ESCRITÓRIO DE PROJETOS E AS ESTIMATIVAS E MÉTRICAS EM TEMPOS DE BIG DATA

Delermando BRANQUINHO Filho

**RESUMO**

# **INTRODUÇÃO**

Para uma empresa, ou mesmo para um empreendedor, a melhor coisa que pode acontecer é acertar as estimativas de seus projetos. Como essas estimativas são, na sua maioria, *a priori* e divididas em paramétricas e análogas, sobra muito pouco para ser feito sobre o assunto.

Na segunda guerra, com o objetivo de simular os melhores cenários ou com os menores riscos, o Método de Monte Carlo (MMC) ajudou o Projeto Manhattam. O objetivo naquele momento era resolver problemas de natureza probabilística (HAMMERSLEY, 1964). Os tempos mudaram, os avanços tecnológicos chegaram e novos algoritmos preditivos surgiram.

O objetivo geral desse trabalho é demonstrar que algoritmos em *Big Data* ou em *Data Science* (do inglês Ciência de Dados) podem resolver problemas em estimativas descritas no PMBoK® (PMBoK, 2015, p.338). As ferramentas e técnicas, como por exemplo, Análise de Sensibilidade, Análise do Valor Monetário Esperado e a Modelagem e Simulação serão abordados de acordo com as técnicas inovadoras da Ciência de Dados (*Data Science*).

Dentre os objetivos específicos deste trabalho podemos destacar a demonstração de resultados usando-se outras técnicas de simulação, como por exemplo *Bootstrap*, *Jacknife*, Árvore de Decisão, *Radom Fores*t e Regressão não Linear. O objetivo aqui é demonstrar outras técnicas mais simples para simulação de estimativas e métricas, pois o MMC pode criar essas estimativas, então é fundamental exercitar tais práticas. A simulação de Monte Carlo é pouco usada por causa de sua complexidade (JUNIOR 2001; SABBAG 2015). Em tempos de *Big Data* podemos substituir a complexidade pela simplicidade das novas tecnologias de estimação.

Na Seção dois abordamos as estimativas e métricas do PMBoK para a gestão de projetos. Na seção três iremos abordar alguns algoritmos e modelagens que cobrem alguns dos conhecimentos quantitativos, análises gráficas do monitoramento e controle dos resultados do trabalho de projeto.

# **CONDEITOS BÁSICOS EM GESTÃO DE PROJETOS**

## **Estimativas em Gestão de Projetos**

As estimativas mais comuns em gestão de projetos são as estimativas análogas, ou *Top-Down* (do inglês de cima para baixo), faz uso da similaridade entre projetos anteriores para determinar os valores para o projeto atual. Muito usada pela maioria dos Gerentes de Projeto por sua facilidade e rapidez. Em contrapartida, ela ocorre quando não temos muitos detalhes e/ou quando a principal restrição é temporal.

Outra estimativa *Top-Down* é a Estimativa por Ordem de Grandeza. Nesse caso, as informações que temos são mínimas e temos que nos basear na experiência de um profissional (Opinião Especializada). Segundo o PMBoK, essa estimativa é realizada durante o processo de iniciação e pode representar de -50% até +100% do esforço necessário. Essa margem é muito ampla e deixam muitas possibilidades, o que pode, invariavelmente, aumentar o risco.

As Estimativas Paramétricas são, em sua essência, uma relação estatística entre os dados históricos e outras variáveis do projeto atual, possibilitando assim, a criação de parâmetros, como por exemplo, a probabilidade do custo de uma atividade ficar entre determinados valores. Esse tipo de estimativa pode suportar um alto nível de precisão quando os pacotes de trabalho da EAP (Estrutura Analítica de Projetos), ou as atividades desses pacotes de trabalho são estimados, pois é mais fácil estimar o custo, tempo e recursos de uma atividade quando comparamos estimativas de elementos mais complexos. Aqui podemos perceber que as atividades ou pacotes de trabalho estimados agregam valor nas instâncias superiores da EAP, criando assim as estimativas desses níveis. Os gerentes de projetos, no entanto, devem equilibrar a precisão com o desequilíbrio das restrições do projeto na relação entre Escopo, Qualidade, Cronograma, Orçamento, Recursos e Risco (PMBoK, 2013).

A estimativa PERT (do inglês Program Evaluation and Review Technique), que quer dizer literalmente Avaliação do programa e técnica de revisão, é uma das mais usadas e assertivas técnicas de estimativa. Na realidade essa estimativa nasceu em 1950 de forma independente da CPM (do inglês Critical Path Method), conhecido como Método do Caminho Crítico. Ambas foram muito usadas em conjunto, o que permitiu dizer por muitos anos que seriam apernas uma técnica e não duas. Neste trabalho faremos um estudo apenas da PERT no sentido de melhorar o seu desempenho com algoritmos de Aprendizado de Máquina.

Se por um lado os projetos ajudam a garantir o sucesso dos empreendimentos os quais ele suporta, boas estimativas ajudam a garantir o sucesso dos projetos em que as mesmas suportam. A hipótese aqui é: se temos boas estimativas e que elas se aproximam de valores reais, então temos a vantagem competitiva de errar menos e ter mais sucesso em nossos projetos.

## **Métricas em Gestão de Projetos**

Quanto as Métricas, podemos destacar algumas onde o monitoramento é fundamental. A Produtividade é uma dessas métricas, onde o esforço planejado deve ficar próximo ao esforço realizado. Todas as métricas atuam nas seis restrições e podem afetar a produção.

A Linha de Base do Escopo é, de certa forma, uma métrica empírica definida nos primeiros momentos do planejamento do projeto. Como os projetos são planejados em Ondas Sucessivas (PMBoK, 2015, p.45), temos a vantagem de conhecermos mais sobre ele, mas temos também a desvantagem de sofrer alterações por conta de restrições externas não previstas entre outras.

A Qualidade sem dúvida é uma das mais controladas e monitoradas dentro dos projetos, pois é ela é uma métrica importante que ajuda a garantir as entregas. A correta definição e monitoramento dessa métrica permite detectar erros e anomalias que podem comprometer as seis restrições em projetos.

O Cronograma é a última métrica que iremos abordar neste trabalho. Essa métrica tem como base a comparação da data alvo com a data planejada. Podemos expandir as datas de conclusão planejadas para pacotes de trabalho, atividades ou qualquer outro ponto de término. Podemos ainda acrescentar em nossa entrada de dados, ou colunas de nosso dataset (arquivo de trabalho dos cientistas de dados), qualquer nível de granularidade.

Para finalizar a lista de métricas, abordaremos o Custo e a Margem Bruta. O primeiro podemos dividi-lo em Custo Real (CR), Valor Agregado (VA) e Valor Planejado. Já o segundo (VA) é uma constante monitorada para medir o real desempenho do projeto.

# **DATA SCIENCE NA GESTÃO DE PROJETOS**

## **Estimativas e Métricas em Big Data**

Medir é, além de obrigatório em gestão de projetos, algo comum na administração de modo geral, pois conforme a célebre fala de Eduard Deming:

“Não se ***gerencia*** o que não se ***mede***,não se ***mede*** o que não se ***define***,não se***define***o que não se ***entende***,e não há***sucesso***no que não se ***gerencia***”

**(William Edwards Deming)**

Relacionamos, nas seções anteriores, as estimativas e métricas e a sua importância fundamental para o controle dos projetos, sem esquecer que esse controle passa por medições. Poderíamos ter listado outras estimativas e métricas em gestão de projetos, mas para exemplificar o uso de algoritmos em Data Science ficamos com as principais e elas são suficientes. Em Ciência de Dados para Big Data, a medição são coletas de dados, a definição são os modelos, o entendimento o Aprendizado de Máquina (Machine Learning) entre outros algoritmos (BAEZA-YATES; RIBIERO-NETO 2013).

Data Science é uma expressão que tenta, que sintetiza o processo de transformação de dados (PROVOST; FAWCETT, 2013). Para isso é preciso criar modelos que descrevam as relações existentes entre os dados e suas influências. Sabendo como esse comportamento pode direcionar o sucesso dos empreendimentos, então podemos usaf análises preditivas sobre as nossas estimativas e métricas.

Um bom exemplo de uso de *Data Science* em projetos é a Árvore de Decisão. Na Figura 11-16 do PMBoK (PMBoK 2013. p.338) temos um exemplo de análise do valor monetário esperado. Essa análise pode ser feita usando a mesma técnica em *Data Science*, na Linguagem R (Linguagem de programação e computadores largamente usada em *data* Science) há vários pacotes onde passamos os parâmetros obtemos a mesma árvore exemplificada no PMBoK já com as melhores opções.

Em qualquer empreendimento, os patrocinadores e partes interessadas estão muito preocupados com as estimativas, principalmente porque essas estimativas são *a priori*, ou seja, a grosso modo podemos compará-las às premissas, pois ainda não se confirmaram. Como estimativas, há uma variação quando ocorre sua confirmação. Isso fica claro quando falamos do custo real (CR) ou do valor agregado (VA). Se pudermos prever os valores futuros dessas variáveis, entre outras, como por exemplo, ROI (Retorno do Investimento) e TCO (Custo Total de Propriedade), então o planejamento estará próximo à realidade.

## **Criando Modelos Preditivos**

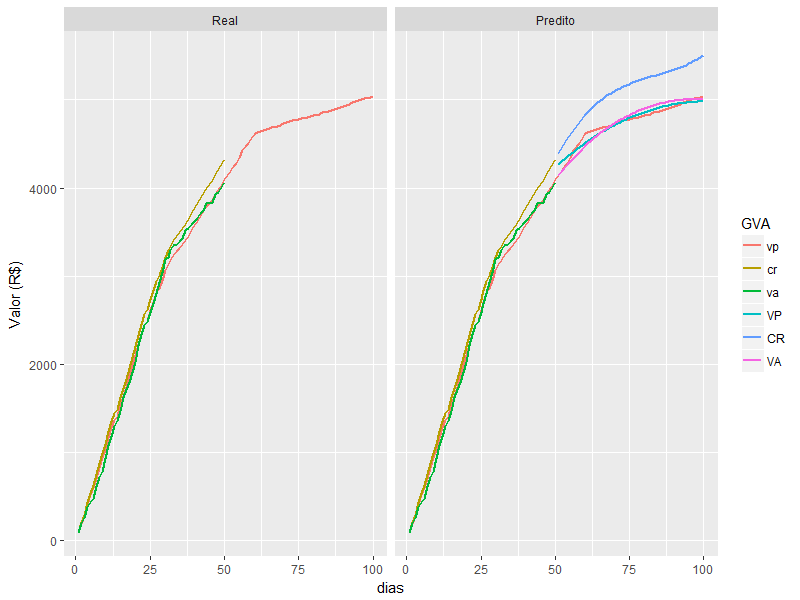
A capacidade de prever eventos futuros é, sem dúvida, uma funcionalidade almejada por todos os gestores em qualquer indústria. A predição de índices de desempenho futuro já está sendo estudada por vários pesquisadores (CHEN;CHEN;LIN, 2016). A ciência tenta explicar eventos do passado ao longo do tempo como forma de predizer esses eventos no futuro. Essa técnica é conhecida como *Time Series* (KIM, 2003). A Simulação de Monte Calo é uma das mais conhecidas, mas como dissemos anteriormente, muito pouco usada no dia-a-dia em projetos por causa de sua complexidade (BARRETO; ANDRADE, 2014).

Para exemplificar a possibilidade de uso de outros modelos, criaremos um cenário para demonstração desse tipo de predição sem envolver a Simulação de Monte Carlo. Usaremos o modelo linear generalizado (GLM do inglês *Generalized Linear Model*). Esse modelo é uma generalização da regressão linear tradicional e permite variáveis de resposta que possuem modelos de distribuição de erros diferentes de uma distribuição normal ou Gaussiana. O GLM generaliza a regressão linear, e em nosso caso, adicionaremos a Regressão Polinomial ou Quadrática, pois esse tipo de regressão, diferente da linear, permite representar os modelos de resposta curvilínea, já que nossa Curva-S (PMBoK, 2013. P.219) obedece esse fenômeno.

Na Figula-1 temos dois gráficos, no lado esquerdo representados os dados reais (Real), do lado direito os valores previstos (Predito). Ambos com 100 dias de duração para um determinado projeto. Os valores desses gráficos foram criados de forma sintética através de distribuições Gaussianas, levamos em consideração a relação ao gráfico da Figura 2-8 da página 38 do PMBok, onde a execução do projeto consome a maioria dos recursos, onde as caldas da curva ficam com as menores frequências.

As representações gráficas abaixo são complementares, sendo que o gráfico à direita completa o gráfico à esquerda com valores preditos em caixa alta na legenda (GVA).

**Figura 1 - Gestão do Valor Agregado (GVA) Real x Predito**



**Fonte: Elaboração própria**

Não se pode deixar de observar que apenas a curva do curso real está distante das demais. As curvas de predição do valor agregado e planejado seguem a tendência das curvas planejadas. Para prever os valores do gráfico Predito usamos a GLM com formulação quadrática na *Linguagem R*, poderia ser *Python, Weka, Rubi on Rails* ou qualquer outra linguagem de programação para Data Science.

A linha de comando na Linguagem R é bastante simples:

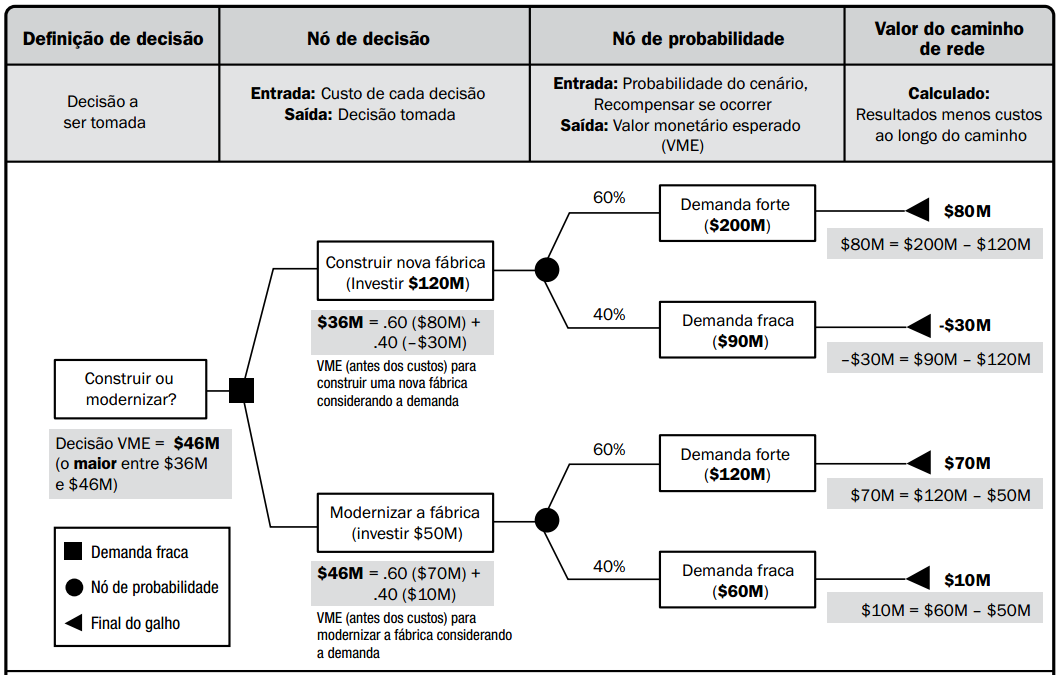
fit <- lm(df$valor ~ dias + I(dias^2))

Com esse modelo de previsão do GVA podemos visualizar diariamente, ou na frequência determinada pelo PMO, como será o desempenho do nosso projeto, além de receber avisos e alertas em caso de problemas. Além do GVA outros valores, lembrando dos índices e métricas, podem ser previstos usando esse modelo (GLM Polinomial), há outros modelos de regressão com finalidades outras que podem ser mesclados e usados para testar o melhor desempenho de um projeto com determinadas características, outro projeto com atributos diferentes pode exigir outros modelos. Será preciso testar alguns e medir a sua acurácia para selecionar o melhor. Isso nos remete ao fato de aprimorar os estudos e sistemas em *Big Data* para a Gestão de Projetos, especialmente quando falamos de Sistemas de Gerenciamento de Projetos (SGP).

Outro modelo para predição está intimamente ligado com a Gestão de Projetos descrito no PMBoK. Aliás o PMBoK está repleto de exemplos (PMBoK 2013, p.239) onde podemos usar *Machine Learning* (do inglês Aprendizado de Máquina) entre outros campos de estudo da Ciência da Computação e Inteligência Artificial, bem como a Estatística. Podemos destacar, em relação à Estatística, que gráficos como Histogramas, Diagramas de Dispersão, Diagramas de Pareto entre outros gráficos de controle, são oriundos de estudos estatísticos. O Aprendizado de Máquina evoluiu do estudo de reconhecimento de padrões e da teoria da aprendizagem computacional (MICHALSKI; CARBONELL; MITCHELL, 2013).

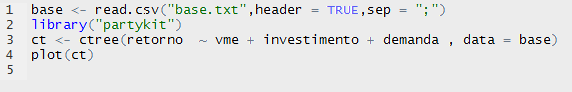
Na página 145 do PMBoK destacamos o gráfico do programa do processo de decisão (GGPD), além do gráfico de árvore, que em *Machine Learning* chamamos pelo mesmo nome em inglês de *Decision Tree*. Poderíamos exemplificar cada cálculo, cada ferramenta e técnica do PMBoK e associá-los a soluções, Modelos e Algoritmos em Inteligência Artificial. Para esse artigo vamos nos restringir a apenas mais um. A Análise do Valor Monetário Esperado (VME), conforme o próprio PMBoK “[...] é um conceito estatístico que calcula o resultado médio quando o futuro inclui cenários que podem ocorrer ou não (ou seja, análise em situações de incerteza).”. A Figura-2 mostra esse Diagrama da Árvore de Decisão.

**Figura-2 - Diagrama da Árvore de Decisão**



Fonte: Extraído do PMBoK Figura 11-16 (2013, p.338)

Apesar do exemplo didático do PMBoK ajudar o entendimento, na vida real as decisões comtemplam um número maior e variado de atributos a serem considerados nas decisões. Vamos usar esse mesmo exemplo usando *Machine Leaning*. O Algoritmo é o mesmo *Decision Tree*. As linhas de comando abaixo geram o diagrama da Figura-3.

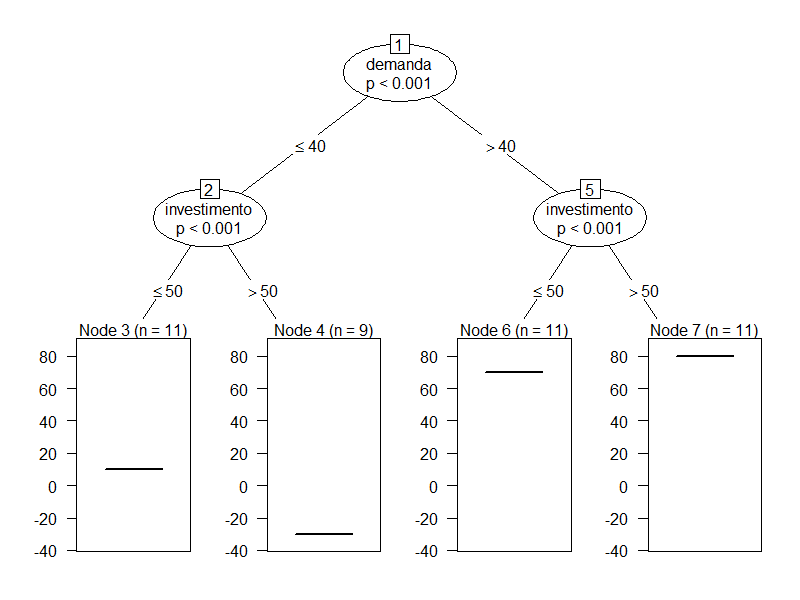


As variáveis *investimento* e *demanda* foram retiradas originalmente do problema descrito no diagrama anterior da Figura-2. A indicação do *p-value,* na Figura-3 abaixo*,* menor do que 0,001 é um indicativo estatístico clássico, também chamado de nível descritivo ou probabilidade de significância (ARSHA, 1988). Quando esse valor é igual ou superior a 0,05 significa que você não tem evidências suficientes para suportar suas crenças. Em nosso caso o *p-value* é muito pequeno, corroborando assim nossos resultados.

Os Valores Monetários Esperados no diagrama de árvore apresentado no PMBoK ficam entre $36 e $46, esses dados foram inseridos em um *dataset* (configuração de dados para processamento em *data* Science) contendo as probabilidades, os valores de modernização e construção, além dos valores demanda.

Em uma situação real, outros atributos podem estar presentes podendo passar de dezenas ou centenas (PROVOST; FAWCETT, 2013).

**Figura 3 - Diagrama em *Decision Tree* na Linguagem R**



**Fonte: Elaboração própria**

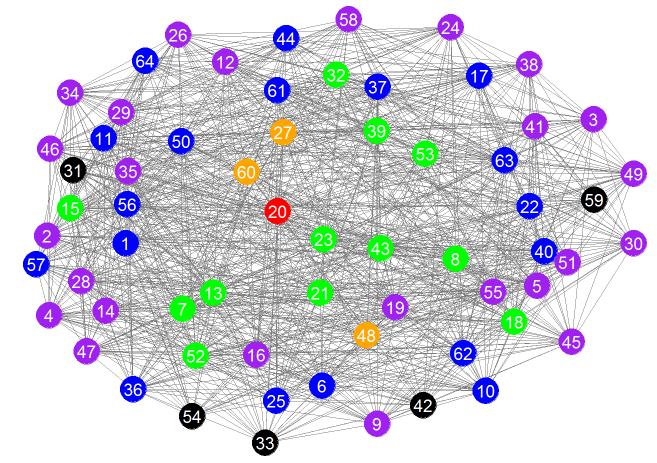
No diagrama acima observa-se os mesmos resultados no diagrama anterior, mas com uma facilidade de interpretação melhor. Com demanda maior do que $40 teremos um retorno de $80 quando o investimento for maior que $50, o inverso é: para uma demanda menor ou igual a $40, teremos prejuízo se o investimento for maior que $50.

Alguns podem pensar que o gerenciamento das partes interessadas poderia ficar de fora, mas há estudos avançados que agregam valor nesse gerenciamento com o uso de modelos e algoritmos em *Data Science*, muito além da Matriz RACI e das técnicas analíticas (CASTRO-HERRERA; CLELAND-HUANG, 2009; BARAN; JANTUNEN, 2004). É possível medir o nível de engajamento das partes interessadas e isso não é recente (BOHUS; HORVITZ, 2009).

Suponha que estamos procurando fazer uma análise de nossas partes interessadas para potenciais problemas oriundos da análise de risco. Precisamos entender os vários modelos classificatórios, como por exemplo, o grau de poder, interesse, influência e impacto. Segundo o PMBoK o “Modelo de Relevância descreve os tipos de partes interessadas com base no seu poder (capacidade de impor sua vontade), na urgência (necessidade de atenção imediata) e na legitimidade (seu envolvimento é apropriado) (PMBoK 2013. p.395).

Em *Data Science q*ueremos extrair uma métrica chamada centralidade de proximidade (*Closeness Centrality*). As métricas de centralidade tentam aproximar uma medida de influência de um indivíduo dentro de uma rede. A distância entre dois vértices é o caminho mais curto. A *Farness* (do inglês fadiga) de um determinado vértice *v* é a soma de todas as distâncias de cada vértice para *v*. Finalmente, a proximidade de um vértice *v* é o inverso da fadiga. Em outras palavras queremos saber quem influência quem dentro da rede.

**Figura 4 - Gráfico representativo do interesse, influência e poder dos *Stakeholders***



**Fonte: Elaboração própria**

No gráfico da Figura-4 a parte interessada de número 20 é aquela com o maior poder e influência para o projeto em questão. Deve ser usada em parceria com o gráfico da Figura 13-4 do PMBoK (PMBok, 2013. p.397) onde a informação de gerenciar com atenção deve ser anotada. É possível usar esse mesmo gráfico para determinar quais partes interessadas devem ser mantidas satisfeitas, informadas ou apenas monitoradas.

# **CONCLUSÃO**

## **Considerações Finais**

Conforme demonstrado, poderíamos suportar todas as métricas e estimativas, bem como a exploração, amostras, estatísticas dos gráficos de controle, *Benchmaking*, o Método do Caminho Crítico (MCC) contendo as variáveis de início mais cedo, mais tarde e término mais cedo e mais tarde, todos sendo submetidos a algoritmos e modelos com o objetivo de se obter o melhor cenário para o monitoramento e o controle do desempenho do trabalho.

Se levarmos em consideração que o MCC pode variar muito dentro de projetos de médio e grande porte, criar múltiplos cenários com suas dependências pode ser uma tarefa árdua. Podemos usar técnicas e modelos dentro das melhores práticas do PMBoK através de técnicas de amostragem, como por exemplo *Bootstrap* ou *Jacknife*, podemos gerar esses cenários para uma análise de riscos, sem ter que a necessidade de usar a Simulação de Monte Carlo, bem mais complexa e com os mesmos resultados (SNOEK; LAROCHELLE; ADAMS, 2012).

A diversidade de soluções em *Data Science* que resolve os problemas em gerenciamento de projetos é grande e vem aumentando a cada dia. Novos algoritmos estão sendo testados e construídos. Devemos pensar que a cerca de 20 anos atrás a Inteligência Artificial estava restrita a estudos acadêmicos, nas hoje ela invadiu as empresas, alterando suas estratégias e alterando os conceitos de tecnologias maduras, como por exemplo BI (*Business Intelligence*, do inglês inteligência de negócios) que agora se transforma em *Business Analytics* (LARSON; CHANG, 2016; CHEN; CHIANG; HL; STOREY 2012).

## **Trabalhos Futuros**

Com o avanço do *Big Data*, do *Data Science* e da Inteligência Artificial, os projetos serão mais desafiadores, pois o volume de dados e variáveis a serem controlados aumentará sistematicamente. Dados históricos agora ganham um papel fundamental para a melhoria das estimativas preditivas, pois não podemos confiar apenas nas medidas de tendência central como a média. Precisamos de uma maior assertividade nas decisões.

O ideal é que as empresas fornecedoras de sistemas para gerenciamento de projetos comecem a inserir em suas funcionalidades, elementos e dispositivos para predição, a exemplo do que foi feito neste artigo. Quanto maior for o número de funcionalidades e opções para o monitoramento e controle dos projetos, maior será o ROI do empreendimento.

Não podemos esquecer de mencionar os profissionais envolvidos no gerenciamento de projetos, e não estamos limitados apenas a equipe do projeto, mas para esse artigo consideramos que o treinamento é parte fundamental do amadurecimento da Ciência de Dados nas organizações, do entendimento à solução dos problemas, uma face nova colaborando para o aprimoramento e melhoria do desempenho.

**BIBLIOGRAFIA**

ARSHA, H Kuiper's. **P-value as a Measuring Tool and Decision Procedure for the Goodness-of-fit Test**, Journal of Applied Statistics, Vol. 15, No.3, 131-135, 1988.

BAEZA-YATES; RIBEIRO-NETO. ***Modern information retrieval*: The concepts and technology behind search**. New York: ACM Press, 2013.

BARAN, Eric; JANTUNEN, Teemu. **Stakeholder consultation for Bayesian decision support systems in environmental management**. Forest, v. 27, n. 35.6, p. 37.1, 2004.

BARRETO, Guilherme de A; ANDRADE, Marinho de A. **Estimação Paramétrica de Modelos Auto regressivos via Estatística Beyesiana e Simulação de Monte Carlo.** São Carlos. USP2014. Disponível em 04/07/2017

BOHUS, Dan; HORVITZ, Eric. **Learning to predict engagement with a spoken dialog system in open-world settings.** In: Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Association for Computational Linguistics, 2009. p. 244-252.

CASTRO-HERRERA, Carlos; CLELAND-HUANG, Jane. **A machine learning approach for identifying expert stakeholders.** In: Managing Requirements Knowledge (MARK), 2009 Second International Workshop on. IEEE, 2009. p. 45-49.

CHEN, Hong Long; CHEN, Wei Tong; LIN, Ying Lien. **Earned value project management: Improving the predictive power of planned value**. International Journal of Project Management, v. 34, n. 1, p. 22-29, 2016.

CHEN, Hsinchun; CHIANG, Roger HL; STOREY, Veda C**. Business intelligence and analytics: From big data to big impact**. MIS quarterly, v. 36, n. 4, 2012.

HAMMERSLEY, John Michael; HANDSCOMB, David Christopher. **Monte Carlo Methods**. Methuen; J. Wiley, 1964.

http://www.academia.edu/download/43503075/ESTIMAO\_PARAMTRICA\_DE\_MODELOS\_AUTO-REGRE20160308-12567-1yxvb1n.pdf

JORION, P. Value at Risk: **A nova fonte de referência para o controle do risco de mercado**. São Paulo: Bolsa de Mercadorias e Futuros, 1997.

JÚNIOR, DUARTE; MARCOS, Antonio. AM **Risco:** **definições, tipos, medição e recomendações para seu gerenciamento. Gestão de risco e Derivativos**. São Paulo: Atlas, 2001.

KIM, Kyoung-jae. **Financial time series forecasting using support vector machines**. Neurocomputing, v. 55, n. 1, p. 307-319, 2003.

LARSON, Deanne; CHANG, Victor. **A review and future direction of agile, business intelligence, analytics and data science**. International Journal of Information Management, v. 36, n. 5, p. 700-710, 2016.

MICHALSKI, Ryszard S.; CARBONELL, Jaime G.; MITCHELL, Tom M. (Ed.). **Machine learning: An artificial intelligence approach.** Springer Science & Business Media, 2013.

PMI. **PMBOK (®) Guide.** Ed. 5ª. Project Management Institute, 2015.

PROVOST, Foster; FAWCETT, Tom. **Data science and its relationship to big data and data-driven decision making.** Big Data, v. 1, n. 1, p. 51-59, 2013.

PROVOST, Foster; FAWCETT, Tom. **Data science and its relationship to big data and data-driven decision making.** Big Data, v. 1, n. 1, p. 51-59, 2013.

SABBAG, Omar Jorge; COSTA, Silvia Maria Almeida Lima. **Análise de custos da produção de leite: aplicação do método de Monte Carlo.** Extensão Rural, Santa Maria, v. 22, n. 1, p. 125-145, 2015.

SNOEK, Jasper; LAROCHELLE, Hugo; ADAMS, Ryan P. **Practical bayesian optimization of machine learning algorithms.** In: Advances in neural information processing systems. 2012. p. 2951-2959.