

Engenharia Web

IoT Aula 02

Método Científico

Professor Mestre: Adilson Lopes Khouri

3 de novembro de 2019

Método Científico

Cronograma

Aula	Conteúdo
Aula 01	Introdução IoT
Aula 02	Método Científico
Aula 03	EDA
Aula 04	Modelos
Aula 05	Seminários
Aula 06	Seminários
Aula 07	Seminários
Aula 08	Seminários

Qual o objetivo de uma equipe de dados?

- ▶ Extrair valor dos dados!
- ▶ O valor são: idéias, previsões, padrões
- ▶ Com esses insumos acima é possível para as empresas tomarem melhores decisões

Exercício

- ▶ Como tomar melhores decisões?

Como tomar melhores decisões?

- ▶ Usando dados e idéias obtidos anteriormente
- ▶ Mas de que forma? Há um processo para isso?

Método Científico

- ▶ É uma forma para extrair dados, analisar, criar hipóteses e tomar decisões
- ▶ Evita que vc cometa erros comuns, não garante que não haverá erros
- ▶ Citar exemplo do 6 sigma ou outras normas de controle de qualidade.

Método Científico

- ▶ Existe há aproximadamente 300 anos nas áreas científicas
- ▶ É um processo com passos para a validar uma determinada hipótese
- ▶ Garante um determinado rigor no processo
- ▶ É usado por cientistas do mundo inteiro para descobrir novos conhecimentos

Processo

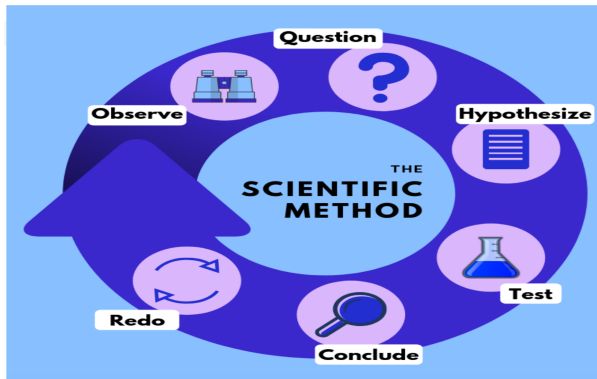


Figura: Exemplo de método científico - processo

Processo

- ▶ Observe: Queremos analisar algo... estudar algo... para tal, devemos observar esse evento. Tipicamente no mercado isso é um problema de negócio, no meio acadêmico um problema de pesquisa
- ▶ Question: Após observar/estudar um problema inicial devemos elaborar hipóteses sobre ele. (O que são hipóteses?)
- ▶ Hypothesize: Uma pergunta/comportamento de negócio que queremos validar/explicar

Processo

- ▶ Test: Devemos testar a hipótese e as possíveis predições, realizadas usando experimentos reproduzíveis
- ▶ Conclude: Analisar os resultados e tirar conclusões sobre eles, devemos validar a hipótese
- ▶ Redo: O experimento pode e deve ser refeito para garantir a consistência da teoria e validação da hipótese

Exemplo Simples

- ▶ Vamos supor que você tenta ligar a lâmpada de sua cozinha mas nada acontece (a luz não acende!)
- ▶ Como aplicar o método científico aqui?
- ▶ Podem descrever as etapas do processo?

Solução lâmpada

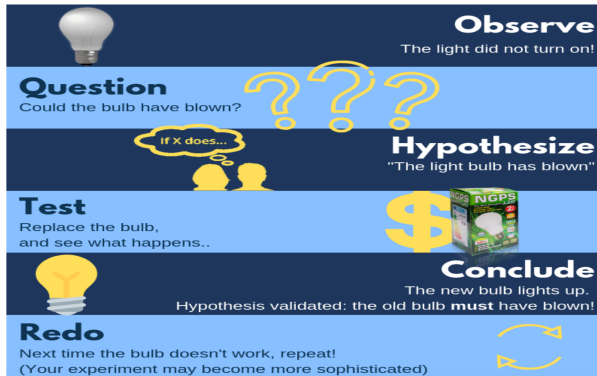


Figura: Solução lâmpada

Exemplo Real

- ▶ Vamos supor que você trabalhe em uma área de análise de crédito de um banco.
- ▶ Como aplicar o método científico aqui?
- ▶ Podem descrever as etapas do processo?

Exemplo Real

- ▶ Vamos supor que você trabalhe em uma área de análise de risco de transações de cartão de crédito.
- ▶ Como aplicar o método científico aqui?
- ▶ Podem descrever as etapas do processo?

Exemplo Real

- ▶ Vamos supor que você trabalhe seja um médico.
- ▶ Como aplicar o método científico aqui?
- ▶ Podem descrever as etapas do processo?

Método Cinetífico

- ▶ A ciência estuda como aumentar o conhecimento da humanidade (óbviamente em áreas que não conhecemos ;)
- ▶ Método científico é um método iterativo que padroniza o processo de conduzir experimentos, para ter resultados mais precisos, valiosos e confiáveis
- ▶ Para data science, IoT, podemos adaptar esse método para as empresas. Onde nosso objetivo é localizar informação de negócio.

Processo Adaptado

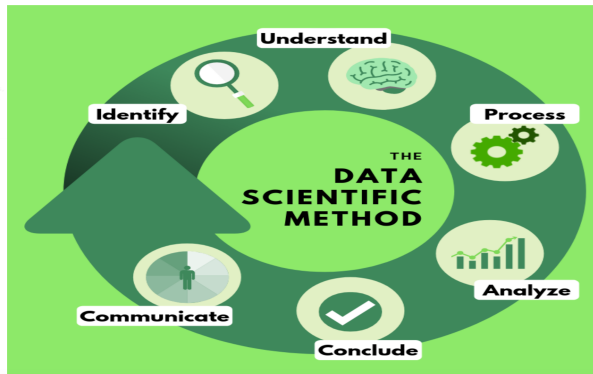


Figura: Exemplo de método científico - processo adaptado

Processo Adaptado

- ▶ Identificar: Nessa etapa identificamos a necessidade do negócio.
- ▶ Que tipo de questões devemos responder?
 - ▶ Que tipo de decisões devem ser tomadas com esses dados?
 - ▶ Podemos formular hipóteses sobre esses dados? Quais?
 - ▶ Quanto tempo temos para explorar?
 - ▶ Quais decisões o cliente pretende tomar com esses dados?
 - ▶ Qual seria o resultado ideal?
 - ▶ Como apresentar os resultados?

Processo Adaptado

Ferramentas

- ▶ Entrevista com o cliente
- ▶ Brainstorm

Processo Adaptado

- ▶ Entendimento: Ter um contato inicial com o conjunto de dados
- ▶ Que tipo de questões devemos responder?
 - ▶ Tamanho dos dados? Linhas x colunas; Gb;..
 - ▶ Quantos arquivos temos?
 - ▶ Há quantas fontes de informação?
 - ▶ Os dados de diferentes fontes tem formatação/tipagem igual?
 - ▶ Necessita de limpeza?
 - ▶ O que os campos (variáveis) significam?
 - ▶ Há outliers?
 - ▶ Há dados nulos?

Processo Adaptado

Ferramentas

- ▶ Workshops/brainstorming
- ▶ Jupyter Notebook
- ▶ Numpy and Pandas
- ▶ R
- ▶ seaborn/matplotlib

Processo Adaptado

- ▶ Processamento: Preparação dos dados, limpeza, padronização, tipagem, reshape de dados
- ▶ Que tipo de operações sobre os dados devemos realizar?
 - ▶ Criar um dataset final para apresentar ao modelo
 - ▶ Remoção de variáveis não úteis para o modelo (Como eu defino o que é ou não útil?)
 - ▶ Remoção de duplicações
 - ▶ Normalização
 - ▶ Tipagem de dados deve ser unificada
 - ▶ Decidir o que fazer com outliers
 - ▶ Tratar dados null

Processo Adaptado

Ferramentas

- ▶ Workshops/brainstorming
- ▶ Jupyter Notebook
- ▶ Numpy and Pandas
- ▶ R
- ▶ seaborn/matplotlib
- ▶ NLTK (Natural Language Processing Toolkit — another Python library)

- ▶ Análise: Estudos dos dados, entendimento de padrões, gráficos e relações.
- ▶ Aqui avaliamos valores de variáveis e suas distribuições, variáveis tomadas par a par, padrões sobre o tempo
- ▶ Validamos hipóteses definidas anteriormente, criamos modelos para prever valores.
- ▶ Atividades típicas nesse estágio
 - ▶ Se há dados temporais identificar sazonalidade e padrões no decorrer do tempo
 - ▶ Se há dados geográficos identificar padrões espaciais em conjunto com outras variáveis
 - ▶ Correlação
 - ▶ Variância
 - ▶ Estatística descritiva (correlação, desvio padrão, box plot)
 - ▶ Classificação de texto usando processamento de linguagem natural
 - ▶ Usar técnicas de machine learning (SVM, Redes Neurais, Xgboost, árvores, Random Forest)
 - ▶ Técnicas de redução de dimensionalidade

Processo Adaptado

Ferramentas

- ▶ Mysql/Postgresql/Cassandra/Mongo/Neo4J/Vertica/arquivos
- ▶ Jupyter Notebook
- ▶ JetBrains DataGrip (Pycharm IDE)
- ▶ R
- ▶ seaborn/matplotlib
- ▶ NLTK (Natural Language Processing Toolkit — another Python library)
- ▶ Scikit-Learn
- ▶ Tensor Flow
- ▶ Keras

Inspiração

“If it disagrees with experiment, it’s wrong. In that simple statement is the key to science.” — Richard P. Feynman

Processo Adaptado

- ▶ Conclusões: Nessa fase devemos definir quais foram as conclusões que chegamos na fase de análise.
- ▶ Comprovamos as hipóteses?
- ▶ Quais planos de ação podemos definir?
 - ▶ Validar se os achados permitem responder as questões originais
 - ▶ Aceitar ou rejeitar hipóteses
 - ▶ Definir as conclusões mais importantes para os clientes
 - ▶ Descrever tudo em linguagem que o cliente possa entender
 - ▶ Encontrar relações de pareto sobre os planos de ação
 - ▶ Definir recomendações para o cliente

Processo Adaptado

Ferramentas

- ▶ Powerpoint
- ▶ Apresentação para o cliente
- ▶ Pareto

Processo Adaptado

- ▶ Comunicação: O cliente precisa te entender!
- ▶ Cliente não sabe o que é TCP/IP!
- ▶ Cliente quer resultados/informação e planos de ação para ganhar dinheiro
 - ▶ Criar gráficos muito bonitos e simples para séries temporais (e.g. Grafana, Superset)
 - ▶ Criar gráficos muito bonitos e simples para dados espaciais (e.g. Leaflet JS, Plotly ou Superset)
 - ▶ Gráficos estatísticos (D3.js, Matplotlib or Seaborn)

Processo Adaptado

Ferramentas

- ▶ Grafana
- ▶ Apache Superset
- ▶ Psycopg2
- ▶ Angular
- ▶ Microsoft Office
- ▶ D3.js

Processo Adaptado

- ▶ Plano de ação: Certo, o cliente entendeu.. mas ele está usando?
 - ▶ Se o cliente está usando ótimo... cso contrário o que podemos mudar?

Processo

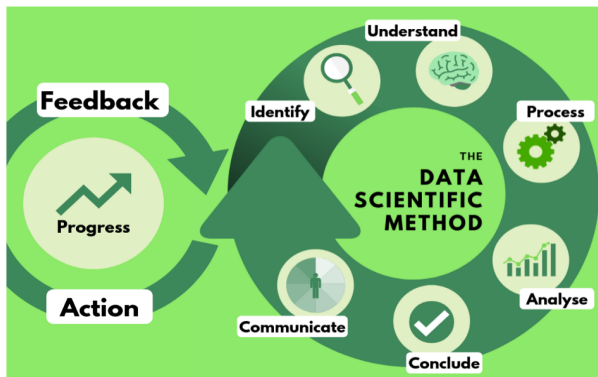


Figura: Exemplo de método científico com feedback - processo

Inspiração

“Without data, you’re just another person with an opinion.” — W. Edwards Deming

Exercício

- ▶ Kaggle
- ▶ UCI
- ▶ Posso usar outro dataset? SIM ;)

Dúvidas...

Alguma dúvida?

Contato

- ▶ E-mail: 0800dirso@gmail.com (alunos SENAC)
- ▶ E-mail: adilson.khoury.usp@gmail.com
- ▶ Phone: +55119444 – 26191
- ▶ [Linkedin](#) do professor
- ▶ [Lattes](#) do professor
- ▶ Slides no [GitHub](#) do professor