



By @kakashi_copiador



Estratégia
Concursos

//estratégia tech



ESTRATÉGIA
CONCURSOS



<http://anchor.fm/estrategia-tech>



http://t.me/estrategia_ti

<http://t.me/lacerdaph>



Estratégia
Concursos



<http://instagram.com/estrategiaconcursosti>

https://t.me/kakashi_copiador



DATA MINING

Prof. Raphael Lacerda

<https://www.instagram.com/p/B74PyOzj5V3/>
<https://www.estrategiaconcursos.com.br/blog/prazer-um-eterno-estudante/>

https://t.me/kakashi_copiador



ESSE CURSO É PARA VOCÊ?

Prof. Raphael Lacerda

Questão de alinhamento

A descoberta de novas regras e padrões em conjuntos de dados fornecidos, ou aquisição de conhecimento indutivo, é um dos objetivos de data mining.

Questão de alinhamento

CRISP-DM é uma metodologia proprietária, baseada em 6 fases, entre elas a Business Understanding e Data Understanding, com o objetivo de implantar um projeto de data mining.

Questão de alinhamento

No que se refere aos diversos tipos de algoritmos utilizados para minerar dados, a técnica utilizada em tarefas de classificação, regressão e segmentação de dados em tipos de dados heterogêneos é denominada

A algoritmos genéticos.

B redes neurais.

C banco objeto relacional.

D classes privadas.

E redes privadas.

INDO MAIS
FUNDO!



SE VOCÊ ERROU...

Prof. Raphael Lacerda



@estrategiaconcursosti



Estratégia
Concursos

SEU LUGAR É AQUI!!!



Estratégia
Cast



@estrategiaconcursosti



Estratégia
Concursos

AGORA EM ÁUDIO TBM!



MINERAÇÃO

Prof. Raphael Lacerda



PROCURANDO POR DADOS!



Prof. Raphael Lacerda

Partiu minerar então?



1 Bitcoin equals

49,149.12

Brazilian Real

Jun 29, 14:38 UTC - Disclaimer

1

Bitcoin

49149.12

Brazilian Real

Data provided by Morningstar for Currency and Coinbase for Cryptocurrency

1D 5D **1 M** 1 Y 5 Y Max



1 United States Dollar equals

5.43 Brazilian Real

Jun 29, 14:52 UTC - Disclaimer

1

United States Dollar

5.43

Brazilian Real

Data provided by Morningstar for Currency and Coinbase for Cryptocurrency

https://t.me/kakashi_copiador

1D 5D **1 M** 1 Y 5 Y Max



A close-up, high-angle shot of a large, dense pile of white dice. The dice are scattered in various orientations, showing different faces with black pips. The lighting is bright, creating soft shadows and highlighting the texture of the dice. A semi-transparent purple rectangle is overlaid in the center of the image, containing the text "Dados AGAIN" in white, bold, sans-serif font.

Dados AGAIN

Data Mining





Cambridge
Analytica

facebo



Padrões





**I GIVE PRIVATE
INFORMATION ON
CORPORATIONS TO
YOU FOR FREE,
AND I'M THE VILLAIN**



**I GIVE YOUR
PRIVATE INFORMATION
TO CORPORATIONS
FOR MONEY.
AND
I'M MAN OF THE YEAR**



keep calm



**KEEP CALM AND
REMEMBER:**

**IF YOU ARE GETTING A SERVICE FOR FREE
YOU ARE NOT THE CUSTOMER
YOU ARE THE PRODUCT**

Exemplos

**DATA
MINING**



Exemplos

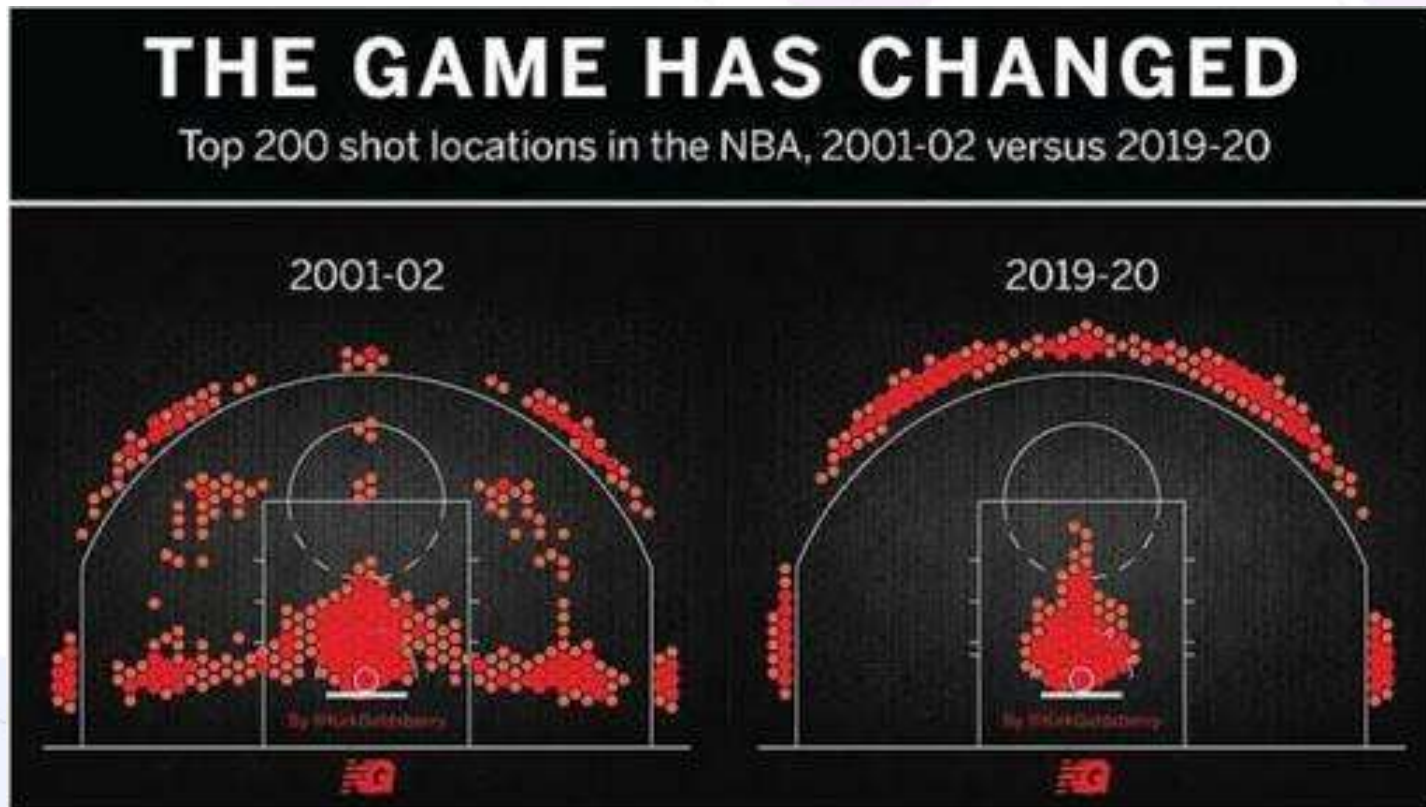


Exemplos



<https://economia.uol.com.br/noticias/redacao/2019/06/07/midia-e-marketing-publicidade-digital-verbas-iab-brasil.htm>

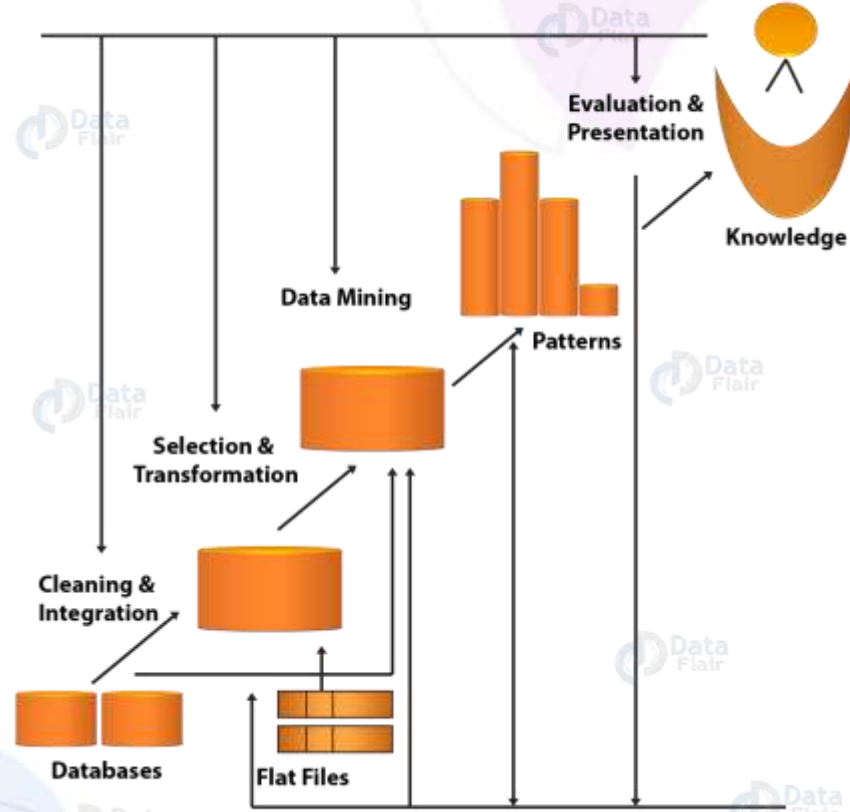
Exemplos



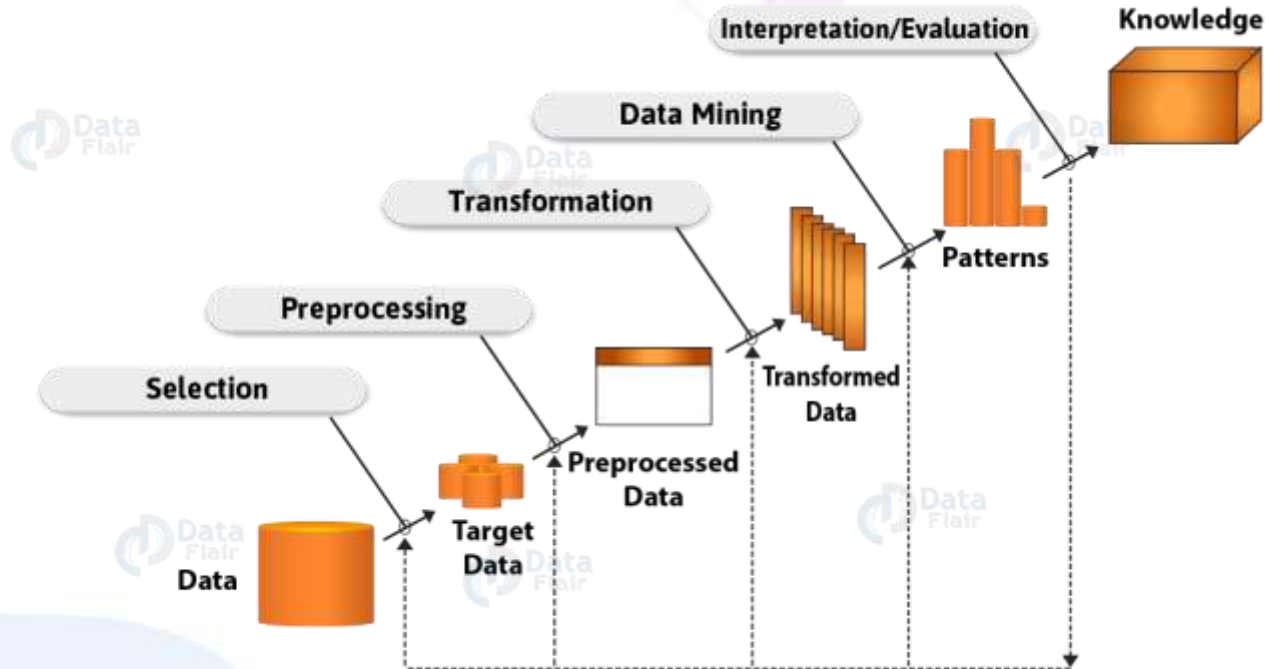
KDD



KDD



KDD



*São
Paulo
Tri
Mund iiiiiiiiall*

#DicaDoProf



Objetivos



Objetivos

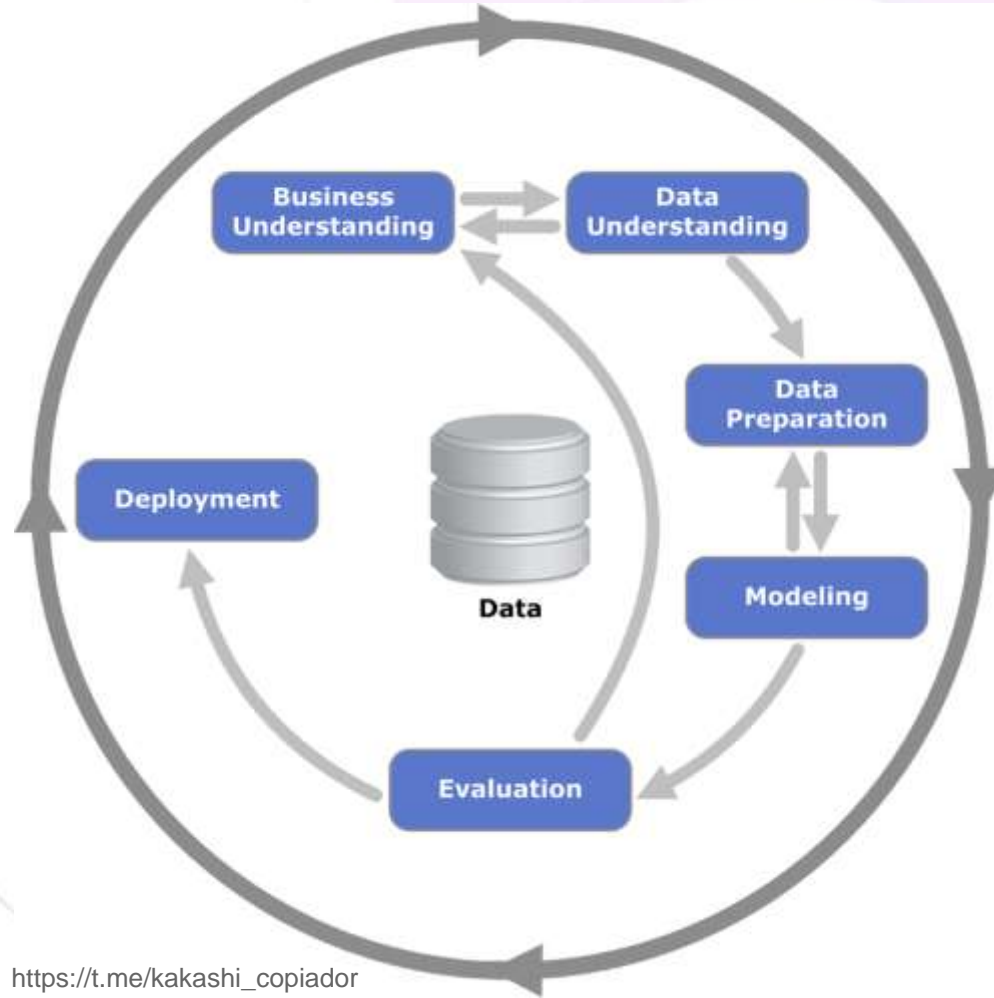


P i c o

PREVISÃO IDENTIFICAÇÃO CLASSIFICAÇÃO OTIMIZAÇÃO

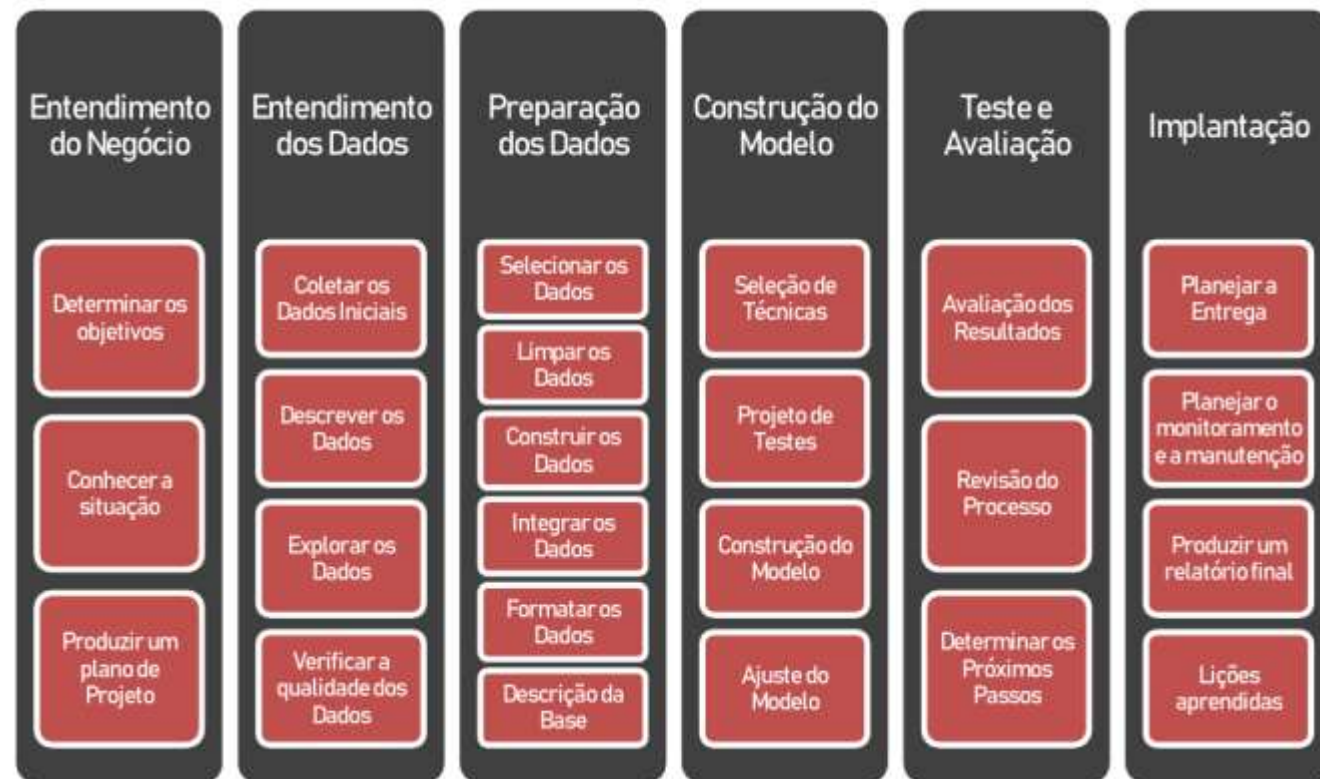


CRISP-DM



Atividades das fases

que homem

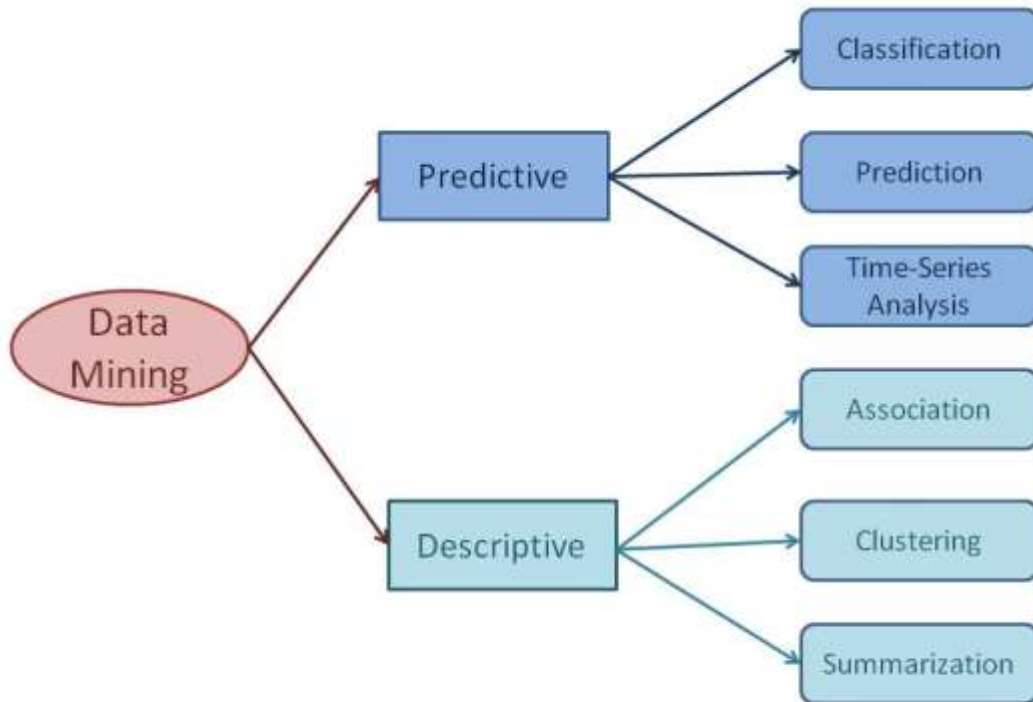




agrupando...

<https://www.wideskills.com/data-mining-tutorial/05-data-mining-tasks>

<https://academic.csuohio.edu/fuy/Pub/pot97.pdf>



Data mining techniques

Classification

Clustering

Regression

Outer

Sequential
Patterns

Prediction

Association
Rules

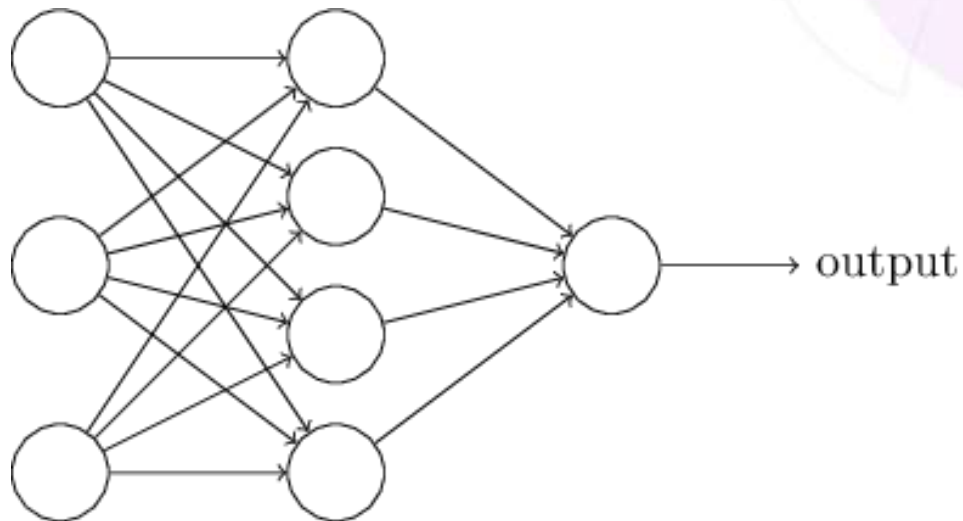
<https://www.guru99.com/data-mining-tutorial.html#11>



<https://data-flair.training/blogs/data-mining-techniques/>

<https://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>

Redes Neurais



<https://www.sciencedirect.com/science/article/abs/pii/S0957417408002005>



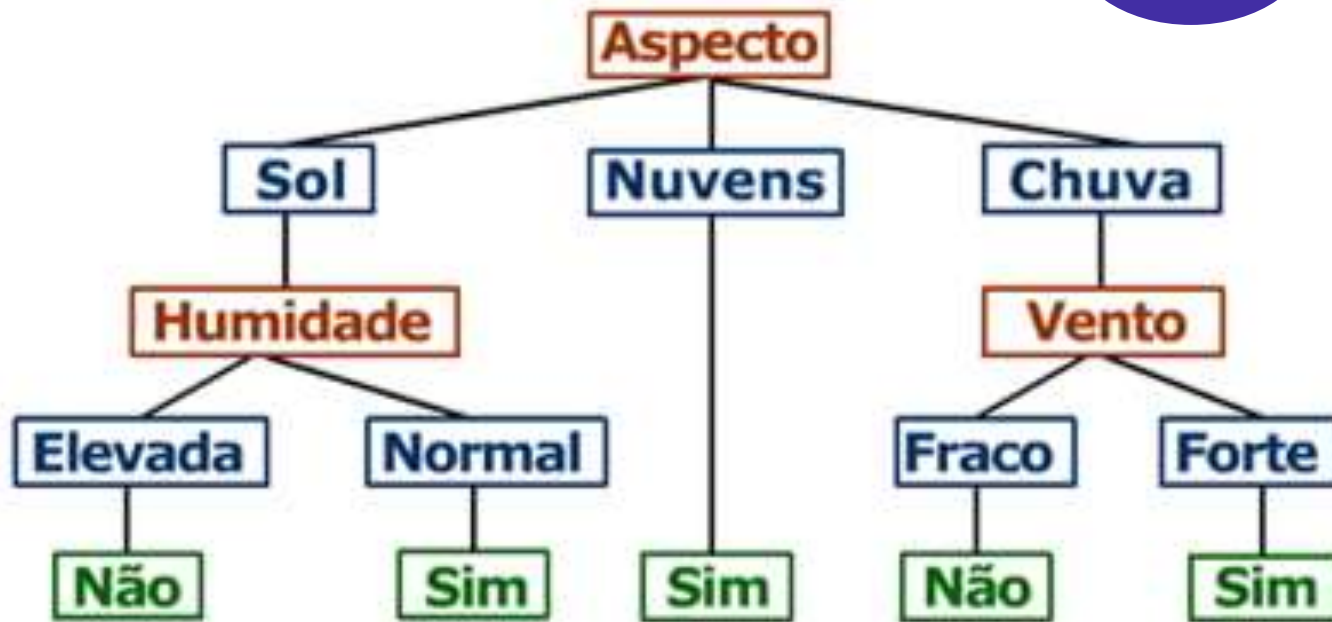
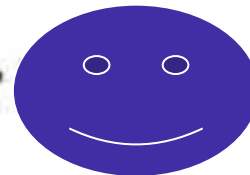
opennn

neural networks

Prof. Raphael Lacerda

Árvores de Decisão

Árvore de Decisão para Jogar



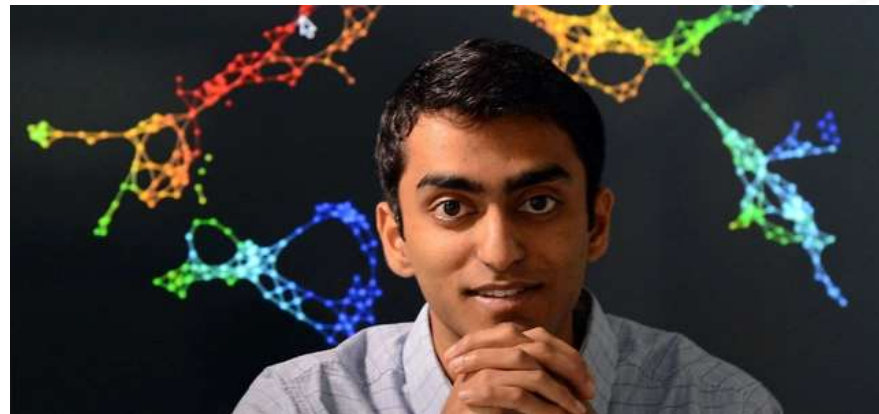
Classificação



<https://www.youtube.com/watch?v=E-gpSQQe3w8>

*Muthu Alagappan at
TEDxSpokane*

TED



Chris Paul



Defensive Stopper
Mid-range Shooter

Steve Nash

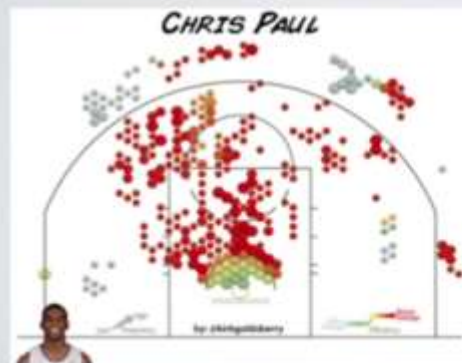


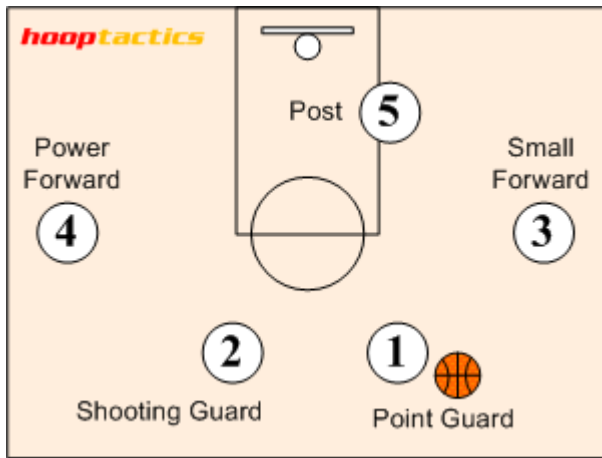
Skilled Passer
Pick and Roller

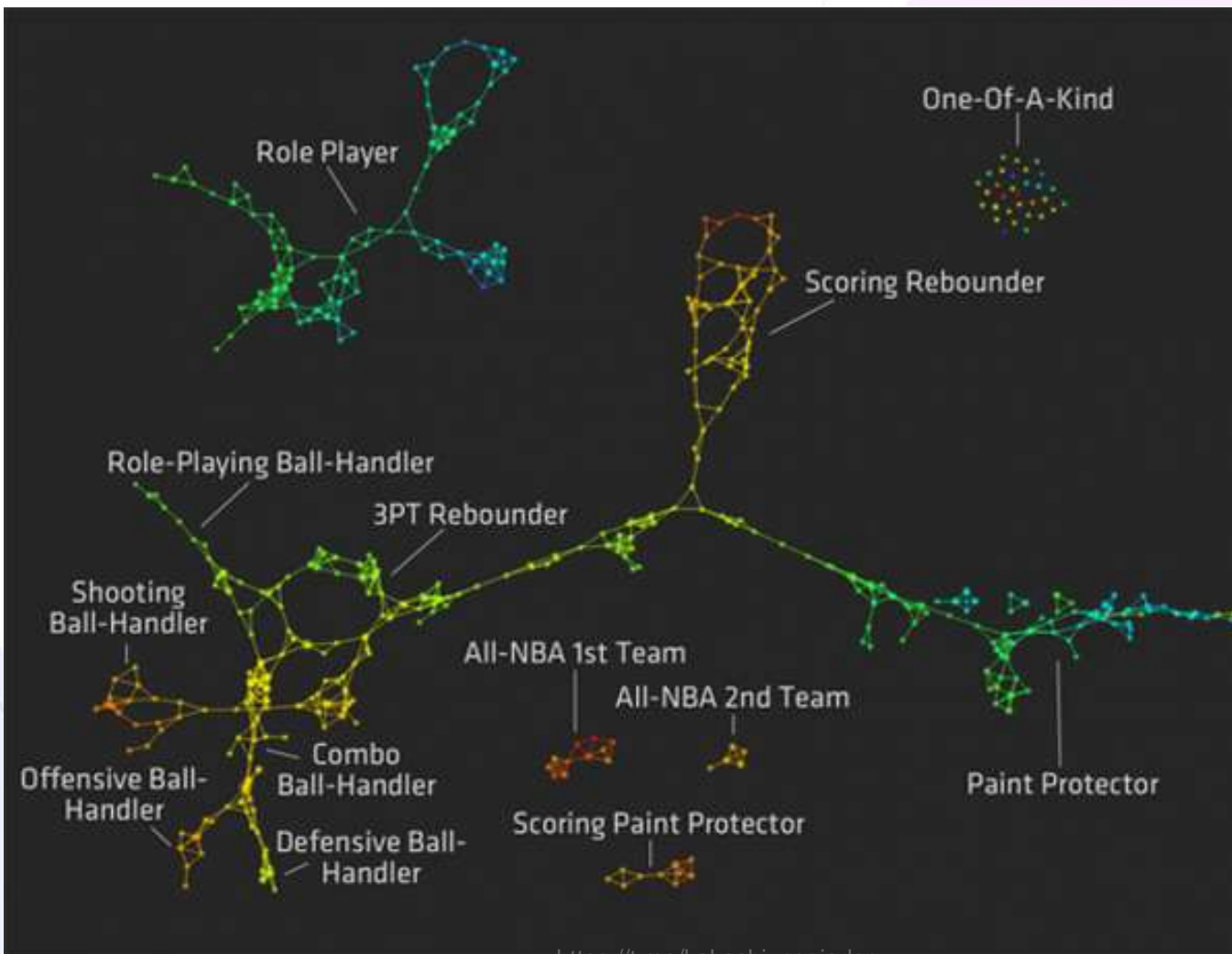
Jason Kidd



Rebounder
Post-up Ability

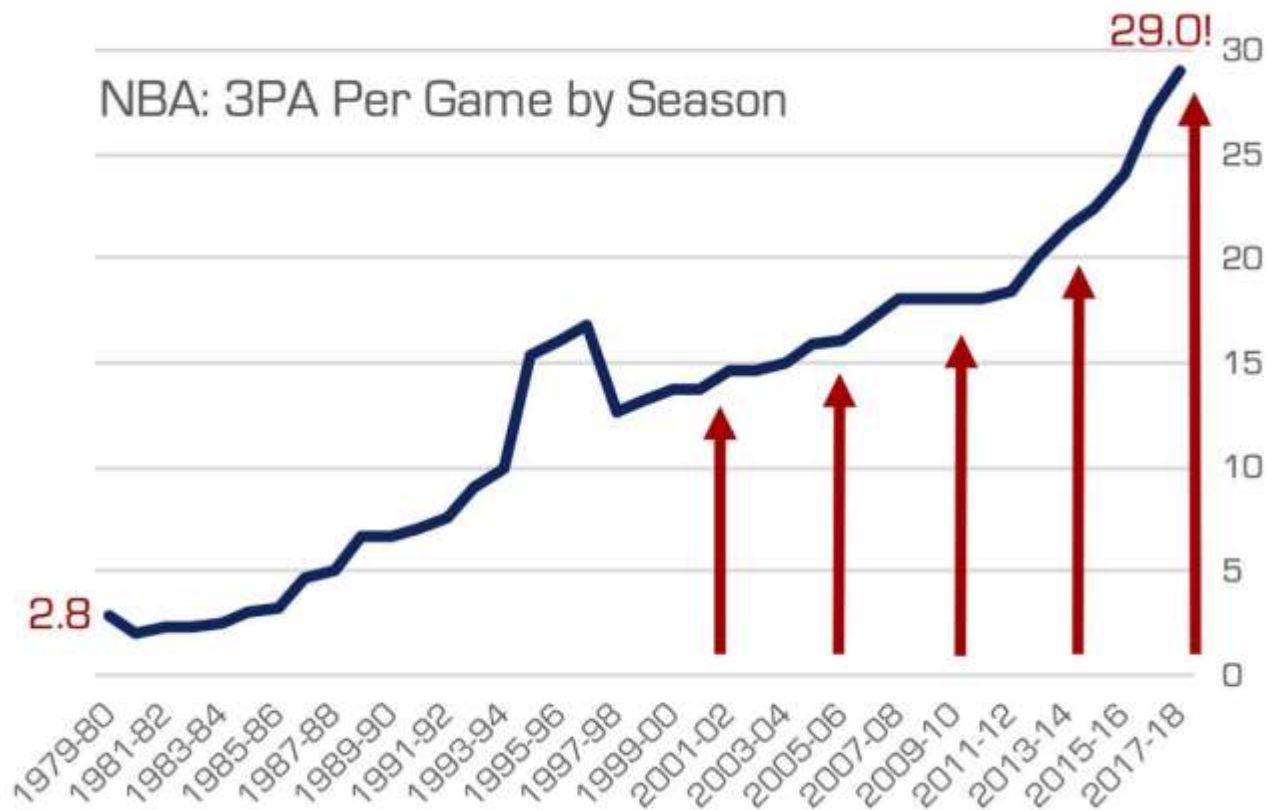






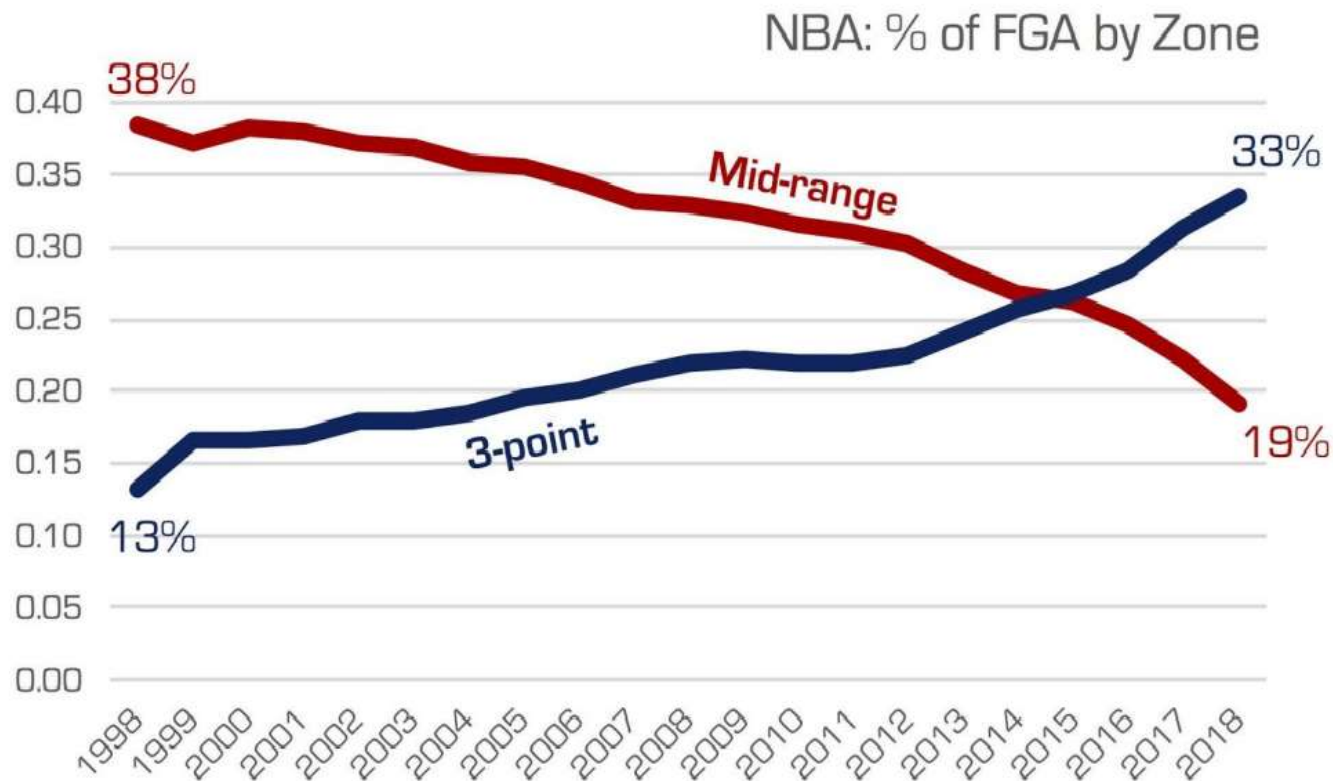
<https://bolapresa.com.br/as-13-posicoes-do-basquete/>

Regressão / Predição



<https://shottracker.com/articles/the-3-point-revolution>

Regressão / Predição



Associação



Agrupamento



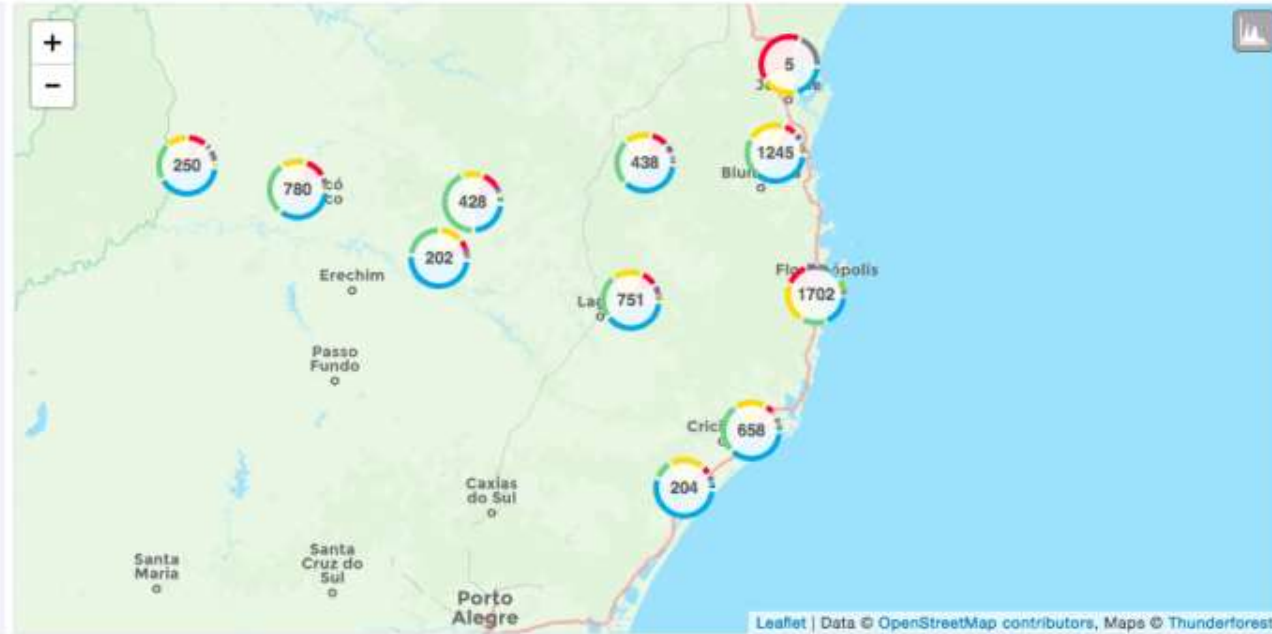
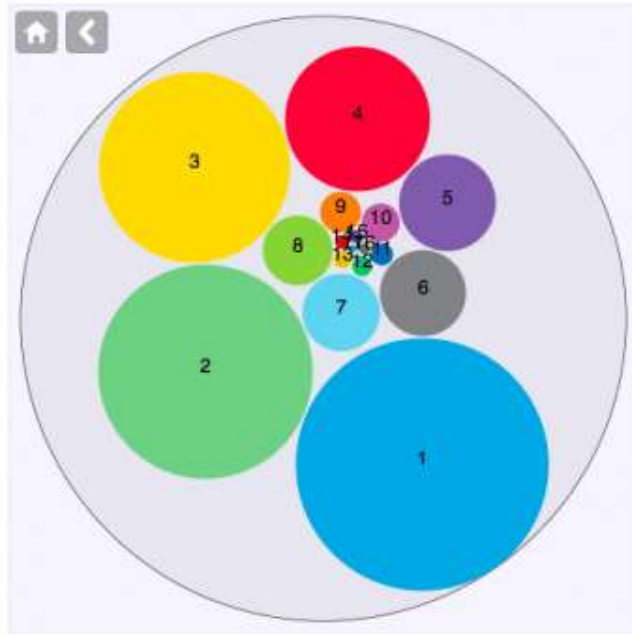
clusterização em SC

Analysis: Fazenda SC

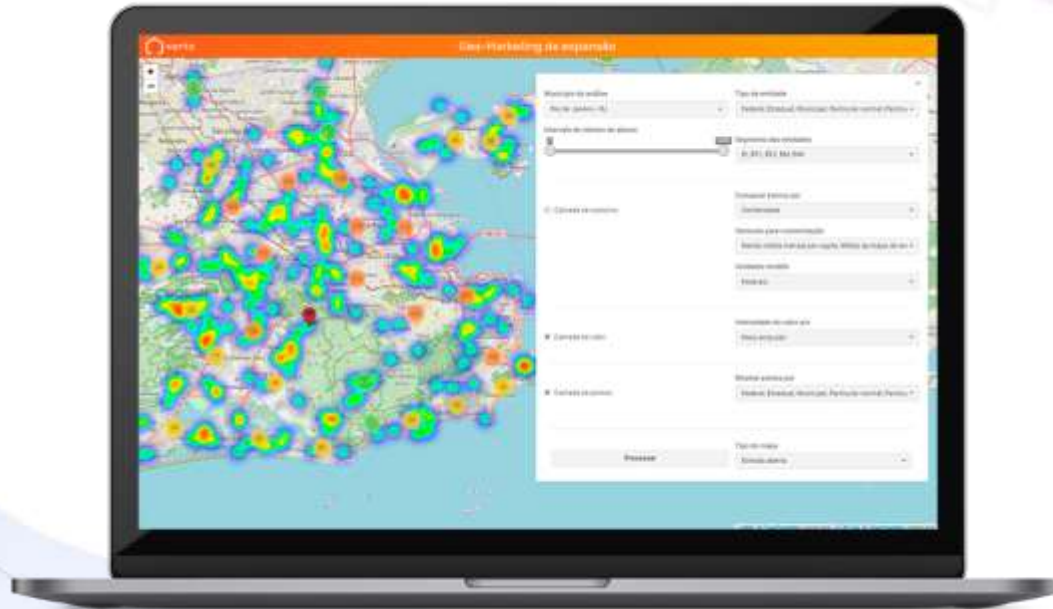
Segmentation Sharpness 25.35%

main

Size 100.00% - Distinctiveness 25.35%

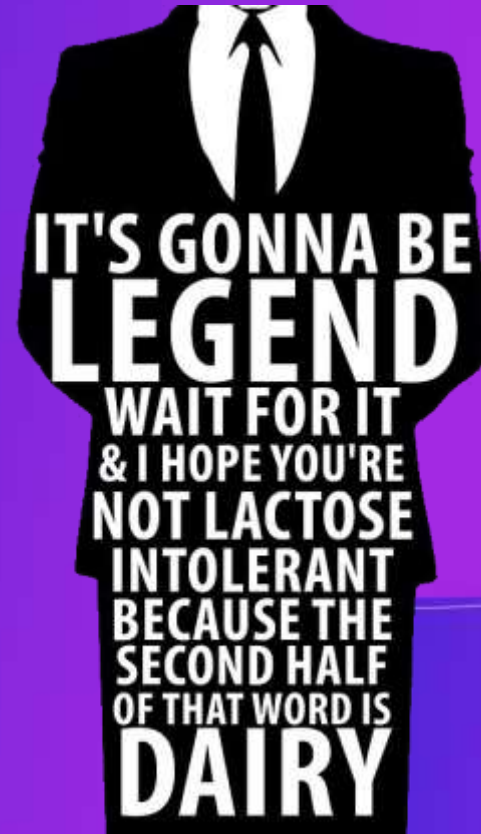


clusterização em SC

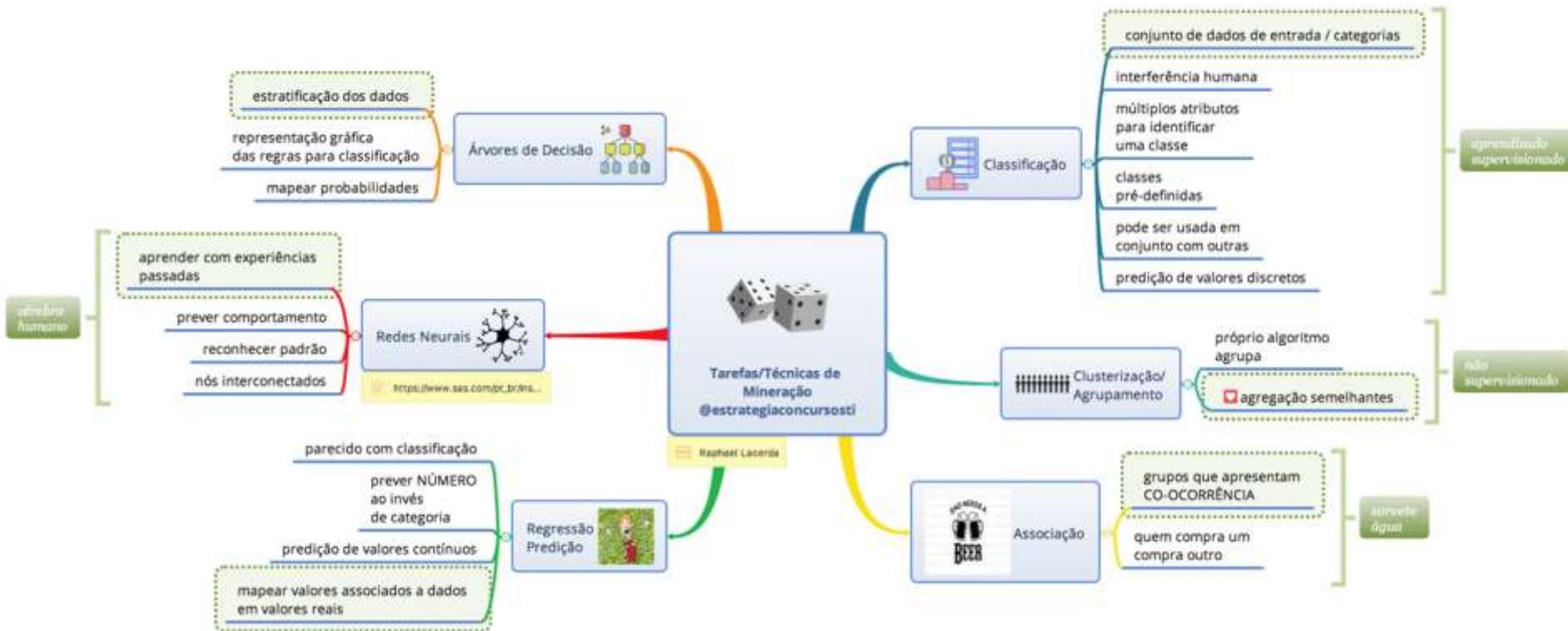


era pra entender?





Prof. Raphael Lacerda





Anomalias



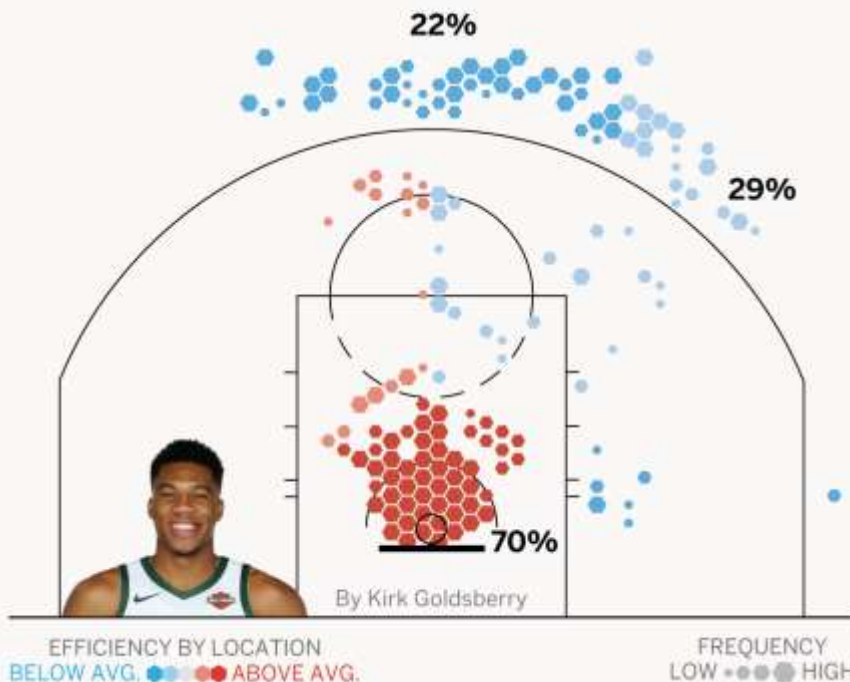
Most Unassisted Dunks

PLAYER	SEASON	DUNKS
Giannis Antetokounmpo	2018-19	116
Dwight Howard	2006-07	95
Shaquille O'Neal	2004-05	93

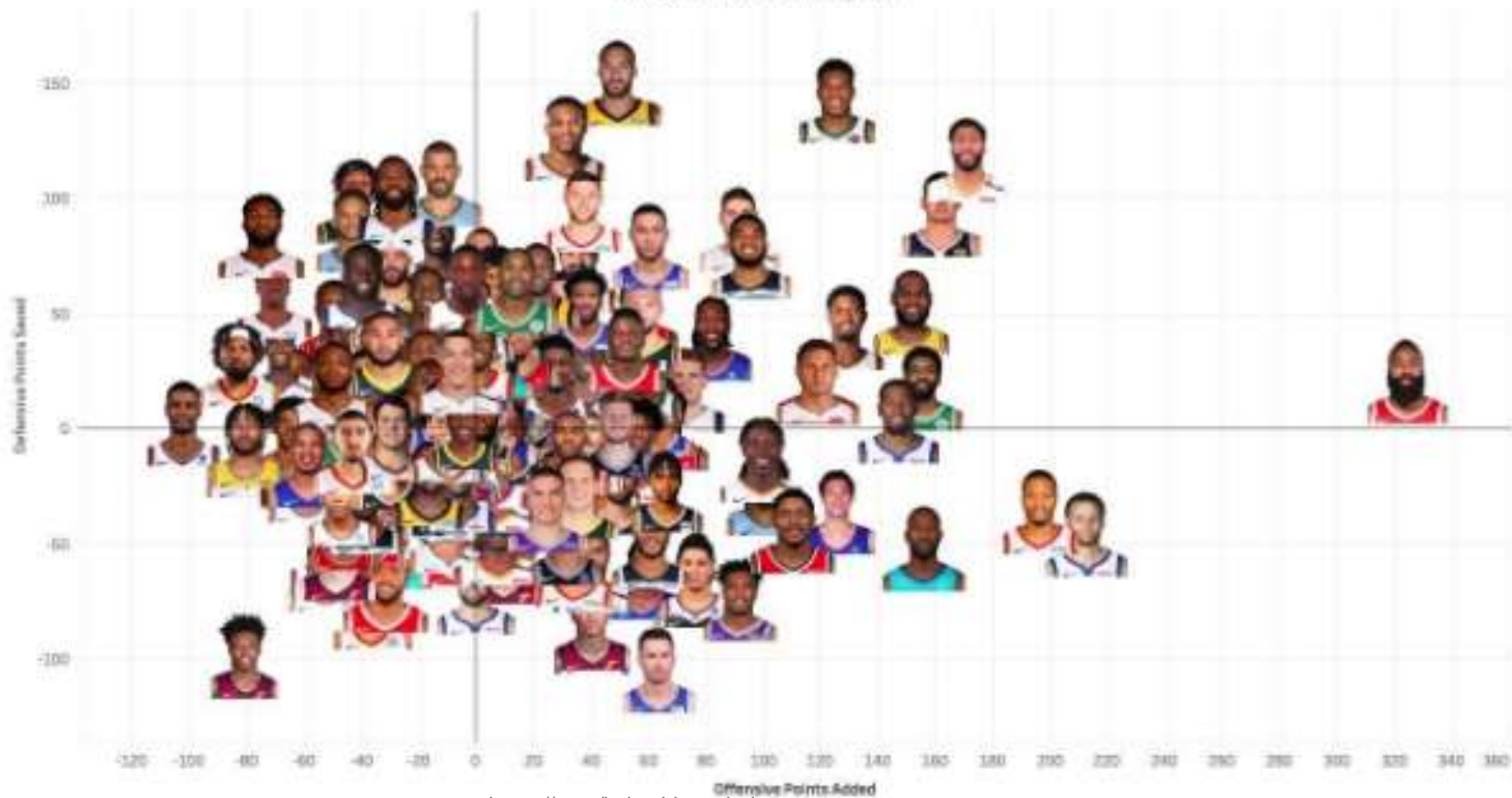
Since 1996-97

GIANNIS ANTETOKOUNMPO, 2018-19

The most active and efficient interior scorer in the NBA

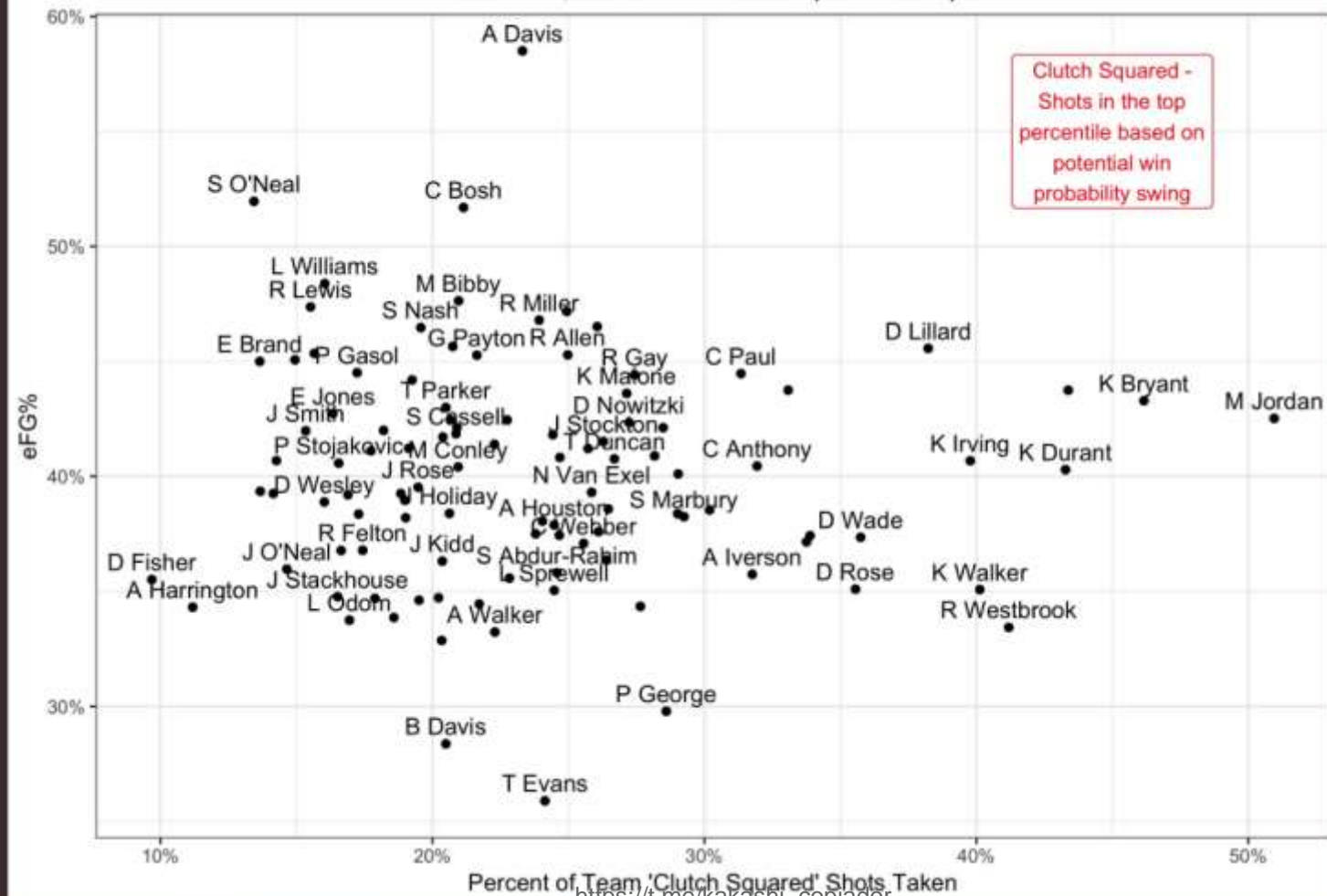


2018-19 TPA, All Players

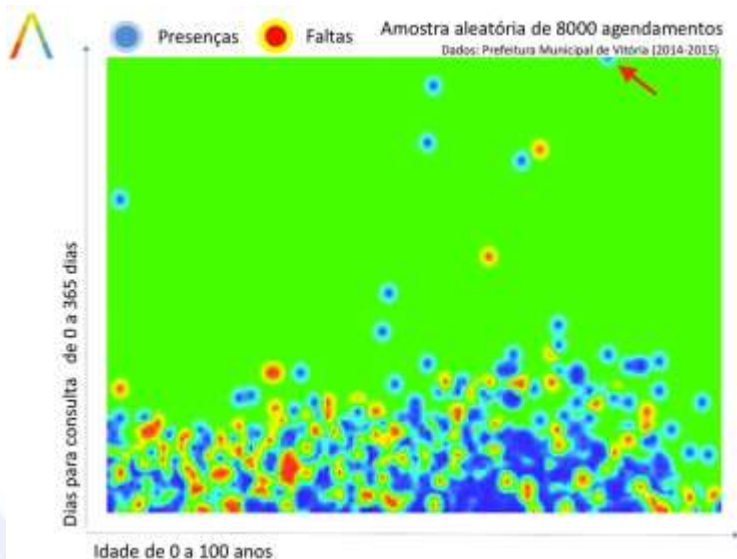


Shot Success and Usage in 'Clutch Squared' Situations

Since 1996, minimum 100 'clutch squared' attempts



clusterização em SC



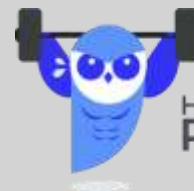
Cod	Nome Completo	Idade	Etnia	Valor da compra
1	José Carlos	33	branca	5500,00
2	Manoel da Silva	57	negra	2500,00
3	Maria Leite	27	indígena	3700,00
4	Antônio Silveira	470	branca	2900,00
5	Pedro Lemos	44	pardo	3300,00





MINERAÇÃO

QUESTÕES



HORA DE
PRATICAR!

<https://questoes.estrategiaconcursos.com.br/cadernos/d945a572-827e-4d6a-a1bf-305153785768>

Prof. Raphael Lacerda

EBSERH/2020

Dado os três conceitos técnicos abaixo, assinale a alternativa que corresponda respectivamente à tecnologia referente a cada um desses conceitos.

1. processo de explorar grandes quantidades de dados à procura de padrões consistentes.
2. refere-se ao processo de coleta, organização, análise, compartilhamento e monitoramento de informações que oferecem suporte a gestão de negócios.

EBSERH/2020

3. depósito de dados digitais que serve para armazenar informações detalhadas relativamente a uma empresa.

a) 1.Data Warehouse - 2.Business Intelligence - 3.Data Mining

b) 1.Data Mining - 2.Data Warehouse - 3.Business Intelligence

c) 1.Business Intelligence - 2.Data Warehouse - 3.Data Mining

d) 1.Data Mining - 2.Business Intelligence - 3.Data Warehouse

e) 1.Business Intelligence - 2.Data Mining - 3.Data Warehouse

rapha/2020

CRISP-DM é uma metodologia proprietária, baseada em 6 fases, entre elas a Business Understanding e Data Understanding, com o objetivo de implantar um projeto de data mining.

TCE-RO/2019

No que se refere aos diversos tipos de algoritmos utilizados para minerar dados, a técnica utilizada em tarefas de classificação, regressão e segmentação de dados em tipos de dados heterogêneos é denominada

A algoritmos genéticos.

B redes neurais.

C banco objeto relacional.

D classes privadas.

E redes privadas.

SANASA/2019

As ferramentas de Exploração identificadas como Ferramenta 1 e Ferramenta 2 na imagem, dentro do contexto a que se aplicam, são, dentre outras,

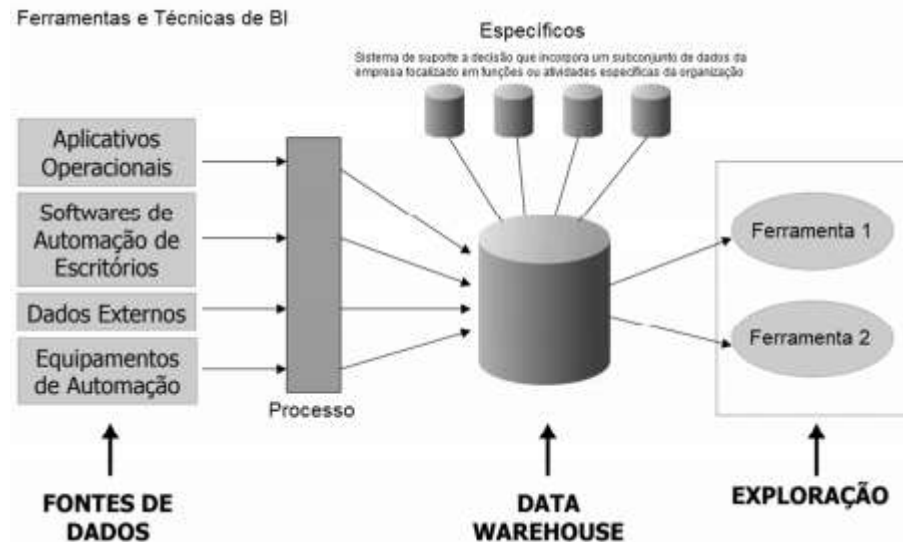
A Staging Area e Data Mining.

B OLAP e Data Mining.

C Snowflake e Staging Area.

D OLAP e Near Line Storage.

E Near Line Storage e Star Schema.



SEFAZ-BA/2019

Além dos indicadores reativos que, uma vez implantados, automaticamente detectam as ocorrências com base nos indicadores mapeados, existem também os controles proativos, que requerem que os gestores os promovam periodicamente. Uma das técnicas que os gestores podem usar requer que sejam selecionadas, exploradas e modeladas grandes quantidades de dados para revelar padrões, tendências e relações que podem ajudar a identificar casos de fraude e corrupção. Relações ocultas entre pessoas, entidades e eventos são identificadas e as relações suspeitas podem ser encaminhadas para apuração específica.

SEFAZ-BA/2019

As anomalias apontadas por esse tipo de técnica não necessariamente indicam a ocorrência de fraude e corrupção, mas eventos singulares que merecem avaliação individualizada para a exclusão da possibilidade de fraude e corrupção e, no caso da não exclusão, uma investigação.

O texto se refere à técnica de

A data mart. B data warehousing. C big data. D OLAP. E data mining.

TCM-BA/2018

A respeito das técnicas e(ou) métodos de mineração de dados, assinale a opção correta.

A O agrupamento (ou clustering) realiza identificação de grupos de dados que apresentam coocorrência.

B A classificação realiza o aprendizado de uma função que pode ser usada para mapear os valores associados aos dados em um ou mais valores reais.

ESTA É
DIFÍCIL!

TCM-BA/2018

C A regressão ou predição promove o aprendizado de uma função que pode ser usada para mapear dados em uma de várias classes discretas definidas previamente, bem como encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados.

D As regras de associação identificam grupos de dados, em que os dados têm características semelhantes aos do mesmo grupo e os grupos têm características diferentes entre si.

E Os métodos de classificação supervisionada podem ser embasados em separabilidade (entropia), utilizando árvores de decisão e variantes, e em particionamento, utilizando SVM (support vector machines).

TCM-BA/2018

Assinale a opção correta a respeito do CRISP-DM.

A

CRISP-DM é uma suíte de ferramentas proprietárias que vem se tornando um padrão da indústria para mineração de dados, uma vez que fornece um plano completo e tecnologias para a realização de um projeto de mineração de dados.

B

A verificação da qualidade dos dados é uma atividade da fase de entendimento dos dados.

TCM-BA/2018

C Durante a fase de preparação dos dados, é realizado um inventário de requisitos, suposições e restrições de recursos.

D Na fase de avaliação dos dados, são realizadas as atividades de identificar valores especiais dos dados e catalogar seu significado.

E Na fase de preparação dos dados, são realizadas as atividades de analisar o potencial de implantação de cada resultado e estimar o potencial de melhoria do processo atual.

ESTÁ É
DIFÍCIL!

IPHAN/2018

Na busca de padrões no data mining, é comum a utilização do aprendizado não supervisionado, em que um agente externo apresenta ao algoritmo alguns conjuntos de padrões de entrada e seus correspondentes padrões de saída, comparando-se a resposta fornecida pelo algoritmo com a resposta esperada.

TJ-AM/2019

Em sistemas de suporte à decisão, uma das abordagens da árvore de decisão é a predição, em que são criadas regras com base em eventos já ocorridos para aplicação em eventos futuros similares.



D_2P_2

TCM-MG/2018

Uma empresa, ao implementar técnicas e softwares de big data, deu enfoque diferenciado à análise que tem como objetivo mostrar as consequências de determinado evento.

Essa análise é do tipo

A preemptiva.

B perceptiva.

C prescritiva.

D preditiva.

E evolutiva.

SEFAZ-SC/2018

A aplicação de técnicas de mineração de dados (data mining) pode ser de grande valia para o Auditor. No caso das pesagens, por exemplo, uma ação típica de mining, que é passível de ser tomada com o auxílio de instrumentos preditivos, é

A quantificar as ocorrências de possíveis pesagens fraudulentas ocorridas durante todo o trimestre que antecede a data da análise, em alguns postos selecionados, mediante parâmetros comparativos preestabelecidos.

B realizar uma abordagem surpresa em determinado posto, com probabilidade significativa de constatar ocorrência fraudulenta.

SEFAZ-SC/2018

C reportar ao escalão superior as características gerais das pesagens e permanências de todos os caminhões, nos cinco maiores postos do Estado, no mês que antecede a data de análise.

D analisar o percentual de ocorrências das menores permanências de caminhões nos postos, no último ano, em relação ao movimento total.

E relacionar os postos onde ocorreram, nos últimos seis meses, as menores permanências das empresas suspeitas e informar o escalão superior para a tomada de decisão.

PF/2018

Situação hipotética: Na ação de obtenção de informações por meio de aprendizado de máquina, verificou-se que o processo que estava sendo realizado consistia em examinar as características de determinado objeto e atribuir-lhe uma ou mais classes; verificou-se também que os algoritmos utilizados eram embasados em algoritmos de aprendizagem supervisionados. Assertiva: Nessa situação, a ação em realização está relacionada ao processo de classificação.

TCE-RS/2018

O modelo de referência CRISP-DM tem seu ciclo de vida estruturado nas seguintes 6 fases:

A

Estruturação do Negócio, Limpeza dos Dados, Indicação das Métricas, Modelagem, Estimativa e Exportação dos Dados.

B

Otimização do Negócio, Redução dos Dados, Replicação dos Dados, Modelagem, Importação dos Dados e Backup.

TCE-RS/2018

C Entendimento do Negócio, Entendimento dos Dados, Preparação dos Dados, Modelagem, Avaliação e Implantação.

D Preparação do Negócio, Replicação dos Dados, Indexação dos Dados, Diagramação do Negócio, Estimativa e Organização.

E Otimização do Negócio, Entendimento dos Dados, Indexação dos Dados, Exportação dos Dados, Organização e Importação dos Dados.

EBSERH/2018

A descoberta de novas regras e padrões em conjuntos de dados fornecidos, ou aquisição de conhecimento indutivo, é um dos objetivos de data mining.

STJ/2018

O processo de mineração de dados está intrinsecamente ligado às dimensões e a fato, tendo em vista que, para a obtenção de padrões úteis e relevantes, é necessário que esse processo seja executado dentro dos data warehouses.

ESTA É
DIFÍCIL!

PF/2018

1 - Pode-se definir mineração de dados como o processo de identificar, em dados, padrões válidos, novos, potencialmente úteis e, ao final, compreensíveis.

2 - Descobrir conexões escondidas e prever tendências futuras é um dos objetivos da mineração de dados, que utiliza a estatística, a inteligência artificial e os algoritmos de aprendizagem de máquina.

DPE-AM/2018

Dentre os algoritmos utilizados em data mining, há um algoritmo que visa o estabelecimento de categorias, a partir do conjunto de dados, bem como a distribuição dos dados nas categorias estabelecidas. Essa descrição corresponde aos algoritmos de

A classificação.

B sumarização.

C visualização.

D evolução.

E detecção de desvios.

TCE-PE/2017

Durante a fase de entendimento do negócio, busca-se descrever claramente o problema, fazer a identificação dos dados e verificar se as variáveis relevantes para o projeto não são interdependentes.

CIASC/2017

Árvores de decisão, redes neurais e modelos de regressão empregam aprendizado supervisionado para criar a função de mapeamento entre um conjunto de campos de dados de entrada e uma variável de destino.

CIASC/2017

Assinale a alternativa que contém as principais fases do processo de Data Mining CRISP-DM.

A Amostragem; Exploração; Modificação; Modelagem; Execução; Avaliação.

B Amostragem; Exploração; Modelagem; Modificação; Avaliação; Implementação.

C Compreensão do negócio; Compreensão dos dados; Preparação dos dados; Modelagem; Avaliação; implementação.

D Compreensão dos dados; Amostragem; Preparação dos dados; Implementação; Avaliação.

E Compreensão do negócio; Exploração dos dados; Modificação dos dados; Implementação; Avaliação.

CRF-SP/2017

Acerca dos métodos tradicionais de Mineração de Dados, um desses métodos “produz Árvores de Decisão a partir de uma abordagem recursiva de particionamento de um conjunto de dados, e utiliza conceitos e medidas da Teoria da Informação”. Trata-se do método:

A C4.5.

B Apriori.

C Máquinas de Vetores Suporte.

D Classificador Bayesiano Ingênuo.

AL-MS/2016

Um famoso site de vendas sempre envia ao cliente que acabou de comprar um item X, ou o está analisando, a seguinte frase: Pessoas que compraram o item X também compraram o Y. Para isso, o site deve estar aplicando a técnica de Data Mining denominada

A profiling.

B coocorrência.

C regressão múltipla.

D regressão logística.

E classificação.

ANAC/2016

São objetivos da Mineração de Dados:

A Distribuição, Identificação, Organização e Otimização.

B Previsão, Priorização, Classificação e Alocação.

C Previsão, Identificação, Classificação e Otimização.

D Mapeamento, Identificação, Classificação e Atribuição.

E Planejamento, Redirecionamento, Classificação e Otimização.

TCE-PA/2016

CRISP-DM é uma metodologia proprietária que identifica as fases Business Understanding e Data Understanding na implantação de um projeto de data mining.

CNMP/2015

As ferramentas de Data Mining permitem ao usuário avaliar tendências e padrões não conhecidos entre os dados. Esses tipos de ferramentas podem utilizar técnicas avançadas de computação como redes neurais, algoritmos genéticos e lógica nebulosa, dentre outras.

Data Mining é o processo de descobrir conhecimento em banco de dados, que envolve várias etapas. O KDD – Knowledge Discovery in Database é uma destas etapas, portanto, a mineração de dados é um conceito que abrange o KDD.

TCU/2015

- 1 - Devido à quantidade de informações manipuladas, a (cloud computing) computação em nuvem torna-se inviável para soluções de big data.
- 2 - No ambiente organizacional, devido à grande quantidade de dados, não é recomendado o emprego de data mining para atividades ligadas a marketing.
- 3 - A finalidade do uso do data mining em uma organização é subsidiar a produção de afirmações conclusivas acerca do padrão de comportamento exibido por agentes de interesse dessa organização.
- 4 - O uso prático de data mining envolve o emprego de processos, ferramentas, técnicas e métodos oriundos da matemática, da estatística e da computação, inclusive de inteligência artificial.

METRÔ-DF/2014

A mineração de texto consiste basicamente na extração de informação de qualidade a partir de textos em linguagem natural. Esse processo possui normalmente cinco fases principais. Com relação à fase que permite a recuperação da informação minerada, assinale a alternativa correta.

A Coleta.

B Pré-processamento. C Indexação.

D Algoritmo. E Análise de resultados.

Correios/2011

BI é o processo de coleta, transformação, análise e distribuição de dados, coletados em informações estratégicas, para tomada de decisões nas empresas, incorporando o conceito de gerenciamento de dados e permitindo extrair dados arquivados em vários sistemas, identificá-los, definir padrões, detectar tendências e fazer previsões.

DATAPREV/2009

“Mining é parte de um processo maior de conhecimento, que o processo consiste, fundamentalmente, na estruturação do banco de dados; na seleção, preparação e pré- processamento dos dados; na transformação, adequação e redução da dimensionalidade dos dados; e nas análises, assimilações, interpretações e uso do conhecimento extraído do banco de dados.”

O processo maior citado no início do texto é denominado:

A Data mining; B Data mart; C Data warehouse;

D KDD; E Segmentação de dados.

várias

Na técnica de árvore de decisão em data mining, é empregada a abordagem denominada

A análise de volumetria.

B combinação de variáveis.

C estratificação.

D avaliação de dados.

E percepção.

<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#2d551d266686>

várias

Em DataMining, as árvores de decisão podem ser usadas com sistemas de classificação para atribuir informação de tipo.

Com o uso da classificação como técnica de Data Mining, busca-se a identificação de uma classe por meio de múltiplos atributos. Essa técnica também pode ser usada em conjunto com outras técnicas de mineração de dados.

FUB/2018

No data mining, uma regra de associação relaciona a presença de um conjunto de itens com outra faixa de valores de um outro conjunto de variáveis.

FUB/2018

Agrupar registros em grupos, de modo que os registros em um grupo sejam semelhantes entre si e diferentes dos registros em outros grupos é uma maneira de descrever conhecimento descoberto durante processos de mineração de dados.

Pref. PR/2019

Os métodos de agrupamento visam reunir objetos similares e dissimilares a objetos pertencentes a outros grupos.

ITAIPU/2019

Castro e Ferrari (2016) mencionam que as funcionalidades da mineração de dados são utilizadas para especificar os tipos de informação a serem obtidos nas tarefas de mineração. Sobre esse tema, considere as seguintes funcionalidades:

1. Análise de grupos: tarefa supervisionada que separa um conjunto de objetos em grupos.
2. Classificação: tarefa supervisionada para predição de valores discretos.
3. Regressão (ou estimação): tarefa supervisionada para predição de valores contínuos.
4. Associação: tarefa que descobre relações entre atributos por meio da detecção de anomalias nos conjuntos de dados.

É/São funcionalidade(s) da etapa de mineração:

ITAIPU/2019

Os algoritmos de Mineração de Dados podem ser classificados quanto a seus objetivos, sendo alguns a classificação, o agrupamento e a identificação de regras de associação. A respeito dessas classificações e seus algoritmos, assinale a alternativa correta.

A Algoritmos de agrupamento podem ser utilizados para classificação não supervisionada.

B Algoritmos de agrupamento são também chamados de algoritmos supervisionados.

ITAIPU/2019

C

Algoritmos de classificação têm como resultado um modelo descritivo dos dados de entrada.

D

Algoritmos de identificação de regras são também conhecidos como algoritmos preditivos.

E

Algoritmos de agrupamento são equivalentes a algoritmos de identificação de anomalias.

SANASA/2019

Considere que a SANASA busca realizar a gestão de recursos hídricos subterrâneos com base em parâmetros conhecidos que determinam a poluição das águas subterrâneas. Um desses parâmetros, para exemplificar, seria o nitrato, um indicador de poluição difusa de água subterrânea. Criando-se regras para realizar o aprendizado supervisionado do sistema de Data Mining utilizando-se uma certa técnica, chegar-se-á a um resultado que considera os diversos parâmetros para se descobrir se um certo aquífero tem água potável ou não, comparando-se com uma definição conhecida.

SANASA/2019

Nesse cenário, a técnica aplicada é denominada

A Associação.

B Classificação.

C Clustering.

D Regressão.

E Prediction.

TRF4/2019

Um Tribunal pretende analisar fatos (fatores ambientais e perfis profissionais, entre outros) que esclareçam por que alguns colaboradores se destacam profissionalmente enquanto outros não se desenvolvem e acabam por se desligar do órgão. Para facilitar essa análise, o Tribunal solicitou um auxílio tecnológico que indique quais características nos fatos apresentam razões positivas que justifiquem investimentos mais robustos no treinamento de colaboradores que tendem a se destacar a médio e longo prazos.

TRF4/2019

Para tanto, o Analista implantará um processo de análise científica preditiva com base em dados estruturados, que consiste na obtenção de padrões que expliquem e descrevam tendências futuras, denominado

A snowflake. B drill over. C star schema. D slice accross.

E data mining.

SEFAZ-BA/2019

“A Secretaria da Fazenda realizou procedimento que visa estimular os contribuintes paulistas a, voluntariamente, regularizarem as obrigações tributárias que devem ser transmitidas ao Fisco. Em uma etapa da operação foram selecionados 48 contribuintes do Regime Periódico de Apuração com R\$ 143 milhões em débitos de ICMS.

SEFAZ-BA/2019

Esta ação tem caráter orientador, com o objetivo de alertar as empresas sobre divergências em suas declarações e indicar a regularização, porém, a ação pode resultar em indicações de empresas de fachada que apresentaram um conjunto de indícios que sugerem tratar-se de documentos fiscais inidôneos ou de simulação de operações para gerar créditos falsos de ICMS na apuração mensal do tributo a recolher.

SEFAZ-BA/2019

Para a elaboração da operação autorregularização foram selecionados contribuintes com divergências nas notas fiscais emitidas entre janeiro a dezembro de 2016 em comparação com as informações declaradas nas Guias de Informação e Apuração do ICMS do mesmo período."

(Baseado em: <https://portal.fazenda.sp.gov.br>)

SEFAZ-BA/2019

Ao ler esta notícia, uma Auditora Fiscal da área de TI concluiu, corretamente, que o cruzamento e a análise simultânea de diversas informações, como Nota Fiscal Eletrônica (NF-e), inadimplência, regularidade no cumprimento de obrigações, porte da empresa, composição do quadro societário, autos de infração anteriores, localização geográfica e atividade econômica, entre outras,

A caracterizam-se como operações MOLAP, que usam a base de dados relacional para acessar os dados do cubo.

B caracterizam-se como operações HOLAP, que não usam a base de dados relacional para acessar os dados do cubo.

SEFAZ-BA/2019

C podem ter sido realizados com a ajuda de ferramentas de Data Marts integrados, que não requerem a construção de um DW.

D podem ter sido realizados com ferramentas de Data Mining em tempo real, uma vez que os dados do DW são constantemente atualizados a partir da chave de tempo que indica o dia no qual os dados foram extraídos dos sistemas transacionais.

E podem ter sido realizados com a ajuda de ferramentas de Data Mining, que permitem a exploração de grandes volumes de dados para identificar padrões de comportamento e relacionamentos.

BASA/2018

As ferramentas e técnicas de mineração de dados (data mining) têm por objetivo

A preparar dados para serem utilizados em um “data warehouse” (DW).

B permitir a navegação multidimensional em um DW.

C projetar, de forma eficiente, o registro de dados transacionais.

D buscar a classificação e o agrupamento (clusterização) de dados, bem como identificar padrões.

E otimizar o desempenho de um gerenciador de banco de dados.

Transpetro/2018

Classificação é uma importante tarefa utilizada na etapa de mineração de dados, que tem como uma de suas características básicas

A construir seus modelos de enquadramento, a partir de um conjunto de dados contínuos.

B poder ser implementada por algoritmos estáveis e de significativa eficácia, tais como C4.5, classificadores bayesianos ou K-Prototypes.

Transpetro/2018

C ser um método de aprendizado de máquina não supervisionado, observando o teorema NFL – No Free Lunch.

D ter a sua eficácia avaliada por uma métrica denominada suporte, que indica quantas vezes um item de dado foi corretamente classificado.

E ter como seu primeiro processo o aprendizado de uma função de mapeamento $y = f(X)$, que associa uma ocorrência de dados X em uma classe y .

Transpetro/2018

Um desenvolvedor recebeu um conjunto de dados representando o perfil de um grupo de clientes, sem nenhuma informação do tipo de cada cliente, onde cada um era representado por um conjunto fixo de atributos, alguns contínuos, outros discretos. Exemplos desses atributos são: idade, salário e estado civil. Foi pedido a esse desenvolvedor que, segundo a similaridade entre os clientes, dividisse os clientes em grupos, sendo que clientes parecidos deviam ficar no mesmo grupo. Não havia nenhuma informação que pudesse ajudar a verificar se esses grupos estariam corretos ou não nos dados disponíveis para o desenvolvedor. Esse é um problema de data mining conhecido, cuja solução mais adequada é um algoritmo

A de regressão B não supervisionado C por reforço D semissupervisionado
E supervisionado

SABESP/2018

O conceito de Data Mining descreve

A o uso de teorias, métodos, processos e tecnologias para organizar uma grande quantidade de dados brutos para identificar padrões de comportamentos em determinados públicos.

B o conjunto de métodos, tecnologias e estratégias para atração voluntária de visitantes, buscando a conversão consistente de leads em clientes (realização de compra).

SABESP/2018

C as atividades coordenadas de modo sistemático por uma determinada organização para relacionamento com os seus distintos públicos, bem como com outras organizações, sejam públicas, privadas ou não governamentais.

D o conjunto de tarefas e processos, organizados e sistematizados, normalmente como uso de uma plataforma tecnológica (hardware e software, ou até mesmo em cloud computing) para a gestão do relacionamento com clientes.

SABESP/2018

E

o trabalho de produzir levantamento sobre os hábitos de consumo de mídia de um determinado público, identificando horários, tempo gasto etc., associando ao perfil socioeconômico, potencial de consumo, persuasão etc.

PRODEB/2018

Classificação é o processo de encontrar um modelo que descreva classes diferentes de dados. As classes são predeterminadas, por exemplo, em uma aplicação bancária, clientes que possuam um cartão de crédito podem ser classificados como “risco baixo”, “risco justo”, ou “risco alto”. Esse tipo de atividade é também chamada, em modelagem de dados, de

A estratégia B planejamento C esquema D data mining

E aprendizado supervisionado

CRF-SP/2018

“A etapa de Mineração de Dados compreende a busca efetiva por conhecimentos úteis no contexto da aplicação de KDD (Knowledge Discovery in Database), ou Descoberta do Conhecimento em Bases de Dados. É a principal etapa do processo de KDD.”

Acerca de algumas das tarefas do KDD, analise a assertiva a seguir:

“compreende a busca por uma função que mapeie os registros de um banco de dados em um intervalo de valores reais”.

CRF-SP/2018

Assinale a alternativa que apresenta esta tarefa.

A Regressão.

B Classificação.

C Sumarização.

D Agrupamento.

CLDF/2018

No contexto de Business Intelligence (BI) pode-se descobrir interesses de cunho político dos cidadãos pelo perfil de cada um, categorizando-os por faixa etária, escolaridade, poder aquisitivo ou outras particularidades típicas. A captura dos perfis pode ser feita, por exemplo, por meio de uma aplicação Web que desperte o interesse de acesso ao conteúdo do site pelos cidadãos em troca do fornecimento de seus dados pessoais, tornando possível realizar cruzamentos e associações desses dados e então tomar decisões a respeito. Esse processo refere-se a uma aplicação típica da tecnologia denominada

A data mart. B dashboard C OLTP D data mining. E star schema

UFRGS/2018

A análise de _____ permite estudar a relação entre dois conjuntos de valores e quantificar o quanto um está relacionado com o outro, no sentido de determinar a intensidade e a direção dessa relação. Isto é, essa análise indica se, e com que intensidade, os valores de uma variável aumentam ou diminuem enquanto os valores da outra variável aumentam ou diminuem.

Assinale a alternativa que completa corretamente a lacuna do texto acima.

A correlação B dispersão C classificação D agrupamento E regressão



**UFA..
CHEGA NÉ?
ACABOU?**



Prof. Raphael Lacerda