INTERFACE TECNOLÓGICA

APRENDIZAGEM DE MÁQUINA E MINERAÇÃO DE DADOS: avaliação de métodos de aprendizagem

MACHINE LEARNING AND DATA MINING: evaluation of learning methods

Lucas Oukus Corcovia – lucas.corcovia1@gmail.com
Renato dos Santos Alves – renato.jcam@gmail.com
Faculdade de Tecnologia de Taquaritinga (FATEC) – SP – Brasil

RESUMO

Pode se conceituar mineração de dados o processo de explorar grandes quantidades de dados à procura de padrões consistentes, já o aprendizado de máquina é um método de análise de dados que automatiza a construção de modelos analíticos. Portanto, o objetivo deste artigo é apresentar conceitos da mineração de dados e suas fases, mostrar abordagens e técnicas de aprendizado de máquina, esclarecer os principais algoritmos de classificação e agrupamento e demostrar aplicações práticas. A metodologia do trabalho consiste em levantamento bibliográfico, utilização de conjuntos de dados e operação de *softwares* para testes. Como resultado serão feitos experimentos utilizando conjuntos de dados, a fim de identificar as diferenças dos algoritmos estudados abordando aspectos como viés indutivo e sensibilidade a ruído.

Palavras-chave: Aprendizado de Máquina. Mineração de Dados. Inteligência Artificial.

ABSTRACT

Data mining can be conceptualized as the process of exploring large amounts of data in search of consistent patterns, since machine learning is a method of data analysis that automates the construction of analytical models. Therefore, the objective of this article is to present concepts of data mining and its phases, to show approaches and techniques of machine learning, to clarify the main classification and clustering algorithms and to demonstrate practical applications. The methodology of the work consists of bibliographic survey, use of data sets and operation of software for tests. As a result, experiments will be performed using data sets in order to identify the differences of the studied algorithms, addressing aspects such as inductive bias and noise sensitivity.

Keywords: *Machine Learning. Data Mining. Artificial Intelligence.*

1 INTRODUÇÃO

A todo o momento dados são armazenados, formando grandes volumes de dados. Os dados armazenados contêm informações ocultas que quando reveladas são de grande

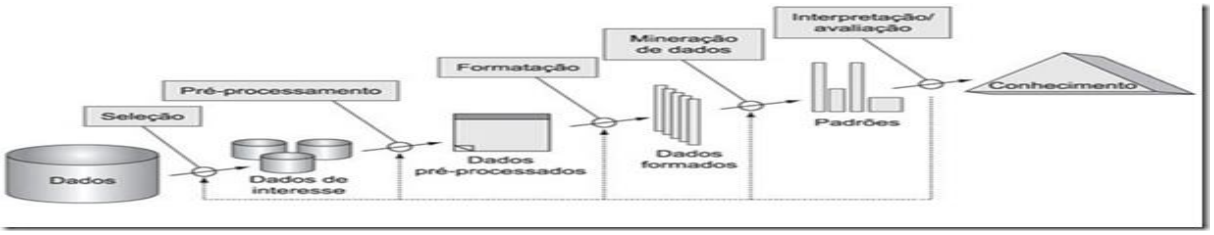
importância para tomada de decisão. Devido ao grande volume de dados, a extração destas informações não é uma tarefa trivial, é necessário o de teorias e ferramentas para o auxílio na extração e análise de informações úteis. O aprendizado de máquina está diretamente ligado à mineração de dados e a estatística, mas foca nas propriedades dos métodos estatísticos, assim como sua complexidade computacional. Sua aplicação prática inclui o processamento de linguagem natural, motores de busca, diagnósticos médicos, bioinformática, reconhecimento de fala, reconhecimento de escrita, visão computacional e locomoção de robôs.

1.1 Mineração de Dados

As ferramentas e técnicas empregadas para análise automática e inteligente destes imensos repositórios são os objetos tratados pelo campo emergente da descoberta de conhecimento em bancos de dados (DCBD), da expressão em inglês *Knowledge Discovery in Databases* (KDD). Mineração de dados é a etapa em KDD responsável pela seleção dos métodos a serem utilizados para localizar padrões nos dados, seguida da efetiva busca por padrões de interesse numa forma particular de representação, juntamente com a busca pelo melhor ajuste dos parâmetros do algoritmo para a tarefa em questão.

Descoberta de conhecimento em bancos de dados é o processo não trivial de identificar em dados padrões que sejam válidos, novos (previamente desconhecidos), potencialmente úteis e compreensíveis, visando melhorar o entendimento de um problema ou um procedimento de tomada de decisão. (FAYYAD; PIATETSKYSHAPIRO; SMYTH, 1996). Na Figura 1 é possível visualizar o processo de KDD:

Figura 1: Fases do Processo de KDD



Fonte: Fayyad, Piatetskyshapiro e Smyth (1996)

1. **Seleção:** Criação de um conjunto de dados alvo, selecionar um conjunto de dados, ou focar num subconjunto, onde a descoberta deve ser realizada. Selecionar ou segmentar os dados de acordo com critérios definidos.

2. **Pré-processamento:** Operações básicas tais como, remoção de ruídos quando necessário, manipular campo de dados ausentes, formatação de dados de forma a adequá-los à ferramenta de mineração.

3. **Transformação**: Redução de dados e projeção, com a utilização de características úteis para representar os dados dependendo do objetivo da tarefa, com o objetivo de reduzir o número de variáveis e/ou instâncias a serem consideradas para o conjunto de dados.

4. **Mineração de Dados:** Escolha e execução do algoritmo de aprendizagem de acordo com a tarefa a ser cumprida. É a verdadeira extração dos padrões de comportamento dos dados.

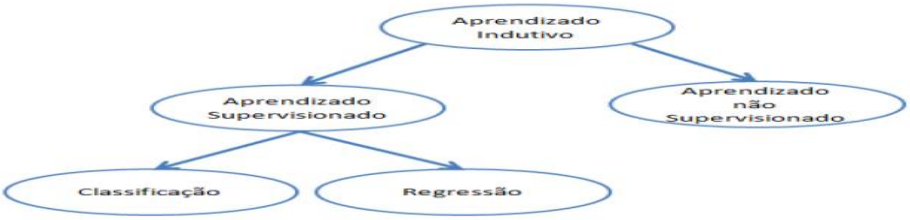
5. **Interpretação:** Interpretação dos resultados, com possível retorno aos passos anteriores; consolidação; incorporação e documentação do conhecimento e comunicação aos interessados; identificado os padrões estes são interpretados, e os mesmos darão suporte a tomada de decisões.

1.2 Abordagens para o Aprendizado de Máquina

O aprendizado de máquina ou aprendizagem automática é um subcampo da inteligência artificial dedicado ao desenvolvimento de algoritmos e técnicas que permitem um programa de computador aprender a partir de uma experiência E com respeito a uma classe de tarefas T e com medida de desempenho P, se seu desempenho na tarefa T melhorar com a experiência E. (MITCHELL, 1997). Enquanto na inteligência artificial existem dois tipos de raciocínio - o indutivo, que extrai regras e padrões de grandes conjuntos de dados, e o dedutivo - o aprendizado de máquina só se preocupa com o indutivo.

O raciocínio indutivo é dividido, em dois dos principais métodos de aprendizagem, que são o supervisionado (as instâncias estão rotuladas, a classe é conhecida) e o não supervisionado (instâncias não rotuladas não existe classe associada). Quando o aprendizado é supervisionado, e os rótulos assumem valores discretos, o foco da aprendizagem está na classificação, mas quando os rótulos assumem valores contínuos, o foco da aprendizagem está na regressão. A hierarquia do raciocínio indutivo pode ser vista na Figura 2.

Figura 2: Hierarquia do Aprendizado Indutivo



Fonte: Monard et al. (1997)

A aplicação prática do AM (Aprendizado de Máquina) inclui o processamento de linguagem natural, motores de busca, diagnósticos médicos, bioinformática, reconhecimento de fala, reconhecimento de escrita, visão computacional e locomoção de robôs.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Paradigmas de Aprendizagem de Máquina

Alguns paradigmas de AM estão sendo estudados constantemente, tais como o paradigma simbólico, estatístico/probabilístico, *instance-based*, conexionista e evolucionista. Nesse artigo é dado ênfase ao paradigma simbólico, tratando a descoberta de conhecimento com o uso dos métodos e algoritmos "Árvores de Decisão" e K-Means. Mas para construir um texto mais completo, no princípio, será feita uma breve descrição de todos esses paradigmas de acordo com Monard et al., (1997), como segue:

- **Paradigma Simbólico:** Os sistemas de aprendizado simbólico buscam aprender construindo representações simbólicas de um conceito através da análise de exemplos e contraexemplos desse conceito. As representações simbólicas estão tipicamente na forma de alguma expressão lógica, árvore de decisão, regras de produção ou redes semânticas.
- **Paradigma Estatístico:** Pesquisadores em estatística têm criado vários métodos de classificação, muitos deles empregados em AM. Destaca-se neste caso o Aprendizado Bayesiano segundo Duda, Hart e Stork, (2000), que utiliza um modelo probabilístico baseado no conhecimento prévio do problema utilizando exemplos de treinamento para determinar a probabilidade final de uma hipótese.
- ***instance-based*:** Realiza novas classificações com base em casos similares cuja classe é conhecida e assume que o novo caso terá a mesma classe. Esta filosofia exemplifica os

sistemas *instance-based*, que classificam casos nunca vistos usando casos similares conhecidos (QUINLAN, 1993).

- **Conexionista:** O nome conexionismo é utilizado para descrever a área de estudo que estuda as redes neurais. Redes neurais são construções matemáticas relativamente simples que foram inspiradas no modelo biológico do sistema nervoso. Alguns autores têm considerado redes neurais como métodos estatísticos paramétricos uma vez que treinar uma rede neural geralmente significa encontrar valores apropriados para pesos (parâmetros) e viés pré-determinados.
- **Evolucionista:** Este paradigma possui uma analogia direta com a teoria de Darwin, onde sobrevivem os que melhor se adaptaram ao ambiente. Um classificador genético consiste em uma população de elementos de classificação que competem para fazer a predição. Elementos que possuem uma performance fraca são descartados, enquanto os elementos mais fortes proliferam, produzindo variações de si mesmos.
- **Particionamento:** Esse método de clusterização aplicado em aprendizagem não-supervisionada, constrói várias partições de valor *k* que otimiza um critério de particionamento escolhido.

2.2 Aprendizado Supervisionado

Esta categoria de algoritmos possui esta denominação porque a aprendizagem do modelo é supervisionada, ou seja, é fornecida uma classe à qual cada amostra no treinamento pertence. Estes algoritmos são preditivos, pois suas tarefas de mineração desempenham inferências nos dados com o intuito de fornecer previsões ou tendências, obtendo informações não disponíveis a partir dos dados disponíveis. Para o aprendizado supervisionado serão utilizados os seguintes algoritmos em que a ferramenta Weka possui implementação:

Algoritmo ID3: O algoritmo ID3 introduzido por Quinlan, (1986), para indução de modelos de classificação, mais conhecidos por árvores de decisão. O algoritmo ID3 foi um dos primeiros algoritmos de árvore de decisão, tendo sua elaboração baseada em sistemas de inferência e em conceitos de sistemas de aprendizagem. Ele constrói árvores de decisão a partir de um dado conjunto de exemplos, sendo a árvore resultante usada para classificar amostras futuras. O ID3 separa um conjunto de treinamento em subconjuntos, de forma que estes contenham exemplos de uma única classe. A divisão é efetuada através de um único atributo,

que é selecionado a partir de uma propriedade estatística, denominada ganho de informação, que mede quanto informativo é um atributo. (QUINLAN, 1986).

Algoritmo C4.5: O algoritmo C4.5 é um aprimoramento do algoritmo ID3, isto pelo fato de trabalhar com valores indisponíveis, com valores contínuos, podar árvores de decisão e derivar regras. (QUINLAN, 1986). Trabalhar com registros que possuem valores indisponíveis na construção de uma árvore da decisão, pode ser considerado um problema. Os registros que possuem valores desconhecidos podem ser simplesmente descartados do conjunto de treinamento, ou podem ser classificados pela estimativa da probabilidade dos vários valores possíveis. O tratamento de atributos com valores contínuos envolve a consideração de todos os valores presentes no conjunto de treinamento, fazendo com que estes valores sejam ordenados de forma crescente e, após esta ordenação, seja selecionado o valor que favorece à na redução da informação necessária. O uso de valores contínuos pode tornar-se lento se o número de valores for muito elevado, demandando grande tempo de ordenação. O resultado do particionamento recursivo, utilizado na construção de árvores de decisão, pode ser uma árvore muito complexa, de acordo com o seu conjunto de treinamento. O método de podar árvores de decisão é realizado substituindo uma subárvore por um nodo folha. Este método é realizado se uma regra de decisão estabelecer que a taxa de erro prevista na subárvore é muito grande, em relação à utilização de um único nodo folha. A substituição de partes da árvore é realizada considerando que estas não contribuem à exatidão da classificação em determinados casos, produzindo algo menos complexo e assim mais compreensível. (QUINLAN, 1993).

2.3 Aprendizado Não-Supervisionado

Nestes algoritmos o rótulo da classe de cada amostra do treinamento não é conhecido, e o número ou conjunto de classes a ser treinado pode não ser conhecido a priori, daí o fato de ser uma aprendizagem não-supervisionada. Além disso, são também descritivos, pois descrevem de forma concisa os dados disponíveis, fornecendo características das propriedades gerais dos dados minerados. Neste artigo será utilizado o seguinte algoritmo de clusterização por particionamento para aprendizagem não-supervisionada em que a ferramenta Weka possui implementação:

Algoritmo K-Means: O algoritmo k-means é um método não-supervisionado de classificação que tem como objetivo particionar n registros em k agrupamentos, onde $k < n$. Seu funcionamento é descrito a seguir: Dado um valor inicial de k médias (k-means), os registros

são separados em agrupamentos, onde centroides representam o centro de cada agrupamento. Normalmente, as coordenadas iniciais desses centroides são determinadas de forma aleatória. Em seguida, cada registro é associado ao cluster cujo centro está mais próximo, seguindo uma métrica de distância. Existem diversas métricas de distância, como a Euclidiana (Na matemática, Geometria euclidiana é a geometria sobre planos ou objetos em três dimensões baseados nos postulados de Euclides de Alexandria) e a de Manhattan (é uma forma de geometria em que a usual métrica da geometria euclidiana é substituída por uma nova métrica em que a distância entre dois pontos é a soma das diferenças absolutas de suas coordenadas). Quando todos os registros estiverem classificados, os *k* centros são recalculados como as médias aritméticas dos registros de cada cluster. Então, os registros são novamente associados a um agrupamento segundo sua distância à média do cluster e os centros são novamente calculados. Esse passo se repete até que as médias dos clusters não se desloquem consideravelmente.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 Datasets e Testes Realizados

De modo a poder efetuar os testes com cada um dos classificadores que foram propostos, foi necessária a utilização de três *datasets* distintos, esses *datasets* foram retirados do *UCI Machine Learning Repository1* de Dua e Graff, (2019), que é um repositório de banco de dados usado pela comunidade de aprendizado de máquina para análise empírica de algoritmos de aprendizado de máquina. Para realização desse estudo foram utilizados os seguintes *datasets*:

Dataset Balões: Esse conjunto de dados foi utilizando anteriormente no experimento de psicologia cognitiva de Pazzani, (1991), ele possui quatro atributos, uma classe e dezesseis instâncias, sem valores ausentes.

Dataset Censo de Renda: Esse *dataset* é uma extração do banco de dados censo de 1994, feita por Barry Becker, a fim de prever, se a renda de um indivíduo é superior a $50K/ano. Composto por quatorze atributos, uma classe e 32561 instâncias, possui também uma base de teste com 16277 instâncias, valores desconhecidos nas instâncias foram substituídos por “?”. O *dataset* é composto de atributos nominais, contínuos e discretos. Esse é um conjunto de dados associado a tarefas de classificação. (DUA e GRAFF, 2019).

Dataset Sementes: Esse conjunto de dados foi utilizado em Kulczycki e Charytanowicz, (2011), com o objetivo de avaliar um algoritmo de agrupamento gradiente completo, onde edições das propriedades geométricas dos grãos pertencentes a três variedades diferentes de

trigo, sendo 70 grãos cada, selecionados aleatoriamente para o experimento. Esse conjunto de dados pode ser usado para tarefas de classificação e análise de cluster.

Para avaliação dos algoritmos de classificação e agrupamento, foram utilizadas as implementações do Weka software de código aberto com uma coleção de implementações de algoritmos de aprendizagem de máquina. (HALL, FRANK, HOLMES, PFAHRINGER, REUTEMANN E WITTEN, 2009).

Para avaliação dos algoritmos de classificação ID3 e C4.5 e foram utilizados os *datasets* Balões e Censo de Renda, e para avaliação do algoritmo de agrupamento K-Means foi utilizado o *dataset* Sementes.

Na execução dos algoritmos de classificação, em um primeiro momento foi utilizado o *dataset* Balões sem um conjunto de dados de teste, em seguida foi utilizado o *dataset* Censo de Renda fornecendo um conjunto de dados de teste, a fim de avaliar os algoritmos em cenários diferentes.

Enquanto na execução do algoritmo de agrupamento K-Means foi utilizado apenas o *dataset* Sementes, pois este conjunto de dados possui as características necessárias para o uso de algoritmos de agrupamento.

3.2 Experimento Utlizando o Algoritmo K-Means

No artigo “A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização” proposta por Pimentel, França e Omar (2003), é apresentada uma experiência de categorização de alunos através de clusterização. Nesta experiência o objetivo foi encontrar alunos com perfis semelhantes, separá-los em grupos homogêneos, e tentar ajudar os alunos com dificuldade de aprendizado. (FONSECA; BELTRAME, 2010).

No referido trabalho são utilizados os algoritmos K-means e Self-Organizing Maps (SOM).

Segundo Pimentel, França e Omar (2003), como estudo de caso foram escolhidas turmas de graduação em cursos de Informática de duas instituições de ensino superior, e como entrada de dados foi utilizado um questionário com 68 questões, onde os alunos atribuíam uma nota de 0 a 5 em relação ao próprio conhecimento no tópico, para assim identificar o grau de confiança dos alunos em cada tópico das disciplinas utilizadas.

Para obtenção dos resultados do experimento foram utilizados dois tipos de conjuntos de dados de entrada, um envolvendo cada um dos itens do questionário e outro contendo as

médias das respostas por grupos de acordo com uma ontologia que foi previamente criada utilizando o Protege. (FONSECA; BELTRAME, 2010).

Figura 3: Visão dos grupos formados pelo K-means

Identificação do aluno	Conceitos (Questionário)						GRUPO
	A	B	C	D	E	F	
77167	3,4	3,9	0,0	0,0	0,0	0,0	0
77155	2,4	4,0	2,5	1,1	0,0	0,0	0
77140	3,4	4,2	4,0	4,1	3,3	4,5	1
77146	3,3	4,6	4,0	3,5	4,5	4,0	1
77114	3,2	4,4	5,0	4,2	4,6	4,5	1
77134	4,5	4,2	5,0	2,9	4,8	4,5	1
77115	4,1	4,6	5,0	4,8	5,0	5,0	1
77143	1,9	3,0	3,0	0,8	0,0	3,5	2
73166	2,2	1,3	2,0	1,1	0,5	1,5	2
77104	1,2	1,7	2,3	0,9	0,6	1,0	2
77168	1,2	1,7	2,3	1,3	0,6	2,0	2
77105	2,4	3,1	2,3	1,4	0,8	2,5	2

Fonte: Pimentel, França e Omar (2003)

A Figura 3 apresenta os grupos encontrados pelo K-means. De acordo com Pimentel, França e Omar (2003), para verificar os resultados, solicitou-se que um professor analisasse os resultados obtidos pelo algoritmo em relação à própria turma, e o professor afirmou que teria feito o mesmo agrupamento, e que os erros encontrados foram causados por respostas equivocadas dos alunos. Foi observado pelos autores que os resultados obtidos utilizando o K-means foram próximos aos resultados humanos.

4 RESULTADOS E DISCUSSÃO

A seguir é apresentada uma avaliação empírica dos algoritmos, bem como a dos os dados reais. Os vários classificadores foram testados em cenários diferentes. Mas não foram testadas todas as possibilidades para totalidade dos classificadores.

Utilizando o *dataset* Balões o algoritmo ID3 classificou as dezesseis instâncias corretamente, motivo pelo qual o algoritmo apresenta um forte viés indutivo, mantendo uma única hipótese válida durante todo o processo de busca, causando assim *overfitting*, ou seja, um sobre-ajuste aos dados de treinamento. Dependendo da característica do experimento *overfitting* pode ser útil na eliminação de *outliers*. Ao fazer uso do mesmo *dataset* no algoritmo J48 que é a implementação do C4.5, o resultado da classificação foi para quinze instâncias classificadas corretamente e uma incorretamente, justificando assim um melhoramento no combate do *overfitting*, utilizando uma estratégia de poda de árvore, o fato do algoritmo não ter classificado

todas as instâncias como 100% corretas, significa que houve uma melhora na generalização do classificador.

O *dataset* Censo de Renda foi o segundo cenário de execução dos algoritmos de classificação em que é usada também uma base de teste. Nesse segundo cenário o algoritmo ID3 não pode ser executado no Weka, pois o ID3 só pode lidar com atributos nominais, e não são permitidas instâncias incompletas, ou seja, todos os atributos da instância devem ser conhecidos. O J48 implementação do C4.5 no Weka é um aprimoramento do ID3 apresentado em um trabalho por Quinlan, (1993), intitulado como "C4.5: Programs for machine learning", tornando possível trabalhar com atributos categóricos e contínuos. Assim na execução do J48 o algoritmo classificou corretamente 85.83% e incorretamente 14.17% das instâncias, levando 2.62 segundos para construir o modelo.

Para geração dos modelos de verificação (clustering) foi utilizado o *dataset* Sementes, esse *dataset* é associado a tarefas de clusterizacão e será utilizado para identificar as diferentes variedades de trigo. Para essa tarefa utilizou-se o algoritmo K-means (implementado no WEKA com o nome SimpleKmeans). A aplicação do algoritmo K-means necessita da determinação de qual número de clusters serão gerados pelo algoritmo.

5 CONCLUSÃO

Este artigo teve por objetivo apresentar os algoritmos de aprendizagem supervisionada e não supervisionada e seus respectivos algoritmos como ID3, C4.5 e K-Means. A fim de avaliar suas características e desempenho com diferentes *datasets*. Além disso, foi apresentado um exemplo prático para demonstrar a utilização do algoritmo K-Means.

Tem-se, assim, uma boa contribuição do artigo para entendimento sobre mineração de dados e aprendizado de máquina com abordagens teóricas, testes e aplicações práticas.

Os resultados obtidos permitiram tomar conclusões sucintas, como o ID3 é um algoritmo com forte viés indutivo. O C4.5 é um aprimoramento do ID3 que torna possível trabalhar com atributos categóricos e contínuos, e utiliza a estratégia de poda de árvore para evitar *overfitting*. Nos algoritmos de agrupamento, o K-Means com elementos pertencentes a grupos rígidos e de formatos lineares, obteve a convergência dos grupos rapidamente em apenas 0.01 segundo.

Como trabalhos futuros poderão ser desenvolvidos novos análises de *datasets* explorando outras abordagens do aprendizado de máquina e mineração de dados, bem como a elaboração de pesquisas em relação aos temas tratados.

REFERÊNCIAS

DUA, D.; GRAFF, C. **UCI Machine Learning Repository.** Irvine, 2019. Disponível em: http://archive.ics.uci.edu/ml. Acesso em: 06 fev. 2019.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification.** Nova Iorque: Wiley-Interscience, 2000. 688 p. ISBN 978-0471056690.

FAYYAD, U.; PIATETSKYSHAPIRO, G.; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases.** AI Magazine, v. 17. p. 37-54, 1996. DOI https://doi.org/10.1609/aimag.v17i3.1230. Disponível em: https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230. Acesso em: 10 fev. 2019.

FONSECA, F. C. S.; BELTRAME, W. A. R. **Aplicações Práticas dos Algoritmos de Clusterização K-means e Bisecting K-means.** Vitória, 2010. 8 p. Disponível em: https://www.researchgate.net/publication/327121358. Acesso em: 07 fev. 2019.

HALL, M. et al. **The weka data mining software: An update.** 11 ed. Hamilton: SIGKDD Explor, 2008.

KULCZYCKI, P.; CHARYTANOWICZ, M. **A Complete Gradient Clustering Algorithm Formed with Kernel Estimators.** International Journal of Applied Mathematics and Computer Science, Zielona Góra, vol. 20, n. 1, p. 123–134, 25 mar. 2010. Disponível em: https://content.sciendo.com/view/journals/amcs/20/1/article-p123.xml. Acesso em: 30 jan. 2019.

MITCHELL, T. M. **Machine Learning.** Nova Iorque: McGraw-Hill, 1997. 414 p. ISBN 978-0070428072.

MONARD, M.C. et al. **Uma introdução ao aprendizado simbólico de máquina por exemplos.** São Carlos: ICMSC-USP, 1997.

PAZZANI, M. J. **Influence of prior knowledge on concept acquisition: experimental and computational results.** Journal of Experimental Psychology, Irvine, ano 416-432, n. 3, p. 17, 1991. Disponível em: https://www.ics.uci.edu/~pazzani/Publications/jeplmc.pdf. Acesso em: 07 fev. 2019.

PIMENTEL, E.P; FRANÇA, V. F.; OMAR, N. **A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização.** In: XIV SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 10., 2003, Rio de Janeiro. Anais [...]. Rio de Janeiro: NCE - IM/UFRJ, 2003. Disponível em: http://www.br-ie.org/pub/index.php/sbie/article/view/280. Acesso em: 10 fev. 2019.

QUINLAN, J.R. **Induction of Decision Trees.** Machine Learning, 1986. p. 81-106. DOI 10.1023/A:1022643204877.

QUINLAN, J.R. C4.5: **Programs for Machine Learning.** Machine Learning, 1993. p. 235-240. DOI https://doi.org/10.1007/BF00993309.