

MINERAÇÃO DE DADOS NA GESTÃO DE RECURSOS HÍDRICOS SUBTERRÂNEOS: Estudo de caso

DATA MINING IN THE GROUNDWATER RESOURCES MANAGEMENT: Case Study

*Pedro Henrique Gerolin,
Especialista em Gerenciamento de Banco de dados,
Universidade Nove de Julho, UNINOVE
01156-050, São Paulo, SP,
E-mail: pedro_gerolin@yahoo.com.br*

*Claudineia Helena Recco, Mestre,
Universidade Nove de Julho, UNINOVE
01156-050, São Paulo, SP,
E-mail: ch_recco@yahoo.com.br*

Resumo

Devido ao grande crescimento da demanda dos recursos hídricos e sua grande escassez nos últimos anos principalmente, uma alternativa viável e lógica é a exploração das águas subterrâneas, principalmente comparando o custo, a qualidade e quantidade das águas superficiais com as águas subterrâneas. Assim, o recurso hídrico subterrâneo vem ganhando um espaço importante na realidade brasileira, portanto deve ter sua gestão melhorada sempre que possível, e graças a evolução tecnológica, principalmente focando no processo de Mineração de Dados isso se torna cada vez mais fácil.

Palavras Chaves: Recursos Hídricos, Águas Subterrâneas, evolução tecnológica, Mineração de Dados

Abstract

Due to the large growth in demand for water resources and their great shortage in recent years, a viable and logical alternative is the exploration of groundwater, mainly when comparing the cost, quality and quantity of surface water with groundwater. This way, the underground water resource has been gaining important space in the Brazilian reality, so it should be improved as much as possible, and thanks to the technological evolution, especially focusing on the Data Mining process, this becomes easier and easier.

Key Words: Water Resources, Groundwater, technological evolution, Data Mining

1. Introdução

A quantidade de água doce existente no mundo corresponde a menos de 1% do total, todo o restante ou é água salgada, cerca de 97,5% ou está congelada nos pólos do planeta, cerca de 1,9%, a grande maioria da água doce está contida em aquíferos subterrâneos. Os recursos hídricos subterrâneos são de extrema importância para toda a população, por isso devemos achar uma maneira de melhorar a gestão da mesma e se possível inovar, descobrir novas maneiras de controlar a qualidade dessas águas. Devemos tomar conta dessa água devido a grande importância da mesma para a existência humana.

Existem diferentes formas de analisar a qualidade da água, no estado de São Paulo a CETESB é responsável por essa tarefa, utilizando tubos de coletas subterrâneos, é possível analisar inúmeros parâmetros da água, sendo eles físicos e químicos.

Apesar de ser possível coletar e analisar esses parâmetros variados, será que é possível obter informações que possam facilitar as ações de melhoria e gestão desses recursos hídricos? De uma maneira mais fácil, tem como saber se quando as amostras de água possuem determinadas características de poluição, podemos definir outras características importantes que necessitam de alguma intervenção ou prevenção?

O objetivo geral desta pesquisa é mostrar que é possível utilizar a mineração de dados para fazer uma análise desses parâmetros coletados de uma forma até então não feita, e conseguir descobrir correlações entre alguns parâmetros antes não conhecidos. A mineração de dados, ou o Data Mining como é mais conhecido, é uma ferramenta de grande importância para análise de um grande volume de dados e com algoritmos diferenciados que nos permitem analisar com técnicas diferentes novas informações.

Este trabalho visa mostrar que é possível através da mineração de dados, especificamente utilizando as técnicas de classificação e associação que é possível obter informações importantes a respeito das amostras coletadas que dificilmente seria possível correlacioná-las manualmente, uma vez que a quantidade de parâmetros e amostras é muito grande, e também não existe uma lógica já conhecida para isso, o intuito da técnica de associação é conhecer a existência dessas novas lógicas, desses correlacionamentos entre os parâmetros.

Metodologicamente, foi adotada neste trabalho a pesquisa descritiva, focando em um estudo de caso para mostrar que é possível realizar a mineração de dados para complementar as informações sobre a gestão hídrica subterrânea no estado de São Paulo.

2. Análise de Águas Subterrâneas

A CETESB - Companhia Ambiental do Estado de São Paulo é o órgão estadual que ficou responsável legalmente pela realização do monitoramento da qualidade das águas subterrâneas desde 1970, a partir do ano de 1990 o monitoramento passou a ser realizado de forma sistemática, quando foi estruturada a rede estadual composta por poços tubulares utilizados para o abastecimento público de água.

O monitoramento de qualidade visa atender os seguintes objetivos: caracterizar as águas subterrâneas brutas; estabelecer Valores de Referência de Qualidade – VRQ para cada substância de interesse.

O relatório do período 1990-1997 apontou que 72% dos municípios do estado de São Paulo eram abastecidos total ou parcialmente por água subterrânea, atendendo a uma população estimada de 5.500.000 pessoas. O monitoramento constatou contaminação das águas por nitrato, potássio, cromo total e parâmetros microbiológicos em 15% dos poços (21), principalmente no Sistema Aquífero Bauru (CETESB, 1998).

Os aquíferos Bauru e Guarani foram os primeiros aquíferos monitorados. Até 1994, foi implementado o monitoramento de qualidade em 74 poços tubulares com determinação analítica de 17 parâmetros. Entre 2013 e 2015 a Rede ampliou o número de pontos, com 252 em 2013, 272 em 2014 e 282 em 2015. Em função da necessidade de substituição de poços desativados, foram monitorados no total 288 pontos.

Os parâmetros analisados para caracterização química e avaliação da qualidade da água subterrânea bruta no período de 2013 a 2015 foram:

- Parâmetros Físicos: temperatura da água e do ar, sólidos dissolvidos totais e sólidos totais.

- Parâmetros Químicos: pH, alcalinidade bicarbonato, alcalinidade carbonato, alcalinidade hidróxido, condutividade elétrica, dureza total, nitrogênio nitrato, nitrogênio nitrito, nitrogênio amoniacal total, nitrogênio kjeldha total, carbono orgânico dissolvido, cloreto, fluoreto, sulfato e as concentrações totais de alumínio, antimônio, arsênio, bário,

berílio, boro, cádmio, cálcio, chumbo, cobalto, cobre, crômio, crômio hexavalente, estanho, estrôncio, ferro, lítio, magnésio, manganês, mercúrio, molibdênio, níquel, potássio, prata, selênio, sódio, titânio, urânio, vanádio e zinco.

3. Mineração de Dados

Segundo KRIVDA (1996), Data Mining é o processo de descobrimento significativos de novas correlações, modelos, e tendências por seleção a partir de grande soma de dados armazenados em bases de dados e através do uso de modelos de tecnologias reconhecidas tanto quanto técnicas estatísticas e matemáticas.

O Data Mining é parte de um processo maior, conhecido como Knowledge Discovery in Databases (KDD). O termo KDD - Knowledge Discovery in Databases, foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento a partir de base de dados. Segundo Fayyad (1996, p. 30) “KDD é um processo, de várias etapas, não trivial, iterativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados”. Segundo Goldschmidt & Passos (2005, p. 4) o termo iterativo sugere a possibilidade de repetições integrais ou parciais do processo de KDD e a expressão não trivial alerta para a complexidade normalmente presente na execução de processos de KDD. Já com relação a expressão padrão válido indica que o conhecimento deve ser verdadeiro e adequado ao contexto da aplicação de KDD e o termo padrão novo deve acrescentar novos conhecimentos aos existentes, para que todo esse processo gere conhecimento útil que pode ser aplicado de forma a proporcionar benefícios ao contexto de aplicação de KDD.

O Data Mining possui diversas técnicas que podem ser utilizadas, dentre elas, pode-se citar: classificação, regressão, segmentação, associação, projeção e detecção de anomalias, segundo Loria (2008).

O Objetivo deste trabalho é mostrar como as duas técnicas de Data Mining podem ser utilizadas na gestão de recursos hídricos subterrâneos para ajudar e melhorar as análises feitas nos dias atuais. A primeira delas é a técnica de Classificação, onde os parâmetros são analisados de acordo com regras já conhecidas, e também a técnica de Associação, para tentar descobrir novas informações a respeito das análises de água subterrânea. Através desta técnica é possível descobrir novas relações que antes eram desconhecidas, por exemplo, quando o

nível de alguns parâmetros químicos for alto, normalmente outros parâmetros também terão seu nível alto, informação essa que passava despercebido.

3.1 Aprendizagem Supervisionada

Aprendizado Supervisionado significa que os algoritmos que serão utilizados na mineração de dados, necessitam de um “treinamento” prévio, ou seja, é necessário que seja informado uma classe de testes, um conjunto de informações para que os algoritmos se baseiem nessas informações e analisem o restante dos dados. Dentro desse método será utilizada a técnica de classificação.

É possível através de algoritmos supervisionados, definir o valor de um parâmetro baseando-se no valor de outros parâmetros previamente definidos no treinamento da regra. Por exemplo, num conjunto de dados de um supermercado deseja-se descobrir qual o perfil dos clientes que consomem determinados produtos. No caso deste artigo é possível saber quando a água não é potável, baseado no alto nível de Nitrato, quais outros parâmetros também se encontram com um valor acima ou abaixo do normal.

Alguns algoritmos geram como resultado árvores de classificação, enquanto que outros geram regras de classificação.

3.2 Aprendizagem Não-Supervisionada

Nos algoritmos não supervisionados, ou seja, onde não possui uma instrução, onde não se possui o conhecimento desejado como nas técnicas supervisionadas. O intuito desses métodos é descobrir informações que não são conhecidas, descobrir informações até então não pensadas, por exemplo, em um supermercado descobrir através da técnica de associação quais produtos tem uma saída “casada”, ou seja, geralmente quando os consumidores compram determinados produtos, também compram outros produtos que possuem uma determinada correlação que os vendedores não imaginavam.

As regras de Associação têm como premissa básica encontrar elementos que implicam na presença de outros elementos em uma mesma transação, ou seja, encontrar relacionamentos ou padrões frequentes entre conjuntos de dados. O termo transação indica quais itens foram consultados em uma determinada operação de consulta.

4. Estudo de Caso usando a técnica de classificação

Para utilizar a técnica de classificação na gestão de recursos hídricos subterrâneos é necessário que já se tenha o conhecimento do que buscamos analisar. No caso vamos utilizar o conhecimento de nível de Parâmetros que determinam a poluição das águas subterrâneas. De acordo com as informações contidas no site da CETESB (2015) "Um indicador de poluição difusa de água subterrânea é o nitrato. Sua origem está relacionada a atividades agrícolas e esgotos sanitários. Sendo o nitrato uma forma estável de nitrogênio em condições anaeróbias, esta substância pode ser considerada persistente e sua remoção da água para atender ao padrão de potabilidade que é de 10 mg/L, é onerosa e, por vezes, tecnicamente inviável, prejudicando o abastecimento público e privado.", ou seja, é possível pegarmos todas as amostras recolhidas nos últimos períodos e fazer uma classificação das mesmas. Sempre que o nível de Nitrato for acima do nível desejado a água pode ser considerada não potável.

Para isso é necessário fazer um aprendizado no Data Mining, chamado de aprendizado supervisionado, onde temos que mostrar ao sistema qual a regra na qual o mesmo vai usar os algoritmos de classificação para nos mostrar o resultado esperado, usamos a regra de que sempre que um nível de nitrato for acima do normal a água é considerada não potável, com isso após a análise de classificação poderemos obter um resultado de todos os parâmetros que tem uma certa conformidade quando a água for não potável.

Dessa maneira podemos saber em quais períodos e em quais sistemas aquíferos a água se encontra em estado não potável de uma forma mais assertiva e rápida. Pode ser conhecida em anos de estudos e de análises qual porcentagem de amostras são consideradas potáveis ou não potáveis. Para uma análise mais aprofundada pode ser levado em consideração mais de um parâmetro, além do nível de nitrato pode ser feito um aprendizado com mais de um parâmetro, pode ser criado outras faixas de classificação, além de potável pode ser feito uma classificação de nível de poluição elevado ou moderado, tudo depende do nível de classificação que se quer ter o conhecimento. Como mostra na Tabela 1 alguns exemplos dos valores de substâncias em cada medição, e a transformação destes para níveis de intensidade utilizando valores orientados de prevenção. E dessa maneira fazendo manualmente uma classificação para utilizar como treinamento na mineração de dados.

Tabela 1 - Exemplo de como pode ser feito a classificação através de teste supervisionado

ID Medicao	ALC.BICARBONATO	ALC.HIDRÓXIDO	Alumínio Total	Cádmio Total	N. Nitrato	
1	59,00	2,00	0,040	0,003	6,02	
2	120,00	1,00	0,020	0,003	3,10	
3	80,00	2,00	0,015	0,003	4,00	
4	94,00	2,00	0,120	0,003	8,20	
5	114,00	0,00	0,090	0,003	7,20	
ID Medicao	ALC.BICARBONATO	ALC.HIDRÓXIDO	Alumínio Total	Cádmio Total	N. Nitrato	Classificação
1	BAIXO	ALTO	NORMAL	NORMAL	ALTO	Não Potável
2	ALTO	NORMAL	BAIXO	NORMAL	NORMAL	Potável
3	NORMAL	ALTO	BAIXO	NORMAL	NORMAL	Potável
4	NORMAL	ALTO	ALTO	NORMAL	ALTO	Não Potável
5	ALTO	BAIXO	ALTO	NORMAL	ALTO	Não Potável

Fonte: Autoria própria

5. Estudo de Caso usando a técnica de Associação

Quando a classe de uma tarefa de mineração não é determinada, como no caso da classificação, onde se tem conhecimento prévio das combinações desejáveis, uma boa opção é o algoritmo de associação APRIORI do WEKA. WEKA é uma ferramenta gratuita e open source que possui vários algoritmos de Data Mining. Esse algoritmo APRIORI é capaz de descobrir novas regras de associação antes não conhecidas.

A CETESB disponibiliza em seu site diversos relatórios com as informações sobre as análises realizadas pelo órgão, onde possui os valores para cada parâmetro, para cada aquífero analisado. Para que os testes fossem de fato feitos, seria necessária uma autorização e disponibilização dos dados da CETESB. Entretanto através dos relatórios disponibilizados pelo órgão, é possível ter um entendimento de como os dados podem ser analisados e usados para esse estudo.

Nesses relatórios estão as várias medições feitas em um determinado período, separados por sistemas aquíferos e separados por parâmetros. Como mostram, por exemplo, a imagem 1 e imagem 2. São relatórios publicados pela CETESB e disponibilizado na internet. Nesses relatórios em questão estão mostradas as medições dos parâmetros no período de 2013 a 2015 especificamente do sistema aquífero de Bauru. O relatório mostra todos os parâmetros

medidos com seus respectivos valores, quantidade de amostras, valores mínimos, máximos e médios. E também os valores máximos permitidos.

Imagem 1 - Síntese dos resultados de qualidade das águas subterrâneas do Sistema Aquífero Bauru no período 2013 a 2015 - Parte 1

Parâmetro	Unidade	Valor Máximo Permitido	Sistema Aquífero Bauru 2013-2015 (89 pontos)					
			Amostras (n°)	< LQ ¹ (%)	Mínimo	Máximo	Mediana	3º Quartil
Alcalinidade Bicarbonato	mg CaCO ₃ L ⁻¹	--	477	1,1	4	261	84	104
Alcalinidade Carbonato	mg CaCO ₃ L ⁻¹	--	302	27,5	0	81	0	< 2
Alcalinidade Hidróxido	mg CaCO ₃ L ⁻¹	--	269	33,1	0	< 2	0	< 2
Carbono Orgânico Dissolvido	mg C L ⁻¹	--	488	73	< 1	19	< 1	1
Cloreto	mg Cl L ⁻¹	250 ^A	487	34,5	0,2	42	< 5	5,6
Condutividade Elétrica	µS cm ⁻¹	--	488	--	7,3	600	189	242
Dureza Total	mg CaCO ₃ L ⁻¹	500 ^A	488	1,4	1,5	307	67,5	101
Fluoreto	mg F L ⁻¹	1,5 ^B	448	66,3	< 0,02	0,7	< 0,3	< 0,5
Nitrogênio Amoniacal	mg N L ⁻¹	1,25 ^C	487	94,7	< 0,05	0,4	< 0,1	< 0,1
Nitrogênio Kjeldhal Total	mg N L ⁻¹	--	487	94,7	< 0,1	0,84	< 0,5	< 0,5
Nitrogênio Nitrato	mg N L ⁻¹	10 ^B	488	28,5	0,04	25,6	1,2	3,3
Nitrogênio Nitrito	mg N L ⁻¹	1 ^B	488	99	< 0,007	< 0,2	< 0,1	< 0,2
pH	--	6,0-9,5 ^D	488	--	5,2	9,9	7,1	7,7
Sólidos Dissolvidos Totais	mg L ⁻¹	1000 ^A	488	3,7	< 25	448	152	188
Sólidos Totais	mg L ⁻¹	--	488	2,9	25	492	166	204
Sulfato	mg SO ₄ L ⁻¹	250 ^A	271	84,9	< 0,5	13,7	< 1	< 1
Temperatura	°C	--	488	--	22,3	32,1	25,4	25,9
Bactérias Heterotróficas	UFC mL ⁻¹	500 ^B	488	28,9	0	> 5700	2	8
Coliformes Totais	NMP 100mL ⁻¹	ausência ^B	488	--	presente em 38 amostras	ausente	ausente	ausente
<i>Escherichia coli</i>	P/A 100mL ⁻¹	ausência ^B	488	--	presente em 6 amostras	ausente	ausente	ausente

Fonte: CETESB (2016)

Imagem 2 - Síntese dos resultados de qualidade das águas subterrâneas do Sistema Aquífero Bauru no período 2013 a 2015 - Parte 2

Parâmetro ¹	Unidade	Valor Máximo Permitido	Amostras (n°)	< LQ ² (%)	Mínimo	Máximo	Mediana	3° Quartil
Alumínio	µg Al L ⁻¹	200 ^{A, E}	488	59	< 2	126	< 2	4,2
Antimônio	µg Sb L ⁻¹	5 ^B	488	93	< 0,01	0,13	< 0,01	< 0,01
Arsênio	µg As L ⁻¹	10 ^B	488	56,1	< 0,2	10,6	< 0,5	0,5
Bário	µg Ba L ⁻¹	700 ^B	482	0,4	< 1	1118	185	290
Berílio	µg Be L ⁻¹	4 ^C	488	95,5	< 0,5	< 5	< 0,5	< 2
Boro	µg B L ⁻¹	2400 ^{D, F}	488	33,2	< 1	74,3	3,2	6,2
Cádmio	µg Cd L ⁻¹	5 ^B	488	75,2	< 0,005	0,1	< 0,005	< 0,005
Cálcio	mg Ca L ⁻¹	--	488	1	< 0,5	102	18,2	27,7
Chumbo	µg Pb L ⁻¹	10 ^B	488	39,5	< 0,02	8,2	0,06	0,17
Cobalto	µg Co L ⁻¹	70 ^D	488	13,5	< 0,01	2,7	0,03	0,12
Cobre	µg Cu L ⁻¹	2000 ^B	488	21,1	< 0,05	70,2	0,43	1,1
Crômio	µg Cr L ⁻¹	50 ^B	488	0	0,45	108	20,4	37,4
Estanho	µg Sn L ⁻¹	--	488	96,9	< 0,05	4,2	< 0,2	< 0,2
Estrôncio	µg Sr L ⁻¹	--	480	0	5,0	3030	220	420
Ferro	µg Fe L ⁻¹	300 ^{A, D}	488	45,5	< 2	1619	2,4	5,7
Lítio	µg Li L ⁻¹	--	488	0,2	< 0,1	15,9	3,4	6,2
Magnésio	mg Mg L ⁻¹	--	487	3,9	0,1	23,3	5,5	8,1
Manganês	µg Mn L ⁻¹	100 ^{A, H}	488	37,1	< 0,05	53,9	0,15	0,67
Mercúrio	µg Hg L ⁻¹	1 ^B	488	100	< 0,1	< 0,1	< 0,1	< 0,1
Molibdênio	µg Mo L ⁻¹	30 ^{D, I}	488	66,8	< 0,01	1,4	< 0,01	0,01
Níquel	µg Ni L ⁻¹	70 ^B	488	4,9	< 0,1	12,1	0,74	1,5
Potássio	mg K L ⁻¹	--	488	1,2	< 0,1	9,8	3,5	5,1
Prata	µg Ag L ⁻¹	50 ^{D, J}	488	99,8	< 0,03	< 0,1	< 0,1	< 0,1
Selênio	µg Se L ⁻¹	10 ^B	488	75,2	< 0,1	1,2	< 0,1	< 0,1
Sódio	mg Na L ⁻¹	200 ^A	488	0	0,15	74,9	7,6	15,9
Titânio	µg Ti L ⁻¹	--	488	0	0,62	18	2,5	3,3
Urânio	µg U L ⁻¹	30 ^B	488	6,4	< 0,002	1,9	0,08	0,19
Vanádio	µg V L ⁻¹	50 ^C	488	6,4	< 0,02	127	6,4	18,9
Zinco	µg Zn L ⁻¹	5000 ^{A, K}	488	7,2	< 0,2	840	3,4	8,4

Fonte: CETESB (2016)

Com o Data Mining podemos usar a técnica de associação e descobrir regras que não são exploradas até então. O conhecimento dessas novas associações podem ser gigantescos e mudar a forma com que as análises são feitas atualmente. Podem ser descobertas combinações de parâmetros que não eram relacionados anteriormente e com poucas análises de amostras coletadas já possibilitar em menos tempo tomar ações que ajudem na gestão hídrica subterrânea.

Para que a análise dos dados seja feito no WEKA, seriam necessárias algumas modificações, uma vez que o WEKA não saberia dizer se o nível de um parâmetro é alto, normal ou baixo por exemplo. Então seria necessária uma conversão dos dados numéricos

para o seu grau de intensidade, para isso podemos utilizar valores que são disponibilizados pela CETESB.

A CETESB estabelece alguns Valores fixos para classificar ou quantificar os parâmetros coletados, esses valores são derivados de critérios numéricos e valores já existentes em publicações científicas internacionais, esses valores visam justificar as medidas que devem ser tomadas na gestão dos recursos hídricos. São três valores citados pela CETESB, sendo eles o VRQ (Valor de Referência de Qualidade), VP (Valor de Prevenção) e VI (Valor de Intervenção). No nosso caso vamos focar no Valor de Prevenção, que é a concentração de um determinado parâmetro que se encontra acima do normal, e podem causar resultados prejudiciais à qualidade da água subterrânea. Este valor indica que deve ser tomada uma ação de prevenção para que a qualidade do solo e das águas subterrâneas seja mantida em conformidade com o meio ambiente. Conforme mostra as imagens 3, 4 e 5, cada substância possui os valores definidos, VRQ, VP e VI, este último ainda separado para áreas agrícolas, residenciais e industriais.

Imagem 3 - Valores orientadores para solo e água subterrânea no estado de São Paulo 2014 - Parte 1

Substância	CAS Nº	Solo (mg kg ⁻¹ peso seco)					Água Subterrânea (µg L ⁻¹)
		Valor de Referência Qualidade	Valor de Prevenção	Valor de Intervenção (VI)			
		(VRQ)	(VP)	Agrícola	Residencial	Industrial	
INORGÂNICOS							
Antimônio ⁽¹⁾	7440-36-0	<0,5	2	5	10	25	5
Arsênio ⁽¹⁾	7440-38-2	3,5	15	35	55	150	10
Bário	7440-39-3	75	120	500	1300	7300	700
Boro	7440-42-8	-	-	-	-	-	2400
Cádmio	7440-43-9	<0,5	1,3	3,6	14	160	5
Chumbo	7439-92-1	17	72	150	240	4400	10
Cobalto ⁽¹⁾	7440-48-4	13	25	35	65	90	70
Cobre ⁽²⁾	7440-50-8	35	60	760	2100	10000 ^(A)	2000
Crômio total ⁽¹⁾	7440-47-3	40	75	150	300	400	50
Crômio hexavalente	18540-29-9	-	-	0,4	3,2	10	-
Merúrio	7439-97-6	0,05	0,5	1,2	0,9	7	1
Molibdênio	7439-98-7	<4	5	11	29	180	30
Níquel ⁽²⁾	7440-02-0	13	30	190	480	3800	70
Nitrato (como N)	14797-55-8	-	-	-	-	-	10000
Prata ⁽¹⁾	7440-22-4	0,25	2	25	50	100	50
Selênio	7782-49-2	0,25	1,2	24	81	640	10
Zinco	7440-66-6	60	86	1900	7000	10000 ^(A)	1800
HIDROCARBONETOS AROMÁTICOS VOLÁTEIS							
Benzeno	71-43-2	-	0,002	0,02	0,08	0,2	5
Estireno	100-42-5	-	0,5	50	60	480	20
Etilbenzeno	100-41-4	-	0,03	0,2	0,6	1,4	300
Tolueno	108-88-3	-	0,9	5,6	14	80	700
Xilenos	1330-20-7	-	0,03	12	3,2	19	500
HIDROCARBONETOS POLICÍCLICOS AROMÁTICOS							
Antraceno	120-12-7	-	0,3	2300	4600	10000 ^(A)	900
Benzo(a)antraceno	56-55-3	-	0,2	1,6	7	22	0,4
benzo(b)fluoranteno	205-99-2	-	0,7	2	7,2	25	0,4
Benzo(k)fluoranteno	207-08-9	-	0,8	27	75	240	4,1
Benzo(g,h,i)perileno ⁽²⁾	191-24-2	-	0,5	-	-	-	-
Benzo(a)pireno	50-32-8	-	0,1	0,2	0,8	2,7	0,7
Criseno	218-01-9	-	1,6	95	600	1600	41
Dibenzo(a,h)antraceno	53-70-3	-	0,2	0,3	0,8	2,9	0,04
Fenantreno ^(2,4)	85-01-8	-	3,6	15	40	95	140
Indeno(1,2,3-c,d)pireno	193-39-5	-	0,4	3,4	8	30	0,4
Naftaleno	91-20-3	-	0,7	1,1	1,8	5,9	60

Fonte: CETESB (2016)

Imagem 4 - Valores orientadores para solo e água subterrânea no estado de São Paulo 2014 - Parte 2

Substância	CAS Nº	Solo (mg kg ⁻¹ peso seco)					Água Subterrânea (µg L ⁻¹)
		Valor de Referência Qualidade	Valor de Prevenção	Valor de Intervenção (VI)			
		(VRQ)	(VP)	Agrícola	Residencial	Industrial	VI
BENZENOS CLORADOS							
Clorobenzeno (Mono)	108-90-7	-	0,3	1,6	1,3	8,3	120
1,2-Diclorobenzeno	95-50-1	-	0,7	9,2	11	84	1000
1,3-Diclorobenzeno ⁽²⁾	541-73-1	-	0,4	-	-	-	-
1,4-Diclorobenzeno	106-46-7	-	0,1	0,3	0,6	2,1	300
1,2,3-Triclorobenzeno	87-61-6	-	0,01	0,4	1,1	6,1	20 ⁽²⁾
1,2,4-Triclorobenzeno	120-82-1	-	0,01	0,4	1	8,4	
1,3,5-Triclorobenzeno ⁽²⁾	108-70-3	-	0,5	-	-	-	
1,2,3,4-Tetraclorobenzeno ⁽²⁾	634-66-2	-	0,003	-	-	-	-
1,2,3,5-Tetraclorobenzeno ⁽²⁾	634-90-2	-	0,006	-	-	-	-
1,2,4,5-Tetraclorobenzeno	95-94-3	-	0,01	0,3	0,6	3,6	1,8
Hexaclorobenzeno	118-74-1	-	0,02	0,2	1,3	3,4	0,2
ETANOS CLORADOS							
1,1-Dicloroetano	75-34-3	-	0,02	0,1	0,6	1,7	53
1,2-Dicloroetano	107-06-2	-	0,001	0,01	0,03	0,09	10
1,1,1-Tricloroetano	71-55-6	-	0,2	140	120	690	2000
ETENOS CLORADOS							
Cloreto de vinila	75-01-4	-	0,0002	0,001	0,01	0,03	2
1,1-Dicloroeteno	75-35-4	-	0,04	2,8	3,8	22	30
1,2-Dicloroeteno - cis	156-59-2	-	0,01	0,08	0,2	1,1	50 ⁽²⁾
1,2-Dicloroeteno - trans	156-60-5	-	0,03	0,7	1	5,4	
Tricloroeteno - TCE	79-01-6	-	0,004	0,03	0,04	0,2	
Tetracloroeteno - PCE	127-18-4	-	0,03	0,6	0,8	4,6	40
METANOS CLORADOS							
Cloreto de Metileno (diclorometano)	75-09-2	-	0,02	0,1	0,4	2,1	20
Clorofórmio	67-66-3	-	0,06	0,1	0,8	4,5	300
Tetracloreto de carbono	56-23-5	-	0,004	0,03	0,1	0,4	4
FENÓIS CLORADOS							
2-Clorofenol (o)	95-57-8	-	0,06	0,6	1,7	9,4	30
2,4-Diclorofenol	120-83-2	-	0,03	0,5	1,5	8,5	18
3,4-Diclorofenol ^(2, A)	95-77-2	-	0,05	1	3	6	10,5
2,4,5-Triclorofenol	95-95-4	-	0,1	68	170	960	600
2,4,6-Triclorofenol	88-06-2	-	0,1	0,6	1,6	9,6	200
2,3,4,5-Tetraclorofenol ^(A, A)	4901-51-3	-	0,09	7	25	50	10,5
2,3,4,6-Tetraclorofenol	58-90-2	-	0,01	34	85	480	180
Pentaclorofenol (PCP)	87-86-5	-	0,01	0,07	0,6	1,9	9

Fonte: CETESB (2016)

Imagem 5 - Valores orientadores para solo e água subterrânea no estado de São Paulo 2014 - Parte 3

Substância	CAS Nº	Solo (mg kg ⁻¹ peso seco)					Água Subterrânea (µg L ⁻¹)
		Valor de Referência Qualidade	Valor de Prevenção	Valor de Intervenção (VI)			
		(VRQ)	(VP)	Agrícola	Residencial	Industrial	
FENÓIS NÃO CLORADOS							
Cresóis totais	1319-77-3	-	0,2	14	33	190	600
Cresol-p	106-44-5	-	0,005	-	-	-	-
Fenol	108-95-2	-	0,2	24	65	370	900
ÉSTERES FTÁLICOS							
Dietilxil ftalato(DEHP)	117-81-7	-	1	36	250	730	8
Dietil ftalato	84-66-2	-	0,5	33	100	550	4,8
Dimetil ftalato ⁽¹⁾	131-11-3	-	0,25	0,5	1,6	3	14
Di-n-butil ftalato	84-74-2	-	0,1	44	140	850	600
PESTICIDAS ORGANOCLORADOS							
Aldrin	309-00-2	-	0,02	0,4	0,8	6	0,03 ⁽²⁾
Dieldrin	60-57-1	-	0,01	0,3	0,8	5,9	0,6
Endrin	72-20-8	-	0,001	0,8	2,5	17	7 (B)
Carbofuran	1563-66-2	-	0,0001	0,3	0,7	3,8	20 ⁽³⁾
Endossulfan	115-29-7	-	0,7	4,7	12	66	1 ⁽²⁾
DDD	72-54-8	-	0,02	1	7,5	23	0,05
DDE	72-55-9	-	0,01	1,2	8,5	25	0,17
DDT	50-29-3	-	0,01	5,5	22	82	2
HCH alfa	319-84-6	-	0,0003	0,002	0,02	0,04	0,05
HCH beta	319-85-7	-	0,001	0,01	0,06	0,2	0,17
HCH – gama (Lindano)	58-89-9	-	0,001	0,008	0,06	0,2	0,09
OUTROS							
PCBs indicadores ⁽⁵⁾	NA	-	0,0003	0,01	0,03	0,12	3,5
TBT e seus compostos ⁽⁶⁾	NA	-	0,24	16	1,7	270	0,09
Anilina	62-53-3	-	0,023	0,15	0,7	3,2	42

Fonte: CETESB (2016)

Dadas essas informações, pode ser construída uma base de dados que ao invés de valores das medições, sejam colocados a intensidade dos mesmos como, por exemplo, baixo, médio e alto e dessa maneira ao ser feito o Data Mining serão descobertas correlações de quando alguns parâmetros estão acima do normal ou abaixo do normal outros parâmetros tendem a se comportar de determinadas maneiras também. Com essa alteração feita, já seria possível gerar um arquivo de extensão ARFF (extensão utilizada pelo sistema WEKA para ler dados) e colocar o algoritmo de associação para analisar os dados. Abaixo a Tabela 2 mostra um exemplo de como poderia ser feita a relação entre os valores das medições e seus respectivos níveis de intensidade.

Tabela 2 - Exemplo de transformação de dados passando de valores para níveis de intensidade

ID Medicao	ALC.BICARBONATO	ALC.HIDRÓXIDO	Alumínio Total	Cádmio Total	N. Nitrato
1	59,00	2,00	0,040	0,003	6,02
2	120,00	1,00	0,020	0,003	3,10
3	80,00	2,00	0,015	0,003	4,00
4	94,00	2,00	0,120	0,003	8,20
5	114,00	0,00	0,090	0,003	7,20
ID Medicao	ALC.BICARBONATO	ALC.HIDRÓXIDO	Alumínio Total	Cádmio Total	N. Nitrato
1	BAIXO	ALTO	NORMAL	NORMAL	ALTO
2	ALTO	NORMAL	BAIXO	NORMAL	NORMAL
3	NORMAL	ALTO	BAIXO	NORMAL	NORMAL
4	NORMAL	ALTO	ALTO	NORMAL	ALTO
5	ALTO	BAIXO	ALTO	NORMAL	ALTO

Fonte: Autoria própria

Os algoritmos de Associação poderão mostrar nessas análises, por exemplo, quando uma água não é considerada potável conforme descrito anteriormente onde foi abordada a técnica de classificação, e os níveis de nitrato estão acima do normal, o que mais pode estar correlacionado a isso? Quais outros entre tantos parâmetros analisados formam tem a tendência de estarem sempre apontando para um mesmo nível. E baseado nessa informação pode ser tomado uma ação presumindo que sempre que uma água não for potável, provavelmente outros parâmetros estarão se comportando também de uma determinada maneira, sem que sejam necessários outros estudos para descobrir essas informações.

Considerações Finais

A tecnologia nos permite evoluir em muitos aspectos e com o Data Mining podemos evoluir em diversas áreas, seja ela um negócio, saúde, financeiro e até mesmo no meio ambiente. Como um controle mais efetivo dos recursos hídricos subterrâneos, podemos melhorar ainda mais a qualidade da água em todo o país. Seja nos ajudando a classificar a informações que já possuímos de uma maneira mais rápida ou descobrir novas correlações até então desconhecidas. O objetivo desse estudo de caso é mostrar que é possível através do Data Mining analisar os dados já existentes de uma maneira diferente e antecipar ações de gestão hídrica subterrânea.

Referências Bibliográficas

CETESB - **Águas Subterrâneas** 2015. Disponível em <<http://aguassubterraneas.cetesb.sp.gov.br>> - Acesso em 06/02/2017.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **From data mining to knowledge discovery: An overview. In: Advances in Knowledge Discovery and Data Mining**, AAAI Press/The MIT Press, England, 1996, p.1-34.

GOLDSCHIMIDT, R e PASSOS, E. **Data mining: Um guia prático**. Rio de Janeiro: Campus, 2005.

KRIVDA, Cheryl D. **Unearthing Underground Data**. LAN – The Network Solutions Magazine. Cidade, v. 11, n. 5, p. 42-48, mai.1996.

LORIA, J. **Data Mining no SQL Server 2008**. 2008. Disponível em: <http://www.microsoft.com.br/kitbi/arquivos/pt-br%5CDataMining_BRZ.pdf>. Acesso em: 15 janeiro. 2010.

TERMO DE AUTORIA E RESPONSABILIDADE

D E C L A R O para os devidos fins que eu, Pedro Henrique Gerolin, portador da cédula de identidade R.G. nº. 32.750.062-1, e inscrito no CPF nº. 229.891.678-03, residente e domiciliado na Rua Antonio de Macedo Soares, 1471, bairro Campo Belo, cidade São Paulo, estado de São Paulo, sou o(a) autor(a) da MONOGRAFIA sobre o tema DATA MINING com enfoque em gestão hídrica subterrânea, referente à conclusão do Curso de Especialização Lato Sensu em Gerenciamento de Banco de dados, realizado na Universidade Nove de Julho – UNINOVE.

Declaro, ainda, que o trabalho é inédito, não contendo cópias bibliográficas e da rede mundial de computadores (Internet), tendo seguido as normas da ABNT e da Portaria UNINOVE, estando ciente que a infração destas normas levará a reprovação, bem como, incorrerá nas responsabilidades civis e criminais da legislação vigente – Lei nº 9.610/98 – Direitos Autoriais. Isento completamente a Universidade Nove de Julho (UNINOVE) e os professores indicados sendo convidado ou orientador para compor a ata de defesa presencial (do trabalho indicado e anexado) de toda e qualquer responsabilidade pelo conteúdo e idéias expressas na presente monografia.

São Paulo, Março, 2017