



Como começar a aprender Estatística e Ciência de dados.

Autor: Thiago Marques



Introdução

Existem muitos cursos gratuitos e muito bons no Coursera, edX, Datacamp, Data Science Academy, assim como pagos também, comece por aí.

Há pouco tempo no LinkedIn eu falei sobre a grade de Estatística e Data Science que montei para estudar de casa com cursos totalmente gratuitos, assim que eu pedi demissão do meu primeiro emprego.

Grade data Science/estatística:

<https://www.linkedin.com/feed/update/urn:li:activity:6470289860669304832>

Nesse outro Post, eu falo um pouco como foi essa transição depois da faculdade, indico alguns caminhos possíveis a seguir:

Segue o link : <https://www.linkedin.com/pulse/reflex%C3%A3o-thiago-marques/>

Explicando um pouco como vocês vão usar a Estatística na Ciência de dados:

A Estatística por si só, não é nada, contudo aliada a uma área de conhecimento ou alguma ciência, se torna uma ferramenta poderosíssima de tomada de decisões!

Amostragem, Estatística Descritiva e Probabilidade:

Você precisa aprender o básico, como coletar os dados, de forma que essa amostra seja representativa da sua população, como resumir os dados, como se estuda o comportamento deles, que distribuição de probabilidade esses dados são provenientes, para que você consiga elaborar hipóteses que serão confirmadas ou refutadas lá na frente, na chamada inferência Estatística.

Inferência Estatística:

Você irá expandir os resultados que obteve por meio de uma amostra, por isso é importante que a amostra seja representativa da população, ou seja resguarde em termos de proporções as características da população, que nada mais é que a coleção de dados , que possuem determinadas características, as quais você quer estudar.

Modelagem Estatística:

Você formulará modelos, que nada mais são que uma redução da realidade, os quais poderão estimar valores futuros(fazer predições), baseados no histórico dos seus dados, observados alguns pressupostos.

Resumindo, você precisará de:

Estatística descritiva, Amostragem, Probabilidade, Inferência Estatística, Modelos lineares, Regressão logística, Árvores de decisão, Random forest, gradient boosting machine, entre outras técnicas.

Um ponto importante é casar a linguagem de programação com Estatística, siga estudando R ou Python na minha opinião, até hoje, são as melhores e com maiores comunidades para usar em Data Science, aliada as soluções Spark e Hadoop .

Focar em cursos de R, Python, SQL, NoSQL, Hive, Pig, Spark, Hadoop, sempre voltados a análise de dados.

Assim como se aprofundar em TI e Estatística é necessário, o bom Cientista de dados, tem que ser um verdadeiro camaleão! Se adaptar a área de negócios a qual ele está envolvido, de modo que a mesma pode variar constantemente, principalmente se você trabalha com consultoria, estudar as regras de negócio é fundamental!

Não só isso, como se comunicar com profissionais de outras áreas, que sacam mais que você do negócio!