

APRESENTAÇÃO PARA O GRUPO DE SMC 15/05/2018

Leonardo Morales Belluzzi

Trocando Ideias

O objetivo da apresentação
é trocar experiências, um
bate papo.

Um pouco sobre mim

Academia, profissional,
ideias e projetos.



Os tópicos principais da apresentação

Justificativas

Data-minig

Aplicações

80/

Go

MAR

MAY

JUN

01

2005

2006

2007



facebook

[login](#) [register](#) [help](#)

Welcome to Facebook!

Email:

Password:

[Login](#)

[Register](#)

Secure | **Standard**

Facebook is an online directory that connects people through social networks.

The site is open to a lot of places, but not everywhere yet. We're working on it.

You can use Facebook to:

- Look up people at your school.
- See how people know each other.
- Make groups and events with your friends.

[Login](#)

[Register](#)

[about](#) [jobs](#) [advertise](#) [press](#) [terms](#) [privacy](#)

a Mark Zuckerberg production

Facebook © 2006

Connect with friends and the world around you on Facebook.



See photos and updates from friends in News Feed.



Share what's new in your life on your Timeline.



Find more of what you're looking for with Facebook Search.

Sign Up

It's free and always will be.

Birthday

Why do I need to provide my birthday?



Female



Male

By clicking Sign Up, you agree to our [Terms](#), [Data Policy](#) and [Cookies Policy](#). You may receive SMS Notifications from us and can opt out any time.

Sign Up

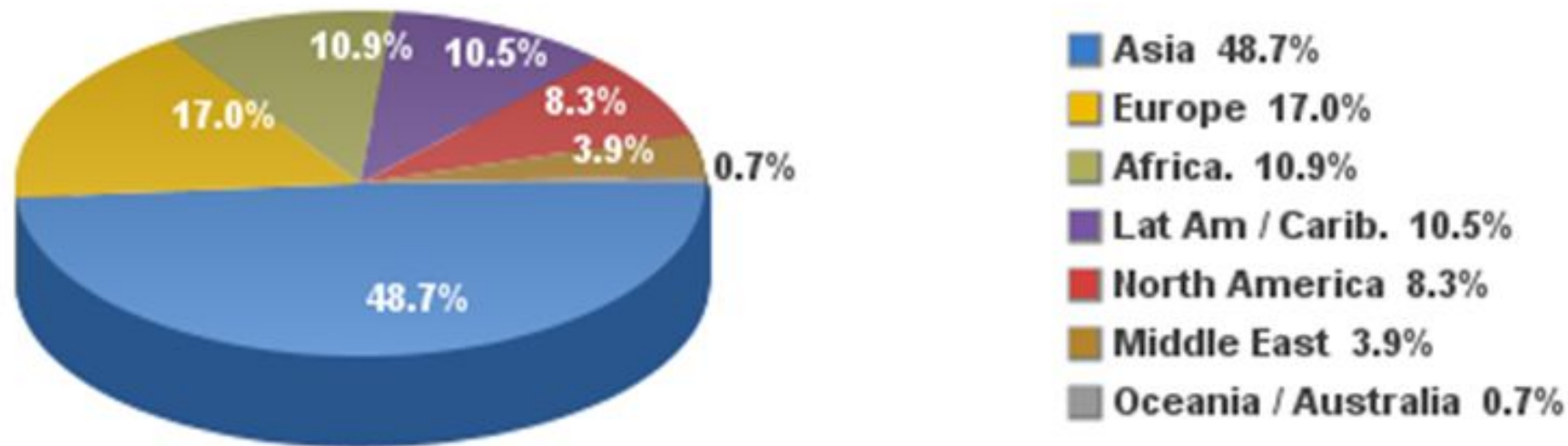
[Create a Page](#) for a celebrity, band or business.

Tamanho é documento?

- 14.3 Trilhão de webpages disponíveis na internet.
- 48 Bilhão webpages indexada pelo Google.
- 14 Bilhão - webpages indexada pelo Bing.
- 672 Exabytes - 672,000,000,000 Gigabytes(GB) de dados disponíveis para download.
- 43,639 Petabytes - Total World-wide tráfego de dados, em 2013.
- Mais de 9,00,000 Servers - do Google, o maior servidor do mundo.
- Over 1 Yotta-byte - Total de dados armazenado na Internet.
- 1 Yotta-byte = 1,000,000,000,000,000,000,000,000 Bytes!

Dados estimados do cite The Archive (ARCHIVE, 2016).

Internet Users in the World by Regions - December 31, 2017



Source: Internet World Stats - www.internetworldstats.com/stats.htm

Basis: 4,156,932,140 Internet users in December 31, 2017

Copyright © 2018, Miniwatts Marketing Group



Quem?

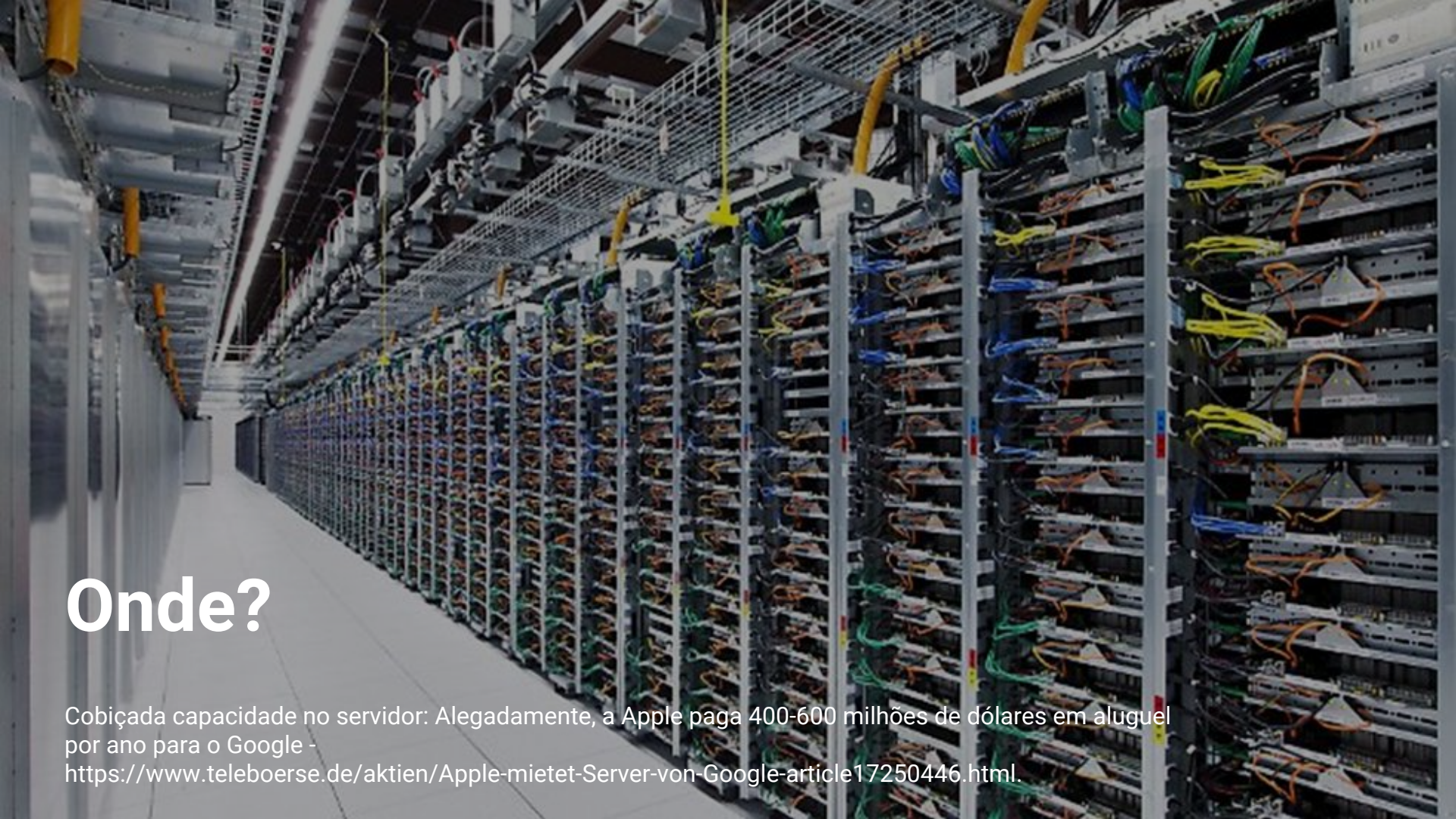
Moby & The Void Pacific Choir - Are You Lost In The World Like Me? -
<https://www.youtube.com/watch?v=VASywEuqFd8>.

A wide-angle photograph of a vast data center. The room is filled with rows of server racks, some of which are illuminated with blue and green lights. The ceiling is high and features a complex network of steel beams and pipes. The floor is a light-colored, polished surface. The overall atmosphere is one of a large, modern, and technologically advanced facility.

Onde?

Data center -

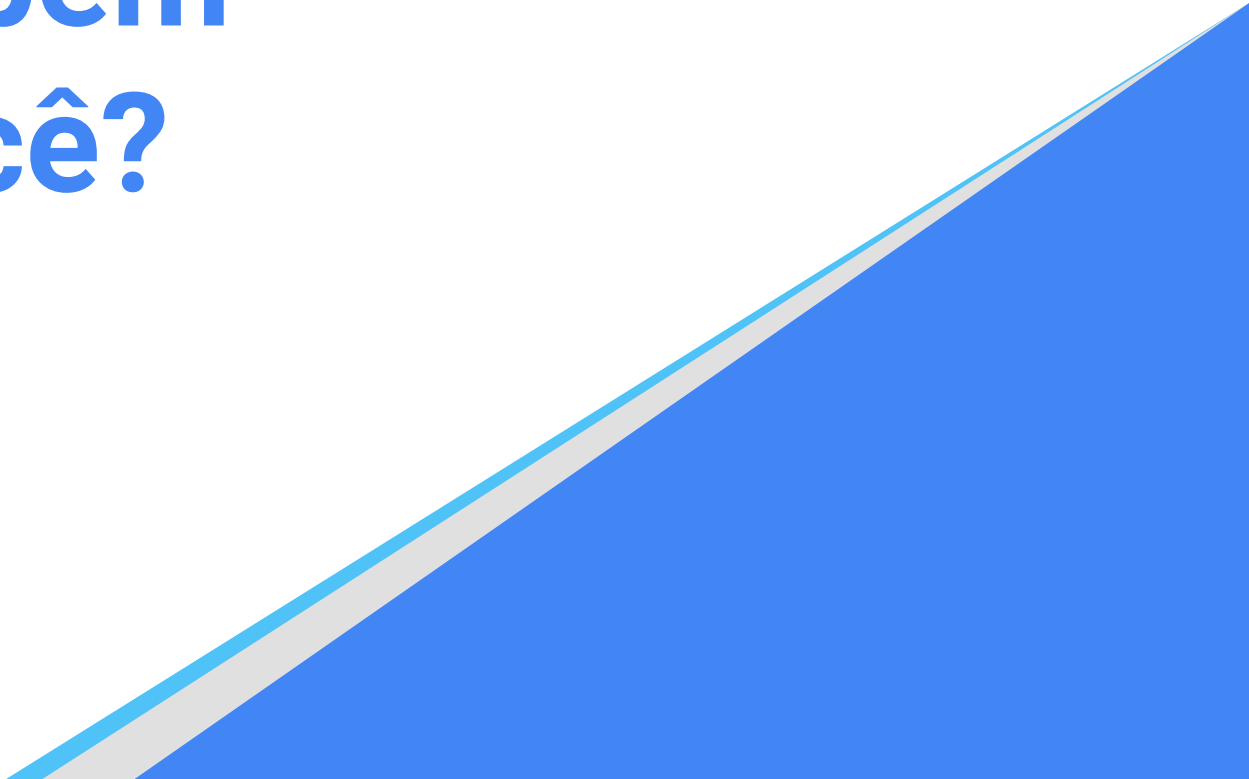
<http://www.dailymail.co.uk/sciencetech/article-2219188/Inside-Google-pictures-gives-look-8-vast-data-centres.html>.

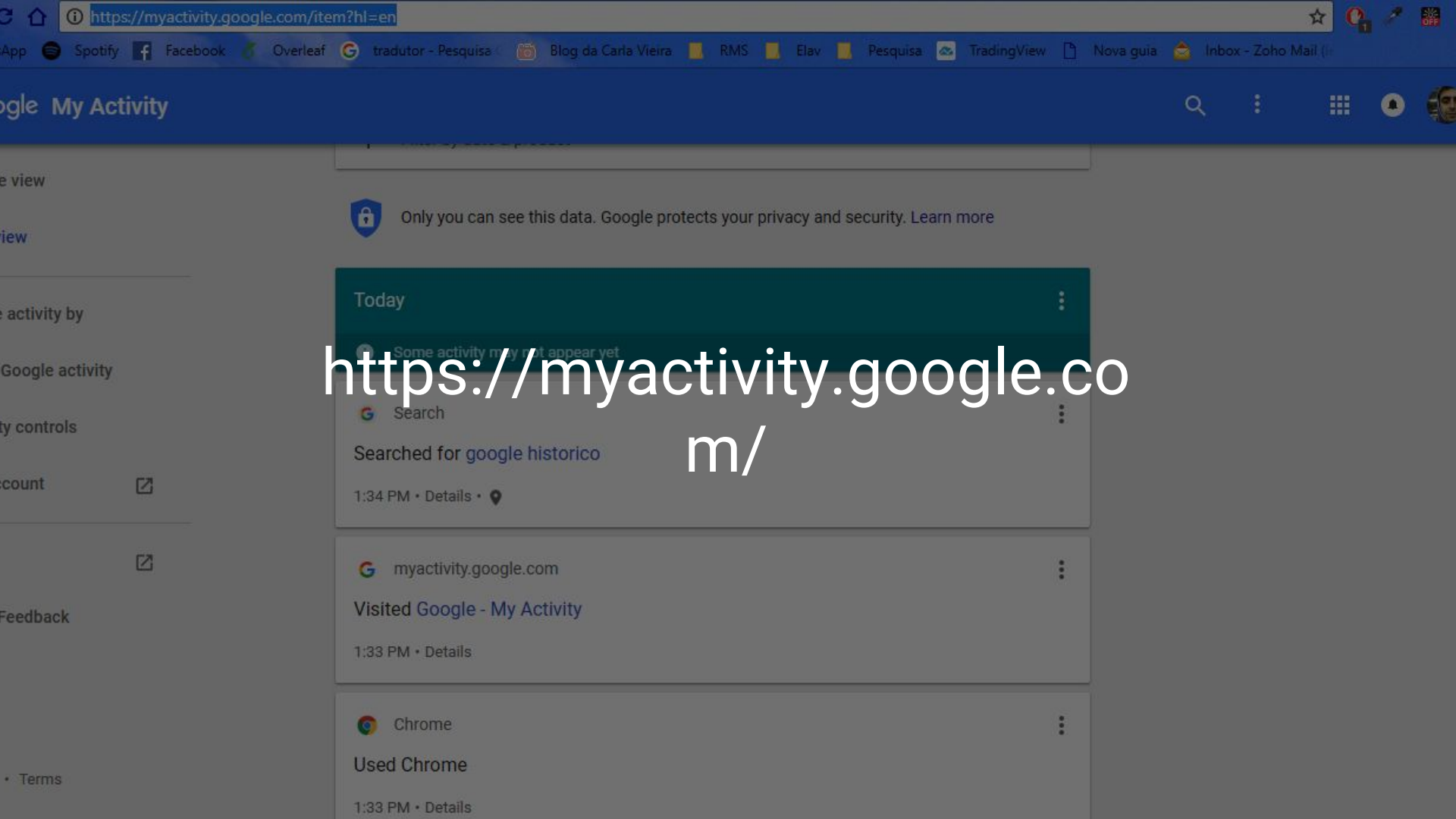


Onde?

Cobiçada capacidade no servidor: Alegadamente, a Apple paga 400-600 milhões de dólares em aluguel por ano para o Google - <https://www.teleboerse.de/aktien/Apple-mietet-Server-von-Google-article17250446.html>.

**O que sabem
sobre você?**

A decorative graphic element in the bottom right corner of the slide. It consists of a large blue triangle pointing upwards and to the right, with a thin grey line running parallel to its hypotenuse.



https://myactivity.google.com/

Google My Activity

view

view

activity by

Google activity

activity controls

account

Feedback

Terms



Only you can see this data. Google protects your privacy and security. [Learn more](#)

Today



Some activity may not appear yet



Search

Searched for [google historico](#)

1:34 PM • Details •



[myactivity.google.com](#)

Visited [Google - My Activity](#)

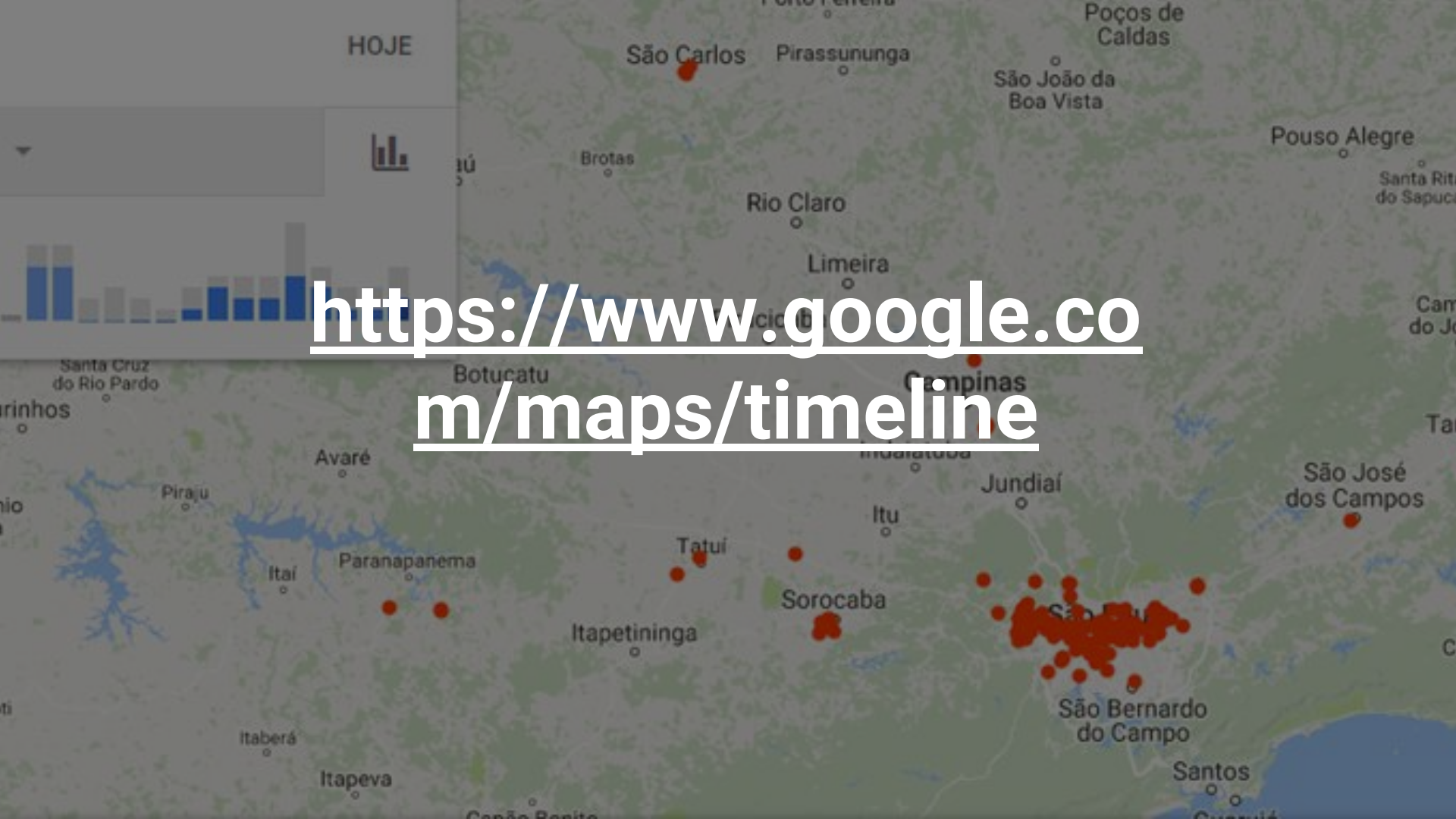
1:33 PM • Details



Chrome

Used Chrome

1:33 PM • Details



<https://www.google.com/maps/timeline>

o tempo

HOJE

27



Escola de Artes, Ciências
e Humanidades da Universidade de

de Abril De 2018

3,4 km
1h 3m



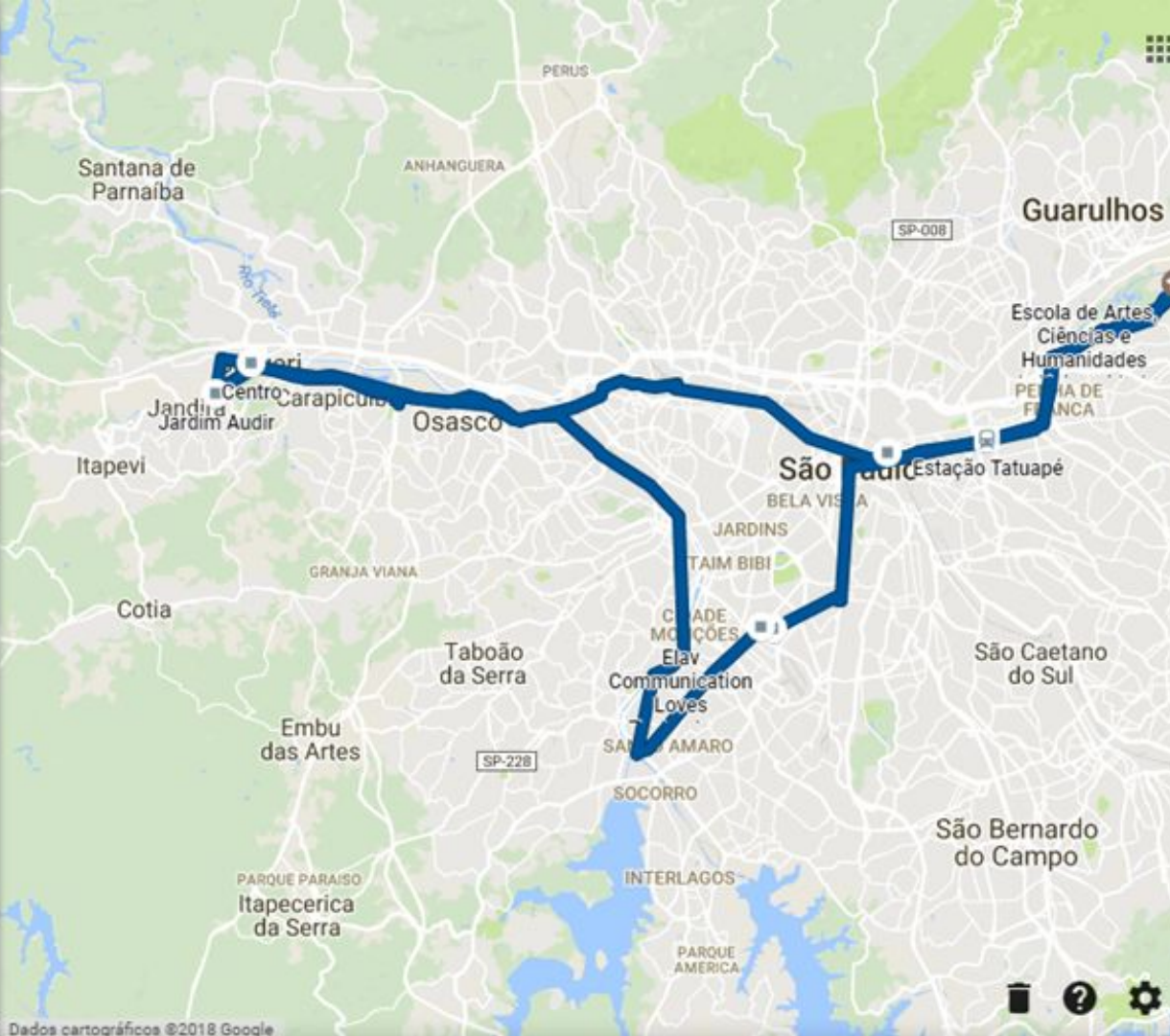
11,4 km
23m

Adicionar um lugar

Dirigindo - 2,8 km

7 min

07:51 - 08:17



Dados cartográficos ©2018 Google

Maiores marcas do mundo

- Apple - US\$ 184,15 bilhões
- Google - US\$ 141,70
- Microsoft - US\$ 79,99 bilhões
- Coca-Cola - US\$ 69,73 bilhões
- Amazon - US\$ 64,79 bilhões
- Samsung - US\$ 56,24 bilhões
- Toyota - US\$ 50,29 bilhões
- Facebook - US\$ 48,18 bilhões
- Mercedes-Benz - US\$ 47,82 bilhões
- IBM - US\$ 46,82 bilhões

Forbes - 2016.

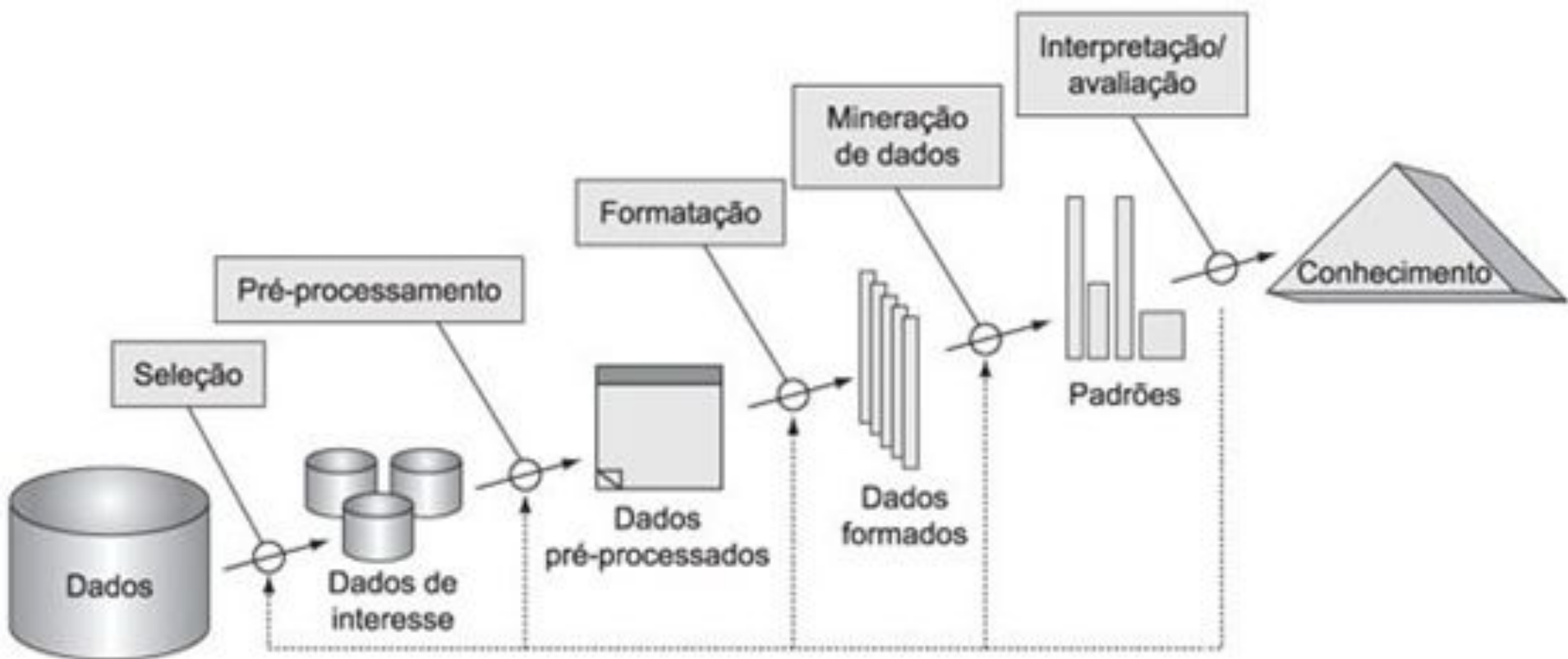
Maiores marcas do mundo

- **Apple** - US\$ 184,15 bilhões
- **Google** - US\$ 141,70
- **Microsoft** - US\$ 79,99 bilhões
 - Coca-Cola - US\$ 69,73 bilhões
- **Amazon** - US\$ 64,79 bilhões
- **Samsung** - US\$ 56,24 bilhões
 - Toyota - US\$ 50,29 bilhões
- **Facebook** - US\$ 48,18 bilhões
 - Mercedes-Benz - US\$ 47,82 bilhões
- **IBM** - US\$ 46,82 bilhões

Forbes - 2016.

Como o Google ganha dinheiro?





LEONARDO MORALES BELLUZZI

APLICAÇÃO DO K-MEANS CLUSTERING PARA
CLASSIFICAR PADRÕES OFENSIVOS NO
BASQUETEBOL

Mineração de Dados - Data Mining

The background of the slide features two diagonal stripes. A wide, solid blue stripe runs from the bottom-left towards the top-right. Overlaid on this is a thinner, light grey stripe that follows a similar path but is slightly offset, creating a layered effect.

Mineração de Dados - Data Mining

Data Mining define o processo automatizado de captura e análise de grandes conjuntos de dados para extrair um significado, sendo usado tanto para **descrever** características do **passado** como para prever tendências para o **futuro**.

From Data Mining to Knowledge Discovery in Databases

Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth

■ Data mining and knowledge discovery in databases have been attracting a significant amount of research, industry, and media attention of late. What is all the excitement about? This article provides an overview of this emerging field, clarifying how data mining and knowledge discovery in databases are related both to each other and to related fields, such as machine learning, statistics, and databases. The article mentions particular real-world applications, specific data-mining techniques, challenges involved in real-world applications of knowledge discovery, and current and future research directions in the field.

Across a wide variety of fields, data are being collected and accumulated at a dramatic pace. There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data. These theories and tools are the subject of the emerging field of knowledge discovery in databases (KDD).

At an abstract level, the KDD field is concerned with the development of methods and techniques for making sense of data. The basic problem addressed by the KDD process is one of mapping low-level data (which are typically too voluminous to understand and digest easily) into other forms that might be more compact (for example, a short report), more abstract (for example, a descriptive approximation or model of the process that generated the data), or more useful (for example, a predictive model for estimating the value of future cases). At the core of the process is the application of specific data-mining methods for pattern discovery and extraction.¹

This article begins by discussing the historical context of KDD and data mining and their intersection with other related fields. A brief summary of recent KDD real-world applications is provided. Definitions of KDD and data mining are provided, and the general multistep KDD process is outlined. This multistep process has the application of data-mining algorithms as one particular step in the process. The data-mining step is discussed in more detail in the context of specific data-mining algorithms and their application. Real-world practical application issues are also outlined. Finally, the article enumerates challenges for future research and development and in particular discusses potential opportunities for AI technology in KDD systems.

Why Do We Need KDD?

The traditional method of turning data into knowledge relies on manual analysis and interpretation. For example, in the health-care industry, it is common for specialists to periodically analyze current trends and changes in health-care data, say, on a quarterly basis. The specialists then provide a report detailing the analysis to the sponsoring health-care organization; this report becomes the basis for future decision making and planning for health-care management. In a totally different type of application, planetary geologists sift through remotely sensed images of planets and asteroids, carefully locating and cataloging such geologic objects of interest as impact craters. Be it science, marketing, finance, health care, retail, or any other field, the classical approach to data analysis relies fundamentally on one or more analysts becoming

Conceitos e Aplicações de Data Mining

Data Mining Concepts and Applications

HELOISA HELENA SFERRA

Universidade Metodista de Piracicaba (Piracicaba, Brasil)
hsferra@uol.com.br

ÂNGELA M. C. JORGE CORRÊA

Universidade Metodista de Piracicaba (Piracicaba, Brasil)
ajcorrea@unimep

RESUMO Atualmente, muito se fala em *Data Mining*, encontrando-se na literatura significativa variedade de estudos sobre o tema. Este artigo tem como objetivo introduzir conceitos básicos dessa tecnologia a interessados que ainda estão iniciando o estudo de *Data Mining*. Nesse contexto, o presente texto pretende apresentar alguns desses conceitos sobre as técnicas que envolvem a descoberta de conhecimento em grandes conjuntos de dados, além de registrar algumas características de um software específico para mineração de dados, o *Clementine*, da SPSS, bem como algumas aplicações realizadas nessa ferramenta.

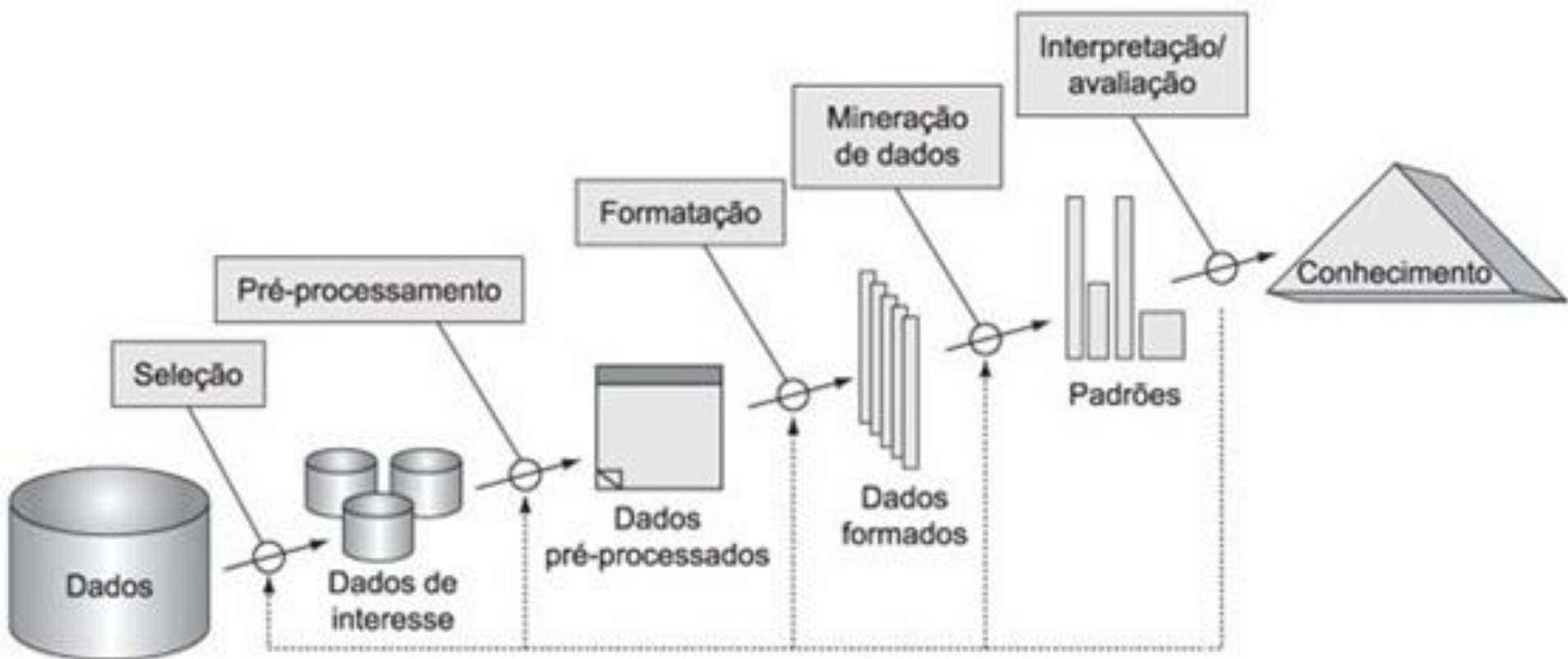
Palavras-chave MINERAÇÃO DE DADOS – DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS – ANÁLISE EXPLORATÓRIA DE DADOS – MODELOS ESTATÍSTICOS DE RELACIONAMENTO ENTRE VARIÁVEIS – CLEMENTINE/SPSS.

ABSTRACT Much is said about Data Mining nowadays and there is a significant variety of studies on the subject. This paper's aim is to introduce some of the technology's basic concepts to those who are beginning their studies on Data Mining. In such context, the present article presents some of the concepts related to the techniques involved in knowledge discovery within large databases. Moreover, it presents some features of a specific software for Data Mining: *Clementine*, from SPSS. The paper also indicates some applications for this tool's use.

Keywords DATA MINING – KNOWLEDGE DISCOVERY IN DATABASES – EXPLORATORY ANALYSIS – STATISTICAL MODELS OF RELATIONSHIP BETWEEN VARIABLES – CLEMENTINE/SPSS.

Knowledge Discovery in Databases – KDD

O processo de extração de conhecimento em bases de dados *Knowledge Discovery in Databases* – **KDD** foi proposto em 1989, refere se a tarefa de produzir conhecimento a partir de dados, transformando dados em informação através das seguintes etapas : seleção de dados; processamento e limpeza; transformação; mineração de dados *Data Mining*; e interpretação dos resultados.



Knowledge Discovery in Databases – KDD

Seleção

A seleção de dados tem como objetivo identificar as origens internas e externas da informação, extraíndo um subconjunto de dados necessário para a aplicação da mineração de dados, selecionando apenas atributos relevantes aos objetivos do processo de extração de conhecimento.

Knowledge Discovery in Databases – KDD

Pré-Processamento

O processamento assegura a qualidade dos dados selecionados. Como o resultado do processo de extração possivelmente será utilizado no processo de tomada de decisão, sendo a qualidade dos dados é um fator extremamente importante.

Knowledge Discovery in Databases – KDD

Formatar

Formatar os dados tem por objetivo é converter o conjunto bruto de dados em uma forma padrão de uso, tornando os dados úteis para a mineração, eliminando redundâncias e valores indesejáveis a aplicação da mineração dos dados.

Knowledge Discovery in Databases – KDD Minerar

A mineração dos dados *data mining* refere-se em descobrir informações valiosas ou conhecimentos de grandes conjuntos de dados, através de métodos e técnicas computacionais.

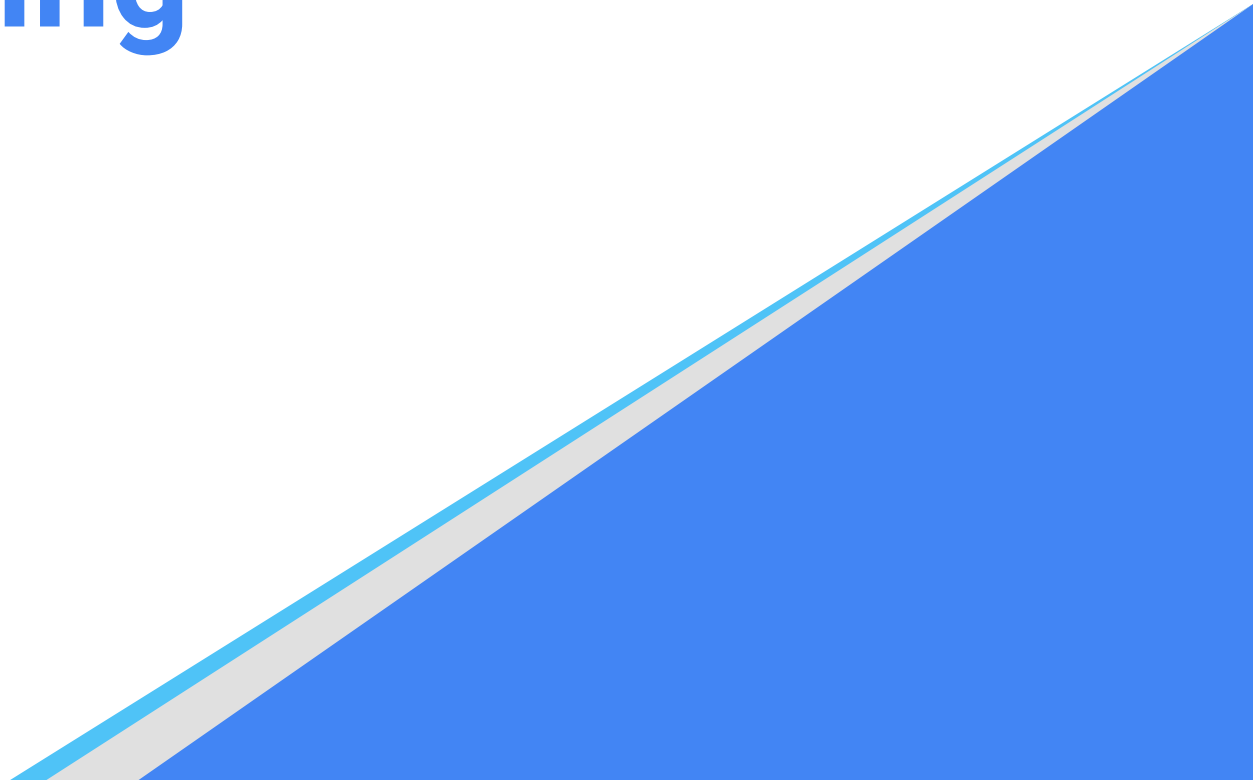
Knowledge Discovery in Databases – KDD

Interpretação

As técnicas de mineração de dados podem ter gerado uma quantidade enorme de padrões, dos quais podem não ser relevantes ou interessantes ao usuário então a interpretação faz com que o usuário possa compreender e utilizar o conhecimento descoberto.

Data-Mining

Métodos



Algorithms for Clustering Data

Anil K. Jain

Richard C. Dubes

Michigan State University



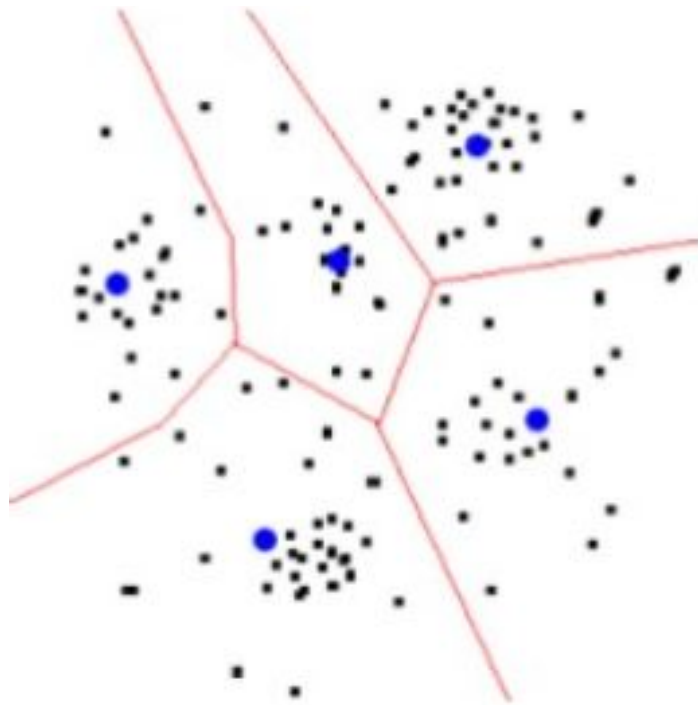
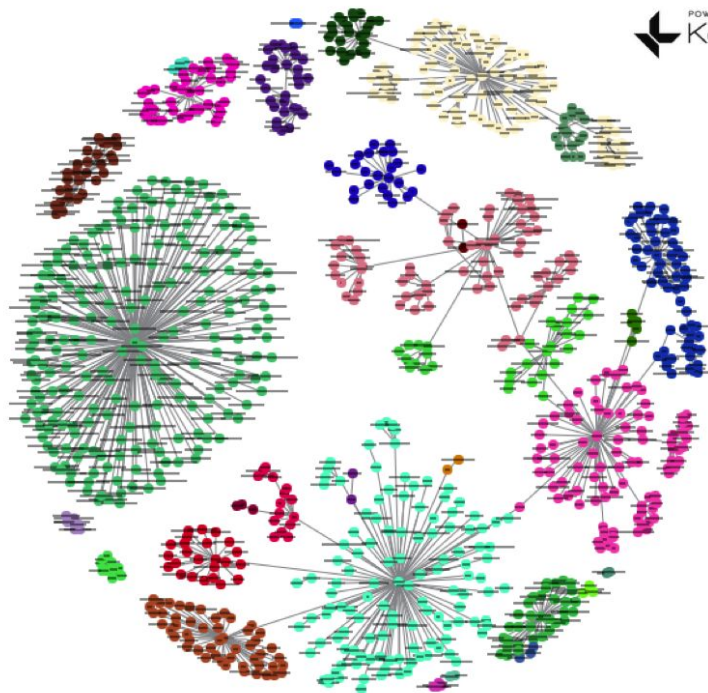
Mineração de Dados - Data Mining

Data Mining define o processo automatizado de captura e análise de grandes conjuntos de dados para extrair um significado, sendo usado tanto para **descrever** características do **passado** como para prever tendências para o **futuro**.

Data-Mining - Agrupamento

Analise de **agrupamento** ou ***cluster analysis*** é o processo de agrupar um conjunto de objetos físicos ou abstratos em classes de objetos similares.

Um cluster é uma coleção de objetos que são similares uns aos outros de acordo com algum critério de similaridade e dissimilares a objetos pertencentes a outros clusters.



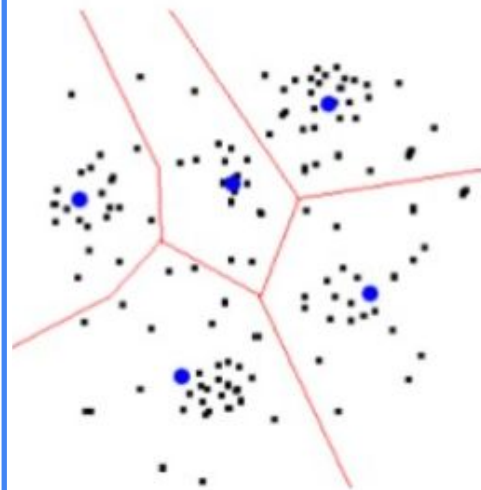
Data-Mining - Agrupamento

- Métodos de **particionamento** ou *partitioning methods*.
- Métodos **hierárquicos** ou *hierarchical methods*.
- Os métodos baseados em **modelos** ou *model-based methods*.

Particionamento

Métodos de particionamento ou partitioning methods trabalham com um conjunto D de dados com n registros e k o número de agrupamentos desejados.

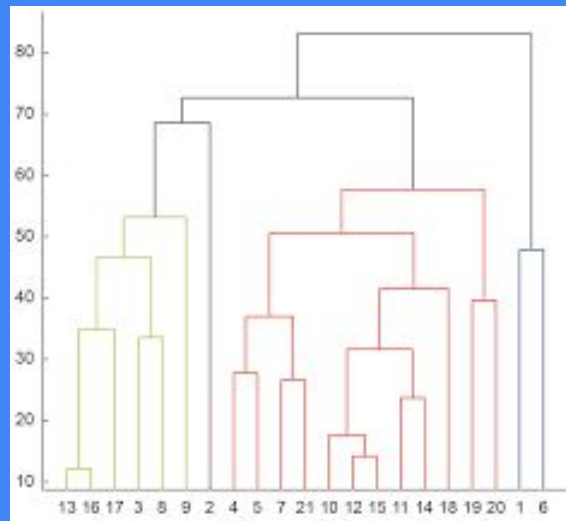
Figura 2 – *K-menas*, com $k=5$



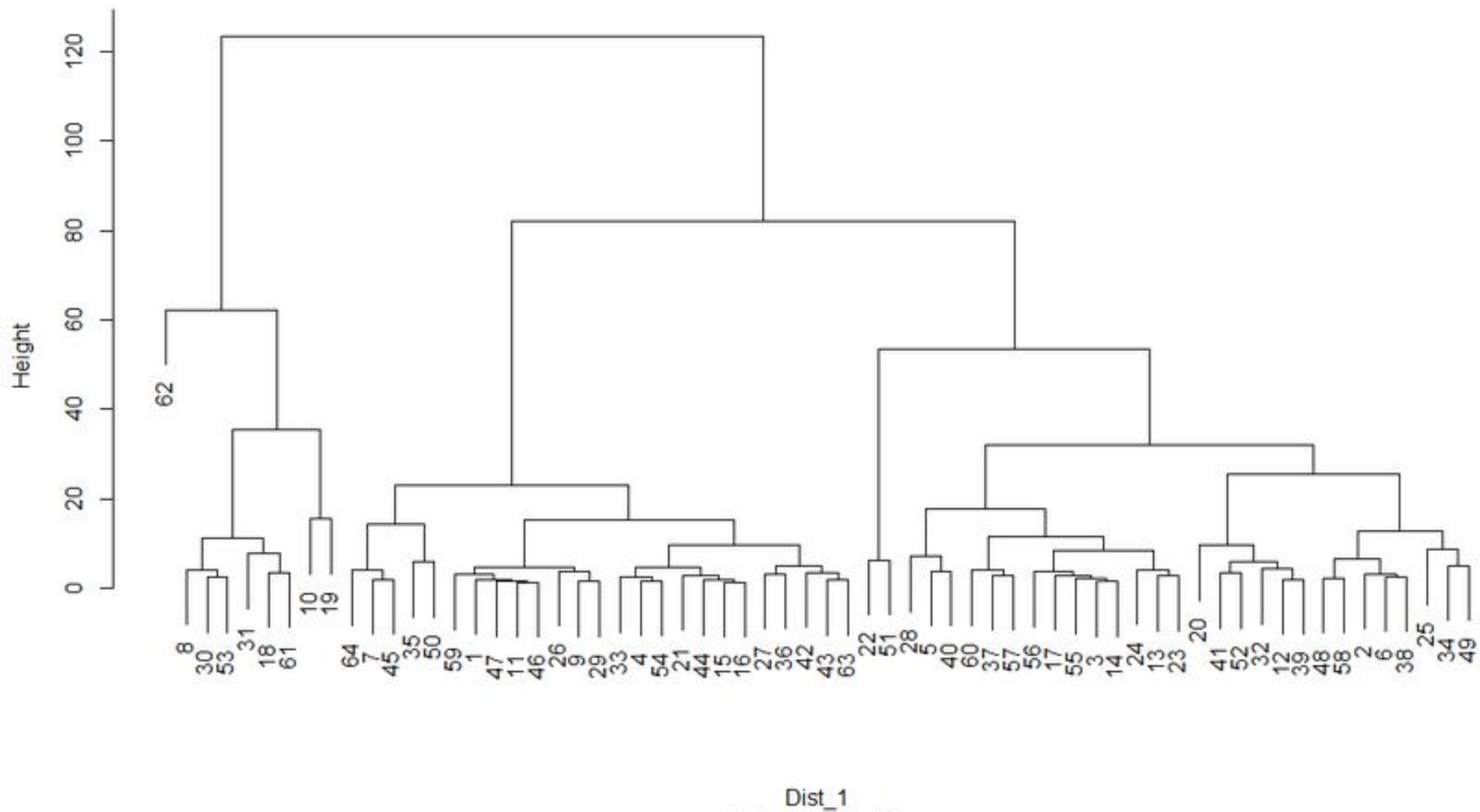
Fonte: (SLOMAN, 2012)

Hierárquicos

- Métodos **hierárquicos** ou *hierarchical methods* idéia básica dos métodos hierárquicos é criar o agrupamento por meio da aglomeração ou da divisão dos elementos do conjunto. A forma gerada por estes métodos é um dendrograma.

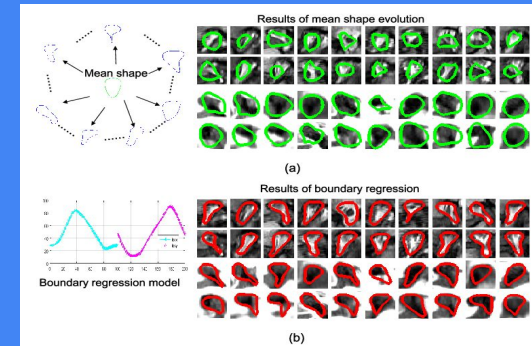


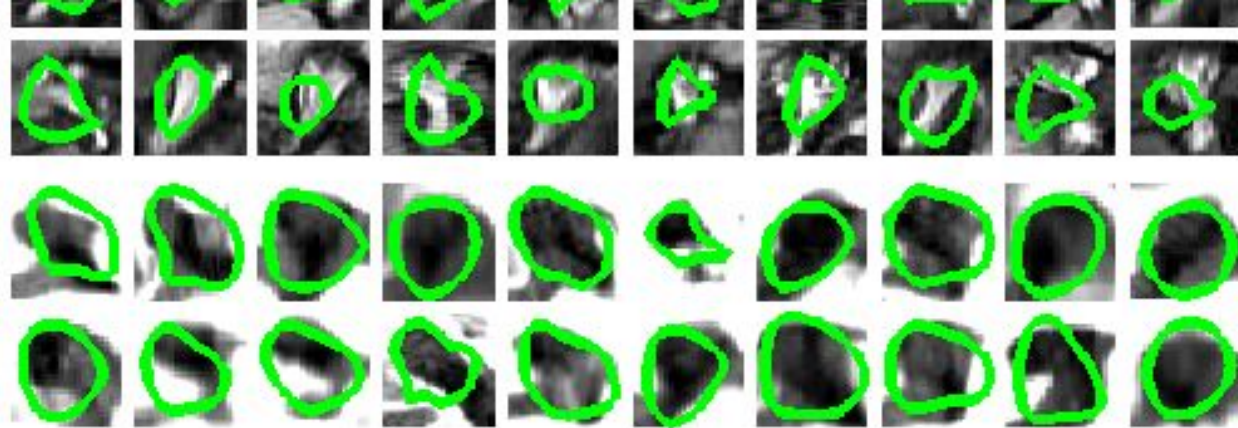
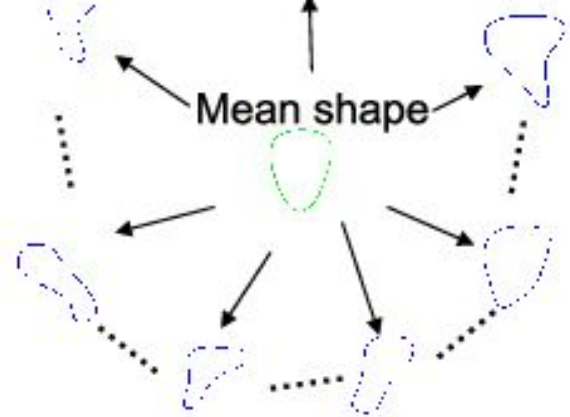
Cluster Dendrogram



Modelos

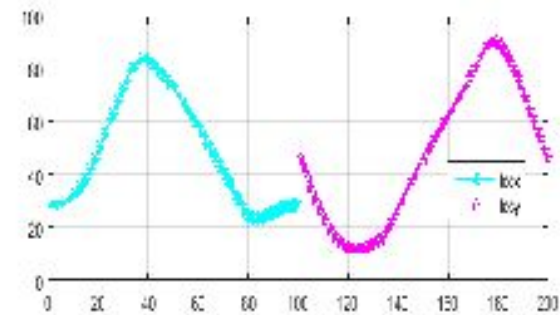
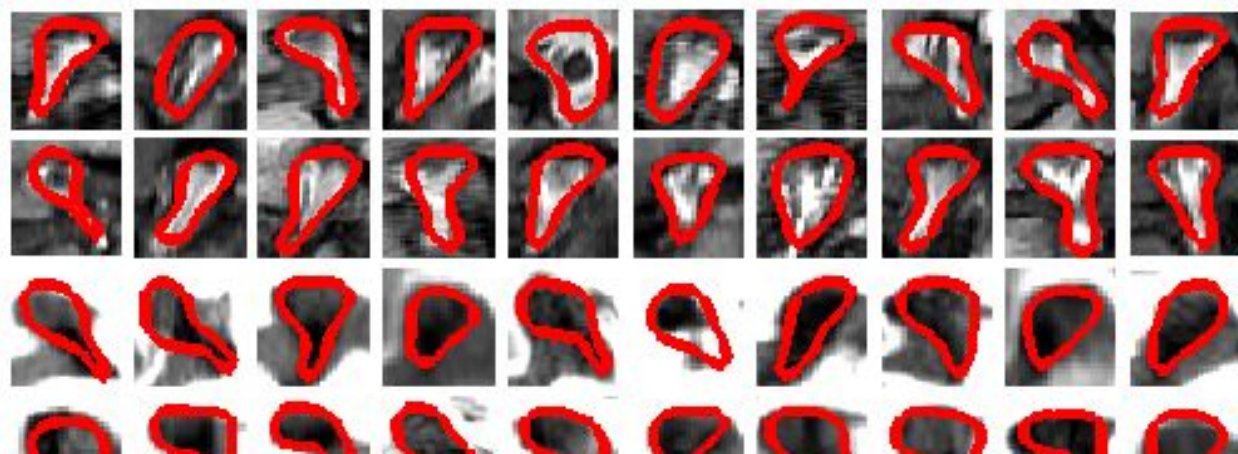
- Os métodos baseados em **modelos** ou *model-based methods* criam um modelo para cada agrupamento e tentam identificar o melhor modelo para cada objeto.





(a)

Results of boundary regression

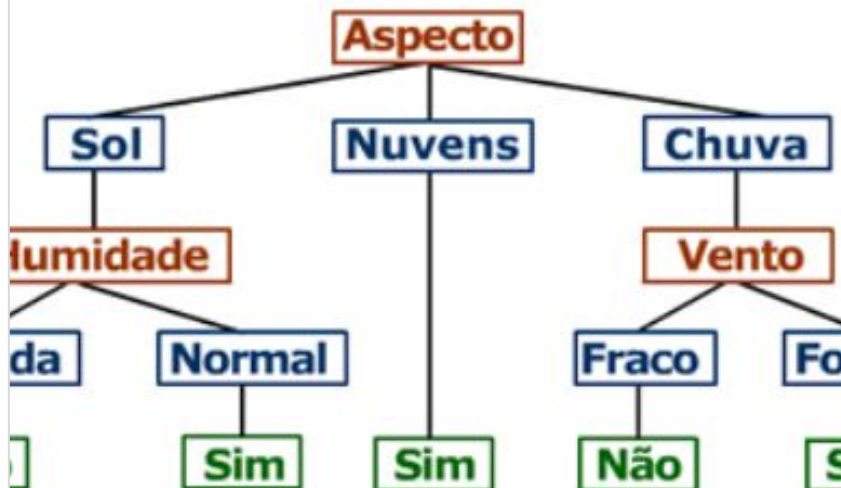


Boundary regression model

Exemplos de Treino

Dia	Aspecto	Temp.	Humidade	Vento	Jogar Ténis
D1	Sol	Quente	Elevada	Fraco	Não
D2	Sol	Quente	Elevada	Forte	Não
D3	Nuvens	Quente	Elevada	Fraco	Sim
D4	Chuva	Ameno	Elevada	Fraco	Sim
D5	Chuva	Fresco	Normal	Fraco	Sim
D6	Chuva	Fresco	Normal	Forte	Não
D7	Nuvens	Fresco	Normal	Fraco	Sim
D8	Sol	Ameno	Elevada	Fraco	Não
D9	Sol	Fresco	Normal	Fraco	Sim
D10	Chuva	Ameno	Normal	Forte	Sim
D11	Sol	Ameno	Normal	Forte	Sim
D12	Nuvens	Ameno	Elevada	Forte	Sim
D13	Nuvens	Quente	Normal	Fraco	Sim
D14	Chuva	Ameno	Elevada	Forte	Não

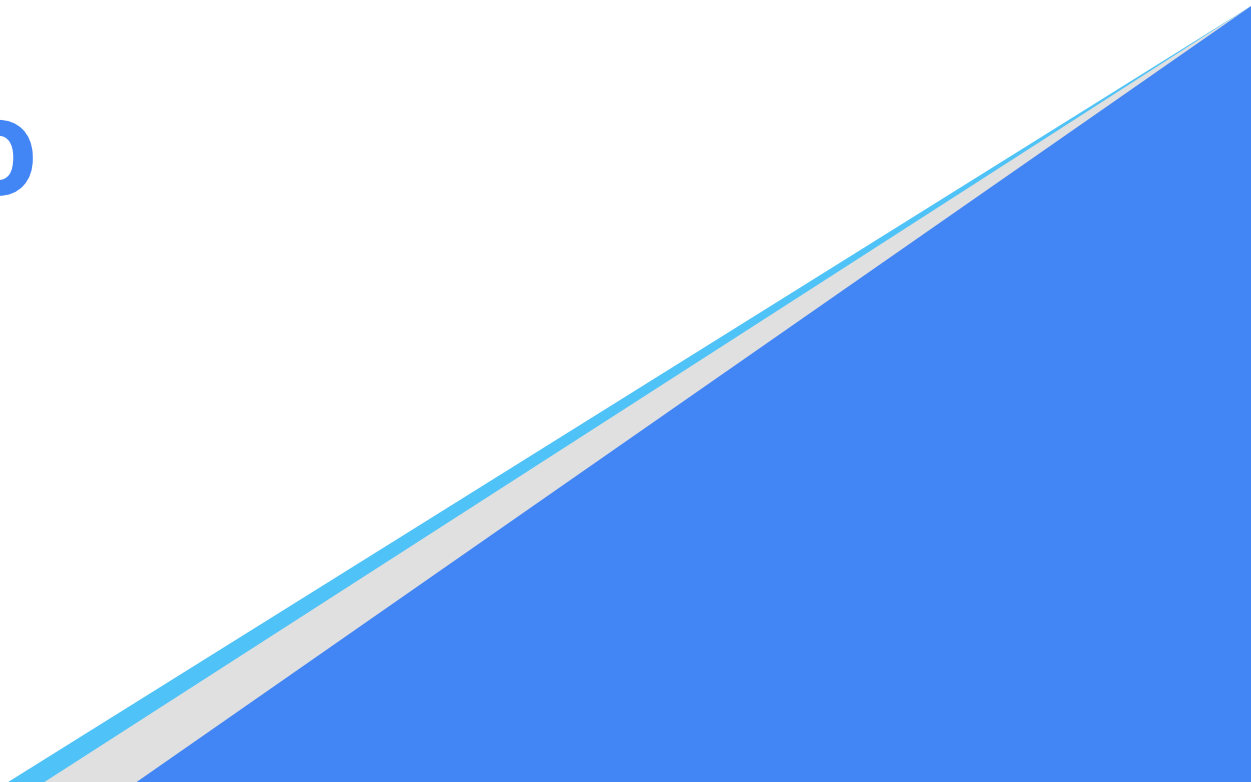
Árvore de Decisão para Jogar Ténis



Árvores de decisão

Árvores de decisão são modelos estatísticos que utilizam um treinamento supervisionado para a classificação e previsão de dados. Em outras palavras, em sua construção é utilizado um conjunto de treinamento formado por entradas e saídas.

Utilização





Carros autônomos já começam a virar realidade



Populares na Netflix



Em alta

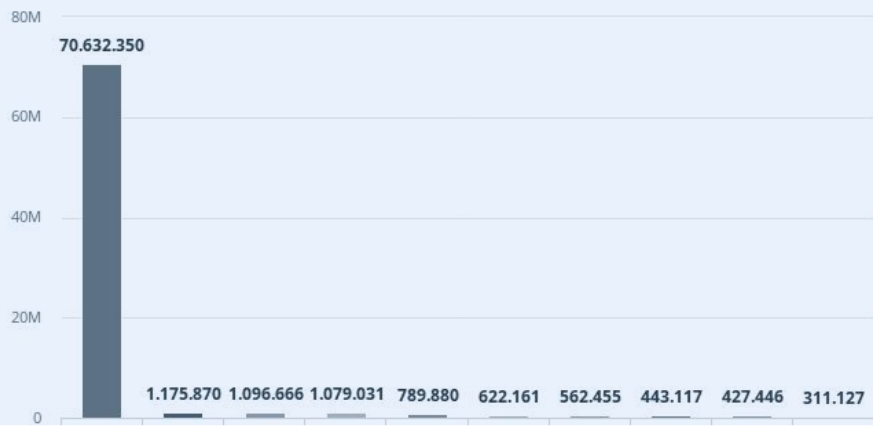


Sistemas de recomendações

Smarthome



Sistema de análise em redes sociais





MACHINE
LEARNING



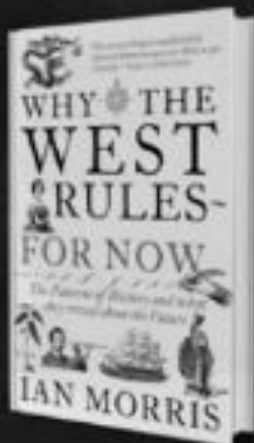
DEEP
LEARNING



Machine learning - Deep learning



IAN MORRIS



ÍNDICE DE DESENVOLVIMENTO SOCIAL

100
90
80
70
60
50
40
30
20
10
0

DESENVOLVIMENTO SOCIAL

POPULAÇÃO MUNDIAL

POPULAÇÃO EM BILHÕES

- COMIDA
- SAÚDE
- INFRA-ESTRUTURA
- OUTROS BENEFÍCIOS



https://www.youtube.com/watch?v=CRy2g_-SqKY - Nerdologia



CONTROLE DE TRÁFEGO AÉREO MANUAL

ESCRITÓRIO DE ARQUITETURA

