



# Causa & Efeito

## DIGITAL

APRIMORAMENTO DE PROCESSOS | QUALIDADE EM SERVIÇOS | DATA ANALYTICS | PESQUISAS DE SATISFAÇÃO

## EDITORIAL

**QUANDO VEMOS** a repercussão que as técnicas de Data Analytics para RH estão tendo no mercado, ficamos muito felizes, pois as utilizamos há muito tempo. Em 1994, uma empresa de grande porte nos chamou porque tinham feito uma pesquisa de clima organizacional e acharam que os resultados da pesquisa não forneciam uma base adequada para a tomada de decisão, pois só a avaliação de cada questão não é suficiente se não estiver associada a satisfação. Aplicamos sobre os dados da pesquisa várias análises estatísticas para modelar o comportamento dos funcionários e definir padrões que permitiram um melhor gerenciamento do capital humano desta organização. Ficou evidente que as questões de avaliação mais baixas não eram as mais relevantes na satisfação do colaborador.

De lá para cá, as técnicas preditivas evoluíram e nós também, temos usado técnicas de regressão mais sofisticadas com resultados fantásticos, que permitem não só uma análise mais apurada, mas também prever o que irá ocorrer. Seja para analisar os dados relacionados aos colaboradores, aos clientes ou ao mercado.

## HR ANALYTICS

### A revolução dos dados chega ao RH!

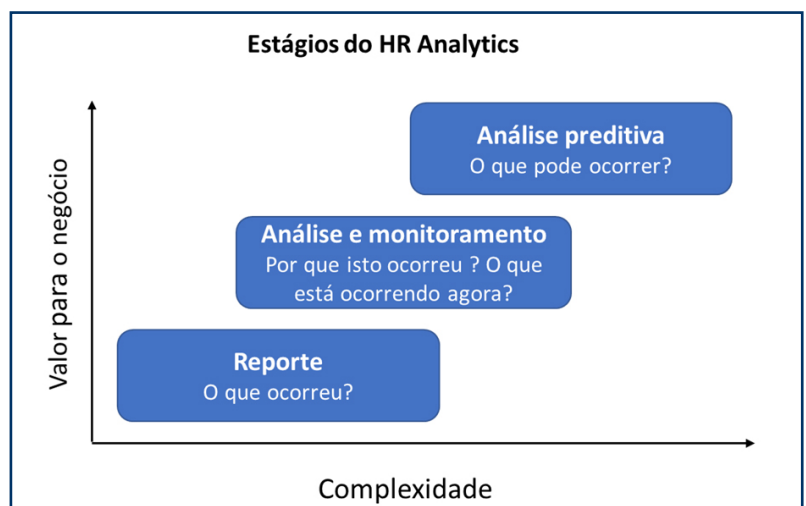
**ENTRE TODAS** as áreas de negócios, o RH ainda é a menos disruptiva, ou melhor era. A entrada do data analytics no RH está fazendo uma verdadeira revolução. Tornando o RH direcionado por dados, permitindo atuar de forma mais assertiva e estratégica, entendendo melhor a força de trabalho, como também prevendo e se antecipando a eventos como a perda de um talento importante.

HR analytics não é ferramenta ou software, trata-se de uma metodologia cuja abordagem é a utilização de dados aplicada a gestão de pessoas, tendo uma visão mais estratégica do papel do colaborador dentro da empresa. Novas ferramentas estatísticas e computacionais possibilitam uma nova abordagem na gestão de pessoas. **CONTINUA** ▣



Analytics não é mais uma boa ideia para o RH – agora é mandatório.

Josh Bersin



# HR ANALYTICS

**PODEMOS DIVIDIR** a maturidade de atuação do HR Analytics em diversos estágios:

- **O reporte:** Se descreve o que ocorreu em relatórios e dashboards, usualmente as ferramentas utilizadas são o Excel, alguma ferramenta de BI como o MS Power BI ou Tableau e algum ERP.

**Análítica:** Neste estágio, se trabalha com estatística descritiva, análise exploratória de dados e gráficos (Representação dos dados, medidas de tendência central, medidas de variabilidade e correlação).

Aqui vemos, por exemplo, que o absenteísmo vem aumentando no último semestre.

- **Análise e monitoramento:** O objetivo aqui é entender o que está ocorrendo e por que. Ferramentas utilizadas são softwares estatísticos como o R e ferramentas de pesquisa.

**Análítica:** Inferência estatística e análise de regressão (amostragem, inferência estatística, intervalo de confiança, teste de hipóteses, análise de regressão simples, análise de regressão múltipla, regressão logística e análise de network – grafos).

Aqui, respondemos qual o fator de maior impacto na última pesquisa de clima.

- **Análise preditiva:** O objetivo aqui é prever o que vai ocorrer, podendo gerar ações de antecipação de problemas.

**Análítica:** Análise de regressão simples, análise de regressão múltipla, regressão logística, regressão baseada em regras/árvores (Machine Learning), análise de fator e análise de cluster.

Ferramentas utilizadas são softwares estatísticos como o R.

Aqui respondemos a seguinte questão: Quais os funcionários com maior chance de deixar a empresa?

Aplicando os conceitos de HR Analytics nas informações de sua equipe, uma nova dimensão será disponibilizada, você irá se surpreender com a assertividade de suas decisões.

A Siqueira Campos atua no treinamento, consultoria e construção de modelos em HR Analytics.

## VOZ DO CLIENTE



Material e didática do instrutor. Trouxe situações comuns do nosso cotidiano e também situações específicas para nossa área de atuação que é validação analítica.

**Maybi Cristina Silva**

Pfizer – Métodos Estatísticos para Validação Analítica



O que mais gostei foram material, conhecimento do instrutor, exemplos dos exercícios.

**Iracema Malta**

Schneider Electric – Lean Seis Sigma

## ANÁLISE DE REGRESSÃO

### Novos modelos, resultados impactantes

**A ANÁLISE** preditiva é uma das estrelas da revolução dos dados, pois auxilia na previsão e tomada de decisão. Atualmente, existem novas famílias de técnicas de regressão que tem aumentado sensivelmente a precisão. Aqui temos um exemplo da comparação entre estas famílias.

O gráfico box-plot, apresentado a seguir, é o resultado de um trabalho acadêmico, com a participação de nosso diretor técnico\*, que tinha como objetivo criar um modelo preditivo utilizando regressão. Foram comparados 12 modelos distintos de regressão para prever o pH em um processo industrial.

Os dados do box-plot referem-se à comparação do erro de diversos modelos de regressão (linear, não-linear e baseados em árvores/regras) em um o processo com 32 variáveis. **CONTINUA** ▣

# ANÁLISE DE REGRESSÃO

**OS DADOS** originais foram divididos em dois grupos, dados de treino e dados de teste, os modelos foram gerados baseados nos dados de treino e validados no conjunto de dados de teste. Foi utilizado o método de validação cruzada com 10 repetições, com isto foram gerados 10 valores distintos para cada modelo.

Dois modelos baseados em árvores tiveram o melhor desempenho geral: RF (Random Forest) e Cubist. Os modelos Random Forest criam um conjunto de árvores de decisão – usando uma infinidade de árvores para corrigir o overfitting (overfitting, é um problema quando um modelo estatístico se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados).

Modelos cubistas criam árvores de decisão e, em seguida, as colapsam em regras. As regras ajudam a definir os valores previstos. No nosso caso, essa abordagem teve os menores erros medidos pelo RMSE, raiz do erro quadrático médio, que mede a diferença entre o valor real e o valor previsto. Quanto menor, melhor.

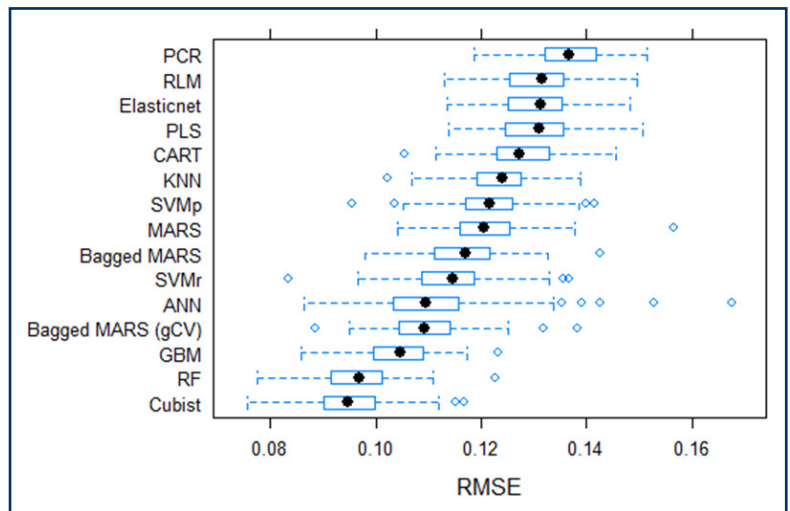
Os modelos baseados em árvores chamados também de Machine Learning, tem propiciado resultados, de um modo geral, muito superiores aos modelos de regressão linear. Todo o barulho feito com isto não é à toa, pois tem gerado impacto em várias áreas, como medicina, indústria, bancos, mercado financeiro, agricultura, recursos humanos, propiciando previsões muito mais assertivas como mostra este box-plot.

A Siqueira Campos atua no desenvolvimento de modelos preditivos, possibilitando, por exemplo, prevenção de churn, previsão de vendas, análise de mercado, modelagem de crédito entre outros.

\* Marco Siqueira Campos, Sharon Morris, Jeff Nieman, Derek Notes e Jose Zuniga

## MODELOS TESTADOS

- PCR – Principal Components Regression
- PCM – Pair Correlation Method
- Elasticnet – Elastic network
- PLS – Partial Least Squares
- ANN – Average Neural Networks
- KNN – K-Nearest neighbors
- SVMp – Support Vector Machines com kernel polinomial
- SVMr – Support Vector Machines com kernel radial
- MARS – Multivariate Adaptive Regression Spline
- Bagged MARS – Bagged Multivariate Adaptive Regression Spline
- Bagged MARS (gCV) – Bagged Multivariate Adaptive Regression Spline com gCV pruning
- CART – Classification and Regression Tree
- GBM – Stochastic Gradient Boosting Machine
- RF – Random Forest
- Cubist – Cubist



Modelos: ■ lineares, ■ não lineares e ■ baseado em regras e árvores.

## SIPOC

# Estabelecendo as fronteiras do processo

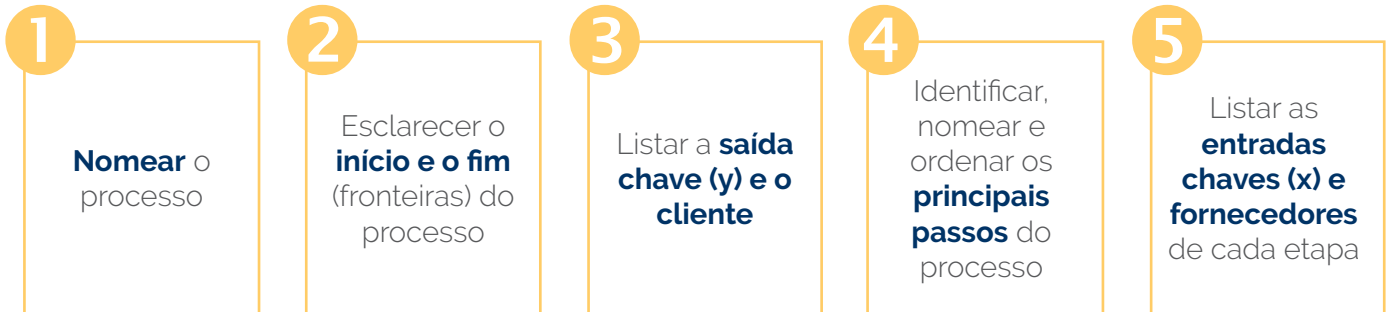
**PARA APRIMORAR** processos, um passo importantíssimo é o mapeamento do processo em questão, mas por mais incrível que parece, nem sempre as pessoas envolvidas no processo sabem claramente todos os elementos chave como, por exemplo, as fronteiras do processo, onde o processo inicia e onde termina.

Um mapa que facilita este entendimento, é o SIPOC, nome originado das iniciais em inglês Fornecedores, Entradas, Processo, Saída e Cliente. O propósito deste mapa é estabelecer as fronteiras do processo, ou seja, onde inicia e onde acaba o processo, e outros elementos críticos do processo como os principais passos, saída e o cliente.

- **Suppliers (Fornecedores):** quem fornece os recursos para o processo.
- **Inputs (Entradas):** recursos/informação que afetam o processo.
- **Process (Etapas do Processo):** principais etapas de um processo específico (usualmente, 5 a 7 etapas).
- **Outputs (Saídas):** entregas/resultados de cada etapa do processo.
- **Clients (Clientes):** quem recebe a saída do processo. **CONTINUA** ▣

O **SIPOC** fornece um ótimo resumo do processo, propiciando um entendimento do seu funcionamento e com isto possibilitando a melhoria do mesmo.

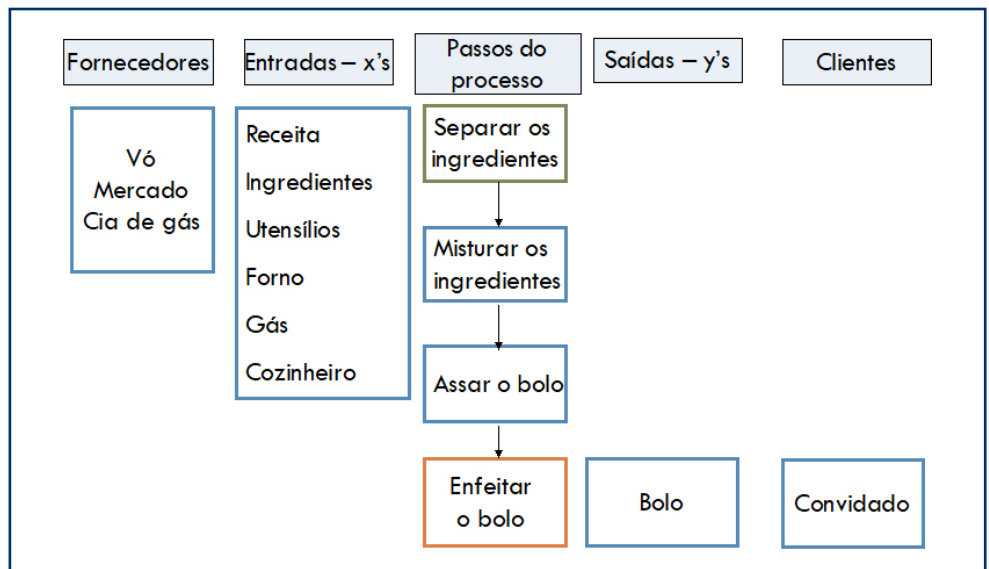
## PASSOS PARA A ELABORAÇÃO DE UM SIPOC



**ESTE MAPA** também permitirá uma melhor compreensão do todo e onde buscar dados para analisar o processo (variáveis de entrada e de saída).

Comece olhando seu processo através de um SIPOC, você vai se surpreender com sua simplicidade e utilidade.

*Exemplo de SIPOC – Bolo*



## ARTIGO

### Dados faltantes (missing data) e agora?

**DADOS FALTANTES** é um problema que afeta a todos que atuam com dados, de pesquisa científica a construção de modelos em organizações financeiras, estes nossos artigos descrevem alternativas da sua falta.

Continue lendo o artigo no LinkedIn, em duas partes:

- [PARTE 1](#)
- [PARTE 2](#)

