

# MINERAÇÃO DE DADOS

Thiago Marzagão<sup>1</sup>

<sup>1</sup> marzagao.1@osu.edu

## VISÃO GERAL DA DISCIPLINA

# ROTEIRO

- aplicações
- mercado de trabalho
- logística do curso
- preview do conteúdo

# APLICAÇÕES

- recommender systems
- detecção de fraude
- inteligência artificial
- mercado financeiro
- lingüística

# RECOMMENDER SYSTEMS



NETFLIX

# RECOMMENDER SYSTEMS

www.amazon.com

Apps news programming dataScience posts academia random stuff entertainment CADE CÉREBRO

Recommendations for You, Thiago

The screenshot shows a row of recommended books for Thiago Marzagão. From left to right, the titles are:  
1. "Statistical Analysis & Data Mining Applications" by Robert Nisbett, John Elder, Gary Miner  
2. "Learning pandas" by Michael Hyatt  
3. "Head First Java" by Bruce Harreld, David Cobb, et al.  
4. "Python Data Science Essentials" by Alfonso Gravano, Fabio Zuliani  
5. "STATISTICS DONE WRONG: THE MONTY HALL PROBLEM" by Alan Beaulieu  
6. "UNDERSTANDING MACHINE LEARNING: FROM THEORY TO PRACTICE" by Sanjoy Dasgupta, Herbert A. Hauptmann  
7. "Test-Driven Development with Python" by Harry J. N. Percival  
8. "Foundations of Machine Learning" by Mehdi Mohri, Afshin Rostamizadeh, and Yishay Mansour

See all your recently viewed items >

Back to top

Get to Know Us

Waiting for www.amazon.com...

Make Money with Us

Amazon Payment Products

Let Us Help You

# RECOMMENDER SYSTEMS

www.netflix.com/browse

Apps news programming dataScience posts academia random stuff entertainment CADE CÉREBRO

NETFLIX Browse DVD Search Thiago

### Top Picks for Thiago

### Trending Now

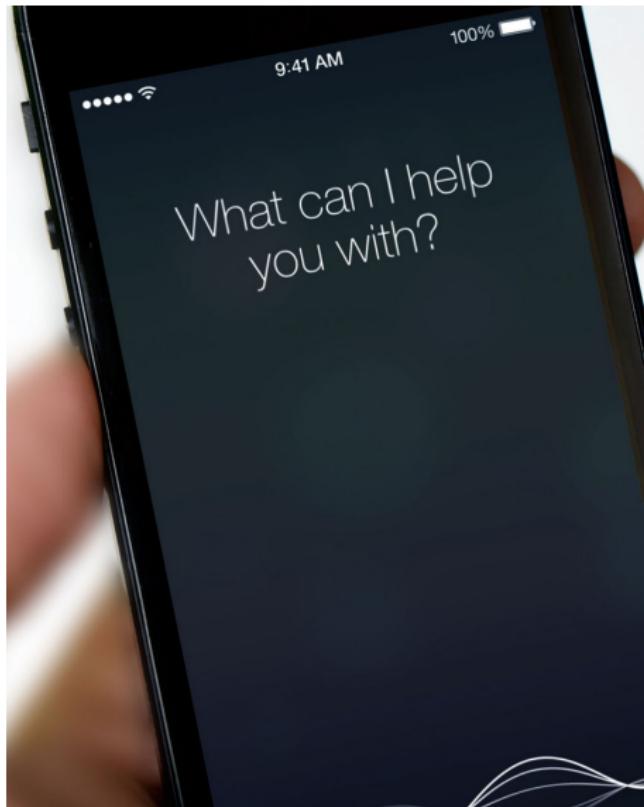
### Watch It Again

Waiting for cdn0.netfliximg.net...

# DETECÇÃO DE FRAUDE



# INTELIGÊNCIA ARTIFICIAL



# INTELIGÊNCIA ARTIFICIAL



## MERCADO FINANCEIRO

**Página inicial | Comunidades e Coleções | Data | Autor | Orientador | Título | Assunto | Tipo de documento | Ajuda**

**BDM >**  
**Monografias de Graduação >**  
**Administração >**

Please use this identifier to cite or link to this item: <http://bdm.unb.br/handle/10483/10732>

Files in This Item:			
File	Description	Size	Format
<a href="#">2014_SarahSabinodeFreitasMarcelino.pdf</a>		1,8 MB	Adobe PDF <a href="#">View/Open</a>

**Título:** Formação de portfólio por meio de Máquinas De Suporte Vetorial  
**Autor(es):** [Marcelino, Sarah Sabino de Freitas](#)  
**Orientador(es):** [Albuquerque, Pedro Henrique Melo](#)  
**Assunto:** [Máquinas de Suporte Vetorial](#)  
[Portfólios \(Ações\) - formação](#)  
[Benchmarking \(Administração\)](#)

**Data de apresentação:** 20-Nov-2014  
**Data de publicação:** 28-Jul-2015  
**Referência:** MARCELINO, Sarah Sabino de Freitas. Formação de portfólio por meio de Máquinas De Suporte Vetorial. 2014. 98 f., il. Monografia (Bacharelado em Administração)—Universidade de Brasília, Brasília, 2014.  
**Resumo:** A presente pesquisa teve como objetivo replicar a metodologia de Máquinas de Suporte Vetorial proposta por Fan e Palaniswami (2001) no contexto brasileiro da formação de portfólio. O SVM foi então utilizado para verificar se o uso de Máquinas de Suporte Vetorial na formação de portfólios de fato contribui para que o retorno seja superior ao de um benchmark do mercado, sendo que o ativo escolhido para tal comparação foi o

# LINGÜÍSTICA

# MERCADO DE TRABALHO

www.kdnuggets.com/2013/02/salary-analytics-data-mining-data-science-professionals.html

Apps news programming dataScience posts academia random stuff entertainment CADE CÉREBRO

**KDnuggets™ Data Mining, Analytics, Big Data, and Data Science** Subscribe to **KDnuggets News** | Follow @KDnuggets | Contact search KDnuggets Search

Data Mining Software | News | Jobs | Academic | Companies | Courses | Datasets | Data Mining Course | Education | Meetings | Polls | Webcasts

**Most Viewed**

- 1. [TensorFlow Disappoints - Google Deep Learning falls shallow.](#)
- 2. [7 Steps to Mastering Machine Learning With Python.](#)
- 3. [5 Best Machine Learning APIs for Data Science](#)
- 4. [R vs Python for Data Science: The Winner is ...](#)
- 5. [A Statistical View of Deep Learning](#)
- 6. [Amazon Top 20 Books in Databases & Big Data.](#)
- 7. [9 Must-Have Skills You Need to Become a Data Scientist](#)

**Most Shared**

- 1. [The different data science roles in the industry](#)

**Table 1: 2013 Annual Income/Salary by Employment type**

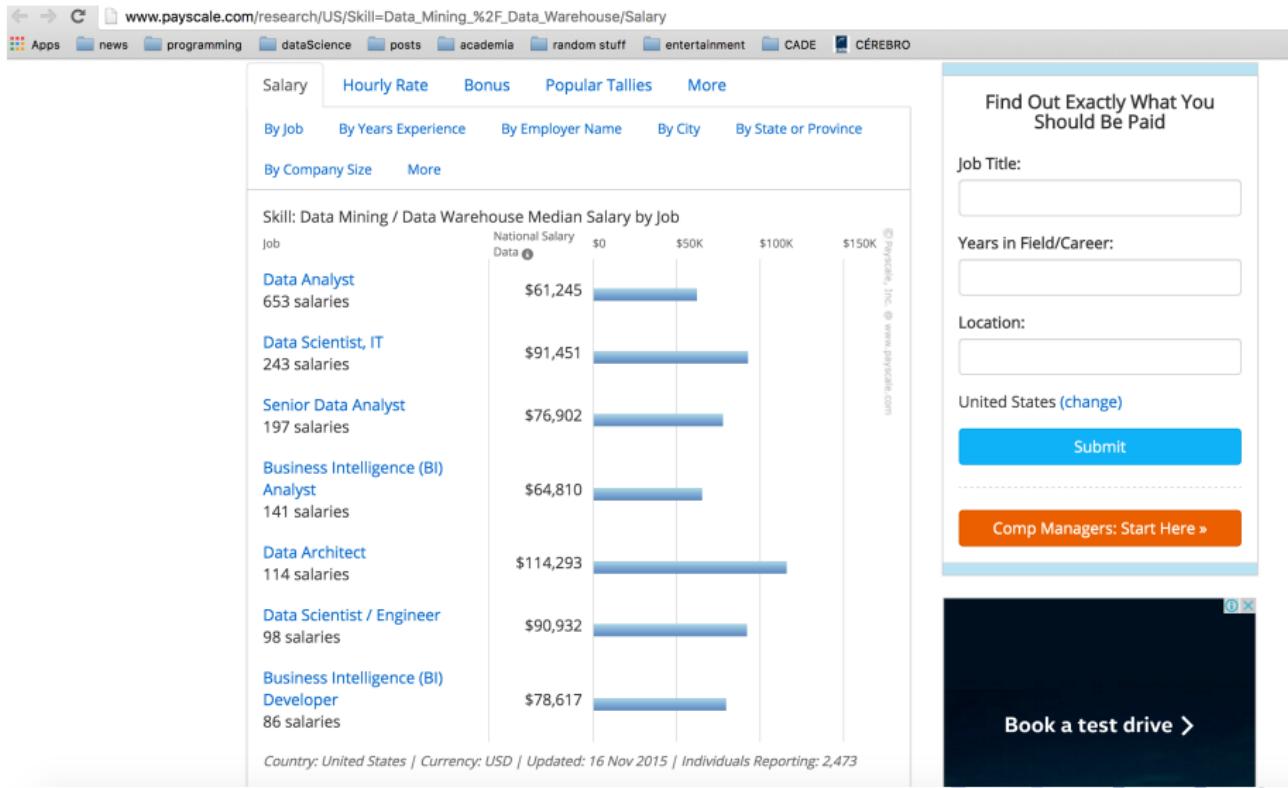
Employment	2013 Avg. Salary	2012 Avg. Salary	% Change	2013 Count	2012 Count
Self-employed	136.4	105.8	29.0%	25	26
Company	111.3	101.6	9.5%	284	375
University/Academia	89.6	63.2	41.9%	42	49
Government	72.5	71.8	1.0%	12	14
Student	33.4	34.1	-2.0%	19	17
Unemployed/Retired	20.0	48.3	na	1	6
All	105.2	94.1	11.9%	383	487

For estimating average salary we used mid-point of the range, 20K for \* \$180" range.

We note also that there was only 1 self-reported Unemployed/Retired, vs 6 in 2012.

In Table 2 we break down the data by region. Since the numbers of self-employed and government data scientists are relatively small, we grouped them with industry and university, respectively. Students and unemployed were excluded.

# MERCADO DE TRABALHO



# PREVIEW DO CONTEÚDO

- algoritmos de mineração
  - regressão linear
  - classificação
    - árvores de decisão
    - máquinas de suporte vetorial (SVM)
  - clusterização
  - mineração de textos

# PREVIEW DO CONTEÚDO

- computação
  - programação
    - Python
    - SQL
  - computação em nuvem
  - trabalhando c/ dados sujos

# LOGÍSTICA DO CURSO

- site do curso: <http://thiagomarzagao.com/teaching/mineracao>
- no site: programa, slides, datasets, links p/ leituras
- comunicação: Slack
- avaliação: trabalho final, individual