

Sabrina Araújo nº USP: 12566182

mac0102 - resumo palestra 29/04

A palestra do dia 29 de abril teve como principal assunto o tema ciência de dados. Inicialmente, o professor Roberto Hirata comentou sobre uma palestra feita pelo Kohavi, no qual era discutido o funcionamento da ferramenta de pesquisa Bing e a relação com dados. A partir disso, os conceitos sobre ciência de dados foram apresentados.

O método no qual um cientista e um engenheiro utilizam em uma pesquisa é diferente. O cientista formula hipóteses, faz experimentos, analisa resultados, se não der certo, reformula as hipóteses. Já o engenheiro, define um problema, analisa os requisitos, faz um brainstorm, faz um protótipo da solução e analisa para fazer o produto.

Nesse contexto, o cientista de dados formula uma questão, obtém os dados, explora os dados, modela os dados e reporta os resultados.

O cientista da computação não "respeita os dados", apenas há a entrada e a saída. "Dados são como carne para passar em um processador". Além de que os cientistas da computação não são acostumados a fazer perguntas e querem imprimir o maior número possível de um 'erro', por exemplo. Já os cientistas "normais" querem apenas o suficiente e significativo para as medidas que serão usadas.

Assim, diante de todos os conceitos passados de ciência de dados, é possível concluir que é importante ser criativo com os dados que estão à disposição.