

# Data Science 101

Os segredos para começar na área

# Quem é Déborah

- Ciência da Computação
- Se dando muito mal nas disciplinas
- Descobre a linha de comando e os mainframes
- Ganhar prêmios e descobre que talvez seja boa nisso

**Terminando a graduação**

# TCC

- "Escolha como tema algo que você quer trabalhar no futuro"
- Gosto de consultoria porque tenho que sempre aprender algo novo, Data Science é a área de Computação mais próximo a isso então vou fazer sobre isso

# TCC

- Simplesmente peguei um dataset e comecei a analisar (sem saber nada)
- Sofri bastante
- No meio disso descobri uma competição de Machine Learning na Microsoft Imagine
- Participei e acabei ganhando o Machine Learning Award :)

# Microsoft Imagine



# Um pouco sobre o Machine Learning Award

- O objetivo principal era divulgar o Azure ML Studio (ferramenta para criar modelos de predição arrastando blocos)
- Isso permite que a gente crie modelos sem saber necessariamente implementar os algoritmos



# Um pouco sobre o Machine Learning Award

- Poder criar os modelos de forma simples me liberou para trabalhar outros pontos do projeto
  - Que dados eu tinha?
  - O que eu poderia extrair desses dados?
  - Como eu poderia entregar valor a partir dos dados?

# Um pouco sobre o Machine Learning Award

- A única restrição era que fosse um jogo, e os dados eram analytics de jogadores fictícios
- O jogo tinha uma arte inicial que parecia ser pra crianças, resolvi focar nisso
- Criei um jogo para ajudá-las a aprender Matemática através do monitoramento do progresso no jogo

**Ganhei o Machine  
Learning Award, já sou  
uma data scientist?**

(pra mim **não**, porque  
ainda **não** tinha um  
cliente)

[1]

em Data Science muitas vezes você mesmo vai ter que  
ser responsável pelo seu aprendizado

[1]

**em Data Science muitas vezes você mesmo vai ter que  
ser responsável pelo seu aprendizado**

e isso aumenta a complexidade das coisas

[1]

**em Data Science muitas vezes você mesmo vai ter que  
ser responsável pelo seu aprendizado**

e isso aumenta a complexidade das coisas  
(mas é legal)

[2]

para progredir a gente precisa se propor a fazer algo  
que não tá 100% apto a fazer



[2]

para progredir a gente precisa se propor a fazer algo  
que não tá 100% apto a fazer

e considerar o cenário onde isso pode dar errado

[2]

**para progredir a gente precisa se propor a fazer algo  
que não tá 100% apto a fazer**

e considerar o cenário onde isso pode dar errado

escolher um tema onde eu era totalmente leiga foi arriscado, mas eu tava  
ok com uma possível falha

**Deu tudo certo,  
apresentei e me formei!**

**Na verdade...**

# O que aconteceu

- Comecei o TCC com um tema relacionado a satisfação de usuários em bancos
- Durante o TCC apareceu o Machine Learning Award e competi
- Ganhei um ML, fui para Seattle, tive mentoria com o pessoal de lá e resolvi mudar o tema do TCC
- Mudei e apresentei o TCC no período seguinte

**Deu tudo certo,  
apresentei e me formei!**

**Ok, e agora?**

# E agora?

- Senti a necessidade de algo mais formal
- Ainda não tinha certeza se queria fazer mestrado ou não
- Precisava de algo pra testar essas duas coisas:
  - Nanodegree de Deep Learning da Udacity



[3]

**é preciso encontrar formas de testar suas hipóteses**

com o Nanodegree descobri que preciso de uma estrutura mas ainda  
gosto de ser responsável pelo meu progresso

**Mestrado!**

# Meu mestrado

- Posso ter essas duas coisas
- Processamento de Linguagem Natural + Visualização de dados <3
- Aumentar a colaboração entre pesquisadores de áreas diferentes

**Tá, mas como eu começo?**

# Como começar?

- Como você gosta de aprender?
- Como foram suas experiências de aprendizado até hoje?
- De que experiência você gostou mais? Por quê?
- *Meu* ponto de vista: tente aprender sozinha, fuçar o que tem de conteúdo na internet, pegar um conjunto de dados e fazer uma análise
  - **Teste** se você quer mesmo entrar na área ou não

**Sei o básico, e agora?**

**Show your work**

# Show your work

- Talvez uma das coisas mais importantes que aprendi até hoje



# Show your work

- Talvez uma das coisas mais importantes que aprendi até hoje
- Qual é seu trabalho dos sonhos? Com o que você gostaria de estar trabalhando hoje?

# Show your work

- Talvez uma das coisas mais importantes que aprendi até hoje
- Qual é seu trabalho dos sonhos? Com o que você gostaria de estar trabalhando hoje?
- Comece a fazer isso por conta própria. **Agora**

# Show your work

- Talvez uma das coisas mais importantes que aprendi até hoje
- Qual é seu trabalho dos sonhos? Com o que você gostaria de estar trabalhando hoje?
- Comece a fazer isso por conta própria. **Agora**
- **Como?**

# Lei de Parkinson

# Lei de Parkinson

"O trabalho se expande de modo a preencher o tempo disponível para a sua realização"

# Lei de Parkinson

- É sério!
- Se você definir que só tem 30 min pra fazer y você VAI fazer y nos 30min

**Fiz y, e agora?**

**Show your work**

MOSTRE PARA AS PESSOAS O QUE VOCÊ FEZ



# Show your work

- Minha experiência
  - Escrevi um texto e mandei pro freeCodeCamp, eles gostaram e o texto foi publicado
  - Continuei escrevendo mais textos :)

# Show your work

- Minha experiência
  - É um mar de rosas?
    - Não
    - Você se sente vulnerável, recebe elogios mas também recebe críticas
    - E uma crítica dói bastante (se não for construtiva)
    - Mas tudo bem, a gente supera :)
    - E fica mais forte

**De novo, atenção:**

MOSTRE PARA AS PESSOAS O QUE VOCÊ FEZ

**Conseguindo clientes**

# As pessoas precisam te encontrar

- Tenha informações sobre você online (de preferência um site seu)

# As pessoas precisam te encontrar

- Tenha informações sobre você online (de preferência um site seu)
- Fala e mostre seu trabalho para outras pessoas

# As pessoas precisam te encontrar

- Tenha informações sobre você online (de preferência um site seu)
- Fala e mostre seu trabalho para outras pessoas
- **Não é garantia** que só isso vai te fazer conseguir clientes, mas as chances aumentam



# Como consegui meu primeiro projeto

# Como as pessoas me encontraram

- Quando eu estava fazendo o Nanodegree não tava entendendo como usar o TensorFlow
- E também não encontrava explicações que ajudassem
- Decidi que ia aprender e escrever um post pra ajudar outras pessoas também

## Featured



Déborah Mesquita in freeCodeCamp

Apr 9 · 12 min read



# Big Picture Machine Learning: Classifying Text with Neural Networks and TensorFlow

Read more...



2.1K

28 responses



# Como as pessoas me encontraram

- O texto acabou viralizando e chamou atenção no Medium
- Do nada começaram a chegar propostas e convites no LinkedIn

# Propostas que recebi

- Fiquei em dúvida entre duas propostas:
  - Projeto para uma empresa de Nova York sobre detecção de fraudes de cartão
  - Projeto para uma empresa de comunicação para situações de emergência

# Propostas que recebi

- Fiquei em dúvida entre duas propostas:
  - Projeto para uma empresa de Nova York sobre detecção de fraudes de cartão
  - Projeto para uma empresa de comunicação para situações de emergência
- Qual escolher?

# O básico sobre o desenvolvimento de habilidades

# Desenvolvimento de habilidades

- Se a gente aceita um desafio muito acima da nossa capacidade atual a gente simplesmente não sai do lugar



# Desenvolvimento de habilidades

- Se a gente aceita um desafio muito acima da nossa capacidade atual a gente simplesmente não sai do lugar
- Em vez de simplesmente decidir fazer algo fora da sua zona de conforto, escolha fazer algo um pouco acima do seu nível atual

# Desenvolvimento de habilidades

- Se a gente aceita um desafio muito acima da nossa capacidade atual a gente simplesmente não sai do lugar
- Em vez de simplesmente decidir fazer algo fora da sua zona de conforto, escolha fazer algo um pouco acima do seu nível atual
- Se a tarefa for muito acima das suas habilidades atuais você vai se sentir sobrecarregada e seu progresso vai ser mínimo

# Fraude de cartão

- Eles já tinham um produto funcionando
- Eles já tinham muitos dados
- O objetivo deles era bem definido (criar um modelo para detectar fraudes de cartão)

# Comunicação de Emergência

- Seria uma prova de conceito, o projeto era uma inovação
- Eu teria que definir que dados coletar
- Eles ainda não sabiam muito bem o que queriam, ia ser minha tarefa dar utilidade aos dados

# Fraude de cartão

- Eles já tinham um produto funcionando
- Eles já tinham muitos dados
- O objetivo deles era bem definido (criar um modelo para detectar fraudes de cartão)

Um pouco acima do meu nível na época

# Comunicação de Emergência

- Seria uma prova de conceito, o projeto era uma inovação
- Eu teria que definir que dados coletar
- Eles ainda não sabiam muito bem o que queriam, ia ser minha tarefa dar utilidade aos dados

Muito acima do meu nível na época

# Detecção de fraudes de cartão no ecommerce

- Escolhi esse projeto (com medo, mas fui)

# Detecção de fraudes de cartão no ecommerce

- Escolhi esse projeto (com medo, mas fui)
- A importância de deixar claro quem você é quais são suas habilidades
  - Eles leram o texto e viram como eu me comunicava

# A importância de conhecer suas habilidades

# 0 meu perfil

- Eu não me vejo ganhando uma competição no Kaggle, por exemplo
  - Meu background em Matemática não é muito bom (nem me interessa)
- Mas me vejo ajudando a implementar soluções utilizando dados
  - Sou boa em escrever
  - Gosto de pensar na estratégia e no projeto como um todo
  - Adoro quando preciso aprender algo novo pra resolver um problema



# T-shaped Data Scientist

**"[...] uma característica definidora dos cientistas de dados é  
a amplitude de suas habilidades – sua capacidade de fazer sozinho ao  
menos versões de nível de protótipo de  
todas as etapas necessárias  
para obter novas idéias ou criar produtos de dados"**

Analyzing the Analyzers by Marck Vaisman, Sean Murphy, Harlan Harris (2013)

"Nós também achamos que os cientistas de dados mais bem sucedidos são aqueles com experiência substancial e profunda em pelo menos um aspecto da ciência dos dados, seja estatística, big data ou business communication"

Analyzing the Analyzers by Marck Vaisman, Sean Murphy, Harlan Harris (2013)

# T-shaped professional

- Termo promovido desde o início da década de 1990
- O "T" representa a amplitude de habilidades, com ênfase em uma área representada pelo traço vertical do T
- Os profissionais em forma de T podem trabalhar mais facilmente em equipes interdisciplinares do que aqueles com menos amplitude

# T-shaped professional

- A ciência dos dados é um campo inerentemente colaborativo e criativo
- Um cientista de dados pode trabalhar com pessoas que possuem um conjuntos de habilidades sobrepostas para que os projetos de dados sejam concluídos de maneiras inovadoras

**"Ao esclarecer suas especialidades  
outros podem entender mais rapidamente a sua experiência"  
(e quais suas maiores contribuições no projeto)**

Analyzing the Analyzers by Marck Vaisman, Sean Murphy, Harlan Harris (2013)

**Entender suas habilidades também pode te ajudar a  
escolher como vai ser seu caminho de aprendizado**

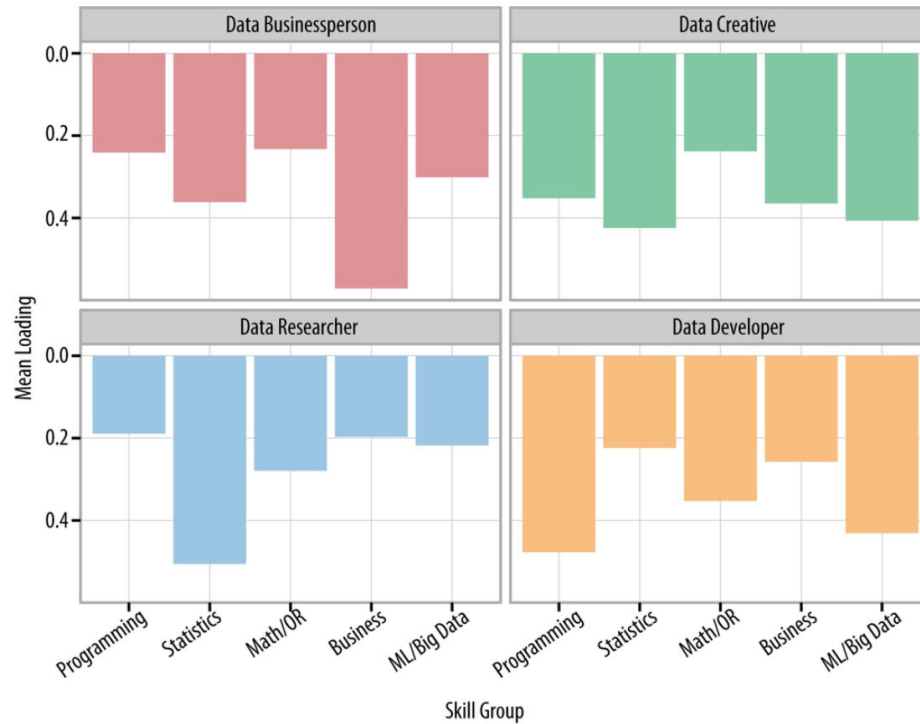


Figure 4-2. Mean Skill Group loadings for survey participants categorized into four Self-ID Groups.



**Voltando pro projeto**

# Como foi participar do projeto

- Estimei que terminaria em 1 mês
- Durou 3 meses :)
- Eles gostaram do resultado e do meu trabalho
- Agora estou trabalhando com eles para implementar o projeto na plataforma

*"Thanks Deborah. It's a pleasure working with you.  
Your quick grasp in understanding our product and what we're  
looking for is truly impressive and is making this a very enjoyable  
project to work on."*

**Resumindo!**

# Resumindo!

- Faça coisas e mostre seu trabalho para outras pessoas
- Em vez de simplesmente decidir fazer algo fora da sua zona de conforto, escolha fazer algo um pouco acima do seu nível atual
- Conheça suas habilidades e deixe isso guiar seu caminho e suas escolhas

**Obrigada!**

# Meu contato

deborahmesquita.com

<https://medium.com/@dehhmesquita>

@dehhmesquita