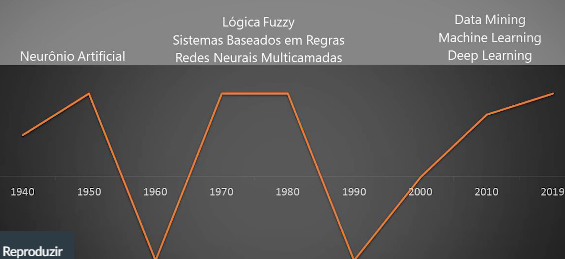
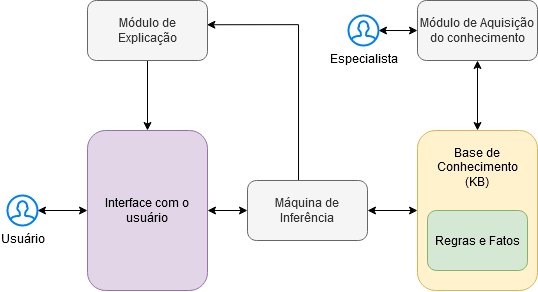
Estudo Individual 22/08/2022

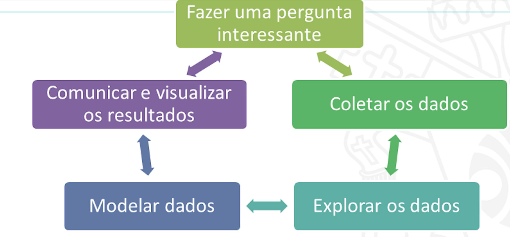
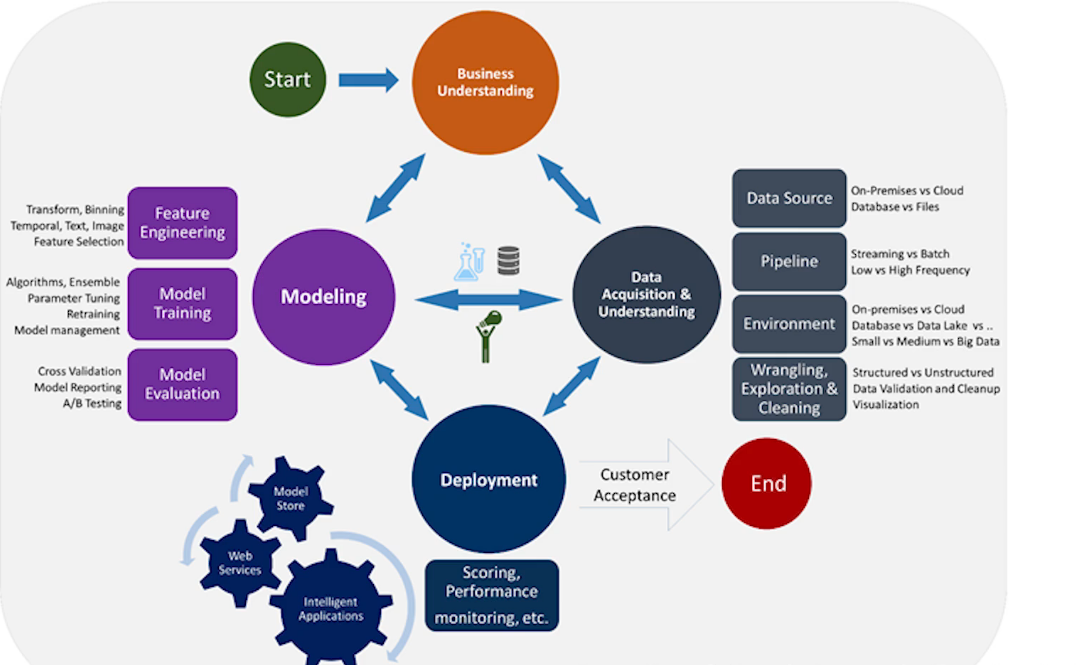
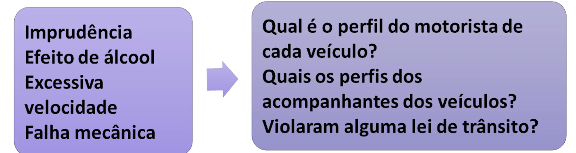
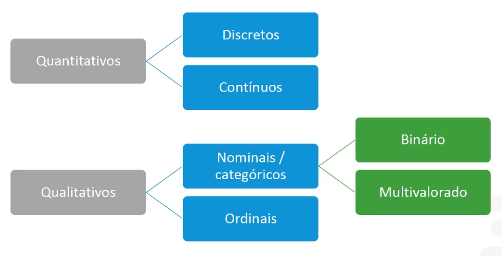
**Unidade I - Inteligência Artificial**

* Década de 1970 – volta da IA por causa da IBM, sistemas baseados em regras.
* Depressão e volta em 2000 com Data Mining, Machine Learning e outros.
* Paradigmas:
  + Pensamento humanamente
  + Pensamento Racionalmente
  + Agindo Humanamente
  + Agindo Racionalmente
* Pesando humanamente
  + Pensamento igual ao do homem
  + Como um computador seria programado para resolver um problema igual ao homem.
* Agindo humanamente
  + Criar máquina que executa de funções dos seres humanos da MESMA forma que o homem, se comportar igual ao homem.
  + Proposto por Alan Turing, teste de Turing
  + Sem intervenção humana
* Pensando racionalmente
  + Faculdades mentais através de modelos computacionais
  + Perceber, raciocinar e agir
  + Rede semântica
  + Raciocínio lógico
  + Construção de redes de conhecimento
* Agindo racionalmente
  + IA que seguem um protocolo racional, independente do pensamento humano.
  + Agente é algo que percebe o ambiente através de sensores e interage com o ambiente através de atuadores
  + Agente tem que ser racional.
  + EX: robô aspirador.

**Aplicações e Técnicas de Inteligência Artificial (IA)**

* Jogos digitais:
  + Máquina de Estados Finitos
    - Comportamento dos fantasmas do pacman
  + Aprendizados em Jogos – aprende e se adapta em jogos.
  + Previsão de comportamento.
* Música
  + Criação de musica
  + Music Information retrival – músicas parecidas para playlist
  + Sistema de reconhecimento de músicas
* Saúde
  + Diagnostico medico através de regras
  + Diagnostico medico por imagem
  + Apoio ao médico
  + Padrões(vigilância) epidemiológicos – chatbot + ia
* Setor financeiro
  + Para que se tenha aprendizado de máquina é necessário ter um padrão.
  + Duas abordagens:
    - Buy and Hold – analise fundamentalista
    - Day trade – analise técnica
  + Robôs para compra e venda de ativos
  + Analise de risco para empréstimos
* Transporte aéreo
  + Segurança
  + Custo de Operação
  + Atrasos e Cancelamentos
  + Conforto e Atendimento
* Setor Jurídico
  + Revisão e análise de documentos
  + Identificação de relacionamentos entre conceitos, empresas, etc...
  + TAR(technology assisted review)
  + Recomendação de custodiante
  + Analise de sentimentos
* Jornalismo e Comunicação
  + Tendencia (reportagem aumentada) – toneladas de documentos achando as breaking News!
  + Reduzir custos do dia
  + StoryTelling – templates
  + Jornalismo orientado a dados(data-driven journalistic Operation)
  + Reportagem orientada a dados
* Computação Evolutiva
  + Técnicas inspiradas na evolução biológica
    - Colônia de Formigas
    - Inteligência de Enxames
    - Enxame de partículas
    - Algoritmos Genéticos
      * População Inicial – conjunto de soluções
      * Cromossomo – conjunto de característica
      * “Seleção natural parecido com Darwin” – processo cíclico
* Aprendizado de Máquina
  + Conjunto de técnicas que visa deixar os computadores mais inteligente
  + Tipos:
    - Online – aprendizado enquanto está em funcionamento
    - Offline – Aprendizado acontece em intervalos
    - Supervisionado – utiliza-se dados para treinar o sistema, perde a capacidade de se adaptar.
    - Não supervisionado – aprendizado por reforço , baseado por feedbacks e recompensas
* Agentes Inteligentes
  + Agentes racionais
    - Racionalidade 🡪 Ser racional significa tomar decisões corretas que nos levem a alcançar nossos objetivos.
    - Percebe o ambiente e age:
    - Percebe por meio de sensores
    - Atua por meio de atuadores
  + Função do agente:
    - Sistemas baseado em regras
* Tipos de agentes racionais
  + Agente reativo simples 🡪 mapear as ações
  + Agente baseado em modelo 🡪armazena modelo probabilístico para tomar decisão mais adequada.
  + Agente baseado em objetivo 🡪 ponto do mundo ou objetivo que o agente quer chegar – sempre em direção do objetivo
  + Agente baseado em utilidade 🡪decisão/ação baseado no qual útil ela é.
  + Agente com aprendizado 🡪 capaz de ajustar suas regras, modelos e ações
  + Sistema multiagente 🡪comunicação entre agentes – capaz de coordenar a ação em conjunto
* Agentes baseados em regras
  + Sistemas Especialistas
  + 
    - A base de conhecimento (KB - Knowledge Base) contém os fatos e regras previamente cadastrados por um especialista no domínio do problema.
    - A máquina de inferência é o motor que busca as informações relevantes na base de conhecimento e as interpreta, de forma a encontra a solução ao questionamento do usuário.
    - Módulo de Aquisição do Conhecimento é o componente que permite ao sistema especialista sempre adicionar novos conhecimentos de diversas fontes à base de conhecimentos.
    - A interface com o usuário possui uma característica particular. Ao invés de ser uma interface baseada em tarefas que precisam ser executadas, ela recebe questionamentos que precisam ser respondidos.
    - O módulo de explicação fornece ao usuário toda a linha de raciocínio que permitiu a um sistema especialista chegar em uma conclusão.
  + Características de um Sistema Especialista
    - Retém o conhecimento de um especialista de forma mais permanente.
    - Ajuda a distribuir a expertise humana.
    - Pode consolidar o conhecimento de vários especialistas.
    - Podem resolver problemas complexos deduzindo novos fatos a partir do conhecimento existente.
  + Vantagens de um Sistema Especialista
    - Baixo custo.
    - Resposta rápida.
    - Raciocínio pode ser explicado.
  + Desvantagens de um Sistema Especialista
    - Difícil formalizar o senso comum.
    - Desenvolvido para um domínio específico.
    - Requer treinamento excessivo.
    - Precisa ser treinado manualmente.
* Democratização da IA
  + O que tornou a IA das grandes empresas disponíveis para público foi a API, ou micro serviços
  + <https://azure.microsoft.com/pt-br/services/cognitive-services/>
* Inteligência artificial na sociedade
  + Desvantagens e perigos:
    - Pode ser usada para construir coisas devastadoras
    - Pode criar erros ocultos
    - Perda de habilidades e pensamento críticos – analista de economia
    - Podem ser hackeadas e manipulados
    - Pode institucionalizar alguns vieses da base de dados – recomendação de currículos
    - Perda de empatia
    - Pode causar perda de controle – carros autônomos
    - Pode implicações sociais muito além do seu negócio
    - Empregados e consumidores ainda precisam ter seus papeis
    - Deve honrar direitos da segurança e privacidade

Unidade 2 - Ciência de dados

* O que é Ciência de Dados?
  + Vem pra tratar o volume enorme de dados que usamos no cotidiano.
  + Permite entender melhor o mundo
  + Computação + Estatística + Domínio.
  + Tendencia alta no mercado
  + Ciências de dados 🡪 engloba um processo:
  + Data Science Lifecycle – modelo de referência da Microsoft
* Definição do problema:
  + A fase de definição de problema é uma das mais importantes.
  + Pergunta – pergunta orientada a dados
  + Pode ser necessário a redefinição do problema.
* Redefinição do problema:
  + Orientado a dados locais 🡪 transforma problemas abstratos em problemas específicos
  + Um problema não pode ser defino com precisão se somente é dado atenção aos dados que mostram diretamente o problema. É necessário fatos e julgamentos🡪 fatos = circunstancias que causam o problema. julgamentos = observações a serem disputadas ou decididos.
  + EX: FATOS(ESQUERDA) E JULGAMENTOS(DIREITA)
* Tipos de Dados
  + É imporante saber o tipo para assim ser possivel criar regras para esse dado
  + 
  + Quantitativos = numéricos
  + Qualitativo = não-numérico
    - Ordinais = possível ordenar
    - Nominais/categóricos = não conseguimos ordenar

Dados Qualitativos:

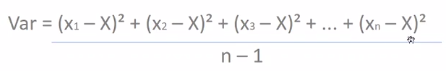
* Binominais/Binários: -- Normalmente são opostos
  + Sim/Não
  + Comprou/Não
  + Simétricos 🡪 Ambos valores possuem a mesma importância EX M/F
  + Assimétricos 🡪Apenas o valor positivo é relevante EX Comprar produto ex: febre em sintomas, ou seja, o sim é importante
  + Em geral, dados esparsos aqueles com muitos zeros, são assimétricos.
* Polinomiais
  + Vários nomes
  + Ex: centro,sul,centro-sul
* Categóricos ordinais
  + Existem uma ordenação natural
  + Ex: criança,jovem,adulto

Dado quantitativos:

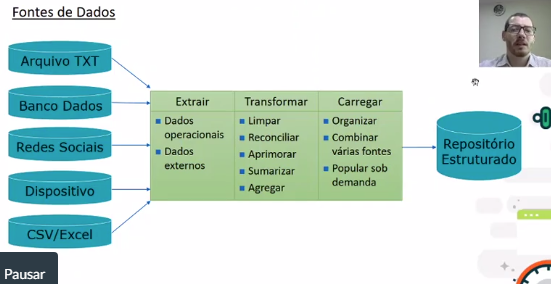
* Normalizados: [0..1]
  + Capacidade ociosa: 70% ociosa (0.7)
  + Andamento da operação: 30% concluída (0.3)
* Não normalizados: [min.max]
  + Idade: [0..120] anos
  + Temperatura: [10 .. 40] graus celsius
* Escala:
  + Intervalar: Define faixas de valores e relação entre eles. Nem sempre permitem definir a razão entre os valores
  + Racional: Números possuem significado absoluto (zero absoluto) exemplo: número de consultas em um hospital, peso.

16/09/2022

Estatísticas Descritivas:

* Medidas e Propriedadas
  + Frequencia de um dado
    - Absoluta e relativa(%)
    - N de vezes que o dado aconteceu no conjunto
  + Média
    - Medida mais comum
  + Mediana
  + Moda – mais frequente
  + Percentil
  + Intervalo 🡪 diferença entre max e min (amplitude)
  + Variância 🡪
  + Desvio padrão 🡪 raiz quadrada da variância
  + Desvio médio absoluto
  + Desvio mediano absoluto
  + Intervalo interqueril

ETL

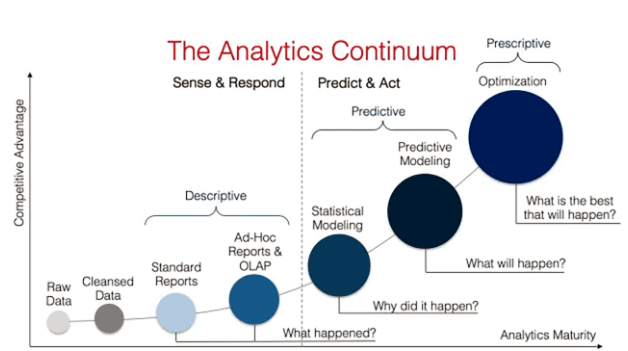
* Extract, Transform and Load
* Businees Intelligence 🡪conjunto de técnicas que permite a empresa a extrair informação útil por meio de dados
* Limpeza de dados:
  + Erros nos dados
  + Tamanhos inconsistentes, e outros erros
  + Tranformação de dados

04/10/2022

**Big Data**

* Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.”
* Cinco v:
  + Volume
  + Velocidade – dados perdem a velocidade bem rápido, viram antigos
  + Variedade
  + Veracidade
  + Valor

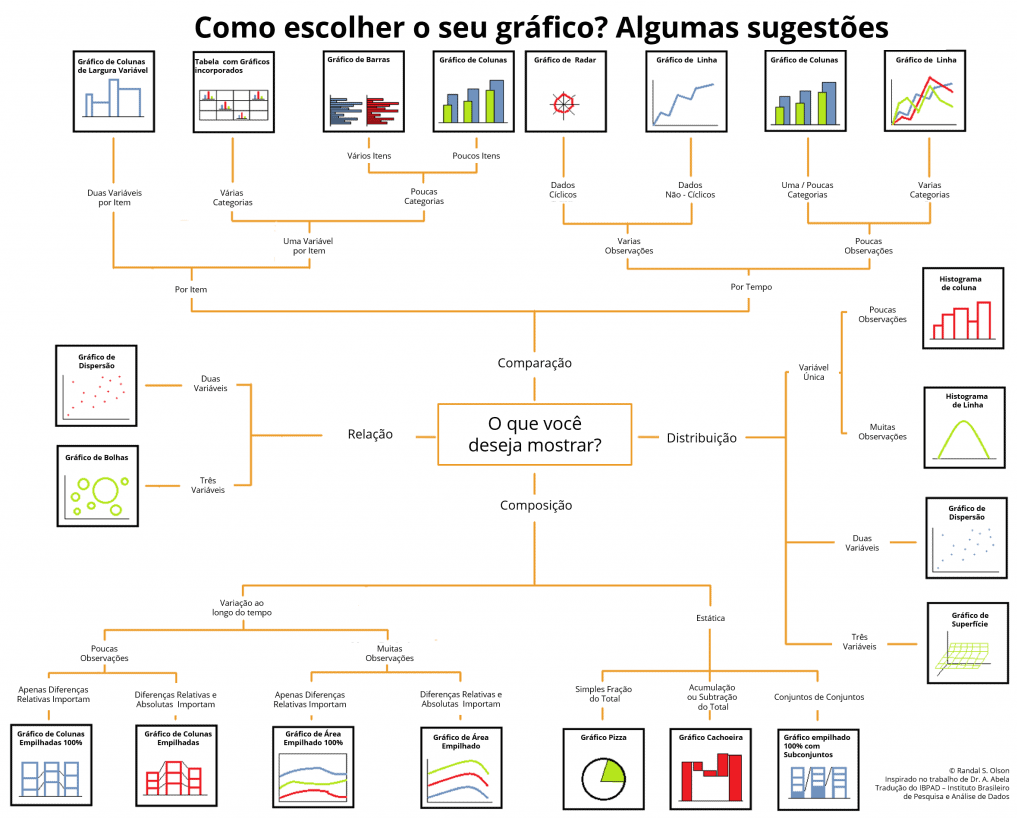
**Big Data Analytics**

* Sem analisar não temos veracidade e valor 🡪 dado cru
* The analytics Continuum:
  + Apenas 0,5% de dados são analisados
  + Gráfico sobre a análise de dados
* Analise descritiva 🡪 utilizando estatísticas
* Analise Diagnostica 🡪 A única diferença está na utilização de comparações e correlações estatísticas para identificar insights (intuições) para compreender a presente correlação de variáveis
* Analise Preditiva 🡪 A análise preditiva cria modelos que permitem prever potenciais resultados e nos permite se preparar para o futuro. Algumas perguntas possíveis seriam: "Qual é a previsão de demanda para o produto X no próximo mês? Qual o número esperado de visitantes do museu na próxima semana? (prever e agir)
* Analise Prescritiva🡪A análise prescritiva avança um pouco além da análise preditiva. Agora, além de tentarmos prever o futuro, utilizaremos a tecnologia para nos sugerir um curso de ação, ou seja, um caminho a ser seguido baseados nos dados analisados. Algumas perguntas possíveis são: "Como eu posso vender mais produtos X? Como eu posso aumentar o engajamento dos meus alunos na sala de aula virtual?"(Modelos de otimização)
* Analise Cognitiva🡪A análise cognitiva está ganhando espaço recentemente e se relaciona com o uso de tecnologias autônomas que aprendem e se adaptam sem a intervenção humana. Isso envolve monitoramento, análise interação e ação em tempo real. A aplicação mais famosa no momento que utiliza este tipo de análise são os carros autônomos, que precisam reagir em tempo real a milhares de dados coletados a cada segundo, entendendo o ambiente e decidindo sobre a melhor ação a ser tomada de forma automática, ou seja, sem que seja aprovado ou validado pelo homem.

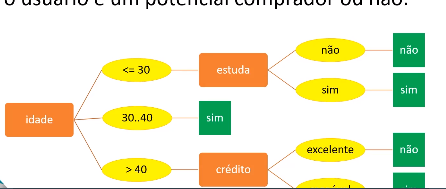
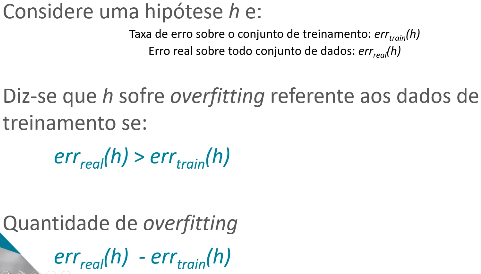
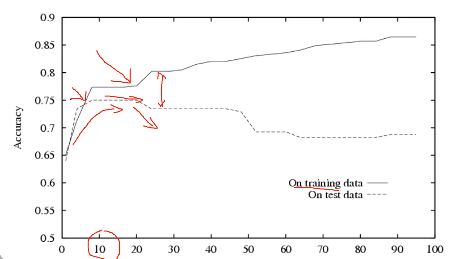
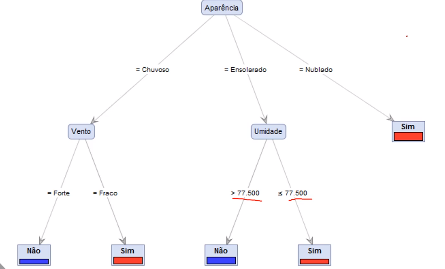
Estratégia de dados do Negocio

* Uma estratégia de dados nos ajuda a entender perguntas que gostaríamos de responder e que tipos de dados eu preciso para responde-la
* EX:
  + Mais clientes
  + Mais engajamentos dos meus clientes
* Tipos de dados coletados:
  + Pessoais 🡪 identificação pessoal + identificação não pessoal (IPS,cookies,tipo do dispositivo)
  + Engajamento🡪detalha como os consumidores interagem com o site.
  + Comportamentais🡪ações concretas, ex histórico de compra e heat maps
  + Atitude🡪métricas de satisfação, compra, conveniência e etc...
* Como utilizar os dados:
  + Para melhorar experiencia do cliente
  + Transformar os dados em fluxo caixa
  + Refinar marketing
  + Proteger mais dados
* Questões legais
  + Lei de Proteção de Dados Links para um site externo. (LGPD)
  + General Data Protection Regulation (GDPR)

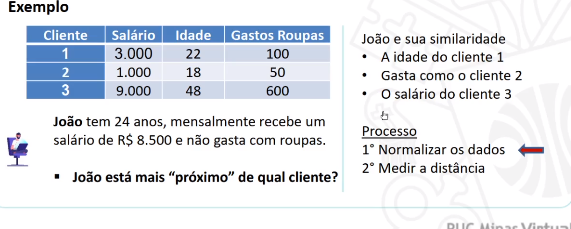
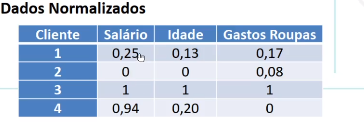
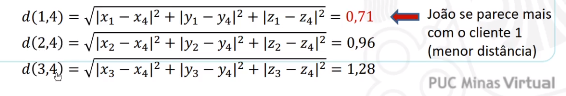
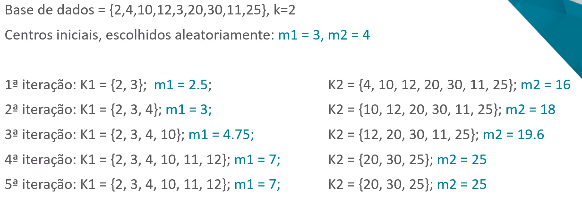
Análise Visual de Dados

* Visualização de dados:
  + Tendencia de crescimento de dados
  + É necessário tecnologias que auxiliam a visualização destes dados
  + Dados visuais atraem mais atenção
  + Vantagens:
    - Decisões mais ágeis
    - Pessoas mais engajadas
    - Melhora compreensão dos fatos
  + Ferramentas:
    - Planilhas eletrônicas
    - Bussines Inteligence
    - Scripts em python
    - Desen softwares
* Tipos de gráficos e recursos visuais:
  + Qual gráfico escolher?:
    - Distribuição
    - Relacionamento
    - Comparação
    - Composição
    - Localização
    - Conexão
  + Metodologia de escolhas:

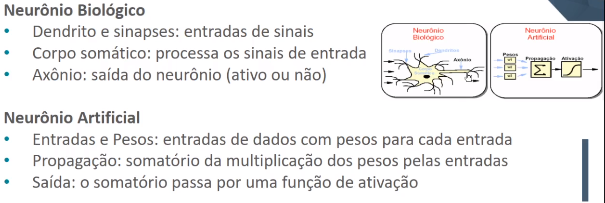
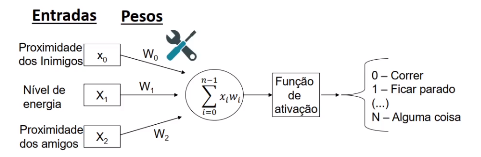
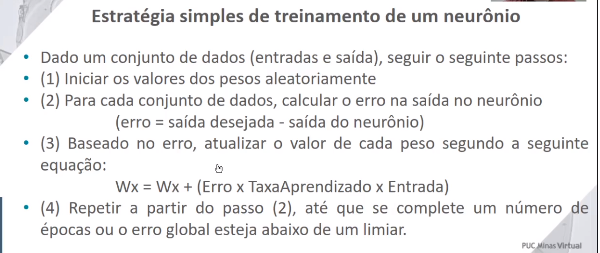
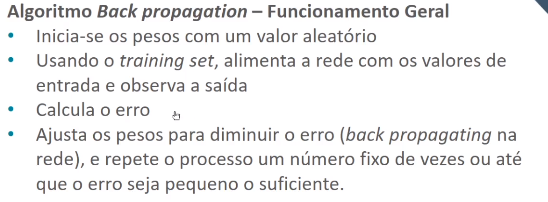
Unidade III – Aprendizado de Máquina

* Machine Learning 🡪 estudo de algoritmos que melhoram automaticamente através da experiencia
* Modelo de aprendizado:
  + Conjunto de regras que tenta representar o conhecimento por meio de dados
  + Tomada de decisão
  + Ciclo de aprendizado
  + Recebe dados, cria modelo, valida modelo e distribui modelo
* Estilos de aprendizado:
  + Supervisionado
    - Recebe dados rotulados
  + Não supervisionado
    - Recebe dados não rotulados
    - Busca padrões que se repetem
    - Elementos que são semelhantes e distintos (agrupamento e segmentação)
  + Semisupervisionado
    - Combina os dois exemplos anteriores pois é muito caro rotular a base de dados
    - Rotula parte da base de dados
    - Com a parte rotulada se cria pequenas regras e modelos, dessa forma, rotulando a base de dados que não está
  + Reforço
    - Trabalha com modelo de exploração do ambiente
    - Tentativa e erro
    - Um agente explorar um ambiente. A cada instante de tempo o agente está em um estado qualquer, executa uma ação, vai para um novo estado e recebe uma recompensa.
    - Problema da aprendizagem por reforço consiste em escolher uma política de ações que maximize o total de recompensas recebidas pelo agente.
  + Deep Learning
    - Baseado no modelo de rede neural
    - O aprendizado profundo também possui aprendizado supervisionado, não supervisionado e de reforço.
    - O objetivo é utilizar modelos de redes neurais extremamente complexos e com muitas camadas para identificar estruturas em problemas onde os dados não são claramente estruturados. Adequado para extrair recursos significativos de dados brutos.
* Principais tarefas do Machine Learning:
  + Grupos:
    - Descritivos – agrupamento e associação
    - Preditivas – classificação e regressão
  + Agrupamento (Clustering):
    - Divide automaticamente em grupos por suas similaridades e diferenças
  + Associação:
    - Encontrar **padrões frequentes**, associações, correlações entre itens
  + Classificação:
    - Tarefa para extrair modelos que descrevem classes de dados e também para predizer tendencias de dados
  + Regressão:
    - Conjunto de técnicas estatísticas para estimar a relação entre variáveis e ambiente.
* Arvore de decisão:
  + Usado para **classificação** ou **regressão** – aprendizado supervisionado
  + Funcionamento parecido com arvore binaria
  + Principais algoritmos para indução de arvore de decisão:
    - ID3
    - C45
    - CART
  + Métodos de seleção de atributos:
    - Ganho de informação (ID3)
    - Taxa de ganho(C45,J48)
    - Incide GINI – impureza (CART)
    - Redução de variância (CART)
  + Algoritmos servem para descobrir atributo que produz os nos mais puros
  + Menor entropia(possível):
  + Overfitting acontece quando ocorre um superajustamento:
  + OverFitting no grafico:
  + Evitando o overfitting:
    - Para de crescer a arvore quando não for estaticamente relevante
    - Gerar a arvore completa e depois poda-la
  + Arvore com atributos contínuos:

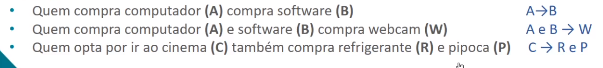
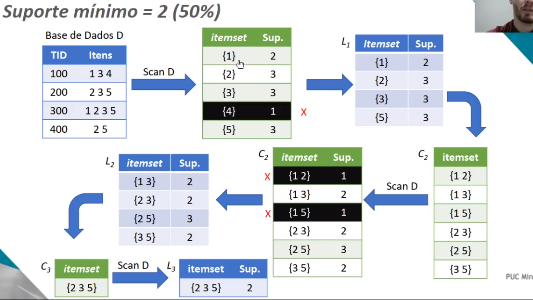
**K-means**

* **Tipo de aprendizado supervisionado**
* Neste caso, temos um problema clássico de aprendizado não supervisionado, onde o objetivo principal é identificar grupos de dados semelhantes em um conjunto de dados.
* Agrupamento ou Clustering
  + Cluster 🡪 conjunto de objetos similares entre si e mais diferentes de outros clusters
  + Clustering🡪Separar os objetos em grupos com base na similaridade, e em seguida atribuir rótulos a cada grupo
    - Qualidade do resultado depende da medida da similaridade.
  + Aplicações: distribuição e processamento de dados, economia, marketing, classificação de documentos
  + Principais métodos:
    - Baseados em particionamento 🡪 proximidade de um ponto de dados com o centróide dos clusters.
    - Baseados em densidade🡪 atribui os pontos em regiões de mesma densidade pertencem ao mesmo cluster, procuram regiões de alta densidade de similaridade
    - Hierarquicos 🡪 Na abordagem aglomerativa, todos os pontos de dados começam em clusters separados e, em seguida, vão sendo agregados pela menor distância. Nos métodos divisórios, todos os pontos de dados são classificados como um único cluster e particionados pela maior distância.
* Medidas de Similaridade
  + Exemplo joão:
    - Normalizar Dados contínuos:
      * Colocar dados entre 0 e 1
      * Maior = 1 e Menor = 0
      * Os demais viram proporcionais , segue formula:
    - Medir distancia:
      * Distancia Euclidiana entre todos os casos – clientes :
* K-means
  + Cada cluster é representado por um ponto central - centroide
  + K = qntd de clusters desejada
  + Variacoes: k-medoides,k-modas,k-medianas
  + Rápido
  + Requer uma medida de distancia, e a possiblidade de calcular media entre os objetos
  + Erros: não é ótimo – solução ramdom resart e looping infinito – solução: limitar n de iterações (para após 10)
  + EX:
  + Obs – OLHAR 4 iteracao
  + Reumo dos passos:
    - Definir o número de clusters K.
    - Selecionar aleatoriamente K pontos no espaço de dados.
    - Alocar os dados ao cluster pertencente ao centróide mais próximo.
    - Computar os centróides de cada cluster a partir dos pontos alocados.
    - Reatribuir cada ponto ao centróide mais próximo.
    - Repetir os passos 4 e 5 até que não haja mais realocações de dados.

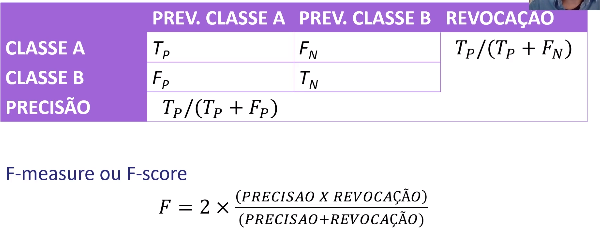
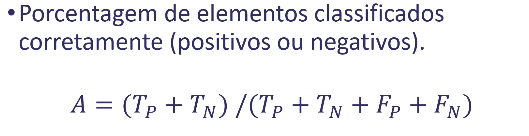
**Redes Neurais Artificiais e suas aplicações**

* Neurônio biológicos:
  + Sinapses/dentritos (trafego de sinais elétricos)
  + Se adapta para dar resposta – devem ser **treinados**
* Neurônio artificial:
  + Sistemas baseados em neurônios biológicos e na estrutura paralela do cérebro , com capacidade de adquirir, armazenar e utilizar conhecimento experimental.
* Comparação:
* Função de ativação: -- função degrau ou função sigmoide
  + Recebe o valor de saída deo neurônio e coloca em um formato desejado
  + EX:
    - 0 ou 1(degrau)
    - 0% e 100%(Sigmoide)
    - ****Sim ou Não(degrau)
* Rede Neural Percepton
  + Possui uma ou mais camadas com vários neurônios artificias que trabalham de forma conjunta, sendo que saída e entrada estão conectadas
  + Dividido em:
    - Treinamento da Rede: Necessidade de treinamento e adaptação
    - Uso efetivo da rede: Resposta
  + Treinamento da Rede:
    - **Taxa de aprendizado**: velocidade na qual o sistema esta caminhando para encontrar a solução (o tamanho do passo do aprendizado)
      * Taxa muito baixa vai necessitar de muitas épocas de treinamento, ou seja, demora mais para cumprir o objetivo
      * Taxa alta pode tornar algoritimo oscilante, demora para convergir
    - **Erro Global:** media dos erros para todo o conjunto de deados de teste
    - **Época:** cada período de treinamento, no qual todo o conjunto de dados já foi usado
    - Estratégia simples:
    - Se a base de treino é incompleta vai gerar um neurônio incompleto
* Perceptron Multicamadas e suas aplicações
  + Uma rede neural perceptron de uma única camada tem dificultidade de representar funções não lineares ou modelos que separem os linearmente (XOR ou potência)
  + Motivações: Superar as limitações das redes de camada unica + aproximações para funções continuas, Separar rede conjunto de dados por uma rede de 3 camadas
  + Algoritmo Back propagaration:
  + Aplicações:
    - Previsão de vendas – regressão rede neural de multicamadas

**Regras de Associação**

* Encontrar padrões entre um conjunto de itens ou objetos de um bd ou qualquer repositório de informação
* Aplicações: cestas de comprar, marketing, portifólio...
* Formato das regras: A🡪B A e B 🡪 C A🡪B e C
* EX:
* Medidas estastisticas:
  + **Suporte**: probabilidade dos elementos acontecerem
  + **Confiança:** confianca em que um elemento implica o outro
* Algoritimo Apriori
  + Fase I 🡪 descobre todos os conjuntos de itens com suporte maior ou igual ao min especificado pelo usuário
  + Fase II 🡪A partir dos conjuntos de itens frequentes, descobre regras de associação com fator de confiança maior ou igual ao especificado do usuário
  + Exemplo:

**Avaliação de Modelos de ML**

* **3 etapas:**
  + Avaliação off line – resultados e validação
  + Avaliação offline de dados operacionais
  + Avaliação online
* Avaliação offline:
  + Verificação cruzada
  + Parâmetros e hiper parâmetro – parâmetros = variáveis ajustadas no processo de aprendizado, hiper = precisam ser ajustados, mas não são aprendidos
* Critérios**:**
  + Matriz de confusão:
  + Acurácia:
  + Acuracia média – acurácia por classe

Tendencias em ML

* Auto ML:
  + Ia que cria os modelos de ML
  + Redução no tempo da criação
  + Diminuem a barra de conhecimento necessário em ML
  + EX: AWS, GOOGLE CLOUD, AZURE
  + Etapas:
    - Features selection/ engineering
    - Optimização de hiper parâmetros
    - Comparação de modelos
* MLOps
  + Práticas de colaboração e comunicação entre os cientistas de dados
  + Visa auxiliar a gestão ciclo de vida de processos de ML
  + Conceito

**Computação Cognitiva:**

* A chave da mente não é sua estrutura física, mas as relações que os sistemas possuem uns com os outros, e com ambiente externo.
* Conceito: criação de modelos mais precisos de como a mente humana percebe, raciocina e reage aos estímulos

****

* Sistemas cognitivos:
  + Adaptativos
  + Interativos
  + Iterativos
  + Contextuais