

Business Analytics

Marcelo Rosano Dallagassa

2024

Business Analytics

Apresentar Conceitos de Business Intelligence, Business Analytics, Data Warehouse, Data Mart, Modelagem Multidimensional, Visualização de Dados, Construção de Dashboard, Ciência de Dados, baseado no processo de aprendizagem de máquina por meio da técnica de KDD (Descoberta de conhecimento de base de dados), apresentando as suas tarefas e algoritmos com cases e exercícios.

Dentro da proposta, exercita-se a construção de um artigo com uma prática na utilização de Analytics.

Estrutura da Disciplina

- Business Intelligence, Business Analytics;
- Visualização de Dados, Introdução a Ciência de Dados e Aprendizagem de Máquina.
- Processo de KDD (Knowledge Discovery in Data Base).
- Tarefa De Classificação, Algoritmo de Árvore de Decisão e Baseado em Regras, Exercícios.
- Redes Neurais Artificiais
- Tarefa de Regra de Associação, Algoritmo Apriori
- Tarefa de Agrupamento - K-Means, Exercícios
- Séries Temporais - Regressão Linear - Redes Neurais Artificiais e Deep Learning
- Mineração de Processos e suas aplicações
- Apresentação de Trabalhos e Avaliação Final

Avaliação do Módulo

Atividade	Pontos
Desenvolvimento e apresentação de Artigo - Experimento de Analytics	40
Avaliação Final	30
Participação em Aula	30

Business Intelligence x Analytics

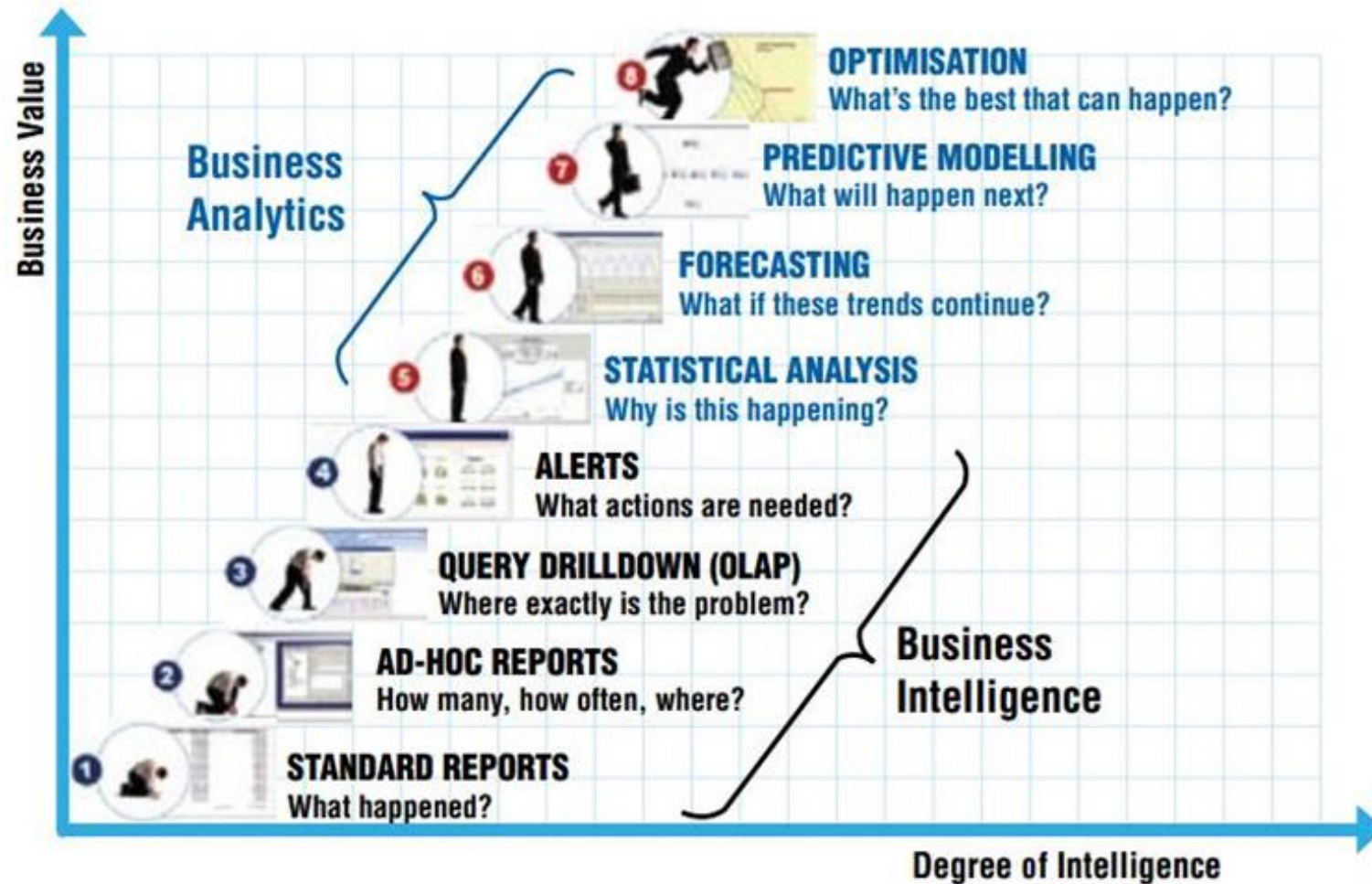


Figura 2- BI x BA - Michael Walken - <http://www.rosebt.com/1/post/2012/09/eight-levels-of-analytics-for-competitive-advantage.html>

Business Intelligence x Analytics

Objeto	<i>Business Intelligence</i>	<i>Business Analytics</i>
Finalidade	Relatórios, KPIs, Dashboards	Identificação de Padrões, correlações, modelos preditivos
Processo	Estático, comparativo	Exploratório, experimental, visual
Qualidade dos dados	Alta	Baixa ou Média (exige processo de limpeza e transformação)
Tipo de Análise	Descritiva, retrospectiva	Preditiva, Prescritiva
Tipo de respostas	O que aconteceu?	O que pode acontecer?

Ferramentas Analytics



Amazon SageMaker

Weka



Plataforma Independente

Software Livre (GPL)

Diversos Algoritmos (ML)

Facilidade de Utilização

Ferramenta de Pré-processamento

Interface Gráfica do Usuário

Interoperabilidade (R, Python, Spark e outras)

Weka Instalação



Weka Website (última versão 3.8 ou 3.9)



Download

<https://www.cs.waikato.ac.nz/ml/weka/>



Escolher ambiente (Windows, MAC ,
Linux)

Orange

Plataforma Independente

Software Livre (GPL)

Diversos Algoritmos (ML)

Facilidade de Utilização

Ferramenta de Pré-processamento

Interface Gráfica do Usuário

Fluxo de construção de análises de dados



Orange



Download








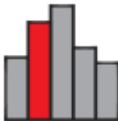
<https://orangedatamining.com/>



Escolher ambiente (Windows, MAC, Linux)

Framework Anaconda

Applications on base (root) Channels

 CMD.exe Prompt 0.1.1 Run a cmd.exe terminal with your current environment from Navigator activated Launch	 JupyterLab 2.1.5 An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture. Launch	 Notebook 6.0.3 Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis. Launch	 Orange 3 3.26.0 Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox. Launch
 Powershell Prompt 0.0.1 Run a Powershell terminal with your current environment from Navigator activated	 Qt Console 4.7.5 PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.	 Spyder 4.1.4 Scientific PYthon Development EnviRnment. Powerful Python IDE with advanced editing, interactive testing,	 Glueviz 1.0.0 Multidimensional data visualization across files. Explore relationships within and among related datasets.

Colab Research

Serviço em nuvem - Google

Gratuito e hospedado Google

Aprendizagem de Máquina e Inteligência Artificial

Facilidade de Utilização

Criação e execução de códigos Python

Interface vinculada com ferramentas Google

Fluxo de construção de análises de dados

Colab Research Google



Colab



Acesso em
colab.research.google.com



Pronto!

A background image showing a group of business professionals in suits walking through a modern office building with large glass windows. The image is partially obscured by a green semi-transparent rectangle containing the title.

Data Mining Ciência de Dados

.....

- Extração não trivial da informação implícita, previamente desconhecida e potencialmente útil a partir dos dados.
- Exploração e análise , por meio semiautomáticos ou automáticos, de grande quantidade de dados a fim de descobrir padrões significativos.

Ciência de Dados



“É um processo não trivial para identificar padrões que sejam válidos e previamente desconhecidos, potencialmente úteis e compreensíveis, visando melhorar o entendimento do problema, possibilitando o auxílio no processo de tomada de decisões” (FAYYAD et al., 1996).

A descoberta de uma **informação preciosa** que pode-se transformar em um conhecimento novo

Ciência de Dados

Justificativas



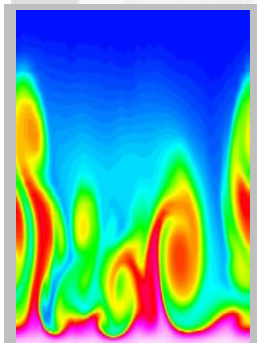
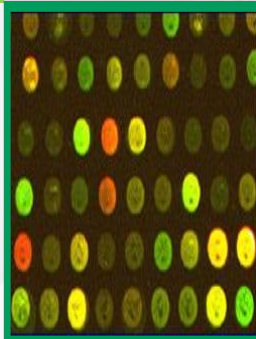
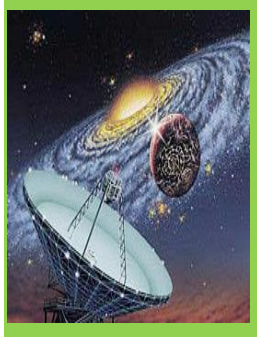
Grande volume de dados armazenados:

- ✓ Web data, e-commerce
- ✓ Compras (departamentos, Supermercados)
- ✓ Bancos, transações com cartão de crédito,...

Outras:

- ✓ Computadores mais acessíveis e mais potentes
- ✓ Mercado mais competitivo
- ✓ Fornecimento de produtos e serviços com mais qualidade, personalizados e com menores custos.

Ciência de Dados

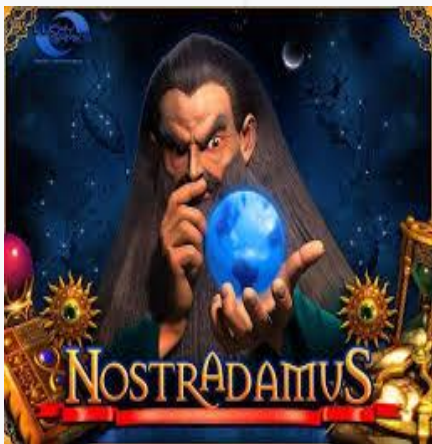


Do Ponto de Vista Científico

- ✓ Dados coletados a enormes velocidades (GB/hora)
- ✓ Sensores remotos em satélites
- ✓ Telescópios investigando o universo
- ✓ Micro arranjos gerando dados de expressão
- ✓ gênica
- ✓ Simulações científicas gerando terabytes de informações

Técnicas tradicionais inviáveis para dados brutos
A Ciência de dados auxilia os cientistas em:

- ✓ Classificação e segmentação de dados
- ✓ Na formulação de hipóteses.




Ciência de Dados

Como serão os custos de produção do próximo ano?

Quais as Doenças que mais acometem os nossos Pacientes?

Será que trata-se de um caso de fraude?

Como será o comportamento desse cliente nos próximos 6 meses?

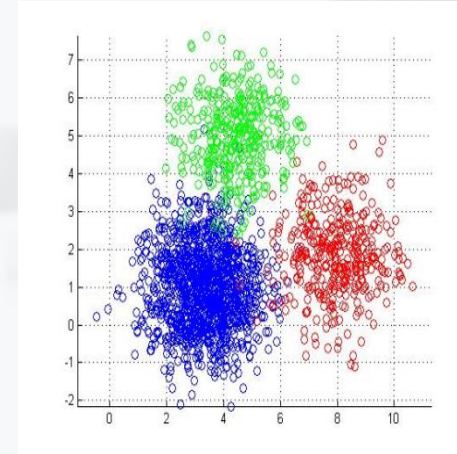
A group of business professionals in suits are walking towards the camera. In the background, there is a modern building with a glass facade. A semi-transparent green rectangular box is overlaid on the image, containing the title text.

Métodos de Aprendizagem de Máquina

Aprendizagem de Máquina - Métodos

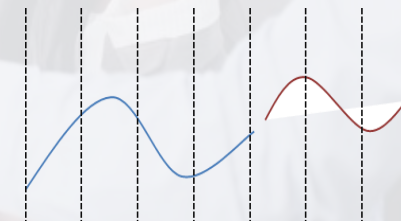
Métodos Descritivos


Encontrar padrões compreensíveis por humanos para descrever os dados



Métodos Preditivos

Usar algumas variáveis para prever valores desconhecidos ou futuros de outras variáveis

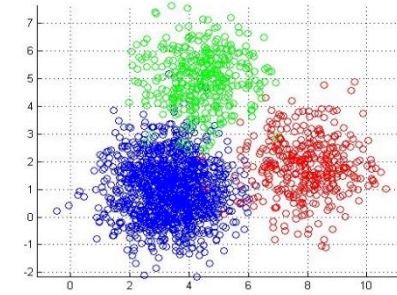
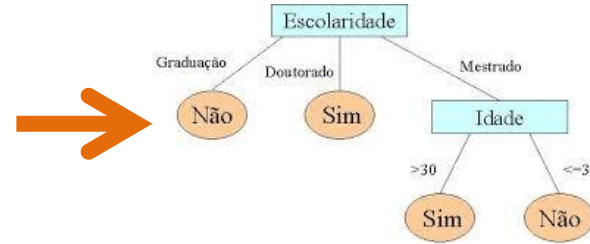


A group of business professionals in suits are walking towards the camera. In the background, there is a modern building with a glass facade. A semi-transparent green rectangular box is overlaid on the image, containing the title text.

Tarefas de Aprendizagem de Máquina

Tarefas de Aprendizagem de Máquina

Classificação (Preditivo)

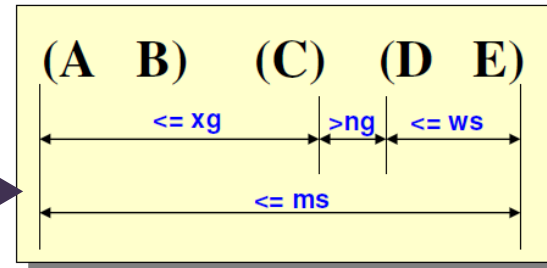


Agrupamento (Descritivo)

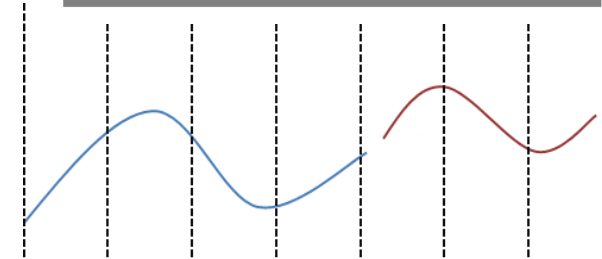
Regra de Associação (Descritivo)

leite, fralda => cerveja

Descoberta de Padrões Sequenciais (Descritivo)



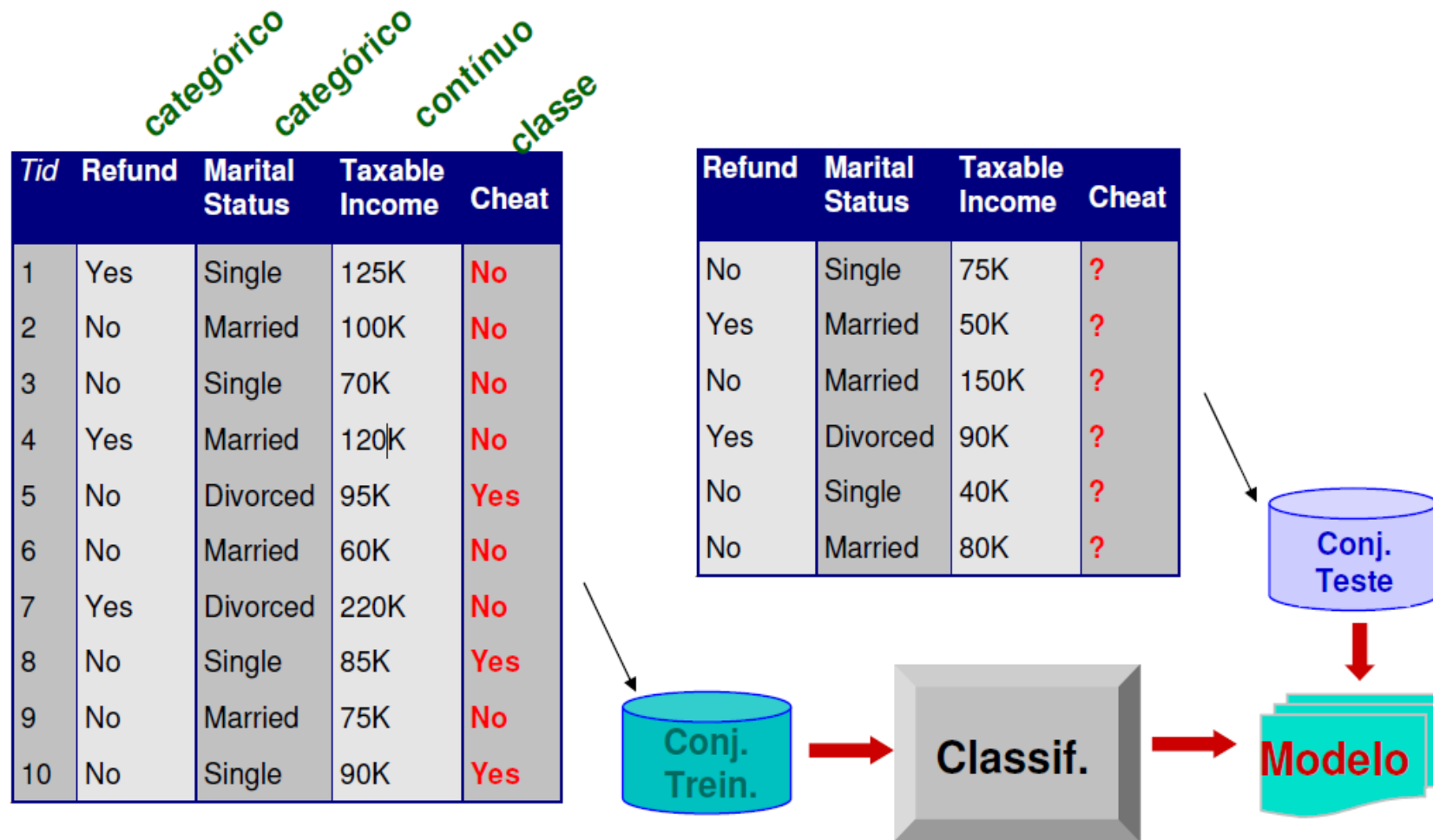
Regressão - Séries Temporais (Preditivo)



Tarefa de Classificação

- Dado um conjunto de registros (conjunto de treinamento)
 - Cada registro contém um conjunto de atributos, um dos atributos é a *classe*.
- Encontrar um *modelo* para o atributo classe em função dos valores dos outros atributos.
- Objetivo: Para registros previamente não rotulados, deve ser assinalada uma classe tão precisa, quanto possível.
- Um *conjunto de teste* é usado para determinar a precisão do modelo.
Geralmente o conjunto de dados é dividido em conjunto de treinamento e de testes, sendo o de treinamento usado para construir o modelo e o de teste para validá-lo.

Tarefa de Classificação



Classificação - Aplicações

- **Marketing Direto**

-Objetivo: Reduzir custo de propaganda escolhendo um conjunto de clientes que provavelmente irão comprar um novo celular.

-Abordagem:

- Usar os dados de maneira similar ao exemplo anterior
- Conhece-se quais clientes decidiram comprar o produto e quais não. Esta decisão {comprar, não-comprar} forma o atributo classe.
- Coletar várias informações demográficas, de estilo de vida, e de interação com a empresa relacionadas a todos os clientes. (exemplo: tipo de negócio, onde estão, quanto recebem, etc)
- Usar esta informação como atributos de entrada para treinar um modelo de um classificador.

Classificação - Aplicações

- **Detecção de Fraude**

- Objetivo: Prever casos fraudulentos em transações de cartão de crédito.

- Abordagem:

- Usar transações de cartão de crédito e a informação sobre os clientes como atributos

- Quando um cliente compra, o que ele compra, quão frequentemente ele paga em dia, etc...

- Rotular as transações passadas como transação do tipo {fraude ou honesta}.
Formação do atributo classe.

- Usar este modelo para detecção de fraudes observando as transações de cartão de crédito sobre uma determinada conta.

Classificação - Aplicações

- **Insatisfação de Clientes:**

- Objetivo: Prever se um cliente tem propensão para migrar para um concorrente.

- Abordagem:

- Usar registros detalhados de transações de cada um dos clientes passados e atuais, para encontrar atributos.

- Quão frequentemente o cliente liga, para qual setor, em qual horário, estado financeiro, estado civil, ocupação profissional, etc...

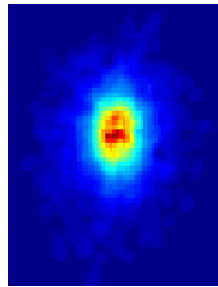
- Rotular o Cliente como tipo {fiel ou não-fiel}.

- Encontrar um modelo adequado.

Classificação - Aplicações

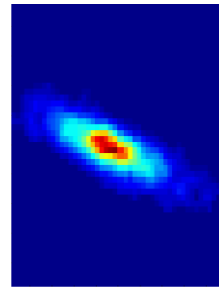
- **Classificando Imagens:**

Inicial



Atributo Classe:
Estágio de Formação

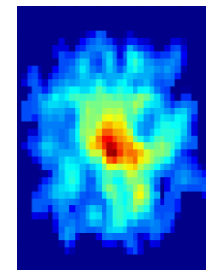
Intermediário



Atributos:

- Característica da imagem
- Característica das ondas de luz recebidas, ...

Final



Quantidade de dados:

- 72 milhões de estrelas, 20 milhões de galáxias
- Catálogo de objetos: 9 GB
- Base de dados das imagens: 150 GB

Tarefas de Agrupamento

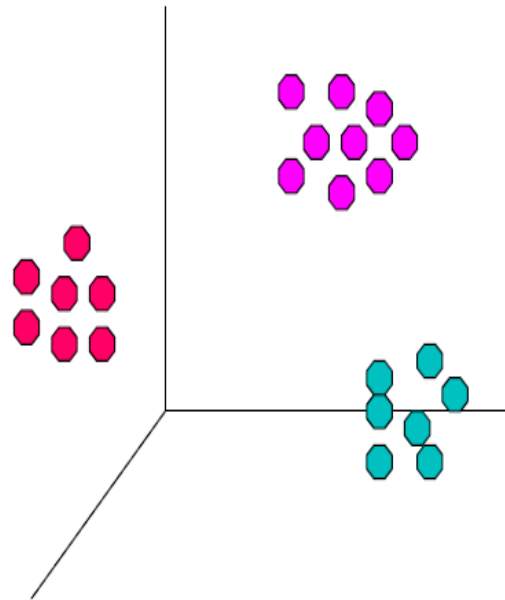
- Dado um conjunto de ponto de dados, cada um tendo um conjunto de atributos e uma medida de similaridade entre eles, encontrar agrupamentos tais que:
 - Pontos de dados em um grupo são mais similares entre si.
 - Pontos de dados em grupos diferentes são menos similares entre si.
- Medidas de Similaridade:
 - Distância euclidiana se os atributos são contínuos
 - Outras medidas dependentes do problema.

Tarefa de Agrupamento

- Agrupamento baseado em distância euclidiana no espaço 3-D

Distâncias intra-grupos
são minimizadas

Distâncias inter-grupos
são maximizadas



Agrupamento - Aplicações

- **Segmentação de Mercado:**

Objetivo: subdividir um mercado em distintos sub-conjuntos de clientes em que cada subconjunto pode ser visto como um mercado-alvo a ser atingido com uma mistura de comunicação (propaganda) distinta.

Abordagem:

- Coletar diferentes atributos de clientes baseado em informação relacionada ao seu estilo e posição geográfica.
- Encontrar grupos de clientes similares.
- Medir a qualidade dos grupos observando padrões de compras dos clientes no mesmo grupo versus aqueles de grupo diferentes.

Agrupamento - Aplicações

- **Agrupamento de Documentos:**

-**Objetivo:** encontrar grupos de documentos que são similares entre si baseado nos termos importantes que aparecem neles.

-**Abordagem:** Identificar termos que ocorrem com frequência em cada documento. Formar uma medida de similaridade baseada na frequência dos diferentes termos. Utilizá-la para agrupar.

- **Resultado:** Recuperação de informações utilizando os grupos para o relacionamento com novos documentos ou termo de pesquisas aos documentos agrupados.

<i>Categoria</i>	<i>Total de Artigos</i>	<i>Corretamente colocados</i>
<i>Financeiro</i>	555	364
<i>Estrangeiro</i>	341	260
<i>Nacional</i>	273	36
<i>Metrô</i>	943	746
<i>Esportes</i>	738	573
<i>Lazer</i>	354	278

Tarefas de Regra de Associação

- Dado um conjunto de registros, cada um dos quais contém certo número de itens de uma coleção;
- Produzir regras de dependência que vincularão a ocorrência de um item baseado nas ocorrências de outros itens.

<i>ID</i>	<i>Items</i>
1	Pão, Refri, Leite
2	Cerveja, Pão
3	Cerveja, Refri, Fralda, Leite
4	Cerveja, Pão, Fralda, Leite
5	Refri, Fralda, Leite

Regras Descobertas:

{Leite} --> {Refri}

{Fralda, Leite} --> {Cerveja}

Regra de Associação - Aplicações

- Marketing e Promoção de Vendas:

-As regras descobertas tem o seguinte formato:

{Pão francês,...} -> {Batata frita}

Batata frita como consequente -> Pode ser utilizado para determinar ações para o incremento das vendas.

Pão francês como antecedente -> Pode ser utilizado para determinar quais os produtos seriam afetados se a empresa descontinuasse a sua comercialização.

Pão francês no antecedente E batata frita no consequente -> Pode ser utilizado para determinar quais os produtos devem ser vendidos associados com o pão francês para promover as vendas de Batata fritas.

Regra de Associação - Aplicações

Disposição de produtos de um Supermercado:

- **Objetivo:** Identificar itens que são comprados por um número suficiente de clientes
- **Abordagem:** Processar os dados coletados no ponto-de-venda com scanners para encontrar a dependência entre os itens.
- **Uma regra clássica:**
 - ✓ Se um cliente compra fraldas e leite, então ele provavelmente comprará cerveja.
 - ✓ Portanto, não se surpreenda se você encontrar engradados de cerveja próximo as fraldas.

Séries Temporais - Regressão

- Prevê um valor de uma dada variável continuamente valorada baseada nos valores de outras variáveis, assumindo um modelo de dependência linear ou não-linear.
- Muito estudado no campo da estatística e redes neurais.
- Exemplos:
 - Prever a quantidade de vendas de um novo produto baseado nos gastos de propaganda.
 - Prever a velocidade dos ventos em função da temperatura, umidade, pressão do ar, etc.
 - Previsão de séries temporais do índice de mercado financeiro, moedas, bolas de valores, etc.

Atividade 01

Responda as seguintes questões:

- ✓ A atividade utiliza-se da aprendizagem de máquina?
- ✓ Qual é o método (descritivo / preditivo)
- ✓ Qual é a tarefa (classificação, agrupamento, associação ou regressão linear)?

Atividade 01

- a. Dividir os clientes de acordo com a sua região.
- b. Cálculo do total de vendas de uma empresa.
- c. Ordenar um banco de dados de alunos baseado nos números de identificação dos alunos.
- d. Compor itens frequentes de utilização dos serviços de saúde de um paciente com determinada doença
- e. Prever o resultado de uma jogada de um par de dados
- f. Prever o futuro preço das ações de uma empresa usando registros históricos
- g. Monitorar a taxa de batimentos cardíacos de um paciente identificando anormalidades
- h. Monitorar ondas sísmicas de atividades de terremotos
- i. Agrupar consumidores em função do estilo de compra, local, forma de pagamento e demais variáveis de consumo.

Espaço aberto para dúvidas e opiniões

