma variável aleatória (DEMEULEMEESTER; HERROELEN, 2002). O RCPS Estocástico, no entanto, não impõe nenhuma limitação sobre qual a distribuição de probabilidade que a duração de uma atividade deve ter. Esta classe de problemas é muito mais realista, visto que no mundo real os recursos são finitos e o tempo para executar uma atividade é incerto.

A função objetivo neste tipo de problema é diferente das comumente usadas nos problemas básicos RCPS. Não faz sentido falar em menor tempo para realização do projeto, porque apesar deste tempo existir, sua probabilidade de ocorrência é zero (tempo é uma variável contínua). O objetivo então é encontrar a distribuição de probabilidade do tempo total para a realização do projeto. A Figura 2.5 mostra um exemplo de distribuição de tempo total de um projeto.

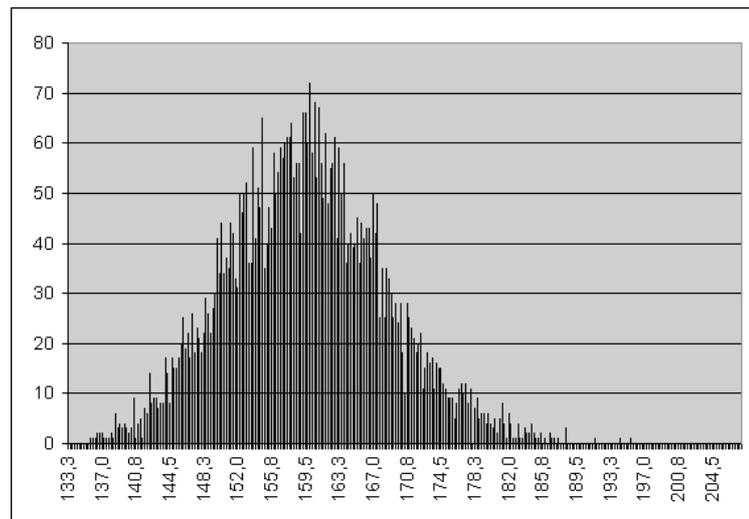


Figura 2.5: Exemplo de uma distribuição dos tempos de um projeto.

A Figura 2.6 mostra a distribuição cumulativa, onde a o projeto tem 80% de chance de terminar em um tempo menor do que 166,8 unidades de tempo.

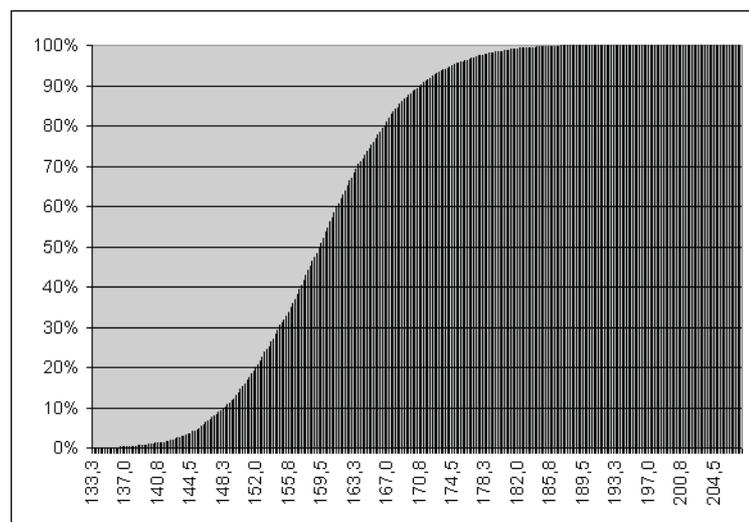


Figura 2.6: Exemplo da distribuição cumulativa dos tempos do projeto.

O método de Monte Carlo (METROPOLIS, 1987) tem sido aplicado com grande sucesso como ferramenta de auxílio a resolução de problemas do tipo RCPS Estocástico (RUBINSTEIN, 1981). Este tipo de problema possui as características ideais para ser resolvido através da simulação, já que se trata de um problema intratável de forma analítica (ARAÚJO; COUTO; SCHMITZ, 1999). Nesta dissertação utilizaremos o modelo estocástico do RCPS, através de um simulador apropriado, para simular os cenários que serão utilizados como entrada para os experimentos com a árvore de classificação.

## 2.4 Valor Presente Líquido - VPL

O valor presente líquido (VPL) é um conceito matemático-financeiro que procura mostrar a contribuição de um projeto para o aumento do valor das receitas de uma empresa, considerando o valor das receitas advindas da implementação do projeto, das despesas decorrentes dos desembolsos para o desenvolvimento do projeto e de uma taxa de juros básica (ALENCAR; SCHMITZ, 2006).

Já que R\$ 1 milhão hoje, não vale R\$ 1 milhão daqui a uma ano, devido ao custo de oportunidade de se colocar tal quantia na poupança ou outra aplicação que possa render juros, o VPL é o resultado da soma de todas as entradas e saídas de capital descontadas (trazidas) ao valor presente (ao dia de hoje). Em termos formais, para um projeto  $P$ :

$$VPL(P) = -I + \frac{FC_1}{(1+k)^1} + \frac{FC_2}{(1+k)^2} + \frac{FC_3}{(1+k)^3} + \dots + \frac{FC_n}{(1+k)^n} = \sum_{t=1}^n \frac{FC_t}{(1+k)^t} - I$$

Onde:

- $VPL(P)$  é o Valor Presente Líquido do projeto  $P$ ;
- $I$  é o investimento de capital no instante zero;
- $t \in [1..n]$  é o período de tempo;
- $FC_1, FC_2, \dots, FC_n$  são os fluxos de caixa nos períodos  $1, 2, \dots, n$ ; e
- $k$  é a taxa de juros (custo do capital).

Por exemplo, suponha a existência de um projeto  $P$  que requer R\$ 120 mil para ser desenvolvido e que depois de sua implementação propicie um fluxo de R\$ 50 mil nos dois primeiros períodos de tempo e R\$ 150 mil no 8 períodos seguintes. Suponha também que a taxa de juros básica seja de 2,5% nos dois primeiros períodos e de 1,5% nos períodos seguintes. Neste caso o  $VPL(P)$  é dado por:

$$\begin{aligned} VPL(P) &= -120 + \frac{50}{(1+2,5\%)^1} + \frac{50}{(1+2,5\%)^2} + \frac{150}{(1+1,5\%)^3} + \dots + \frac{150}{(1+1,5\%)^{10}} \\ &= R\$ 1,06 \text{ milhão} \end{aligned}$$

O VPL é um conceito padrão para o planejamento de investimentos a longo prazo. Usando o VPL um projeto de investimento potencial deve ser empreendido se o valor presente de todas as entradas de caixa menos o valor presente de todas as saídas de caixa for maior que zero. Se o VPL for igual a zero, o investimento é indiferente, pois o valor presente das entradas é igual ao valor presente das saídas de caixa. Se o VPL for menor do que zero, significa que o investimento não é economicamente atrativo, já que o valor presente das entradas de caixa é menor do que o valor presente das saídas de caixa.

Com o VPL é possível se responder tais como:

- É possível obter ganhos ao investir recursos em determinado projeto ?
- Os ganhos de um projeto *A* são maiores que os de um projeto *B*?

A primeira pergunta pode ser respondida se aplicando a fórmula básica o  $VPL > 0$ , o que significa que a expectativa de retorno é positiva. Já a segunda pergunta pode ser respondida comparando-se o VPL do projeto *A* como do projeto *B*. Uma análise abrangente do conceito de VPL pode ser encontrada em (FERREIRA; TEIXEIRA, 2005).

## 3 *Métodos e Ferramentas*

Neste capítulo examinamos os principais métodos e ferramentas utilizados na confecção desta dissertação.

### 3.1 *Árvores de Classificação*

Como um método de inferência, árvores de classificação buscam explicar o comportamento de uma variável *target* combinando os valores de um determinado conjunto de variáveis preditivas. Para atingir este objetivo, os valores das variáveis preditivas são combinados sucessivamente de tal maneira que um espaço n-dimensional é particionado em conjuntos cada vez mais homogêneos de valores em relação à variável *target*.

A maneira como o particionamento é feito permite que o resultado final seja apresentado como um fluxograma, cujo o formato se assemelhe a uma árvore – o que deu origem ao nome da família dos métodos. Além disso, a informação apresentada no fluxograma pode facilmente ser traduzida em um conjunto de regras, que indicam a probabilidade da ocorrência de valores no domínio da variável *target* em diferentes circunstâncias. (BRUN et al., 2001).

As árvores de classificação são métodos não-paramétricos, isto é, nenhuma limitação é imposta sobre a distribuição dos valores das variáveis *target* e preditivas. Em adição, a família de métodos permite que diferentes tipos de relacionamentos possam existir entre as variáveis preditivas. Tudo isso facilita o uso de árvores de classificação para resolver problemas do mundo real, onde a distribuição dos valores das variáveis e os relacionamentos entre elas são, freqüentemente, desconhecidos.

Além disso, árvores de classificação tendem a apresentar resultados satisfatórios mesmo na presença de ruídos e quando poucos dados estão disponíveis; tornando-se uma classe de métodos muito atrativa para ser utilizada em uma grande variedade de situações no mundo real (WITTEN; FRANK, 2005).

### 3.1.1 Um Exemplo de Árvore de Classificação

Um exemplo do uso de árvores de classificação, para análise e predição de informação, pode ser apresentado pelo caso da imobiliária Royal Real State House, em Oxford, Inglaterra. A Royal Real State House aluga e negocia imóveis, muitos deles dos séculos XI, XII e XIII, que se espalham pela cidade e que são extremamente bem preservados. Além disso estes imóveis são próximos a construções com grande importância histórica tais como a Boldean Library e o John Radcliff Hospital, respectivamente, referências europeias em literatura inglesa e no tratamento de uma série de patologias. Além de outras localidades como o centro histórico e o Jardim Botânico.

O impacto destas características é bastante considerável, favorecendo milhares de imóveis alugados e vendidos todos os anos, gerando milhões de libras de taxas de corretagem. Apesar deste grande potencial, a imobiliária, vinha perdendo muitas oportunidades de negócio. O time que prestava o primeiro atendimento ao cliente era incapaz de distinguir clientes com alta probabilidade de realizar algum negócio, dos que provavelmente iriam acabar desistindo da ideia de alugar ou comprar algum imóvel.

Em consequência a imobiliária acabava investindo tempo e dinheiro com clientes que dificilmente iriam trazer algum retorno de investimento. Por outro lado, vários dos clientes com alto potencial de retorno não recebiam o tratamento que mereciam, sendo que muitos negócios acabavam sendo perdidos por falta de esforço adequado por parte dos corretores. Por conta disso, a imobiliária precisava aumentar o número de negócios que eram fechados diariamente, evitando, se possível, que estes negócios fossem transferidos para os concorrentes.

Para atingir este objetivo a imobiliária precisava, primeiro, conhecer o perfil dos clientes com alta probabilidade de fechar um negócio. Por isso a equipe formada para descobrir estes dados, elaborou um questionário com quatro perguntas de modo a entender o perfil dos clientes que, em princípio, se interessam por algum imóvel da imobiliária. A Tabela 3.1 mostra as perguntas do questionário.

As respostas dos clientes nos questionários foram compiladas, junto com a informação de que fecharam ou não negócio, e salvas em um banco de dados criado especialmente para este trabalho. A Tabela 3.2 mostra as respostas das perguntas compiladas, onde as letras "D" e "A", na coluna **Acompanhado**, representam "desacompanhado" e "acompanhado" respectivamente e as letras "S" e "N", na coluna **FechouNeg**, representam os clientes que fecharam e que não fecharam negócio respectivamente. A coluna **ClientID** identifica unicamente cada cliente.

Com estes dados compilados foi escolhido um software de mineração de dados que aplicaria

| AVALIAÇÃO DE POTENCIAL PARA REALIZAÇÃO DE NEGÓCIOS  |
|---|
| <p>A principal razão fornecida pelo cliente para comprar ou alugar um imóvel na cidade é:</p> <p>[1] ascensão profissional dentro da empresa<br/>           [2] obtenção de um novo contrato de trabalho na cidade ou condado<br/>           [3] aumento do núcleo familiar<br/>           [4] casamento ou dissolução do casamento<br/>           [5] outros</p> <p>O cliente pretende utilizar como parte do pagamento na compra de um imóvel ou como garantia de aluguel</p> <p>[1] imóvel(eis) de sua propriedade<br/>           [2] garantia bancária (empréstimo, carta de crédito, etc.)<br/>           [3] garantia pessoal de terceiros (fiador)<br/>           [4] outros</p> <p>O cliente veio acompanhado do cônjuge ou fiadôro?<br/> <input type="checkbox"/> Sim <input type="checkbox"/> Não</p> <p>A faixa etária do cliente é:<br/>           [1] 20-29 anos [2] 30-39 anos [3] 40-49 anos [4] outra</p> |

Tabela 3.1: Perguntas do questionário

| OBS  | ClientID   | Razão | Garantia | Acompanhado | FaixaEtaria | FechouNeg |
|------|------------|-------|----------|-------------|-------------|-----------|
| 1    | RSH-344773 | 1     | 2        | D           | 2           | N         |
| 2    | RSH-202199 | 3     | 3        | D           | 1           | N         |
| 3    | RSH-006279 | 5     | 2        | A           | 3           | S         |
| 4    | RSH-873806 | 3     | 1        | A           | 2           | N         |
| 5    | RSH-860075 | 4     | 3        | A           | 1           | S         |
| 6    | RSH-003090 | 1     | 3        | D           | 1           | N         |
| 7    | RSH-555677 | 1     | 3        | A           | 1           | N         |
| 8    | RSH-517780 | 4     | 2        | A           | 3           | S         |
| 9    | RSH-326331 | 3     | 1        | D           | 2           | N         |
| 10   | RSH-110186 | 1     | 3        | A           | 1           | N         |
| ⋮    | ⋮          | ⋮     | ⋮        | ⋮           | ⋮           | ⋮         |
| 2143 | RSH-346084 | 2     | 3        | A           | 2           | N         |

Tabela 3.2: Respostas do questionário compiladas

a técnica de árvore de classificação de onde a equipe esperava tirar algumas informações que lhes permitisse identificar os clientes com alto potencial para fechar negócio. A Figura 3.1 mostra a árvore de classificação gerada com base nos dados compilados.

Para entendermos as informações apresentadas na árvore, vamos isolar o primeiro nó. A Figura 3.2 apresenta o *Node* 0, que mostra o total de entrevistados que fecharam (S) e não fecharam (N) negócio, com destaque para dois conjuntos de informações. O item 1 mostra a

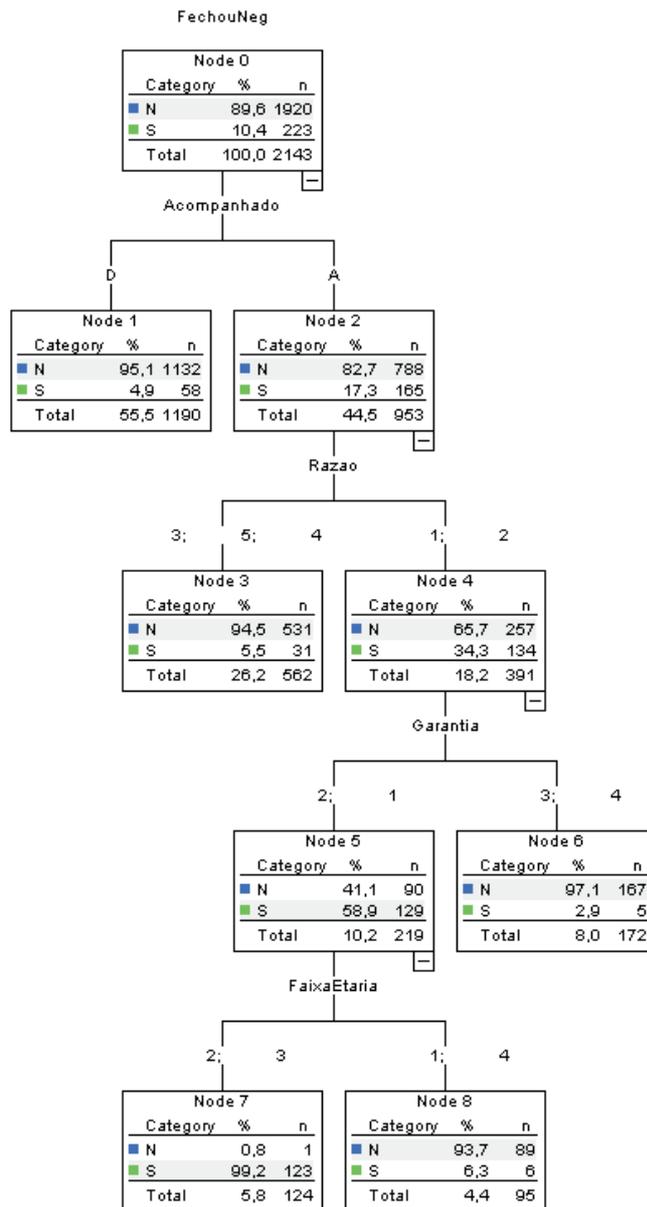


Figura 3.1: Árvore gerada para o caso Royal Real Estate House

porcentagem (%) e o número (n) de casos que foram classificados nas categorias S e N e o item 2 mostra a porcentagem e o número totais destes casos.

FechouNeg

| Node 0   |       |      |
|----------|-------|------|
| Category | %     | n    |
| N        | 89,6  | 1920 |
| S        | 10,4  | 223  |
| Total    | 100,0 | 2143 |

Figura 3.2: Node 0 da Árvore de Classificação isolado

Neste nó temos a primeira informação importante apresentada pela árvore. Apenas 10,4% dos entrevistados fecharam negócio, o que reforça a idéia do número de oportunidades perdidas.

A Figura 3.3 apresenta a primeira divisão realizada pela árvore. Observe que a palavra **Acompanhado** indica que a separação dos casos foi feita em relação à variável que registra se o entrevistado estava acompanhado (A) ou desacompanhado (D) quando respondeu o questionário. De acordo com a árvore, 17,3% dos interessados que vão acompanhados à imobiliária fecharam negócio, ao passo que dos que foram desacompanhados apenas 4,9% fecharam negócio.

Acompanhado

| Node 1   |      |      | Node 2   |      |     |
|----------|------|------|----------|------|-----|
| Category | %    | n    | Category | %    | n   |
| N        | 95,1 | 1132 | N        | 82,7 | 788 |
| S        | 4,9  | 58   | S        | 17,3 | 165 |
| Total    | 55,5 | 1190 | Total    | 44,5 | 953 |

Figura 3.3: Critério de categorização da árvore em relação a variável Acompanhado

Outra informação interessante apresentada na Figura 3.3 diz respeito à razão fornecida pelo interessado para comprar ou alugar um imóvel. Dos que tiveram algum tipo de melhora profissional, ou seja, informaram como razão as opções 1 e 2 do questionário, 34,3% fecharam negócio, ao passo que dos que informaram as razões 3, 4 e 5, apenas 5,5% fecharam negócio.

Seguindo essa mesma linha de raciocínio, dos entrevistados que deram garantias próprias para compra ou locação do imóvel, ou seja marcaram as opções 1 e 2 do questionário, 58,9%

fecharam negócio, ao passo que dos que deram garantias de terceiros ou outros, opções 3 e 4, apenas 2,9% fecharam negócio. Isto nos permite inferir que, em linhas gerais, as pessoas se sentem mais seguras para alugar ou comprar um imóvel, quando têm suas próprias garantias para honrar seus compromissos.

Finalmente, dos entrevistados com idades entre as faixas 30–39 anos e 40–49 anos, ou seja, as opções 2 e 3 do questionário, 99,2% fecharam negócio, ao passo que os entrevistados na faixa etária de 20–29 anos ou outra, ou seja, as opções 1 e 4, apenas 6,3% fecharam negócio.

Combinando todas essas informações pode-se concluir que o perfil dos clientes com alto potencial para fechar negócio é dado pela seguinte regra de negócio: os interessados que foram acompanhados à imobiliária, que tiveram mudanças profissionais positivas, refletidas na escolha das opções 1 e 2 para a razão da compra ou aluguel, que deram garantias próprias para o pagamento, opções 1 e 2 para a garantia, e que são da faixa etária que vai de 30 a 49 anos. Este perfil tem 99,2% de probabilidade de fechar negócio. A Figura 3.4 mostra esta informação graficamente.

As técnicas descritas neste trabalho para analisar modelos estocásticos do RCPS, faz uso extenso de Árvores de Classificação que foi concebida por Morgan and Sonquist na década de 60 (MORGAN; SONQUIST, 1963) e posteriormente aperfeiçoada por outros, tais como: (KASS, 1980), (BREIMAN et al., 1984), (QUINLAN, 1992), (LOH; VANICHESTAKUL, 1988) and (LOH; SHIH, 1997).

### 3.1.2 Formalizando

Em termos formais, dado um conjunto de observações  $\mathcal{O} = \{o_1, o_2, \dots, o_m\}$ , onde cada  $o_j \in [1..m] \in \mathcal{O}$  é uma tupla  $(x_1, x_2, \dots, x_n)$  e cada componente  $x_i \in [1..n]$  assume valores em um domínio diferente, seja  $v_{i \in [1..n], j \in [1..m]}$  o valor do componente  $x_i$  of  $o_j$ . Nestas circunstâncias, uma variável  $w_{i \in [1..n]} \in \mathcal{O}$  é um valor não previamente definido do conjunto  $\{v_{i,1}, v_{i,2}, \dots, v_{i,m}\}$ . Em adição, seja  $w_n$  a variável *target*, cujo comportamento se deseja explicar combinando os valores de  $w_1, w_2, \dots, w_{n-1}$ . Observe que  $w_n$  assume valor no conjunto finito  $J$ .

No paradigma das árvores de classificação, o conjunto de variáveis em  $\mathcal{O}$  é examinado exaustivamente em busca de relações da forma:

$$\begin{aligned}
 & x_i \leq r, && \text{se } x_i \text{ é uma variável contínua} \\
 & \text{e} \\
 & x_i \in \{k_1, k_2, k_3, \dots, k_l\}, && \text{se } x_i \text{ é uma variável discreta.}
 \end{aligned}$$

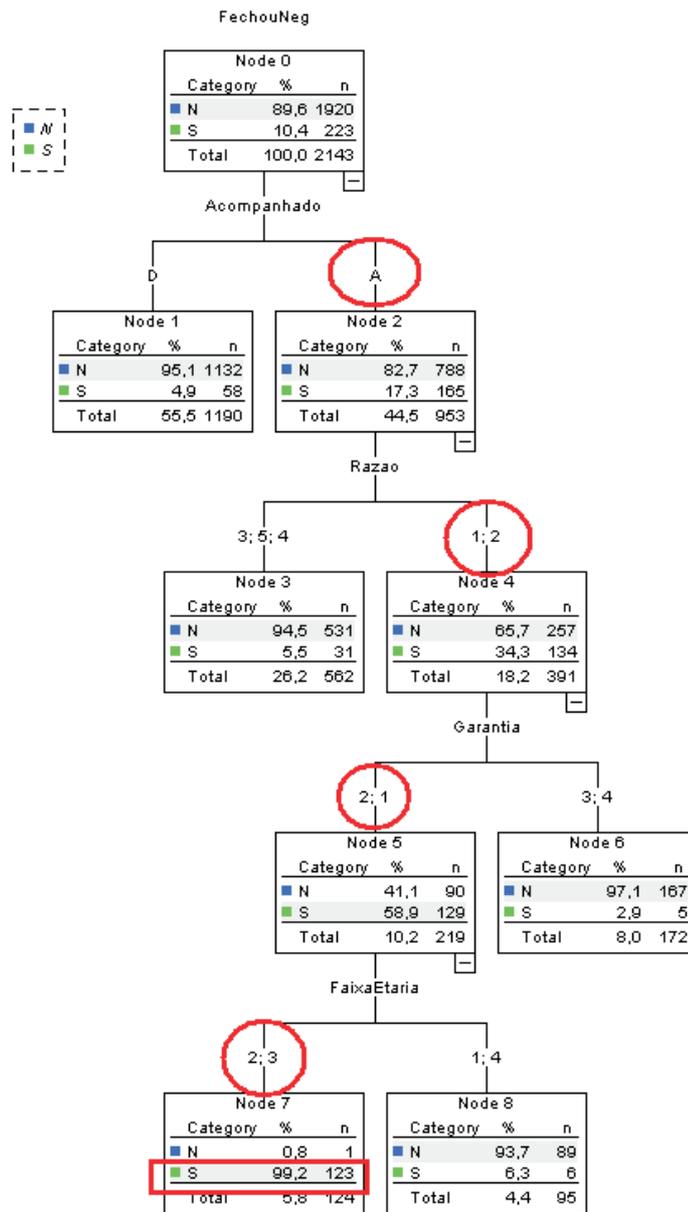


Figura 3.4: Perfil do cliente com alta probabilidade de fechar negócio apresentado pela árvore

onde  $r$  and  $k_{j \in [1..l]}$  pertence ao domínio de  $x_i$ . Cada relação  $R_{x_i \leq r}$  ou  $R_{x_i \in \{k_1, \dots, k_l\}}$  que é encontrada particiona  $\mathcal{O}$  em dois conjuntos diferentes, isto é  $\mathcal{O}_R$  e  $\mathcal{O}_{-R}$ , tal que  $\mathcal{O}_R \cap \mathcal{O}_{-R} = \{\}$  e  $\mathcal{O}_R \cup \mathcal{O}_{-R} = \mathcal{O}$ . Neste caso,  $\mathcal{O}_R$  contém as observações em  $\mathcal{O}$  para as quais  $R$  é válida e  $\mathcal{O}_{-R}$  as observações para as quais a negação de  $R$  é válida.

Uma métrica  $M$  é então utilizada para avaliar quão diverso são os dados em  $\mathcal{O}$ ,  $\mathcal{O}_R$  and  $\mathcal{O}_{-R}$  em relação aos elementos de  $J$ . A diferença entre  $M(\mathcal{O})$  e a média ponderada de  $M(\mathcal{O}_R)$  e  $M(\mathcal{O}_{-R})$  indica a contribuição da relação  $R$  para a redução de diversidade dos elementos em  $J$ , como consequência do particionamento de  $\mathcal{O}$  em  $\mathcal{O}_R$  e  $\mathcal{O}_{-R}$ . Dado que é desejável que a diversidade deve ser reduzida tão rápido quanto possível, a relação  $R$  que provê a maior redução de diversidade é inicialmente utilizada para particionar  $\mathcal{O}$ . O processo é então sucessivamente aplicado a  $\mathcal{O}_R$  e  $\mathcal{O}_{-R}$ , até que nenhum ganho significativo na redução de diversidade seja obtido.

No paradigma das árvores de classificação, quando  $J$  contém somente dois elementos uma das métricas mais freqüentemente utilizadas é o índice de diversidade Gini, proposto inicialmente por Corrado Gini (GINI, 1939) e, mais tarde, adaptado por Breiman *et al.* (BREIMAN *et al.*, 1984) para o desenvolvimento de métodos de classificação. Em termos formais, para um dado conjunto de observações, o índice Gini é dado por:

$$I(\mathcal{O}) = 1 - S, \quad (3.1)$$

onde  $S = \sum_{i=1}^{|J|} P(j_i|\mathcal{O})^2$  para  $j_i \in J$ , e  $P(j_i|\mathcal{O})$  é a probabilidade de ocorrência de objetos  $j_i$  em  $\mathcal{O}$ . Nestas circunstâncias a redução de diversidade propiciada pela relação  $R$  é dada por:

$$\Delta I(\mathcal{O}) = I(\mathcal{O}) - (I(\mathcal{O}_R) \times p_R + I(\mathcal{O}_{-R}) \times p_{-R}) \quad (3.2)$$

onde  $p_R$  e  $p_{-R}$  são respectivamente a proporção de elementos em  $\mathcal{O}_R$  e  $\mathcal{O}_{-R}$ .

## 3.2 O Simulador de cenários

Para obtermos os resultados desta dissertação utilizamos um simulador de cenários para a simulação do RCPS estocástico com heurísticas para o escalonamento das atividades. Este simulador, chamado *Process Simulator*, foi desenvolvido por (KLING, 2005) na Universidade Federal do Rio de Janeiro, como parte da sua dissertação de mestrado. A Figura 3.5 mostra a tela de entrada de dados do *Process Simulator*.

Note que o simulador recebe como entrada uma rede de atividades, onde são definidas as seguintes variáveis:

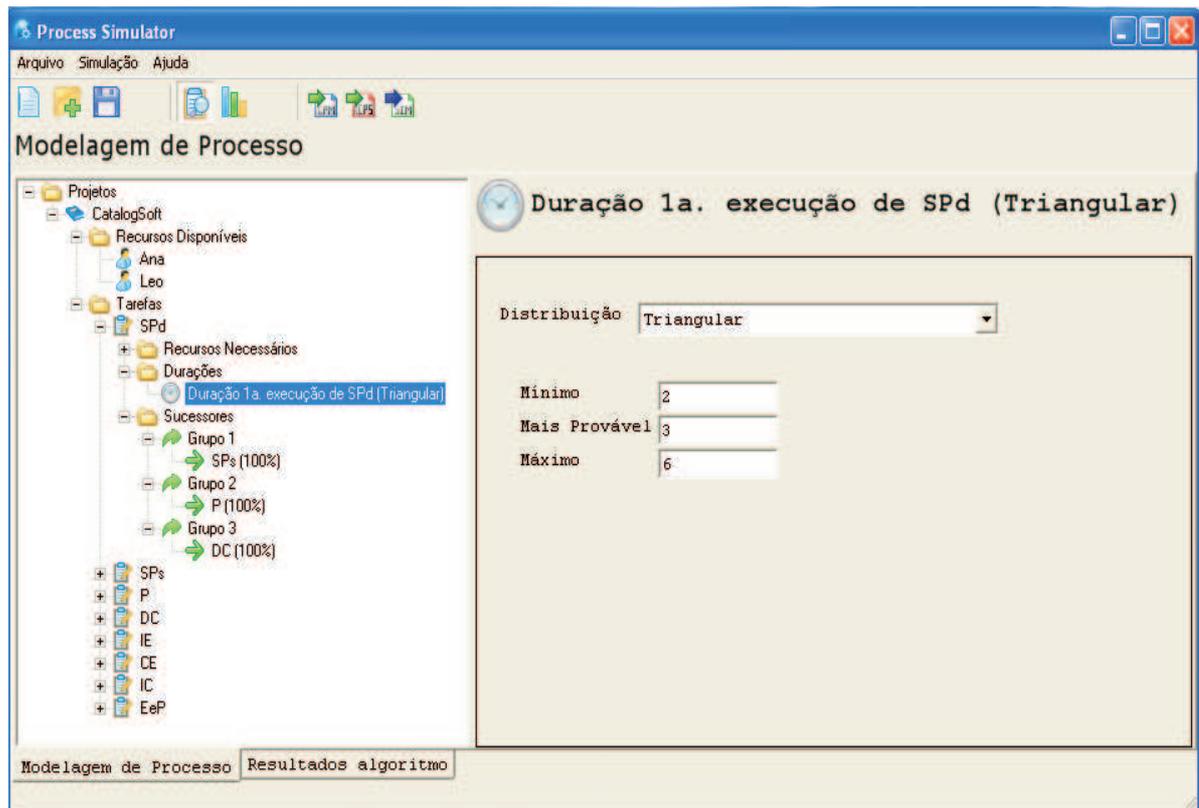


Figura 3.5: *Process Simulator* - Simulador de cenários de teste de um projeto

- **Recursos disponíveis:** Quais tipos de recurso serão alocados para execução do projeto e suas respectivas quantidades disponíveis.
- **Recursos necessários:** Quais recursos, dos disponíveis, são utilizados para a execução de cada atividade e suas quantidades.
- **Durações:** Qual o tipo de distribuição será utilizada para gerar a duração de cada atividade.
- **Sucessores:** Próximas atividades que podem ser iniciadas ao término da atividade em questão.

Nesta proposta utilizamos a distribuição de probabilidade triangular para a geração das durações das atividades nos cenários. As estimativas de três pontos se baseiam na determinação de três valores (SCHERER; POMROY; FULLER, 2003) :

- **Mínimo:** estimativa para a duração da atividade em questão em um cenário otimista,
- **Mais Provável:** estimativa para a duração da atividade baseada em um cenário realista, e
- **Máximo:** estimativa para duração da atividade em um cenário pessimista.

Onde a média  $\mu$  da distribuição triangular é dada por:

$$\mu = \frac{\text{Mínimo} + \text{Mais Provável} + \text{Máximo}}{3},$$

e possui desvio padrão  $\sigma$ :

$$A = \text{Mínimo}^2 + \text{Mais Provável}^2 + \text{Máximo}^2$$

$$B = \text{Mínimo} \times \text{Mais Provável} + \text{Mínimo} \times \text{Máximo} + \text{Mais Provável} \times \text{Máximo}$$

$$C = \text{Mínimo} \times \text{Mais Provável} \times \text{Máximo}$$

$$\sigma = \sqrt{\frac{A - B - C}{18}}$$

### 3.2.1 Heurísticas de Escalonamento

Uma vez que uma rede de atividades de atividades tenha sido definida é necessário escolher a heurística responsável por escaloná-las. Uma heurística é necessária para decidir qual atividade será executada primeiro em casos em que mais de uma atividade solicite o mesmo recurso. Neste caso, a heurística escolhe qual atividade será executada e qual deverá esperar.

Nesta dissertação, estendemos o *Process Simulator* para permitir que mais de uma heurística possa ser utilizada durante o processo de simulação. Para avaliação de cada cenário uma heurística, de uma lista previamente selecionada pelo usuário, é utilizada.

O motivo pelo qual decidimos permitir a utilização de mais de uma heurística durante o processo de simulação é o de fornecer ao gerente mais flexibilidade nas decisões que tenha que tomar com vistas a aumentar a probabilidade de alcançar os objetivos do projetos, já que com a escolha de uma heurística adequada o projeto pode terminar num tempo menor ou pode-se diminuir o problema de concorrência por recursos, por outro lado uma escolha não adequada pode propiciar justamente o oposto. Um estudo mais específico sobre estas heurísticas de escalonamento pode ser visto em (GOERLICH; OLAGUIBEL, 1989).

### 3.2.2 Os Cenários de Teste

Para a geração dos cenários foi escolhido o algoritmo para simulação de estância única do *Process Simulator*. Este nome deve-se ao fato de que cada instância (ou cenário) é executada

em separado, ou seja, em um dado momento da simulação existe sempre uma única instância executando.

O algoritmo para simulação de instância única pode trabalhar tanto com um algoritmo de escalonamento com recursos quanto com um que ignora as restrições de recursos. Para efeito desta dissertação utilizaremos o escalonamento com a consideração dos recursos.

O processo de simulação é demonstrado a seguir.

1. Definir a rede de atividades  $R$
2. Escolher número de cenários  $n$
3. Faça de  $i = 1$  até  $n$ 
  - (a) Sortear quantidade de cada recurso  $r[1..m]$
  - (b) Sortear uma heurística  $h$
  - (c) Executar Algoritmo RCPS com a heurística  $h$
  - (d) Coletar dados da simulação

Note que para cada cenário, além de uma heurística, também é sorteada a quantidade de cada recurso disponível para a execução do projeto, ou seja, à cada cenário a quantidade de cada recurso muda, onde estes sorteios são independentes.

### 3.2.3 Coletando Dados da Simulação

A saída do simulador é um conjunto de cenários de teste gerados com as seguintes informações coletadas:

- O percentual do tempo disponível que foi consumido pela atividade,
- A quantidade de cada recurso disponível para o projeto,
- A heurística utilizada e
- O tempo total de duração do projeto.

A Tabela 3.3 mostra um exemplo das informações coletadas nos cenários de teste, onde as letras de A a E representam atividades, R1 e R2 representam os recursos disponíveis e H1 a H4

as heurísticas utilizadas em cada cenário. A variável *Total* representa o tempo total de duração do cenário.

Posteriormente, as informações apresentadas na Tabela 3.3 foi acrescentada uma variável (*Resultado*) indicando se o projeto terminou ou não dentro do prazo estabelecido pela organização que irá pagar pelo seu desenvolvimento. Nestes caso os valores *D* e *F* foram utilizados para indicar se o projeto terminou dentro ou fora do prazo respectivamente. A Tabela 3.4 apresenta estas informações.

| Variável     | Cenário |       |       |       |     |       |
|--------------|---------|-------|-------|-------|-----|-------|
|              | 1       | 2     | 3     | 4     | ... | n     |
| <i>A</i>     | 0,912   | 0,475 | 0,660 | 0,132 | ... | 0,432 |
| <i>B</i>     | 0,534   | 0,574 | 0,942 | 0,444 | ... | 0,236 |
| <i>C</i>     | 0,160   | 0,856 | 0,650 | 0,270 | ... | 0,260 |
| <i>D</i>     | 0,673   | 0,705 | 0,758 | 0,724 | ... | 0,689 |
| <i>E</i>     | 0,135   | 0,617 | 0,602 | 0,515 | ... | 0,502 |
| <i>R1</i>    | 3       | 1     | 3     | 2     | ... | 1     |
| <i>R2</i>    | 2       | 3     | 2     | 1     | ... | 2     |
| <i>Heu</i>   | H1      | H2    | H3    | H4    | ... | H2    |
| <i>Total</i> | 158,9   | 160,7 | 167,5 | 164,1 | ... | 168,0 |

Tabela 3.3: Cenários de teste gerados pelo *Process Simulator*.

| Variável         | Cenário |       |       |       |     |       |
|------------------|---------|-------|-------|-------|-----|-------|
|                  | 1       | 2     | 3     | 4     | ... | n     |
| <i>A</i>         | 0,912   | 0,475 | 0,660 | 0,132 | ... | 0,432 |
| <i>B</i>         | 0,534   | 0,574 | 0,942 | 0,444 | ... | 0,236 |
| <i>C</i>         | 0,160   | 0,856 | 0,650 | 0,270 | ... | 0,260 |
| <i>D</i>         | 0,673   | 0,705 | 0,758 | 0,724 | ... | 0,689 |
| <i>E</i>         | 0,135   | 0,617 | 0,602 | 0,515 | ... | 0,502 |
| <i>R1</i>        | 3       | 1     | 3     | 2     | ... | 1     |
| <i>R2</i>        | 2       | 3     | 2     | 1     | ... | 2     |
| <i>Heu</i>       | H1      | H2    | H3    | H4    | ... | H2    |
| <i>Total</i>     | 158,9   | 160,7 | 167,5 | 164,1 | ... | 168,0 |
| <i>Resultado</i> | D       | D     | F     | D     | ... | F     |

Tabela 3.4: Dados coletados durante o processo de simulações acrescidos da variável *Resultado*.

## ***4 Estudo de Caso - Sistema para elaboração de catálogos***

### **4.1 Modelo Estocástico do RCPS com Árvores Decoradas**

De acordo com Seneca (4 AC – 65 DC ), o filósofo romano: “as regras fazem o trajeto do aprendiz longo, enquanto os exemplos o fazem curto e bem sucedido”. Em consequência, o método apresentado nesta dissertação é introduzido com ajuda de um exemplo inspirado no mundo real.

Considere uma cadeia de lojas de móveis que utiliza catálogos para aumentar as vendas. Periodicamente esta companhia edita catálogos contendo uma grande variedade de ofertas de produtos que são enviados a grupos de compradores potenciais (*prospects*), selecionados na base de dados da empresa. A condução apropriada deste trabalho requer que oito atividades sejam executadas, eficientemente, dentro de um curto espaço de tempo, a saber:

1. *Seleção de Produtos* - seleciona os produtos que serão ofertados no catálogo;
2. *Seleção de Prospects* - identifica os clientes para quem os catálogos serão enviados;
3. *Precificação* - estabelece o preço promocional de cada produto ofertado no catálogo;
4. *Design do Catálogo* - onde os aspectos gráficos e textuais do catálogo são concebidos e harmonizados;
5. *Impressão de Etiquetas* - onde as etiquetas como o nome e endereço dos *prospects* são impressas e organizadas;
6. *Controle de Estoque* - assegura que os produtos ofertados no catálogo estarão disponíveis para envio quando forem solicitados;
7. *Impressão do Catálogo* - onde a impressão do catálogo é realizada;

8. *Etiquetagem e Postagem* - etiqueta os catálogos e efetua a postagem para os endereços de destino.

Ao longo dos anos a competição acirrada no ramo de móveis tem diminuído as margens de lucro da empresa. Em conseqüência, a cadeia de lojas acredita que sua sobrevivência depende da eficiência de seus processos do negócio. Com vistas a aumentar a eficiência de suas campanhas de marketing, a empresa decidiu desenvolver um sistema de informação com módulos que, trabalhando em conjunto, fornecem o suporte adequado às atividades que fazem parte das campanhas. Como cada atividade é auxiliada por um módulo diferente, oito módulos devem ser construídos de maneira que a informação disponibilizada por um módulo possa ser usada por outro.

Infelizmente, devido à falta de recursos adequados para execução do projeto, somente duas pessoas foram selecionadas inicialmente para trabalhar no desenvolvimento dos módulos do sistema: *Ana*, uma analista de sistemas, e *Leo*, um programador de computador. A Tabela 4.1 mostra os recursos humanos necessários para o desenvolvimento de cada módulo.

| Módulo     |                        | Recurso<br>Necessário |
|------------|------------------------|-----------------------|
| Nome       | Atividade              |                       |
| <i>SPd</i> | Seleção de Produtos    | <i>Ana e Leo</i>      |
| <i>SPs</i> | Seleção de Prospects   | <i>Ana</i>            |
| <i>P</i>   | Precificação           | <i>Ana e Leo</i>      |
| <i>DC</i>  | Design do Catálogo     | <i>Leo</i>            |
| <i>IE</i>  | Impressão de Etiquetas | <i>Ana e Leo</i>      |
| <i>CE</i>  | Controle de Estoques   | <i>Ana</i>            |
| <i>IC</i>  | Impressão do Catálogo  | <i>Leo</i>            |
| <i>EeP</i> | Etiquetagem e Postagem | <i>Ana e Leo</i>      |

Tabela 4.1: Recursos necessários ao desenvolvimento de cada módulo.

A Figura 4.1 mostra a rede de atividades do projeto de desenvolvimento dos módulos do sistema de suporte às campanhas de marketing. Na figura *SPd* é o primeiro módulo a ser desenvolvido e *EeP* o último. Além disso, uma seta conectando duas atividades tais como *SPd*  $\rightarrow$  *P* indica que o desenvolvimento de *P* só pode ser iniciado quando o desenvolvimento de *SPd* estiver terminado e todos os recursos necessários estiverem disponíveis.

#### 4.1.1 Definindo a Duração das Atividades

Para que este projeto possa ser concluído dentro da janela de oportunidade estabelecida pela empresa, é crucial que o gerente de projetos esteja ciente do *makespan* do projeto de acordo com as restrições de recurso corrente e durações das diferentes atividades do projeto.

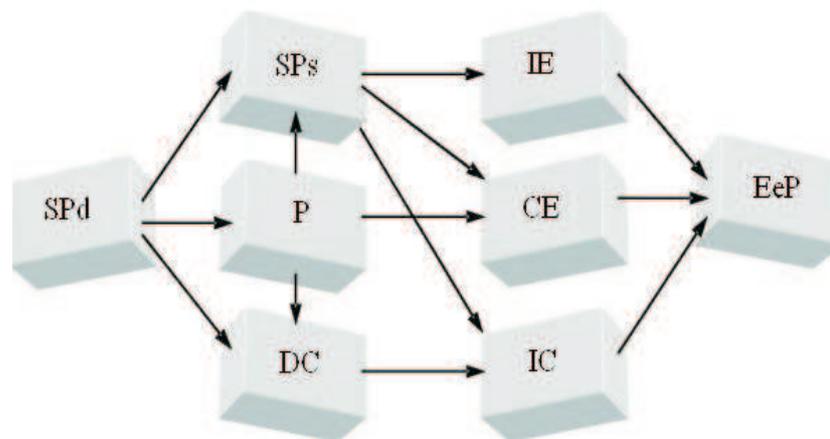


Figura 4.1: Rede de atividades do projeto.

Entretanto, como as atividades ainda não foram executadas, sua duração só pode ser estimada. Neste caso, com o apoio de outros gerentes de projeto experientes e de uma base de dados de projetos previamente executados, estimativas de três pontos foram estabelecidas para cada atividade, indicando sua duração mínimo, mais provável e máxima. A Tabela 4.2 apresenta estes valores.

| Atividades do Projeto | Duração Estimada (unidades de tempo) |               |        |
|-----------------------|--------------------------------------|---------------|--------|
|                       | Mínimo                               | Mais Provável | Máximo |
| <i>SPd</i>            | 2                                    | 3             | 6      |
| <i>SPs</i>            | 3                                    | 6             | 8      |
| <i>P</i>              | 4                                    | 6             | 7      |
| <i>DC</i>             | 1                                    | 8             | 10     |
| <i>IE</i>             | 7                                    | 9             | 11     |
| <i>CE</i>             | 5                                    | 6             | 7      |
| <i>IC</i>             | 4                                    | 5             | 8      |
| <i>EeP</i>            | 1                                    | 2             | 4      |

Tabela 4.2: Durações das atividades

#### 4.1.2 Simulação dos cenários

Considerando que a duração exata de cada atividade é desconhecida e que os recursos disponíveis podem ser insuficientes para assegurar que todas as atividades sejam executadas quando seus precedentes terminarem, um modelo estocástico de simulação por Monte Carlo foi construído para analisar o *makespan* do projeto. Neste modelo, a duração de cada atividade é descrita por uma distribuição triangular de probabilidade, uma das distribuições mais utilizadas para descrever a duração de atividades em cenários de incerteza (CHUNG, 2003), apresentada

no Capítulo 3.

| Variável         | Cenário |       |       |       |     |       |
|------------------|---------|-------|-------|-------|-----|-------|
|                  | 1       | 2     | 3     | 4     | ... | n     |
| <i>SPd</i>       | 0,912   | 0,475 | 0,660 | 0,132 | ... | 0,432 |
| <i>SPs</i>       | 0,534   | 0,574 | 0,942 | 0,444 | ... | 0,236 |
| <i>P</i>         | 0,160   | 0,856 | 0,650 | 0,270 | ... | 0,260 |
| <i>DC</i>        | 0,673   | 0,705 | 0,758 | 0,724 | ... | 0,689 |
| <i>IE</i>        | 0,135   | 0,617 | 0,602 | 0,515 | ... | 0,502 |
| <i>CE</i>        | 0,305   | 0,600 | 0,480 | 0,525 | ... | 0,445 |
| <i>IC</i>        | 0,207   | 0,127 | 0,192 | 0,797 | ... | 0,217 |
| <i>EeP</i>       | 0,383   | 0,596 | 0,156 | 0,456 | ... | 0,560 |
| <i>Leo</i>       | 3       | 1     | 3     | 2     | ... | 1     |
| <i>Ana</i>       | 2       | 3     | 2     | 1     | ... | 2     |
| <i>Resultado</i> | D       | F     | D     | F     | ... | F     |
| <i>Heu</i>       | EFT     | GRD   | LIS   | EFT   | ... | SPT   |

Tabela 4.3: Resultados da simulação.

Subseqüentemente o modelo estocástico foi submetido a um processo de simulação, onde a duração de cada atividade foi registrada como uma porcentagem da sua amplitude de duração<sup>1</sup>, juntamente com a indicação de se o projeto terminou dentro do prazo estipulado ou não. Em adição, para cada cenário foi sorteado uma heurística para escalonamento das atividades concorrentes por recursos, sendo esta informação registrada com os dados da simulação. A escolha da heurística para cada cenário é feita por sorteio que segue uma distribuição normal. A Tabela 4.3 apresenta os valores coletados durante a simulação.

Neste processo de simulação foram observados 6000 cenários. Para definirmos este número de cenários utilizamos o teste KS (*Kolmogorov-Smirnov test*) (BARBOSA et al., 2007) para sabermos o erro entre a distribuição real e a distribuição simulada. Para um número muito grande de observações ( $N$ ) o erro é calculado pela fórmula, de acordo com o grau de confiança  $\varepsilon$ :

$$\bullet K\sqrt{N}$$

Onde  $K$  é uma constante que para um número muito grande de observações ( $N$ ), assume os valores de acordo com o grau de confiança definido:

|                            |                            |                            |                            |
|----------------------------|----------------------------|----------------------------|----------------------------|
| $\varepsilon = 0.20$ (80%) | $\varepsilon = 0.10$ (90%) | $\varepsilon = 0.05$ (95%) | $\varepsilon = 0.01$ (99%) |
| $1.07\sqrt{N}$             | $1.22\sqrt{N}$             | $1.36\sqrt{N}$             | $1.63\sqrt{N}$             |

Portanto, para  $N = 6000$  observações e com um grau de confiança  $\varepsilon = 0.20$  (80%), temos um erro de aproximadamente 0.014 entre a distribuição real e a simulada.

<sup>1</sup>Diferença entre o tempo máximo e mínimo estimado para a duração de uma atividade ou projeto.

A Tabela 4.4 descreve o significado das variáveis apresentadas na Tabela 4.3. Por exemplo, como indicado na pela variável *DC*, no primeiro cenário o desenvolvimento do módulo de suporte ao *design* do catálogo consumiu 67,3% da sua amplitude de duração, isto é,  $1 + 0,673 \times (10 - 1) = 7,05$  unidades de tempo. As seguintes heurísticas de escalonamento de atividades foram utilizadas durante o processo de simulação:

- EFT (*Early Finishing Time*) – prioriza as atividades onde o tempo de término é o menor possível;
- LFT (*Late Finishing Time*) – prioriza as atividades onde o tempo de término é o menor possível;
- GPRPW (*Greatest Proportional Rank Positional Weight*) – prioriza as atividades mais curtas com atividades sucessoras mais demoradas;
- LPRPW (*Least Proportional Rank Positional Weight*) – prioriza as atividades mais longas com atividades sucessoras mais curtas;
- GRPW (*Greatest Rank Positional Weight*) – calcula o somatório da duração da atividade e de todas as suas sucessoras imediatas, priorizando as atividades com o maior somatório;
- LRPW (*Least Rank Positional Weight*) – calcula o somatório da duração da atividade e de todas as suas sucessoras imediatas, priorizando as atividades com o menor somatório;
- GRD (*Greatest Resource Demand*) – prioriza as atividades que demandam a maior quantidade de recursos;
- LRD (*Least Resource Demand*) – prioriza as atividades que demandam a menor quantidade de recursos;
- MIS (*Most Immediate Successors*) – prioriza as atividades com maior número de sucessores imediatos;

| Variável                  | Significado  |
|---------------------------|--|
| <i>SPd, SPs, ..., EeP</i> | Duração das atividades <i>SPd, SPs, ..., EeP</i> expressa como um percentual das suas respectivas amplitudes de duração.                                     |
| <i>Ana and Leo</i>        | Quantidade de analistas de sistemas ( <i>Ana</i> ) e programadores de computador ( <i>Leo</i> ) utilizadas no projeto. Essas quantidades variam entre 1 e 3. |
| <i>Heu</i>                | Heurística utilizada durante o processo de simulação.  |
| <i>Resultado</i>          | O resultado de um determinado cenário, onde <i>D</i> indica que o projeto terminou dentro do tempo permitindo e <i>F</i> caso contrário.                     |

Tabela 4.4: Variáveis de simulação.

- LIS (*Least Immediate Successors*) – prioriza as atividades com menor número de sucessores imediatos;
- LPT (*Longest Processing Time*) – prioriza as atividades mais longas, isto é com maior tempo de processamento; e
- SPT (*Shortest Processing Time*) – prioriza as atividades mais curtas, isto é com menor tempo de processamento.

### 4.1.3 Análise das regras de classificação

A Figura 4.2 mostra a árvore de classificação do modelo de simulação da duração do projeto, construída utilizando o método de Breiman *et al.* (BREIMAN et al., 1984). Na figura a caixa *Node 0* é a raiz da árvore. Ela contém o número total das observações coletadas durante a simulação, isto é, 6.000, e o número e a proporção dos cenários em que o projeto terminou e não terminou no prazo. Por exemplo, inicialmente, em 60,2% dos cenários gerados o projeto terminou no prazo, enquanto em 39,8% não.

O particionamento dos nós, durante a construção da árvore é feito seguindo o método explicado no capítulo 3, ou seja, as variáveis preditivas são examinadas de forma a buscar uma relação que divida este conjunto das variáveis em dois subconjuntos. Em seguida é aplicada uma métrica em cada subconjunto, e no conjunto total para que a diferença entre a média ponderada de cada subconjunto e o conjunto total, mostre qual subconjunto contribui mais para a redução da diversidade deste conjunto total.

Deve-se notar que a relação que é usada particionar o conjunto inicial dos cenários é  $Heu \in \{LIS, LPRPW, LPT, LRD, LPRW, MIS\}$  e, em decorrência, *Node 1* contém as 3.000 observações nas quais esta relação é verdadeira enquanto *Node 2* contém os 3.000 casos restantes onde a relação  $Heu \in \{EFT, GPRPW, GRD, GRPW, LFT, SPT\}$  é verdadeira, isto é, os outros 50,0% dos cenários simulados.

Devido a maneira incremental com que a árvore é construída, todas as relações que são verdadeiras para as observações em um nó também são verdadeiras para as observações nos nós descendentes. Por exemplo, em todos os cenários que são parte do *Node 3* a variável  $Heu \in \{LIS, LPRPW, LPT, LRD, LPRW, MIS\}$  e o número de recursos *Leo* utilizado para executar o projeto é menor ou igual a 1. A Tabela 4.5 mostra as relações que são verdadeiras para as folhas da árvore da Figura 4.2, juntamente com a proporção de cenários no qual o projeto terminou dentro (*D*) e fora (*F*) do prazo, assim como a proporção total de cenários nos quais estas relações são verdadeiras.

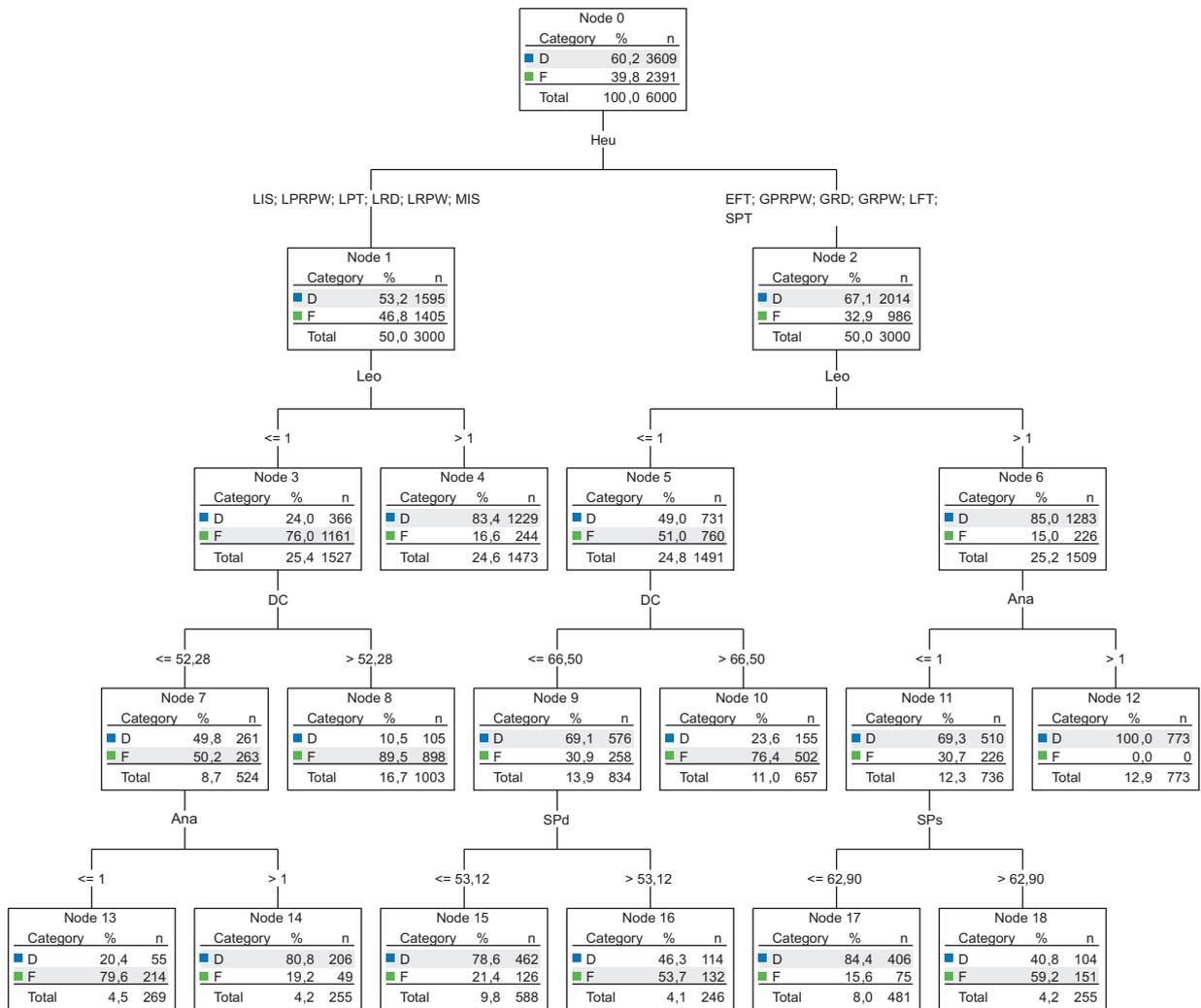


Figura 4.2: O *makespan* da Árvore de Classificação.

Devido a natureza aleatória da simulação de Monte Carlo, estas proporções são, na verdade, estimativas da probabilidade de projeto terminar no prazo em diferentes circunstâncias (CHUNG, 2003). Por exemplo, a proporção de cenários nos quais a relação é dada por  $Heu \in \{LIS, LPRPW, LPT, LRD, LRPW, MIS\} \wedge Leo > 1$  é 24,6% e em 83,4% destes cenários o projeto termina dentro do prazo. Veja o Nó 4 da Tabela 4.5. Em decorrência, a probabilidade de ocorrência de cenários nos quais estas duas relações são verdadeiras é de  $0,246 \times 0,834 = 0,205$ .

De acordo com as informações exibidas na Figura 4.2, a probabilidade projeto do sistema terminar no prazo é 0,602, ou 60.2%. Veja o *Node 0*. Como esta probabilidade está abaixo do padrão de confiança de 80% estabelecido pela empresa, o gerente é compelido à agir. Há diversas ações que podem ser executadas. Por exemplo, o gerente pode tentar convencer as partes interessadas a estender o prazo de término do projeto. Também, com um algum custo extra, pode empregar mais pessoas para trabalhar no projeto ou estender o esforço de trabalho de 5 dias de 6 horas para uma semana de 6 dias de 8 horas durante o desenvolvimento de determinadas ferramentas (um aumento de 60% nas horas de trabalho).

| Nó | Regra  | Resultado (%) |      |                   |
|----|--|---------------|------|-------------------|
|    |  | Dentro        | Fora | Total de Cenários |
| 4  | $Heu \in \{LIS, LPRPW, LPT, LRD, LRPW, MIS\} \wedge Leo > 1$   | 83,4          | 16,6 | 24,6              |
| 8  | $Heu \in \{LIS, LPRPW, LPT, LRD, LRPW, MIS\} \wedge Leo \leq 1 \wedge DC > 52,3$                         | 10,5          | 89,5 | 16,7              |
| 10 | $Heu \in \{EFT, GPRPW, GRD, GRPW, LFT, SPT\} \wedge Leo \leq 1 \wedge DC > 66,5$                         | 23,6          | 76,4 | 11,0              |
| 12 | $Heu \in \{EFT, GPRPW, GRD, GRPW, LFT, SPT\} \wedge Leo > 1 \wedge Ana > 1$                              | 100,0         | 0,0  | 12,9              |
| 13 | $Heu \in \{LIS, LPRPW, LPT, LRD, LRPW, MIS\} \wedge Leo \leq 1 \wedge DC \leq 52,3 \wedge Ana \leq 1$    | 20,4          | 79,6 | 4,5               |
| 14 | $Heu \in \{LIS, LPRPW, LPT, LRD, LRPW, MIS\} \wedge Leo \leq 1 \wedge DC \leq 52,3 \wedge Ana > 1$       | 80,8          | 19,2 | 4,2               |
| 15 | $Heu \in \{EFT, GPRPW, GRD, GRPW, LFT, SPT\} \wedge Leo \leq 1 \wedge DC \leq 66,5 \wedge SPd \leq 53,1$ | 78,6          | 21,4 | 9,8               |
| 16 | $Heu \in \{EFT, GPRPW, GRD, GRPW, LFT, SPT\} \wedge Leo \leq 1 \wedge DC \leq 66,5 \wedge SPd > 53,1$    | 46,3          | 53,7 | 4,1               |
| 17 | $Heu \in \{EFT, GPRPW, GRD, GRPW, LFT, SPT\} \wedge Leo > 1 \wedge Ana \leq 1 \wedge SPs \leq 62,9$      | 84,4          | 15,6 | 8,0               |
| 18 | $Heu \in \{EFT, GPRPW, GRD, GRPW, LFT, SPT\} \wedge Leo > 1 \wedge Ana \leq 1 \wedge SPs > 62,9$         | 40,8          | 59,2 | 4,2               |

Tabela 4.5: Regras que mostram as chances do projeto terminar no prazo.

#### 4.1.4 Análise financeira dos resultados

Os dados coletados durante o processo de simulação (duração total do projeto e a duração do desenvolvimento dos módulos do *DC*, *SPs* e *SPd*, que aparecem na árvore de classificação do *makespan* do projeto), junto com os reais por unidade de tempo cobradas por Leo e por Ana (veja Tabela 4.6), permite a construção da Tabela 4.7 conectando ações e custos extras com as folhas da árvore na Figura 4.3, criando, em decorrência, uma árvore de classificação decorada. Tudo isso permite que o gerente do projeto considere a opção mais apropriada para aumentar a probabilidade de entrega do projeto no prazo. Neste caso, a escolha é simples, dado que a opção G se sobrepõe as demais. Para decisões mais complexas em ambiente de incerteza consulte (WANG, 2002; BAIRD, 2004).

| Semana de Trabalho<br>(Horas) | Reais / Unidade de Tempo |       |
|-------------------------------|--------------------------|-------|
|                               | Ana                      | Leo   |
| Normal (30h)                  | 1.875                    | 1.125 |
| Extendida (48h)               | 2.500                    | 1.500 |

Tabela 4.6: Custos de Ana e Leo

| Ação           | Descrição   | Custo Extra (VPL) |               |
|----------------|---|-------------------|---------------|
|                |   | Min               | Max           |
| A              | Aplicar uma das heurística LIS, LRPW, LPT, LRD, LRPW ou MIS e contratar um outro <i>Leo</i> .   | R\$ 23.231        | R\$ 45.641    |
| B, C, E, H e J | Convencer as partes interessadas a estender o prazo de duração do projeto.  | R\$ 0.00          |               |
| D              | Contratar um outro <i>Leo</i> e uma outra <i>Ana</i> .  | R\$ 61.920        | R\$ 116.580   |
| F              | Aplicar uma das heurística LIS, LRPW, LPT, LRD, LRPW ou MIS, aumentar a jornada de trabalho durante o desenvolvimento do módulo <i>DC</i> e contratar uma outra <i>Ana</i> .  | R\$ 41.073        | R\$ 78.836    |
| G              | Aplicar uma das heurística EFT, GPRPW, GRD, GRPW, LFT ou SPT e aumentar a jornada de trabalho durante o desenvolvimento dos módulos <i>DC</i> e <i>SPd</i> .                  | R\$ 4.587         | R\$ 7.569     |
| I              | Aplicar uma das heurística EFT, GPRPW, GRD, GRPW, LFT ou SPT, contratar um outro <i>Leo</i> e aumentar a jornada de trabalho durante o desenvolvimento do módulo <i>SPs</i> . | R\$ 26.636        | R\$ 43.860,50 |

Tabela 4.7: Ações possíveis e seus respectivos custos com 80% de confiança.

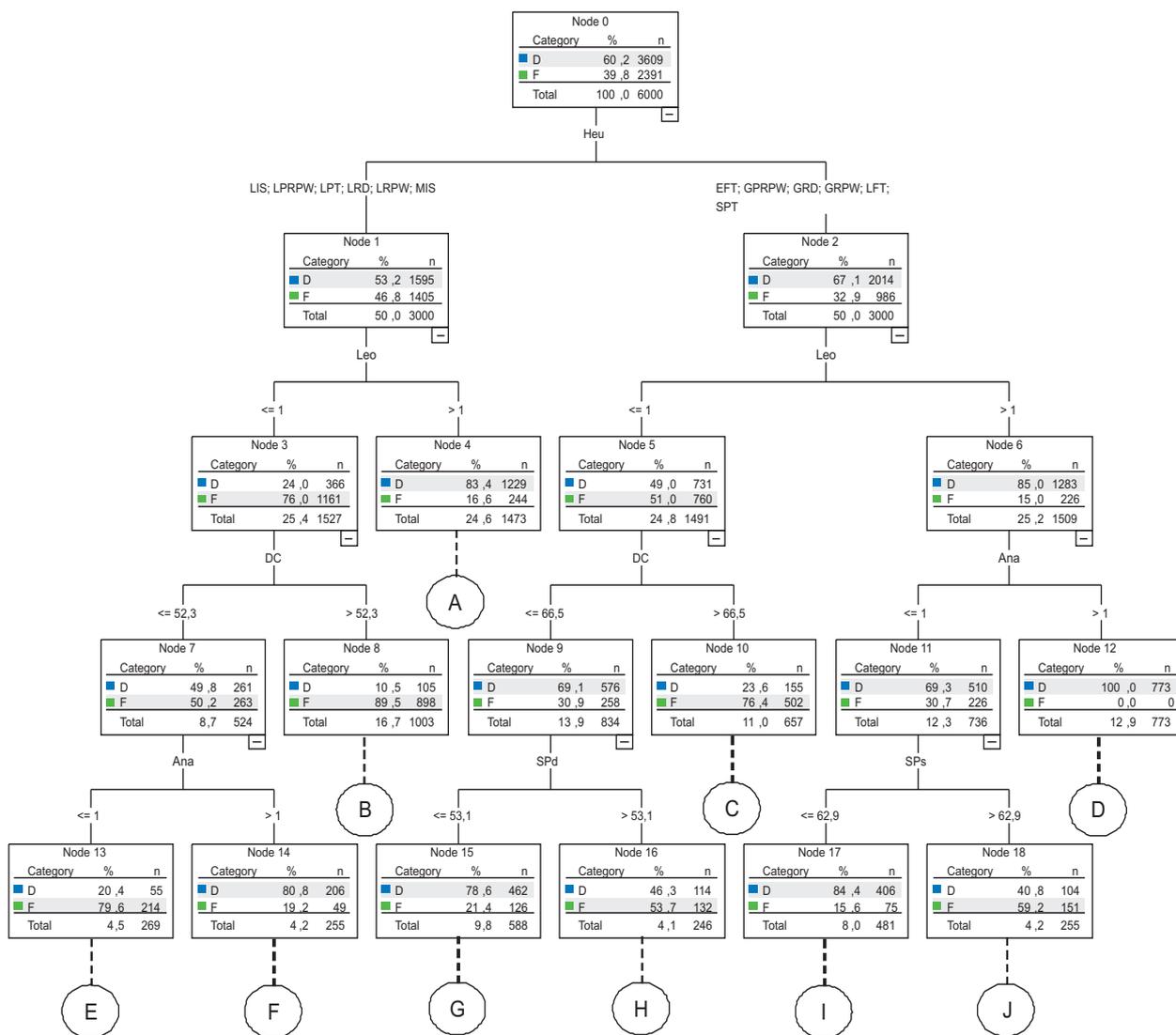


Figura 4.3: Árvore de Classificação Decorada

## 5 *Conclusões*

### 5.1 **Discussão**

Abaixo encontram-se as respostas para algumas questões chaves sobre as implicações das informações fornecidas pelas árvores de classificação decoradas à gerência de projetos.

#### 5.1.1 **Que passos permitem a construção de árvores de classificação decoradas?**

Para ajudar a aumentar as chances de um projeto terminar dentro do prazo podemos:

- (A) Construir um modelo estocástico que represente corretamente as atividades do projeto, suas relações de dependência e as restrições de disponibilidade de recurso a que estão sujeitas;
- (B) Simular a execução do projeto registrando, para cada cenário, o percentual da amplitude de duração das tarefas, a heurística de escalonamento utilizada e se o projeto terminou ou não dentro do prazo;
- (C) Utilizar os dados coletados durante a simulação para construir uma árvore de classificação que tem como variável *target* a variável que indica se o projeto terminou no prazo ou não;
- (D) Examinar a árvore cuidadosamente procurando por oportunidades para melhorar as chances do projeto terminar dentro do prazo;
- (E) Decorar a árvore conectando ações corretivas e custos extras às folhas;
- (F) Selecionar e executar as ações mais adequadas com vistas ao aumento da probabilidade de ocorrência de cenários favoráveis ao término do projeto no prazo;
- (G) Se alguma ação for executada, então o modelo estocástico deve refletir essas mudanças, o processo de simulação deve ser executado novamente e a árvore decorada reconstruída, preparando o caminho para novas ações corretivas que, por ventura, se façam necessárias.

Na verdade, como projetos estabelecem um ambiente de trabalho dinâmico em que diferentes circunstâncias se apresentam no decorrer do tempo, é aconselhável que o acompanhamento do desenrolar das atividades seja feito sempre que alguma informação nova estiver disponível. Por exemplo, ao término de cada tarefa seu tempo de execução se torna conhecido, abrindo vistas para revisão do planejamento. Neste caso, a execução dos passos de (A) a (G) pode se revelar uma ferramenta valiosa para garantir que o projeto tenha boas chances de ser executado conforme o planejado.

### **5.1.2 Como os gerentes de projeto se beneficiam de árvores de classificação decoradas?**

Um dos principais benefícios oriundos da construção de modelos de simulação de RCPS são as informações que eles provêm sobre as relações que existem entre a duração das atividades e o resultado do projeto, isto é, se ele termina no prazo ou não. A compreensão destas relações permite a identificação das atividades que mais influenciam a duração do projeto, propiciando mudanças no planejamento que aumentam as chances de que o projeto seja entregue conforme o combinado. Entretanto, mesmo a análise de modelos de RCPS relativamente pequenos requerem um conhecimento considerável de matemática e estatística. Em consequência, muitos gerentes de projeto têm dificuldade em se beneficiar do uso desta ferramenta, porque não tem o conhecimento técnico adequado.

Em adição, os métodos estatísticos que são mais frequentemente utilizados para estabelecer uma relação de causa e efeito entre as variáveis de um modelo RCPS são correlação e regressão linear. O uso apropriado destes métodos requer que estas relações sejam lineares, o que frequentemente não é o caso. Finalmente, tanto correlação quanto regressão não lidam facilmente com variáveis categóricas. Regressão requer que estas variáveis sejam transformadas em um conjunto ortogonal de variáveis independentes. Correlação linear requer que estas variáveis sejam da escala racional. Mesmo modelos de correlação não paramétricos, tal como correlação de Spearman, requerem que as variáveis sejam pelo menos de escala ordinal. Veja (MCCLAVE; SINCICH, 2005) para uma introdução à correlação e regressão.

O uso de árvores de classificação para analisar modelos de RCPS ajuda a contornar estes problemas. As árvores são modelos não paramétricos, portanto as variáveis podem ter qualquer tipo de relacionamento entre elas. Além disso, a relação entre a duração das atividades e o resultado do projeto são descritas por expressões lógicas que são fáceis de serem lidas e compreendidas, mesmo por aqueles com conhecimento restrito de matemática e estatística. Portanto, embora as árvores de classificação decoradas não sejam capazes de transformar a construção de

modelos de RCPS em uma tarefa mais fácil, elas ajudam na análise dos resultados do processo de simulação, particularmente na obtenção e uso de informações financeiras que favorecem mudanças no planejamento que aumentam as chances de projetos serem entregues dentro do prazo.

### **5.1.3 As árvores de classificação decoradas podem ser usada como uma ferramenta de negociação?**

A despeito de todo o esforço que o gerente de projetos possa fazer, projetos de *software* são repletos de incertezas que fazem com que o planejamento seja uma tarefa razoavelmente complexa. Projetos de *software* dizem respeito a conjunto de atividades que ainda não aconteceram e que jamais serão executadas uma segunda vez da mesma forma. Portanto, estimativas acuradas do tempo de execução destas atividades são sempre difíceis de serem obtidas.

Em adição, não é incomum encontrar em projetos pessoas que estejam trabalhando juntas pela primeira vez, isso sem mencionar que muitos projetos são influenciados por uma grande variedade de fatores políticos, econômicos e sociais, tais como: aumento da taxa de juros, mudanças de requisitos, disponibilidade de novas tecnologias, introdução de leis e procedimentos, ações governamentais de regulamentação dos mercados, etc. Em decorrência, é muito pouco provável que projetos de *software* sejam executados conforme planejado, exigindo que ajustes sejam efetuados no decorrer do tempo. Quanto mais complexo e longo for o projeto, maior a quantidade de ajustes necessários.

Como as regras geradas pelas árvores de classificação decoradas são conjunções de expressões lógicas conectadas a informações financeiras sobre o planejamento de atividades, que são fáceis de serem lidas e compreendidas, e que podem ser apresentadas a pessoal não técnico, elas podem ser utilizadas como uma ferramenta de conscientização da alta gerência para aspectos críticos do projeto, que requerem investimentos adicionais.

Neste caso, a forma incremental com que as regras são apresentadas faz com que o gerente de projetos possa indicar, com precisão, onde os recursos extras devem ser aplicados, descrevendo as conseqüências deste investimento no aumento das chances do projeto ser entregue no prazo. Portanto, não somente as árvores de classificação decoradas tornam a análise dos modelos estocásticos de RCPS mais fáceis, mas podem também ser utilizadas como uma ferramenta de negociação com *stakeholders* para assegurar que o projeto receba os recursos necessários no momento certo.

## 5.2 Conclusão

No mundo extremamente competitivo que surgiu como consequência do advento da Internet e da globalização da economia, organizações, nas mais variadas linhas de negócio, encontram-se sob uma pressão considerável para inovar, aumentando constantemente a qualidade dos produtos e serviços que colocam à disposição do consumidor. Já que a inovação depende da execução de projetos, a quantidade e a complexidade dos projetos que são executados anualmente pelas organizações têm aumentado consideravelmente, especialmente os projetos ligados ao desenvolvimento de *software*.

Uma combinação de ferramentas tais como árvores de classificação decoradas e modelagem estocástica do RCPS, que auxiliam a identificar e executar ações que aumentam as chances de que projetos sejam entregues dentro do prazo, favorecem não somente a gestão efetiva de projetos de *software*, mas também tem impacto positivo sobre o número de projetos que são executados com sucesso. Tudo isso ajuda a criar e manter organizações que são inovadoras e lucrativas, e que se mantêm competitivas por períodos mais longos de tempo.

Nesta dissertação demonstramos a viabilidade de combinar Árvores de classificação decoradas com modelagem estocástica do RCPS com o objetivo de obter informações que favoreçam mudanças no planejamento que aumentam as chances de projetos serem entregues no prazo. Esta combinação de métodos pode ser utilizada para analisar a duração de projetos com muitas vantagens sobre métodos tradicionais.

Árvores de classificação não são difíceis de serem construídas, já que as variáveis preditivas e *target* não precisam seguir nenhuma distribuição de probabilidade, podendo estar correlacionadas de várias maneiras diferentes. Além disso, as regras geradas por estas árvores são fáceis de serem lidas, compreendidas e comunicadas as partes interessadas. Tudo isso ajuda a evitar falhas de comunicação entre membros da equipe e a identificar mudanças no planejamento das atividades que aumentam as chances de que projetos sejam entregues no prazo.

Em adição, quando conectadas a informações financeiras, essas regras favorecem o envolvimento da alta gerência com aspectos críticos do planejamento e execução, tornando mais fácil para a gerência de projetos assegurar que os investimentos adicionais sejam feitos quanto e onde forem mais necessários. Tudo isso torna as árvores de classificação decoradas uma ferramenta muito atrativa para ser usada, em combinação com modelagem estocástica e simulação, no gerenciamento de projetos complexos no mundo real.

Em comparação aos métodos existentes, podemos dizer que árvores de classificação decoradas contribuem com informações que não são abordadas nestes outros métodos. Estas

informações são a probabilidade do projeto estar dentro ou fora do prazo estipulado, como resultado da simulação, um conjunto de regras (ações) que podem ser tomadas pelo gerente de modo a aumentar a probabilidade do projeto estar dentro do prazo e os custos referentes à estas ações. Com mais estas informações os gerentes de projeto podem tomar decisões mais criteriosas com bases estatísticas e não em num processo de tentativa e erro em busca de uma solução.

Os resultados desta dissertação foram publicados na *The Nineteenth International Conference on Software Engineering and Knowledge Engineering*, em um artigo intitulado “Combining Decorated Classification Tree with RCPS Stochastic Models to Gain New Valuable Insights into Software Project Management” (ALENCAR et al., 2007).

### **5.3 Trabalhos Futuros**

Com o grande número de variáveis que o mundo real fornece, o que torna o planejamento de projetos uma tarefa complexa, obviamente não foi possível considerar todas neste método, porém uma destas variáveis que sugerimos de imediato em trabalhos futuros é o tempo de espera por recurso. Esta variável apresenta a quantidade de tempo que uma atividade teve que esperar para iniciar sua execução, uma vez que ela estava apta, por estar sem recursos disponíveis naquele momento. Durante os experimentos desta dissertação esta variável chegou a ser considerada para a elaboração do método, porém não conseguimos utilizá-la de uma maneira adequada de modo a fornecer também informações relevantes ao processo de planejamento. Por esta razão decidimos não utilizá-la. Com um estudo mais aprofundado é possível que ela seja bastante útil para o método e na geração de informações relevantes aos gerentes de projetos.

Uma questão muito importante quando se avalia durações em cadeias de atividades é a correlação entre as variáveis aleatórias. Em nosso modelo assumimos que não existe correlação entre as variáveis aleatórias que representam a duração das atividades, ou seja, assumimos que elas são independentes. No entanto, no mundo real essa regra nem sempre se aplica. Por exemplo, se forem gastos  $x$  dias para desenvolver um programa, é necessário que sejam gastos um número proporcional a  $x$  dias para testá-lo. Um trabalho que estude como adicionar essas correlações ao modelo pode tornar as análises muito mais precisas.

## *Referências Bibliográficas*

- ALENCAR, A. J. et al. Combining decorted classification tree with rcps stochastic models to gain new valuable insights into software project management. In: CHANG, S.-K. (Ed.). *The Nineteenth International Conference on Software Engineering and Knowledge Engineering (SEKE)*. Boston, MA, USA: KSI, 2007. p. 20–31.
- ALENCAR, A. J.; SCHMITZ, E. A. *Análise de Investimentos em Projetos de Tecnologia da Informação*. Rio de Janeiro, Brasil: ExpertBooks, 2006.
- ARAÚJO, R.; COUTO, R.; SCHMITZ, E. Sim project: Uma ferramenta para análise de risco do tempo de realização de projetos. In: NUNES, D. J.; CAMARGO, M. S. de (Ed.). *XIII Simpósio Brasileiro de Engenharia de Software Caderno de Ferramentas*. Florianópolis, SC: Sociedade Brasileira de Computação (SBC), 1999. p. 25–28.
- ARORA, A.; FOSFURI, A.; GAMBARDILLA, A. *Markets for Technology: The Economics of Innovation and Corporate Strategy*. Cambridge, MA, USA: MIT Press, 2001.
- BAIRD, B. F. *Managerial Decisions Under Uncertainty: An Introduction to the Analysis of Decision Making*. Hoboken, NJ, USA: John Wiley & Sons Inc., 2004.
- BARBOSA, V. C. et al. On simulating nondeterministic stochastic activity networks. March 2007.
- BISTARELLI, S.; FIORAVANTI, F.; PERETTI, P. Defense trees for economic evaluation of security investments. In: IEEE. *The First International Conference on Availability, Reliability and Security*. Barcelona, Spain: Computer Society, 2006. p. 416–423.
- BLAZEWICZ, J.; LENSTRA, J.; KAN, A. R. Scheduling subject to resource constraints: Classification and complexity. *Discrete Applied Mathematics*, v. 5, p. 11–24, 1983.
- BOEHM, B. W. *Software Engineering Economics*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1981.
- BREIMAN, L. et al. *Classification and Regression Trees*. Boca Raton, FL, USA: Chapman & Hall/CRC Press, 1984.
- BRUN, M. et al. Multi-resolution classification trees in OCR design. In: GAUTHIER, F. A. O.; WAZLAWICK, R. S.; WILLWRICH, R. (Ed.). Florianópolis, SC: IEEE Computer Society Press, 2001. v. 18, p. 59–66.
- CASTILLO, A. L.; MUÑOZ, D. F. A decision support system to schedule operations in water heater manufacturing. In: THOMPSON, S.; HART, G. (Ed.). *IIE Annual Conference and Exhibition*. Houston, TX, USA, 2004.

CHUNG, C. A. *Simulation Modeling Handbook: A Practical Approach*. Boca Raton, FL, USA: CRC Press, 2003.

COMPTIA. *The CompTIA Project+ Glossary of Project Management Terms*. Oakbrook Terrace, IL, USA: Computing Technology Industry Association, 2006. Informação disponível na Internet em <http://certification.comptia.org/project/docs/Glossary.pdf>. Site visitado pela última vez em 2 de Julho de 2007.

DEMEULEMEESTER, E. L.; HERROELEN, W. S. *Project Scheduling: A Research Handbook*. 1<sup>st</sup> edition. New York, NY, USA: Springer, 2002. (International Series in Operations Research & Management Science).

FERREIRA, A. L.; TEIXEIRA, C. A. W. *Matemática Financeira*. 2<sup>(a)</sup> edição. Rio de Janeiro, Brasil: ExpertBooks, 2005.

FRAME, J. D. *The New Project Management: Tools for an Age of Rapid Change, Complexity, and Other Business Realities*. 2<sup>nd</sup> edition. Hoboken, NJ, USA: Jossey-Bass, 2002. (Jossey Bass Business and Management Series).

FUSSEL, L.; FIELD, S. The role of the risk management database in the risk management process. In: SELVARAJ, H.; MUTHUKUMAR, V. (Ed.). *Proceedings of the 18th International Conference on Systems Engineering*. Las Vegas, NA, USA: Computer Society, 2005. p. 364–369.

GALVÃO, M. Análise quantitativa de riscos com simulação de monte carlo. *Mundo Project Management*, n. 5, p. 6–17, November 2005.

GAO. *Information Technology Investment Management: A Framework for Assessing and Improving Process Maturity*. Washington, DC, USA: United States General Accounting Office, 2000.

GINI, C. Memorie di metodologica statistica. In: \_\_\_\_\_. Milano, Italy: Dott. A. Giuffrè, 1939. I, cap. Variabilità e concentrazione, p. 359–408. Text written in Italian.

GOERLICH, J. M. T.; OLAGUIBEL, R. A. V. *Heuristics Algorithms for Resource-Constrained Project Scheduling: A Review and Empirical Analysis*. Amsterdam, Netherlands: Elsevier Science, 1989. 113–134 p. (Advances in Project Scheduling).

GROUP, I. R. *Effective Project Management: Tools, Templates & Best Practices*. Ontario, Canada: InfoTech Research Group, 2003. 119–157 p. Technical Report.

GU, W.; GERA, S. *The Effect of Organizational Innovation and Information Technology on Firm Performance*. Ottawa, Canada: Minister of Industry, 2004. (The Canadian Economy in Transition Series).

HAMILL, L. The introduction of new technology into the household. *Personal Technologies*, Springer-Verlag, n. 4, p. 1–16, 2000.

HERROELEN, W.; LEUS, R. Project scheduling under uncertainty: Survey and research potentials. *European Journal of Operational Research*, v. 165, p. 289–306, 2005.

- HOSNI, Y. A.; KHALIL, T. *Management of Technology - Internet Economy: Opportunities and Challenges for Developed and Developing Regions of the World*. Atlanta, GA, USA: Elsevier Science, 2004. 227–337 p.
- JONES, A.; KOVACICH, G. L.; LUZWICK, P. G. *Global Information Warfare: How Businesses, Governments, and Others Achieve Objectives and Attain Competitive Advantages*. 1<sup>st</sup> edition. Boca Raton, FL, USA: CRC Press, 2002.
- KAKABADSE, N. K.; KOUZMIN, A.; KAKABADSE, A. From tacit knowledge to knowledge management: leveraging invisible assets. *Knowledge and Process Management*, v. 8, n. 3, p. 137–154, August 2001.
- KASS, G. V. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, v. 29, p. 119–127, 1980.
- KERZNER, H. *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*. 8<sup>th</sup> edition. Hoboken, NJ, USA: John Wiley & Sons, 2003. 449–485 p.
- KERZNER, H. *Using the Project Management Maturity Model: Strategic Planning for Project Management*. 2<sup>nd</sup> edition. Hoboken, NJ, USA: John Wiley & Sons, 2005. 41–47 p.
- KLING, D. V. *Um método para simulação de redes de atividades não-deterministas com recursos escassos*. Dissertação (Mestrado) — Instituto de Matemática, Universidade Federal do Rio de Janeiro, 2005.
- KOLISCH, R.; HARTMANN, S. Experimental investigation of heuristics for resource-constrained project scheduling: An update. *European Journal of Operational Research*, v. 174, n. 1, p. 23–37, October 2006.
- LEWIS, J. P. *Project Planning, Scheduling & Control*. 3<sup>rd</sup> edition. Columbus, OH, USA: McGraw-Hill, 2000.
- LOH, W. Y.; SHIH, Y. S. Split selection methods for classification trees. *Statistica Sinica*, v. 7, p. 815–840, 1997.
- LOH, W. Y.; VANICHESTAKUL, N. Tree-structured classification via generalized discriminant analysis (with discussion). *Journal of the American Statistical Association*, v. 83, p. 715–728, 1988.
- LUCKO, G. Reviving a mechanistic view of cpm schedules in the age of information technology. In: KUHL, M. E. et al. (Ed.). *Proceedings of the Winter Simulation Conference*. Orlando, FL, USA: ACM, 2005. p. 1533–1540.
- MCCLAVE, J. T.; SINCICH, T. *Statistics*. 10<sup>th</sup> edition. Upper Saddle River, NJ, USA: Prentice Hall, 2005.
- MEREDITH, J. R.; MANTEL, S. J. *Project Management: A Managerial Approach*. Hoboken, NJ, USA: John Wiley & Sons, 2002.
- METROPOLIS, N. *The Beginning of the Monte Carlo Method*. Los Alamos, NM, USA: Los Alamos Science, 1987.

- MIT. 2000 *How and Why Machines Work: Lecture 4*. Cambridge, MA, USA: MIT PSDAM LAB, 2002.
- MOORE, L.; CLAYTON, E. *Gert Modeling and Simulation: Fundamentals and Applications*. Melbourne, FL, USA: Krieger Pub Co, 1976.
- MORGAN, J. N.; SONQUIST, J. Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, v. 58, p. 415–434, 1963.
- NICHOLAS, J. N. *Project Management for Business and Engineering: Principles and Practice*. 2<sup>nd</sup> edition. St. Louis, MO, USA: Butterworth-Heinemann, 2004.
- ORR, A. D. *Advanced Project Management: A Complete Guide to the Key*. London, United Kingdom: Kogan Page Limited, 2007.
- PMBOK. *Um Guia do Conjunto de Conhecimentos em Gerenciamento de Projetos*. 3<sup>rd</sup> edition. Newtown Square, PA, EUA: Project Management Institute, 2004.
- QUINLAN, J. R. *C4.5: Programs for Machine Learning*. New York, NY, USA: Morgan Kaufmann, 1992.
- RUBINSTEIN, R. Y. *Simulation and the Monte Carlo Method*. Hoboken, NJ, USA: John Wiley & Sons Inc., 1981.
- SCHERER, W. T.; POMROY, T. A.; FULLER, D. N. The triangular density to approximate the normal density: decision rules of thumb. *Reliability Engineering and System Safety*, v. 82, p. 331–341, August 2003.
- STONEBURNER, G.; GOGUEN, A.; FERINGA, A. Risk management guide for information technology systems. *National Institute of Standards and Technology*, n. 800-30, p. 54, July 2002.
- VALLS, V.; BALLEST, F.; QUINTANILLA, S. Justification and RCPS: A technique that pays. *European Journal of Operational Research*, v. 165, p. 375–386, 2005.
- WAGNER, E. D. Enabling mobile learning. *EDUCAUSE Review*, v. 40, n. 3, p. 40–53, May-June 2005.
- WANG, J. X. *What Every Engineer Should Know About Decision Making Under Uncertainty*. Boca Raton, FL, USA: CRC Press, 2002.
- WIEST, J.; LEVY, L. *A Management Guide to PERT/CPM*. Upper Saddle River, NJ, USA: Prentice Hall, 1969.
- WITTEN, I. H.; FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2<sup>nd</sup> edition. Atlanta, GA, USA: Elsevier Science, 2005.
- WRIGHT, W. F. et al. Information technology, process reengineering and performance measurement: A balanced scorecard analysis of Compaq Computer Corporation. *Communications for The Association of Information Systems*, v. 1, n. 8, p. 1–61, February 1999.
- YANG, B.; GEUNES, J.; O'BRIEN, W. *Resource Constrained Project Scheduling: Past Work and New Directions*. Enschede, The Netherlands, April 2001.

ZAMANI, M. R. A high-performance exact method for resource-constraint project scheduling problem. *Computers and Operations Research*, Elsevier Science, n. 28, p. 1387–1401, 2001.

ZHU, G.; BARD, J. F.; YU, G. Disruption management for resource-constrained project scheduling. *Journal of the Operational Research Society*, Operational Research Society, n. 56, p. 365–381, 2005.