

**UMA ANÁLISE COMPARATIVA DA UTILIZAÇÃO DE MINERAÇÃO DE DADOS EM APOIO AS TOMADAS DE DECISÃO: ANÁLISE DA MINERAÇÃO DE DADOS NA TOMADA DE DECISÕES ENVOLVENDO ESTUDANTES DE PROGRAMAÇÃO COM DIFICULDADES, EDUCAÇÃO A DISTÂNCIA E FERRAMENTAS PEDAGÓGICAS**

**A COMPARATIVE ANALYSIS OF THE USE OF MINING SUPPORTING DATA AS DECISION-MAKING: ANALYSIS OF THE DATA MINING IN DECISION-MAKING INVOLVING PROGRAMMING STUDENTS WITH DIFFICULTIES, EDUCATION A DISTANCE AND PEDAGOGICAL TOOLS**

Lucas de Moura Soares\*  
Leonardo Carlos Balbino\*\*

**RESUMO**

Este trabalho tem por objetivo a realização de um estudo a respeito da utilização da mineração de dados como auxílio na tomada de decisão a partir da análise de outros artigos que utilizaram esta técnica na área da educação, mais especificamente em alunos com dificuldade de programação e metodologias de ensino a distância e então, apresentar uma comparação crítica de todos os casos utilizados. Este artigo, portanto, não só mostra definições envolvendo mineração de dados, mas também mostra que ela pode ser utilizada em situações reais do dia a dia.

**Palavras-chave:** Decisão. Dados. Mineração. Educação.

**ABSTRACT**

This work aims to conduct a study regarding the use of data mining as an aid in decision making from the analysis of other articles that use this technique in the area of education, more specifically in students with programming difficulties and distance learning methodologies and then present a critical comparison of all the cases used. This article, therefore, not only shows definitions involving data mining, but it also shows that it can be used in real day-to-day situations.

**Keywords:** Decision. Dice. Mining. Education.

**Introdução**

Atualmente, a quantidade de dados que uma empresa armazena é gigantesca, e em diversos momentos surgem situações em que é necessário tomar decisões a partir desses

---

\* Graduado do Curso de Sistemas de Informação, da UNDB. [lcsmoura95@gmail.com](mailto:lcsmoura95@gmail.com)

\*\* Graduado do Curso de Análise e Desenvolvimento, da FACAM – Faculdade do Maranhão.  
[leonardocarlos@live.com](mailto:leonardocarlos@live.com)

dados. Assim como em outras áreas, a educacional reúne uma quantidade gigantesca de dados, e portanto, há os dados de menor e maior relevância.

Com a expansão do ensino a distância, surgem os AVA's – Ambiente virtual de aprendizagem, e junto a isso a quantidade de dados aumenta necessitando de outras tecnologias para o controle dessas informações.

A desorganização dos dados pode ser cada vez maior a medida que a instituição cresce, e podem ficar mal distribuídos quando não é dada atenção ao aumento do volume de informações. Então a instituição pode começar a enfrentar problemas além do esperado, e ao invés de resolver o problema com a ajuda dos dados, ela terá primeiro que organizá-los, aumentando o tempo de solução.

As instituições modernas prezam pela qualidade de seus serviços e para isso é necessário estar em constante processo de atualização para acompanhar a evolução tecnológica. Para tal, as instituições de ensino precisam fazer o levantamento dos dados relacionados a cada problemática e a partir das técnicas de mineração de dados, encontrar o melhor caminho para a aplicação de metodologias a partir dos dados levantados.

## **1 Fundamentação Teórica**

A construção deste artigo tem por finalidade analisar como a mineração de dados é utilizada, através da comparação do uso desta em 3 artigos distintos.

A aprendizagem a partir dos dados é extremamente poderosa, e seu uso está transformando a tomada de decisões de empresas dos mais diversos setores em um ritmo acelerado, poupando dinheiro e até mesmo vidas. Este é um momento animador para ser um analista de dados! Para encontrarmos excelência nesse trabalho, precisamos ouvir atentamente com o objetivo de transformar um desafio do mundo real em um problema parecido, mas solucionável (ELDER, 2009).

Antes de iniciar o estudo é necessário que se tenha conhecimento do que é a mineração de dados e onde ela é utilizada. Voltando alguns anos no passado, quando os sistemas de computadores começaram a surgir, as empresas começaram a notar um grande crescimento no volume de dados que precisariam ser armazenados e, com o passar do tempo, a evolução dos hardwares possibilitou que a quantidade de dados armazenados fosse bem maior.

A parte mais importante de qualquer projeto de mineração de dados é definir claramente o problema a ser solucionado. Nenhum modelo conta sozinho a história completa. Não existe uma regra que diz quando você já esgotou os dados (ABBOT, 2017).

A mineração de dados tem por objetivo filtrar um grande volume de dados, para então selecionar os que realmente importam para assim melhor aproveitá-los e conseguir melhores resultados no objetivo que uma organização estiver almejando. O uso desse processo está presente em várias áreas do mundo real, como Educação, Comunicações, Médica e até mesmo pelo governo.

A mineração de dados possui três tipos de modelagem: Descritiva, Prescritiva e Preditiva, cada uma com suas próprias técnicas e métodos que utilizam tipos diferentes de perguntas para poder chegar na decisão.

A modelagem descritiva utiliza agrupamentos para identificar semelhanças com o intuito de ajudar a determinar os motivos de um sucesso ou fracasso de uma situação. As técnicas desta modelagem incluem:

- Clustering – agrupamento de registros que possuem semelhança.
- Detecção de anomalias – Identifica valores diferentes em diversos ambientes (Multidimensionais).
- Associação – Detecta as relações que estão entre os registros.
- Componentes principais (Análise) – Detecta relações entre variáveis.
- Agrupamento por afinidade – Agrupa pessoas pelos seus interesses ou objetivos em comum.

Na modelagem prescritiva o foco é voltado para as variáveis internas, externas, incluindo as restrições para que seja possível indicar uma ação a ser tomada. Dentre as principais técnicas dessa modelagem estão:

- Análise preditiva unida às regras – Desenvolve regras if/Then (Se-então) de padrões utilizados e prevê resultados.
- Otimizar o marketing – Simula a combinação de mídias com mais vantagem para que a empresa possa ter o maior ROI (Return of investment).
- Por fim, a modelagem Preditiva classifica eventos do futuro ou estima resultados ainda desconhecidos. A probabilidade é um exemplo de uma situação em que esta modelagem está presente. Dentre suas técnicas estão:
  - Regressão – Utiliza modelos matemáticos para medir a relação de uma variável dependente em relação a várias outras variáveis independentes.
  - Redes Neurais – São programas de computador responsáveis por detectar padrões e também de fazer previsões.

- Árvores de decisão – São diagramas representativos e avaliativos de problemas envolvendo decisões em sequência. Coloca riscos e resultados em destaque para poder reunir os cursos de ação.
- Máquinas de vetores de suporte – São modelos supervisionados de aprendizagem, responsáveis por analisar dados e reconhecer padrões.

Como auxílio na mineração de dados haverá diversos momentos em que será utilizado o software WEKA, que é do tipo open source e é bastante popular entre os estudantes e professores de instituições de ensino superior.

## **2 Avaliação de Desempenho de Estudantes em Cursos de Educação a Distância Utilizando Mineração de Dados**

Como apoio a atividade educacional, o uso de programas que gerenciam um grande número de dados relacionados aos estudantes tem se tornado uma incitação aos responsáveis pela manipulação das informações.

Nas instituições que oferecem curso a distância, é crucial que ela disponibilize um ambiente exclusivo para esse meio, em português conhecido como AVA. Conseqüentemente estes ambientes irão guardar muitas informações sobre alunos, notas e professores. No presente artigo os professores pretendem fazer uma previsão de desempenho dos alunos, e com isso aumentar a probabilidade de acerto utilizando a mineração de dados.

Foi utilizado a classificação de análise dentre as técnicas de mineração disponíveis. Esta técnica desenvolve modelos de inferência de uma situação particular, chamada: variável a ser prevista; e realiza a junção de outras visões desta variável, chamada: variáveis preditoras.

Foi feita primeiramente a representação dos estudantes no AVA, sendo esta representação, classificada em três dimensões: Na primeira é identificado os dados referentes ao planejamento, organização e o modo do estudante em gerir o tempo na realização do curso; A segunda dimensão foca na interação entre alunos, objetivando observar com qual frequência os estudantes comunicam-se entre si utilizando a plataforma; A terceira dimensão preza pela interação entre alunos e professores para observar como funciona a interação entre eles.

Após a classificação foi apresentado a seleção de uma disciplina, levando em consideração a maior quantidade de alunos concluintes da disciplina, maior quantidade

da oferta da disciplina para diferentes turmas, a disponibilidade do resultado das avaliações e o maior número de funções do AVA que foram utilizadas. A seguir está o modelo utilizado pelos autores:

Tabela 1: Representação das dimensões e variáveis. Fonte: Avaliação de desempenho de estudantes em cursos de educação a distância utilizando mineração de dados, 2012

Dimensão	Atributo	Descrição
Perfil Geral de uso do AVA	nr_acessos	Número total de acesso ao AVA
	nr_posts_foruns	Número total de postagens realizadas em fóruns
	nr_post_resp_foruns	Número total de respostas postadas em fóruns referindo-se a postagens de outros participantes (estudantes, professores, tutores)
	nr_post_rev_foruns	Número total de revisões em postagens anteriores realizadas em fóruns
	nr_sessao_chat	Número de sessões de chat que o estudante participou
	nr_msg_env_chat	Número de mensagens enviadas ao chat
	nr_questoes_resp	Número de questões respondidas
	nr_questoes_acert	Número de questões respondidas corretamente
	freq_media_acesso	Frequência média em que o estudante acessa o AVA
	tempo_medio_acesso	Tempo médio de acesso ao sistema
	nr_dias_prim_acesso	Número de dias transcorridos entre o início do curso e o primeiro acesso do estudante no AVA
Interação Estudante- Estudante	tempo_total_acesso	Tempo total conectado no sistema
	nr_post_rec_foruns	Número de postagens do estudante que tiveram respostas feitas por outros estudantes.
	nr_post_resp_foruns	Número de respostas que o estudante realizou em postagens feitas por outros estudantes.
	nr_msg_rec	Número de mensagens recebidas de outros participantes durante a realização do curso.
Interação Estudante- Professor	nr_msg_env	Número de mensagens enviadas a outros participantes durante a realização do curso.
	nr_post_resp_prof_foruns	Número de postagens de estudantes que tiveram respostas feitas por professores ou tutores do curso
	nr_post_env_prof_foruns	Número de postagens de professores ou tutores que tiveram respostas feitas por estudantes
	nr_msg_env_prof	Número de mensagens enviadas ao professor/tutor durante a realização do curso.
Objetivo da Previsão	nr_msg_rec_prof	Número de mensagens recebidas do professor/tutor durante a realização do curso.
	resultado_final	Resultado final obtido pelo estudante no curso. <u>Representa classe objetivo da técnica de classificação.</u>

Observa-se na tabela 1 que as dimensões trabalham juntamente com os atributos e suas descrições para formar a previsão do resultado final.

Foi utilizado os algoritmos RandomForest e MultiplayerPerceptron no software Weka para selecionar atributos e então obter variáveis mais precisas. Anteriormente foram criadas 3 classes de alunos:

- Alunos que apresentaram desempenho superior – 22 alunos com variação de notas entre 87-97.
- Alunos com desempenho mediano – 109 alunos com variação entre 77-87.
- Alunos com desempenho abaixo da média – 24 alunos com variação entre 67-77.

A partir do resultado foi possível constatar que a situação levantada poderia ser utilizada pelos professores de maneira a acompanhar individualmente os alunos que utilizam o AVA, para assim planejar estratégias buscando minimizar o índice de reprovações.

### **3 Mineração de Dados na Identificação de Grupos de Estudantes com Dificuldades de Aprendizagem no Ensino de Programação**

As salas de aula sempre foram compostas por estudantes que possuem maneiras de aprender diferentes. Isso quer dizer que uma mesma metodologia pode não ser efetiva para todos. O objetivo do artigo 2 é portanto, formar grupos de estudantes com dificuldades em programação para então reunir estratégias de ensino para melhorar o desempenho dos alunos, utilizando a mineração de dados.

No caso apresentado foram utilizados dados oriundos das avaliações da disciplina Programação orientada a objetos, com o total de 33 estudantes matriculados. Foi realizado um escopo e dentro deste os dados foram reunidos com a aplicação de 11 sessões de avaliação: 5 listas de exercício, 4 miniprovas e 2 projetos.

É criado então um quadro para cada aluno contendo informações como id, sessão, conteúdo em que ele irá ser avaliado, nível cognitivo e seu desempenho na avaliação.

Tabela 2: Atributos para o mapeamento do conhecimento do estudante

<b>Atributo</b>	<b>Descrição</b>	<b>Tipo de Dado</b>	<b>Domínio</b>
IdAluno	Código que identifica o aluno participante da sessão de avaliação	Numeric	[1, n]
IdSessao	Código que identifica o número da sessão de avaliação e a sequência temporal	Numeric	[1, n]

IdItem	Código que identifica uma questão numa avaliação formativa	Nominal	[Q <sub>01</sub> , Q <sub>n</sub> ]
Conteúdo	Conteúdo tratado no item avaliado	Nominal	[C <sub>01</sub> , C <sub>n</sub> ]
NívelCognitivo	Nível cognitivo do item avaliado de acordo com a Taxonomia de Bloom	Nominal	LEM – Lembrar ENT – Entender APL – Aplicar ANA – Analisar AVA – Avaliar CRI – Criar
ConceitoItem	Grau de acerto do aprendiz no item avaliado	Nominal	A – 8 ≥ nota ≤ 10 B – 6 ≥ nota < 8 C – 4 ≥ nota < 6 D – 2 ≥ nota < 4 E – 0 ≥ nota < 2 ? – Ausência de informação
ConceitoAvaliacao	Nível de desempenho do aprendiz numa determinada avaliação formativa	Nominal	A – 8 ≥ nota ≤ 10 B – 6 ≥ nota < 8 C – 4 ≥ nota < 6 D – 2 ≥ nota < 4 E – 0 ≥ nota < 2 ? – Ausência de informação
RotuloSomativa	Rótulo da avaliação somativa	Nominal	APR – Aprovado REP – Reprovado

Fonte: Franca (2013)

Durante o tempo de treinamento foram feitos agrupamentos compostos de 2 até 10 grupos, nomeados clusters para identificar os grupos com dificuldades homogêneas presentes na modelagem descritiva.

Com isso foi utilizado o software WEKA junto a um algoritmo de agrupamento (cluster) a partir de 890 registros relacionados a disciplina Programação Orientada a Objetos.

Attribute	Full Data (890)	Cluster#		
		0 (113)	1 (144)	2 (351)
Conteúdo	Arrays	Arrays	Arrays	Estruturas de Repeticao
NívelCognitivo	CRI	ENT	CRI	CRI
ConceitoAvaliacao	?	C	B	?
RotuloSomativa	REP	APR	APR	REP

Figura 1: Agrupamento de Clusters

3 (122)	
Componentes Graficos do Pacote GUI + Objetos e Classes	
APL	
A	
REP	

Figura 2: Agrupamento de clusters

	4	5
	(73)	(87)
<b>Estruturação de Sistemas em Camadas</b>	<b>ENT</b>	<b>ARR</b>
<b>E</b>	<b>E</b>	<b>E</b>
<b>REP</b>	<b>REP</b>	<b>REP</b>

Figura 3: Agrupamento de clusters

No cluster 0 foi o agrupamento relacionado aos alunos que alcançaram rendimento “C” na avaliação sobre arrays e que foram aprovados.

No cluster 1 foram reunidos os estudantes com rendimento “B”, na avaliação sobre arrays e que foram aprovados.

No cluster 2 foram agrupados os estudantes que foram reprovados, e que não responderam as questões da prova a respeito de estruturas de repetição.

No cluster 3 foi agrupado os estudantes que fizeram de maneira satisfatória, questões de componentes gráficos sobre GUI, objetos e classes, mas foram reprovados.

No cluster 4 foram agrupados os alunos que tiveram baixo rendimento em Estruturação de sistemas em camadas e, portanto, foram considerados reprovados.

No cluster 5 mostra os estudantes que obtiveram baixo rendimento na avaliação sobre arrays e os alunos participantes desse grupo foram considerados reprovados.

#### **4 Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente**

Neste artigo é apresentado uma pesquisa que busca, com base na mineração de dados, identificar alunos que podem estar pensando em desistir ou reprovar, dentro do ambiente de ensino à distância, para formar possíveis alterações que permitam a implementação de alertas.

Para o desenvolvimento do artigo, foi realizada a coleta de dados de diferentes maneiras: Entrevistas com professores, acompanhamento da prática docente no ensino a distância juntamente com a análise de frequência e quantidade de acessos e ações realizadas dentro do ambiente virtual, para então relacioná-las com os sintomas de evasão, reprovação e aprovação.

Neste caso foi utilizado um software de mineração WEKA, e tendo como base 161 estudantes, de duas turmas diferentes, mas de uma mesma disciplina. Os alunos



pertencem a três polos diferentes (unidade presencial). Dentro da amostra coletada 78 são homens e 83 são mulheres, sendo que 92 dos alunos possuem idade entre 16 e 25, 52 deles entre 26 e 35, 11 possuem entre 36 e 45 e apenas 6 possuem mais de 46 anos. Foi observado a época do primeiro e último acesso ao AVA (variação de 0 a 80), a quantidade de acesso a biblioteca virtual (variação de 0 a 50), quantidade de postagens no fórum (variação de 0 a 12) e a quantidade de avaliações entregues (variação de 0 a 3).

Os dados reunidos, foram, portanto, relacionados as notas que os estudantes obtiveram durante o período de três meses em que houve o acompanhamento na disciplina.

A partir dos dados foi utilizado a ferramenta RapidMiner para gerar regras no formato Se/Então, objetivando descobrir mais sobre a relação comportamental dos alunos (acesso, entrega de atividades) com informações de ordem demográfica (faixa etária, curso, sexo e polo), para assim extrair as repercussões sobre o desempenho dos alunos para então pensar na ideia da criação do alerta para os professores. Além disso foi aplicado outra técnica da mineração de dados: as Árvore de decisão conforme a figura.

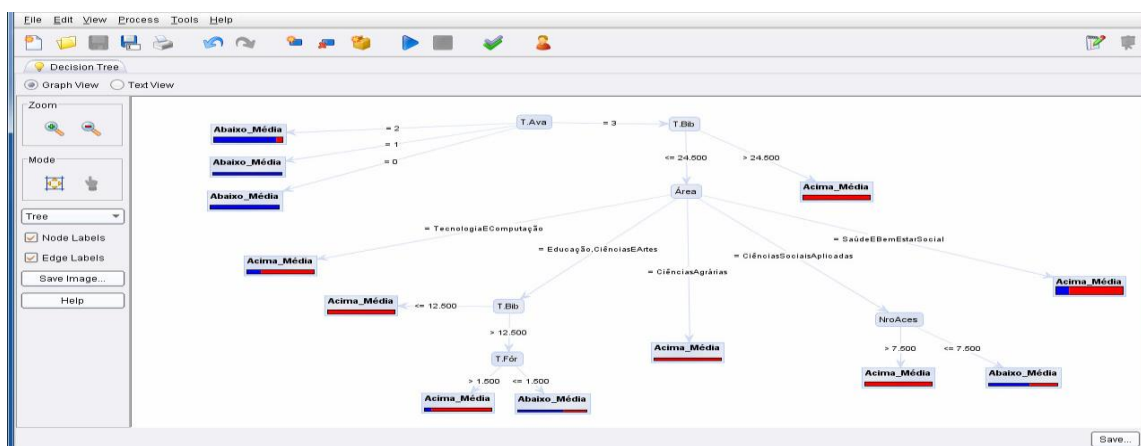


Figura 4 - Árvore de decisão do caso. Fonte: Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio a prática docente, 2008

A partir desta árvore de decisão foi possível constatar que os alunos que não realizaram as avaliações tiveram notas abaixo da média. E todos os que acessaram os materiais presentes no AVA mais de 24 vezes, conseguiram tirar boas notas. Já para alunos de Ciências, Artes e Educação foram categorizados pelo número de postagens nos fóruns, com resultados de  $x \geq 2$ , sendo  $x$  a quantidade de postagens no fórum, e os que cumpriram esse requisito conseguiram tirar nota acima da média.

## **5 Análise Comparativa**

Todos os casos apresentados envolvem uma área: A educação. Porém cada um difere na maneira em que procuram resolver seu problema com a ajuda da mineração de dados.

No trabalho 1 foi utilizado o modelo de classificação de análise presente na modelagem descritiva através da coleta de dados semelhantes, para categorizar um determinado grupo por dimensões além do uso do software Weka para seleção de dados importantes através das técnicas de Witten, Frank e Hall.

No trabalho 2 a técnica utilizada foi a de cluster para formar agrupamentos de estudantes com dificuldades semelhantes através de dados como desempenho e nível cognitivo. A divisão das interpretações possíveis divididas em clusters facilitou o entendimento, classificação e divisão dos alunos.

No trabalho 3 foi utilizado a análise prescritiva, sendo assim tratando de variáveis internas e externas. Especificamente foi desenvolvido uma regra Se/então pertencente a técnica de análise preditiva aliada a regras, aplicando-as na relação comportamental dos alunos com apoio de softwares de mineração de dados (WEKA).

Foi constatado, por tanto que além das técnicas disponíveis há também os softwares de apoio na mineração dos dados que vem representando uma boa opção na solução ágil do problema. Dentre os métodos utilizados o mais fácil de identificar os dados foi o de árvore de decisão.

### **Considerações Finais**

Conforme os resultados dos casos apresentados, confirma-se a importância das técnicas de mineração de dados, que se limitam apenas a educação, mas podem se expandir de forma positiva incrementando a qualidade da tomada de decisões de muitas organizações.

Apesar de envolver um mesmo contexto, observa-se que as técnicas de mineração diferem de acordo como a situação é analisada, e utilizadas de acordo com as necessidades impostas, desde a escolha da modelagem até o momento de escolha da técnica a ser utilizada.

Devido a expansão de cursos e graduações a distância o acúmulo de dados é inevitável e a utilização das técnicas de mineração de dados é de grande importância na reunião de informações relevantes.

### **Referências**

CAMILO, C.; SILVA, J.; **Mineração de dados: conceitos, tarefas, métodos e ferramentas**. Goiás: UFG, 2009.

Disponível em: [http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-09.pdf](http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf). Acesso em: 20 maio 2017.

FRANÇA, R. **Mineração de Dados na Identificação de Grupos de Estudantes com Dificuldades de Aprendizagem no Ensino de Programação**. Universidade Federal de Pernambuco, 2012. Disponível em:

<http://www.seer.ufrgs.br/index.php/renote/article/view/41634/26413>. Acesso em: 15 maio 2017.

GOTTARDO, E.; KAESTNER, C.; NORONHA, R.; **Avaliação de Desempenho de Estudantes em Cursos de Educação a Distância Utilizando Mineração de Dados**.

IFRS, UTFPR, 2012. Disponível em: <http://br->

[ie.org/pub/index.php/desafie/article/view/2773/2426](http://br-ie.org/pub/index.php/desafie/article/view/2773/2426). Acesso em: 20 maio 2017.

KAMPFF, A.; REATEGUI, E.; LIMA, J. Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente. **CINTED-UFRGS Novas Tecnologias na Educação**, v. 6, n. 2, p. 1-8, 2008. Disponível em:

[https://www.researchgate.net/profile/Eliseo\\_Reategui/publication/237652662\\_Minerao\\_de\\_dados\\_educacionais\\_para\\_a\\_construcao\\_de\\_alertas\\_em\\_ambientes\\_virtuais\\_de\\_aprendizagem\\_como\\_apoio\\_a\\_pratica\\_docente/links/53da36820cf2a19eee882d95.pdf](https://www.researchgate.net/profile/Eliseo_Reategui/publication/237652662_Minerao_de_dados_educacionais_para_a_construcao_de_alertas_em_ambientes_virtuais_de_aprendizagem_como_apoio_a_pratica_docente/links/53da36820cf2a19eee882d95.pdf).

Acesso em: 20 maio 2017.

PORTAL SAS, **Mineração de Dados – O que é e por que é importante?** Disponível em: [https://www.sas.com/pt\\_br/insights/analytics/mineracao-de-dados.html#](https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html#). Acesso em: 7 maio 2017.