

Gerenciamento em Supply-chain: Adoção de Ferramentas de Data Mining na indústria de Transformação**Supply-chain Management: The use of Data mining tools in the manufacturing industry**

10.34140/bjbv2n2-011

Recebimento dos originais: 20/01//2020

Aceitação para publicação: 30/03/2020

Alexandre Acácio de Andrade

Doutor em Engenharia Elétrica pela Escola Politécnica da Universidade de São Paulo

Instituição: Universidade Federal do ABC

Endereço: Rua São Paulo, s/nº - Jardim Antares SALA 303 bl Delta

CEP 09606-070 - São Bernardo do Campo - SP – Brasil

E-mail: aacacio@ufabc.edu.br

Júlio Francisco Blumetti Facó

Doutor em Administração de empresas pela EAESP/FGV

Instituição: Universidade Federal do ABC

Endereço: Rua São Paulo, s/nº - Jardim Antares SALA 362 bl Delta

CEP 09606-070 - São Bernardo do Campo - SP – Brasil

E-mail: julio.faco@ufabc.edu.br

Ricardo Reolon Jorge

Doutor em Administração de Empresas pela Fundação Getulio Vargas de São Paulo

Instituição: Universidade Federal do ABC

Endereço: Rua São Paulo, s/nº - Jardim Antares SALA 362 bl Delta

CEP 09606-070 - São Bernardo do Campo - SP – Brasil

E-mail: reolon.ricardo@ufabc.edu.br

Renan Meirelles Horta de Oliveira

Graduado em Engenharia de Gestão pela Universidade Federal do ABC

Gerente de Negócios de Serviços - Siemens Healthcare

Avenida Mutinga, 3800 - Pirituba

CEP 05110-902 - São Paulo - SP - Brasil

E-mail: renan.meirelles@siemens-healthineers.com

RESUMO

Esta pesquisa visa a comprovação prática da influência que as tecnologias de informações geram nas tomadas de decisões e gestões das empresas, bem como em toda a sua cadeia de suprimentos para que estas sobrevivam em um mercado cada vez mais competitivo, justificando assim, os investimentos neste tipo de tecnologia. Para isto, estudou-se detalhadamente conceitos acerca dos temas: Logística – Gerenciamento em *Supply Chain*, Tecnologia da Informação e *Data Mining*. Posteriormente, foi elaborado um questionário e enviado as mais diversas empresas do estado de São Paulo para a coleta de dados. Consequentemente, criou-se um banco de dados e utilizou-se técnicas de associações em mineração de dados, através do algoritmo Apriori e o *software* Weka. Assim, obteve-se resultados que comprovaram, de maneira prática, as vantagens competitivas obtidas por

empresas que optam por investimentos nestas tecnologias (TI). Proporcionando, uma gestão da cadeia produtiva muito mais eficiente e responsiva.

Palavras-chave: Logística, Cadeia de suprimentos, Tecnologia da informação, Mineração de dados, Algoritmos.

ABSTRACT

This research aims at the practical confirmation of the influence that technologies generate information in decision-making and managements of firms as well as throughout its supply chain to enable them to survive in an increasingly competitive market, thus justifying investments in this type of technology. For this, we studied in detail concepts about themes: Logistics - Supply Chain Management, Information Technology, Manufacturing industry and Data Mining Technology. Afterwards, a questionnaire was developed and sent the most diverse companies in the state of São Paulo for the data collection. Consequently, we created a database and used techniques in data mining associations, through the Weka Apriori algorithm and software. Thus, we obtained results proved that, in a practical way, the competitive advantages obtained by companies that opt for investment in these technologies (IT). Providing a management much more efficient and responsive supply chain.

Keywords: Logistics; Supply chain, Information technology, Data mining, Algorithms.

1 INTRODUÇÃO

Atualmente, com a aceleração digital causada pela lei de Moore e a disseminação global da internet, é extremamente difícil atingir excelência em qualquer atividade sem a utilização extensa de recursos de tecnologia da informação (TI). A maioria das vezes as ferramentas de TI são utilizadas como auxiliares e não como centrais na tomada de decisões das empresas. De acordo com WITTEN et al, 2016, “o trabalho do empreendedor é identificar oportunidades – isto é, padrões de comportamento que podem ser transformados em negócios rentáveis e explorá-los”.

Percebe-se uma facilidade crescente em gerar, armazenar e transferir dados (AGRAWAL et al, 1993). Por esse motivo, o tratamento (AGRESTI, 2018) desses dados tem se tornado um grande diferencial para qualquer empresa, pois as informações são valiosas e os dados precisam passar por diversos tratamentos para se tornarem informações úteis, ou seja, se tornarem conhecimento (ver figura 1).

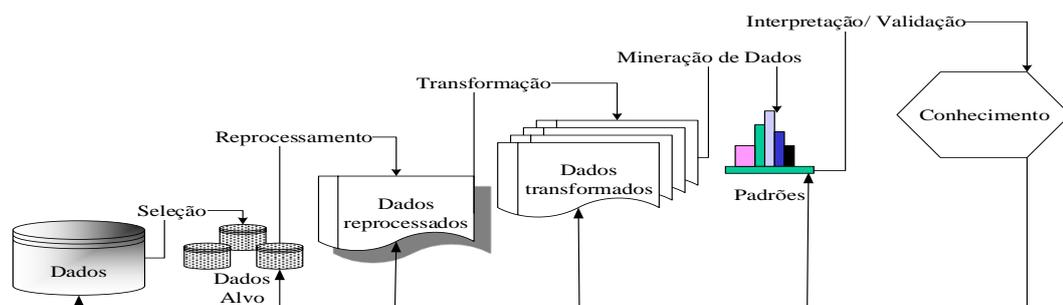


Figura 1 – Método de transformação de dados em conhecimento proposto por FAYYAD (1996) adaptado de ANDRADE (2007)

Podem existir diversas aplicações a bases de dados para uma grande gama de negócios, como citado GIUDICI, 2005 e FARIA, 2020. Muitas vezes existem padrões, difíceis de serem percebidos, que são muito úteis para os negócios, mas que não são trivialmente observados sem a mineração de dados (DUNHAM, 2006) ou alguma ferramenta específica de classificação ou agrupamento similar (FAYYAD et al, 1996). Alguns exemplos simples que podem ser mencionados: bancos de dados de lojas com informações sobre vendas e perfis de clientes; grandes empresas com dados não organizados (PAIVA et al, 2017) sobre os seus distribuidores e compradores; dados da localização da maior parte parceiros de determinada indústria, dentre outros.

O presente trabalho apresenta um estudo de caso utilizando uma ferramenta de mineração de dados, as suas características de uso e implementação, com foco na gestão empresarial estratégica. A ferramenta estudada foi aplicada em um banco de dados real e, com base nos resultados encontrados na mineração de dados, informações relevantes foram encontradas sendo as mesmas com potencial para auxiliar a tomada de decisão (NORVIG & RUSSEL, 2014) em empresas e negócios.

2 PESQUISA DA ATIVIDADE ECONÔMICA PAULISTA - PAEP

O banco de dados utilizado (BD) foi a Pesquisa da Atividade Econômica Paulista (PAEP) que foi realizada a pedido do SEADE (Sistema Estadual de Análise de Dados) dos anos de 1996 e 2001. O BD consiste em respostas de várias empresas à sobre suas atividades econômicas, essas respostas mostram o cenário das empresas de determinados segmentos nos anos propostos.

Em posse de tais bases, iniciou-se o processo de seleção de qual parte da base de dados seria usada. Tais bases reúnem informações de diversas empresas nos anos citados e encontram-se divididas por segmentos (setores: serviços, comércio, bancário, industrial, instituições financeiras e construções) e algumas vezes estavam separadas em versões gerais ou simplificadas. Por razões de facilidade de preenchimento de resposta, as perguntas encontram-se separadas em três blocos (administrativo-financeiro, técnico-produtivo e unidades produtivas industriais) e cada bloco é dividido em capítulos, conforme a Figura 2. Importante ressaltar que 100% das empresas classificadas como de grande porte na época responderam ao questionário.

O sub-banco de dados referente à indústria em geral foi escolhido por representar o cenário mais abrangente e, conseqüentemente, possuir um alinhamento maior com os objetivos propostos pelo trabalho. Além disso, a base utilizada foi a de 2001 por ser a mais recente, no total a base possuía 109 colunas (perguntas) e 11524 linhas (respostas), sendo que isso representa os blocos 1 e 2 do questionário (vide figura 2).

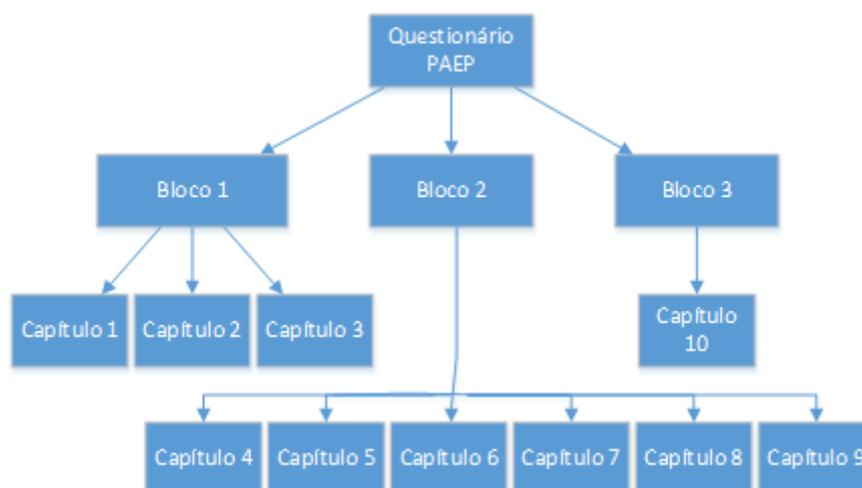


FIGURA 2 - Estrutura de capítulos do questionário da PAEP.

Devido as suas características de múltiplas famílias de algoritmos disponíveis, código aberto e gratuito, grande flexibilidade de utilização com alta usabilidade, neste trabalho optou-se pelo uso sistemático do WEKA (FRANK, 2014). Para os objetivos propostos, destacou-se o uso do algoritmo de associação implementado no WEKA, que foi utilizado no banco de dados, para encontrar relações ou associações entre valores de variáveis categorizadas e essa é a principal função desse algoritmo (NISBET et al, 2009).

3 RESULTADOS

Nesta seção serão apresentados os resultados analíticos extraídos a partir de diversos testes realizados no software WEKA. Os resultados serão apresentados conforme os capítulos mencionados anteriormente (figura 2). Os testes serão comentados de maneira geral, de acordo com os resultados mais relevantes observados. O número de regras variou de acordo com o tipo de teste, sendo que:

- Testes de capítulo simples observaram-se 50 regras;
- Testes por bloco observaram-se 100 regras;
- Teste dos blocos concatenados observaram-se 200 regras.

No entanto, algumas vezes as relações obtidas eram semelhantes, necessitando-se observar mais relações. Tal delimitação foi escolhida para permitir a viabilidade do estudo do banco de dados. Existiu um potencial de realização de estudo mais minucioso, já que a maioria das compilações podem gerar mais de mil respostas.

3.1 TESTE COM PERGUNTAS CORRELACIONADAS

Capítulo 1:

Este capítulo trata de informações patrimoniais, como o capital controlador, origem da empresa e natureza jurídica. Com as regras de associação, encontraram-se relações diretas da origem do capital com principal nacionalidade, principalmente em casos que o capital é nacional e a empresa é brasileira (maioria dos casos).

Um fato interessante observado aqui é que sempre que a empresa possui capital nacional, é brasileira e é independente implica nela não ter participado de uma joint venture nos anos citados. Isso poderia mostrar para um novo investidor no mercado brasileiro que empresas independentes fundamentalmente nacionais não eram abertas a se associar com outras empresas em modalidade joint venture.

Capítulo 2:

Nesse capítulo temos informações econômico-financeiras da empresa, sendo que muitas perguntas referem-se a valores numéricos e não foram consideradas devido ao grande número de diferentes faturamentos de empresas de vários portes. No entanto, o capítulo possui perguntas não numéricas, como se ela presta serviços ou vende produtos, como são suas relações com o mercado externo, como ela se financia e como ela é dividida financeiramente.

Das regras analisadas, algumas se destacaram, como por exemplo, o fato de empresas que não utilizam recursos próprios como forma de se financiar não exportarem e não possuem forma de relacionamento com a comunidade. Isso mostraria que empresas locais e que se financiam externamente não se preocupavam tanto com a comunidade em 2001.

Ainda nesse capítulo empresas que utilizam recursos próprios e empréstimos de bancos não se utilizam nunca de financiamento público em 98% dos casos. Isso mostra uma quebra entre empresas que utilizam dinheiro de um dono (pessoa física) em detrimento às pessoas jurídicas com faturamento alto. Isso é explicado pela modalidade de empréstimo citada (Proex), já que ele só é concedido a empresas com faturamento acima de 600 milhões.

Capítulo 3:

Aqui temos informações sobre recursos humanos e tudo relacionado a esse setor e suas atividades na empresa. Os valores relativos a salários brutos foram retirados, bem como a quantidade de empregados da empresa pela heterogeneidade dos números.

Os critérios que foram considerados nesse capítulo se referem a critérios de promoção dos empregados e quais benefícios são oferecidos a eles. Não foi observada nenhuma regra direta muito relevante, pois os funcionários podem ser promovidos por diversos fatores e os tipos de benefício também tem grande variabilidade.

Capítulo 4:

Esse capítulo trata da utilização de Tecnologia da Informação (TI) nas empresas. Foram colhidos dados sobre quais equipamentos eram utilizados principalmente abrangendo redes, disponibilidade e velocidade da internet.

Um dado interessante que pode ser observado já nos primeiros resultados: todas as empresas da pesquisa que utilizavam computador já possuíam acesso à internet e algum tipo de sistema automatizado. Isso demonstra a tendência já consolidada na época sobre a informatização de processos, o que se seria equivalente hoje em dia ao *cloud computing*.

É digno de nota que nenhuma relação foi encontrada relativa à existência de website da empresa ou não. Hoje em dia é impraticável que uma empresa considerada grande não tenha um site. Atentando-se ao fato de que 100% das empresas categorizadas como grandes em 2001 responderam ao questionário, não existir nenhum indício de relação direta entre a presença de um website entre os cem mil primeiros resultados é curioso. Desta observação, torna-se notável o quanto se tornou importante a presença virtual nos anos que se sucederam ao questionário.

Capítulo 5:

Extremamente ligada ao capítulo anterior, aqui temos informações sobre comércio eletrônico nas empresas e aos possíveis motivos caso ela não utilizasse. Algumas relações óbvias podem ser retiradas sem analisar nenhuma resposta, como por exemplo, o fato dela só realizar esse tipo de comércio caso tenha algum tipo de comunicação eletrônica. Essas relações não foram mostradas aqui devido ao fato da análise estar sendo feita isoladamente para o capítulo.

Além disso, esse capítulo ainda possui algumas outras relações óbvias, já que para uma empresa que não se utiliza de *e-commerce*, todas as perguntas relacionadas a essa prática serão negativas ou nulas, conforme mostrado pelas regras. Esse grande número de trivialidades dificultou a busca por regras valiosas, sendo que nenhuma muito relevante foi encontrada nos primeiros resultados.

Capítulo 6:

As informações contidas aqui são de grande importância porque mostram as inovações tecnológicas que a empresa se utilizava. O termo “inovação tecnológica” se aplica a incrementos em produtos e processos que usem uma tecnologia nova ou uma combinação de tecnologias pré-existentes. As perguntas nesse capítulo se referem a como ocorreu o incremento tecnológico, por quem foi introduzido, se o processo foi patenteado, como foi financiado e algumas outras questões referentes à tecnologia.

A grande maioria das relações encontradas neste capítulo se referia a casos em que a empresa não realizou inovações nem acordos para tal. Por esse motivo, as relações que seriam interessantes de se analisar não foram encontradas, nem nas cinquenta primeiras (conforme citado no começo da seção de resultados) nem nas primeiras cem mil regras.

Para uma empresa focada em tecnologia, seria interessante fazer uma triagem maior do BD, incluindo somente as empresas que fizeram inovações. No entanto, tal manipulação de dados para entendimento de um setor específico do BD foge do escopo inicial do presente trabalho, uma vez que todos os capítulos sofreram o mesmo pré-processamento e só receberiam tratamento especial caso houvesse algum erro operacional.

Capítulo 7:

Esse capítulo trata das estratégias de gerenciamento da empresa. Abrange detalhes como *portfolio* de produtos, estratégia utilizada para alavancar vendas, método de gerenciamento de estoque, de quantos clientes o faturamento da empresa depende, como eles são classificados e qual o principal ramo de atividades da empresa.

A BD deste capítulo, precisou de muitas manipulações para se tornar adequada ao processamento do WEKA. Surpreendentemente, isso facilitou muito a análise de dados do referido capítulo, apesar de algum problema com respostas negativas eventualmente ocorrerem. Uma das regras interessantes é a que empresas com certificação, que utilizam técnicas de produção e têm as vendas pulverizadas sempre utilizam métodos de qualidade e não transferiram processo produtivo (ambos com 99% de confiabilidade). Isso indica que empresas sem dependência de clientes tiram certificações para terem confiabilidade maior no mercado e fazem os processos internamente ao invés de terceirizar. Em 98% dos casos, isso ocorre com empresas atacadistas/varejistas também. Empresas que ampliaram a variedade de produtos oferecidos e aumentaram a escala de produção em 98% dos casos não transferiram processo produtivo.

Capítulo 8:

Este capítulo é pequeno comparado com qualquer outro, contando somente com oito perguntas sobre meio ambiente, sendo que metade é sobre investimentos em reduções dos problemas ambientais e a outra metade sobre as implicações das questões ambientais.

Poucas relações relevantes puderam ser observadas, mas uma delas chamou atenção. Em vários casos que a empresa citava a reutilização ou tratamento de resíduos e/ou efluentes, a resposta para perda de mercados era negativa. Essa relação mostra que ainda não havia grandes implicações mercadológica em ser ou não sustentável (*eco-friendly*). Hoje em dia empresas grandes que não contribuem com questões ambientais são mal vistas, principalmente no cenário internacional.

Capítulo 9:

O capítulo final trata de terceirizações dentro da empresa, basicamente definindo quais setores não são, são total ou parcialmente terceirizados.

Pela análise das respostas, as relações aqui parecem ser dispersas entre as perguntas, mas a maioria das empresas parece preferir terceirizar os serviços de contabilidade. Por outro lado, serviços de recrutamento e seleção de mão de obra tem a tendência de serem feitos internamente. Com exceção desses padrões interempresariais, não foi encontrada nenhuma relação relevante entre as respostas da mesma empresa.

Os resultados alcançados nos capítulos de 1 a 9 são apresentados resumidamente na tabela 1. Os itens foram discutidos individualmente em cada seção do respectivo capítulo.

TABELA1 – Resumo das tendências apresentadas nos capítulos.

| Capítulo 1 | Capítulo 2 | Capítulo 3 | Capítulo 4 | Capítulo 5 |
|--|--|--|---|------------------------|
| Capital nacional + Brasileira + Independente ↓ Sem joint venture (100%) | Não utilizar recursos próprios ↓ Não exporta (100%) + Não se relaciona com a comunidade (100%) | Nenhuma relação | Tendências na utilização de internet, mas empresas sem websites | Muitas regras triviais |
| Capítulo 6 | Capítulo 7 | Capítulo 8 | Capítulo 9 | |
| Grande número de respostas nulas para empresas que não investem em relação tecnológica | Certificação + Técnicas de produção + Vendas pulverizadas ↓ Métodos de qualidade (99%) + Não transferiram processo produtivo (99%) | Reutilização ou tratamento de resíduos e/ou efluentes ↓ Perda de mercado (Falta de tendência de sustentabilidade) | Tendência em terceirizar a contabilidade Tendência de recrutamento e seleção de mão de obra serem feitos internamente | |

3.2 TESTES COM PERGUNTAS DO MESMO BLOCO E COM TODAS AS PERGUNTAS

Observou-se que as regras têm uma forte tendência em ser criadas dentro dos seus respectivos capítulos, sendo que a maioria das relações encontradas já haviam sido apresentadas na seção anterior. O intuito das duas últimas seções era relacionar regras que associassem perguntas de diferentes capítulos e que trouxessem relações não triviais, ajudando as empresas a terem uma vantagem frente a outras empresas.

Algumas das relações que puderam ser observadas que relacionavam capítulos diferentes foram devido a casos negativos. Não importa onde as perguntas estejam, mas se 100% das respostas foram negativas, ela será incluída em uma regra já existente anteriormente (antes de juntar as perguntas). Um exemplo é relativo ao país de origem: caso o país de origem seja o Brasil, o capital controlador será nacional em 100% dos casos. No entanto, há outras perguntas que acontecem à mesma resposta em 100% dos casos, então elas poderiam ser colocadas na regra citada sem causar nenhum impacto. É esse tipo de relação que se observou na maioria dos casos analisados.

4 CONCLUSÃO

Após conhecer melhor as ferramentas presentes em programas de Mineração De Dados e definir o escopo central do projeto, começou-se a utilizar alguns dos programas mencionados nos objetivos específicos e optou-se por utilizar o WEKA para dar continuidade a estudos avançados na Mineração De Dados aplicada. As ferramentas respondem bem e são facilmente ajustáveis. Além disso, o programa é gratuito, fato que faz com que os resultados expostos possam ser reproduzidos por outros interessados no assunto.

A ciência da Mineração De Dados é algo complexo e extenso, que possui muitas aplicações no mercado e muitas delas não muito difundidas no mercado brasileiro. O presente trabalho buscou mostrar como ocorre a mineração de dados e alguns importantes algoritmos, utilizando-se de exemplos de fácil compreensão para qualquer pessoa que se interesse pelas ferramentas apresentadas.

Os algoritmos de classificação se mostraram extremamente úteis em encontrar padrões na base de dados proposta, mas por outro lado requereram um extenso trabalho operacional para viabilizar o seu uso. Além de transformar os dados em *string*, era necessário pensar em qual tipo de resposta faria sentido dentro de uma regra condicional. Esse tipo de raciocínio é o que os especialistas em Mineração De Dados ainda julgam necessário ao se trabalhar com dados: a intuição humana (NISBET et al, 2009). É preciso aprender a desenvolver esse tipo de lógica ao se trabalhar com banco de dados e conseguiu-se atingir esse objetivo.

A última etapa do trabalho saiu do nível operacional e entrou no nível administrativo dos bancos de dados, ou seja, as relações encontradas até então serão questionadas quanto a sua trivialidade, objetividade e utilidade. Algumas regras serão óbvias, mas o trabalho busca encontrar as não triviais e com isso conseguir auxiliar a tomada de decisão de uma empresa em posse desses dados. Além disso, em oposição aos testes de perguntas correlatas, serão realizados testes utilizando a base completa, procurando padrões escondidos, inesperados, e que sejam relevantes na tomada de decisão a nível diretorial ou superior. Os aspectos analisados da base de dados geram um quadro geral da situação externa à empresa, mas ferramentas similares poderiam ser utilizadas no auxílio à tomada de decisão com fatores internos. O foco do trabalho foi mostrar que esses padrões encontrados podem sim ajudar empresas a terem vantagem competitiva se utilizadas da maneira adequada.

Os resultados mostraram quão fundo uma mineração de dados pode ir. Foram encontradas milhões de regras e poucas dela foram relevantes. A relação entre casos negativos é muito mais fácil de ser, uma vez que eles estavam presentes em número muito maior no BD. Para obter vantagem competitiva real, a empresa precisaria filtrar o BD melhor para algum segmento que ela queira entender, conforme foi citado no resultado para o capítulo 6 (ver seção anterior).

Para aprimoramentos futuros, poderia se destacar alguns pontos, como filtragem de dados aplicada a algum tipo específico de empresa (inovação tecnológica, por exemplo), melhor agrupamento de todas as perguntas (como foi feito com o capítulo 7 por uma questão operacional) e aprimoramento de algoritmos do WEKA para alguma ferramenta mais robusta. Era possível filtrar alguns dados no programa, mas o processo em si é uma “caixa preta”, o que não permite escolher o que é mais ou menos importante para o usuário. Com um programa mais refinado, como o Minitab por exemplo, seria possível fazer o algoritmo aprender o que é mais relevante por meio de aprendizagem de máquina ou simplesmente priorizando o tipo de relação que o usuário busca (no caso, relações entre perguntas de diferentes capítulos, por exemplo).

AGRADECIMENTOS

O presente trabalho foi desenvolvido com recursos da UFABC. Os autores agradecem especialmente ao Fórum de Inovação e Competitividade Sustentável da UFABC (FICS/UFABC) pelo apoio nas pesquisas.

REFERÊNCIAS

- AGRAWAL, Rakesh; IMIELIŃSKI, Tomasz; SWAMI, Arun. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data. 1993. p. 207-216.
- AGRESTI, Alan. An introduction to categorical data analysis. John Wiley & Sons, 2018.
- ANDRADE, A.A. Desenvolvimento de Sistema Especialista com Operacionalidade de Aprendizado para Operar em Tempo Real com Sistemas Industriais Automatizados. Tese de Doutorado apresentada à Universidade de São Paulo, 2007.
- FARIA, V. T.; ANDRADE, A. A.; SANTOS, J. P.; FACÓ, J. F. B.; GASI, F. Measuring The Impacts of Database Processing Utilization in Innovation Processes On Companies. INTERNATIONAL JOURNAL OF DEVELOPMENT RESEARCH., v.10, p.34190 - 34194, 2020.
- PAIVA, F. A.; JORGE, R. R.; ANDRADE, A. A.; JUNGER, A. P.; FACO, J. F. B. Utilização Da Teoria Das Opções Reais Em Estudo De Caso Para Avaliação De Troca De Ponto De Venda Para Loja De Varejo Em Shopping Center. Revista de Casos e Consultoria., v.8, p.1 - 24, 2017.
- DUNHAM, Margaret H. Data mining: Introductory and advanced topics. Pearson Education India, 2006.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, Providence, Rhode Island, n. 97, p.37-54, Fall, 1996.
- FRANK, Eibe. Fully supervised training of Gaussian radial basis function networks in WEKA. 2014.
- GIUDICI, Paolo. Applied data mining: statistical methods for business and industry. John Wiley & Sons, 2005.
- NISBET, Robert; ELDER, John; MINER, Gary. Handbook of statistical analysis and data mining applications. Academic Press, 2009.
- NORVIG, Peter; RUSSEL, Stuart. Inteligência Artificial [Ebook Kindle]. Tradução: Regina Célia Simille. Rio de Janeiro: Editora Campus-Elsevier, 2014.
- WITTEN, I.H.; FRANK, E.; HALL, M.A.; PAL C.J. Data Mining - Practical Machine Learning Tools and Techniques, Morgan Kaufmann Series in Data Management Systems, Ed 4, 2016.