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, o que hoje podemos fundir os dois conceitos em uma Startup (do inglês começe), ou seja, de uma empresa iniciante no mercado, a melhor coisa que pode acontecer é acertar as estimativas em 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 são uma realidade e novos algoritmos preditivos surgiram. O objetivo geral desse artigo é demonstrar que outros algoritmos 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). Tentaremos demonstrar como essas ferramentas podem ser inovadas usando ferramentas de Big Data.

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 Forest e Regressão não Linear. O objetivo aqui é demonstrar outras técnicas mais simples para simulação de estimativas e métricas, pois o MMC se utiliza dessas estimativas, então é fundamental exercitar tais práticas. A simulação de Monte Carlo é pouco usada por causa de sua complexidade (JUNIOR 2001; SABBAG 2015).

# **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, 2015).

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 acertivas 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 tempos a vantagem competitiva de errar menos e ter mais sucesso em nossos projetos.

## **Métricas em Gestão de Projetos**

Quanto as Métricas, podemos destacar algumas que você vai precisar monitorar. A Produtividade é uma dessas métricas, onde o esforço planejado deve ficar próximo ao esforço realizado. Todas as estimativas mencionadas anteriormente 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 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 é uma constante monitorada para medir o real desempenho do projeto.

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

Medir é, além de obrigatório em gestão de projetos, algo comum na administração de modo geral, pois conforme 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 (do inglês Ciência de Dados) 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, usando 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 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, são como 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.

# **DATA SCIENCE NA GESTÃO DE PROJETOS**

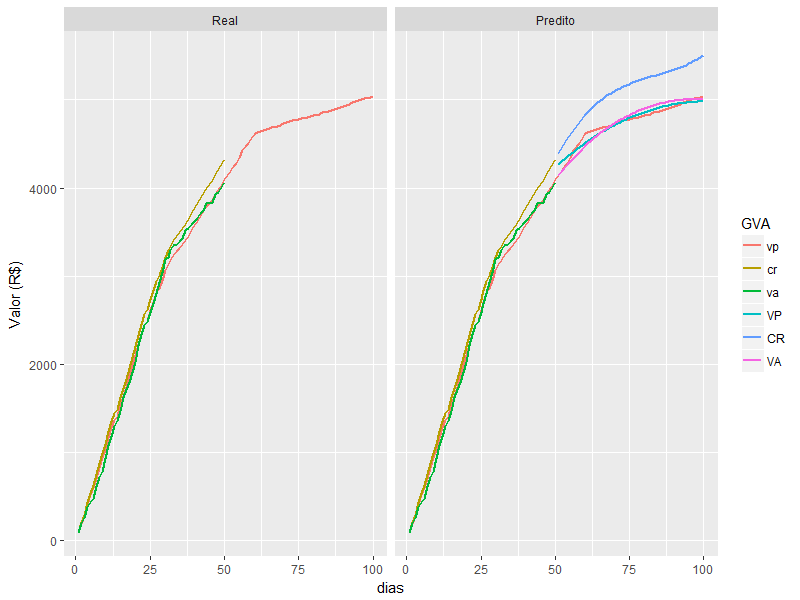
## **Criando Modelos Preditivos**

A capacidade de prever eventos futuros é, sem dúvida, uma funcionalidade almejada por todos os gestores em qualquer nível indústria. A predição de índices de desempenho futuro já está sendo estudada por vários pesquisadores (CHEN;CHEN;LIN, 2016). A ciência tenta explicar eventos ao longo do tempo como forma de predizer esses eventos no futuro. A Simulação de Monte Calo é uma das mais conhecidas, mas como dissemos anteriormente, muito pouco usada 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). Esse modelo é uma generalização da regressão linear tradicional. Esse modelo 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. As representações gráficas abaixo são complementares, sendo que o gráfico à direita completa o gráfico à esquerda com valores preditos.

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



**Fonte: Elaboração própria**

Para prever os valores do gráfico Predito, a GLM com formulação quadrática utilizamos a *Linguagem R*, poderia ser *Python* ou qualquer outra linguagem para Data Science. A linha de comando na Linguagem R é bastante simples:

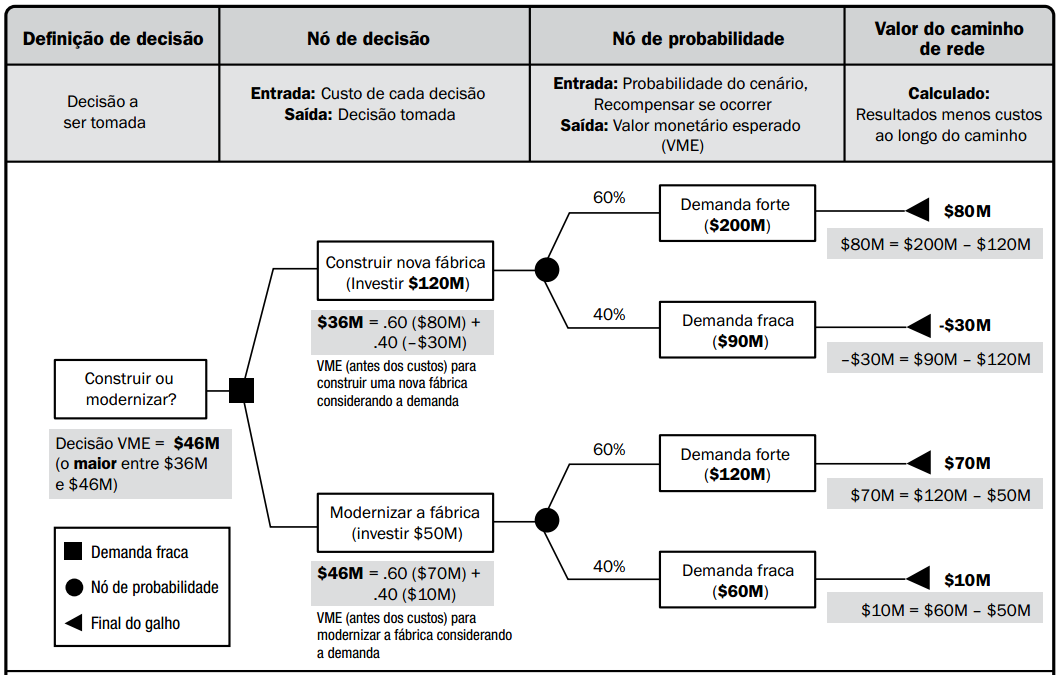
fit <- lm(df$valor ~ dias + I(dias^2))

Com esse modelo de previsão do GVA podemos visualizar diariamente, ou na frequência determinada pelo PMO, como será o desempenho do nosso projeto. Além do GVA, outros valores, índices e métricas podem ser previstos usando esse modelo (GLM Polinomial), há outros modelos de regressão com finalidades outras que podem ser mesclados ou 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.

Outro modelo para predição está intimamente ligado com a Gestão de Projetos e o 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 de *Decision Tree*. Poderíamos exemplificar cada cálculo, cada ferramenta e técnica do PMBoK e associá-los à soluções de Modelos e Algoritmos e 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 calculo 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: 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. Alguns anos atrás estive a frente de um projeto onde vários colaboradores deveriam ser alocados em uma determinada cidade. A premissa era de alugar os imóveis necessários para suportar a equipe, isso implica em dizer que serão vários imóveis. Vamos usar esse mesmo princípio usando *Machine Leaning*. O Algoritmo é o mesmo *Decision Tree*.

**BIBLIOGRAFIA**

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

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

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.

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

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.

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.

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.

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

<http://www.academia.edu/download/43503075/ESTIMAO_PARAMTRICA_DE_MODELOS_AUTO-REGRE20160308-12567-1yxvb1n.pdf>

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