

## Lista 01 IA

**02)** O KDD é um processo utilizado para identificação de padrões em grandes massas de dados, esse processo é dividido em cinco etapas: seleção, pré-processamento, transformação, mineração dos dados e interpretação de resultados.

**Seleção:** nesta etapa que serão decididos quais os conjuntos de dados que serão relevantes para que sejam obtidos resultados com informações úteis.

**Pré-processamento:** limpeza dos dados e seleção de atributos. Informações ausentes, errôneas ou inconsistentes nas bases de dados devem ser corrigidas de forma a não comprometer a qualidade dos modelos de conhecimento a serem extraídos ao final do processo de KDD.

**Transformação:** analisa os dados obtidos da etapa anterior e os reorganiza de uma forma específica para que possam ser interpretados na etapa seguinte.

**Mineração dos dados:** depois de serem transformados, os dados serão lidos e interpretados. A mineração faz com que eles transformados em informações, as quais são indicadas através de força bruta, ou seja, lendo regra por regra e as interpretando.

**Interpretação de resultados:** as regras indicadas pelo processo anterior serão interpretadas e avaliadas. Após a interpretação poderão surgir padrões, relacionamentos e descoberta de novos fatos, que podem ser utilizados para pesquisas, otimização, etc.

**03)** Existem 4 tipos de problemas de aprendizado de máquinas principais: classificação, regressão, agrupamento e regras de associação.

**Classificação:** entradas são divididas em duas ou mais classes, e o aprendiz deve produzir um modelo que vincula entradas não vistas a uma ou mais dessas classes. Exemplo: a filtragem de spam, em que as entradas são as mensagens de emails (ou outros) e as classes são "spam" ou "não spam".

**Regressão:** Semelhante ao problema anterior. Porém o atributo de classe é numérico, não nominal. Exemplo: prever na base do restaurante o valor da comida, com base nos atributos.

**Agrupamento:** um conjunto de entradas é dividido em grupos. De maneira diferente da classificação, não é conhecido o atributo de classificação. Exemplo: identificar perfis de usuários em redes sociais

**Regras de associação:** busca semelhança/associações entre os elementos. Exemplo: quem comprou produto A também comprou produto B.

**04)** Tipos de aprendizagem: aprendizado supervisionado, aprendizado não supervisionado, aprendizado semisupervisionado, aprendizado por reforço e deep learning.

**Aprendizado supervisionado:** são apresentadas ao computador exemplos de entradas e saídas desejadas, fornecidas por um "professor". O objetivo é aprender uma regra geral que mapeia as entradas para as saídas. Exemplo: árvores de decisão.

**Aprendizado não supervisionado:** nenhum tipo de etiqueta é dado ao algoritmo de aprendizado, deixando-o sozinho para encontrar estrutura nas entradas fornecidas. O aprendizado

não supervisionado pode ser um objetivo em si mesmo (descobrir novos padrões nos dados) ou um meio para atingir um fim. Exemplo: redes neurais.

Aprendizagem semisupervisionada: inclui ambos os problemas discutidos anteriormente: ela usa dados rotulados e não-rotulados. Este método permite-nos melhorar significativamente a acurácia, pois permite utilizar dados não rotulados juntamente com uma pequena quantidade de dados rotulados. Exemplo: SVM.

Aprendizado por reforço: um programa de computador interage com um ambiente dinâmico, em que o programa deve desempenhar determinado objetivo. É fornecido, ao programa, feedback quanto a premiações e punições, na medida em que é navegado o espaço do problema. Exemplos: aprender a jogar um determinado jogo apenas jogando contra um oponente; redes neurais.

Deep learning: algoritmos complexos construídos a partir de um empilhamento de diversas camadas de “neurônios”. Exemplo: rede neural.

**05)** Em relação a avaliação da qualidade de modelos de aprendizado de máquina, compreendo que depende do tipo de aprendizado. Problemas de classificação podem ser avaliados pela matriz de confusão, problemas de agrupamento por métricas específicas, etc. Algumas medidas são: acurácia, matriz de confusão, precisão, cobertura, F-measure, etc.