

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

A PRODUÇÃO CIENTÍFICA BRASILEIRA SOBRE APLICAÇÃO DE MINERAÇÃO DE DADOS PARA O AGRONEGÓCIO

José Vinícius Santos Barboza ¹

Sandra Mara Stocker Lago ²

Jerry Adriani Johann ³

RESUMO

O estudo identificou a produção científica brasileira sobre aplicação de Mineração de Dados para o Agronegócio. A pesquisa justifica-se por ser a Mineração de Dados uma ferramenta de grande valia para diversos setores. O Agronegócio tem expressiva participação na economia, gerando desenvolvimento e inovação. O estudo foi de caráter exploratório, através da pesquisa bibliográfica, onde foram analisados periódicos referentes a área Administração, Ciências Contábeis e Turismo, classificados pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) nos estratos A1, A2 e B1 publicados entre janeiro de 2005 até abril de 2015, período em que a pesquisa foi concluída. A busca identificou 30 artigos com o tema Mineração de Dados (*Data Mining*), dos quais foram analisados individualmente. Desses, 11 estudos envolvem o tema Mineração de Dados no Agronegócio. Conclui-se que o tema ainda é pouco explorado, porém com uma grande possibilidade de aplicação em diversas áreas do setor Agronegócio.

Palavras-chave: Mineração de Dados. *Data Mining*. Agronegócio. KDD. *Knowledge Discovery in Databases*.

.

¹ Universidade Estadual do Oeste do Paraná, e-mail: josevinicius@hotmail.com

² Universidade Estadual do Oeste do Paraná, e-mail: smstocker@uol.com.br

³ Universidade Estadual do Oeste do Paraná, e-mail: jerry.johann@hotmail.com

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

1 INTRODUÇÃO

As vantagens competitivas precisam ser modernizadas. O desafio de produzir mais e melhor vai sendo superado pelo constante desafio de criar novos produtos, serviços, processos e sistemas gerenciais (ARAUJO; ZILBER, 2013).

O avanço da tecnologia e da acessibilidade aos recursos computacionais, causou a redução de custos para o uso dessas ferramentas, as quais hoje são fundamentais nas organizações, que recebem diariamente uma quantidade expressiva de dados, seja de seus clientes, fornecedores ou concorrentes.

A gestão de fontes de informações auxilia a organização a combinar informações internas e externas e também a ampliar as práticas inovadoras. Organizações estão desenvolvendo alianças através do uso de fontes externas de informação para inovação, porém ainda carecem de uma estratégia explícita de gestão de informação (GOMES; KRUGLIANSKAS, 2009).

O intenso crescimento de dados tem demandado a necessidade de técnicas e ferramentas inovadoras que de forma inteligente e eficaz, consigam transformar uma grande quantidade de dados em informações significativas e em conhecimento útil. Essas informações geradas são relevantes e fundamentais para o planejamento, gestão e no auxílio a decisão, estando implícitas entre uma vasta quantidade de dados, e não podem ser descobertas ou facilmente identificadas utilizando-se sistemas convencionais de gerenciamento de banco de dados. Para atender essa demanda, surge a ferramenta denominada de *Data Mining* ou Mineração de Dados (SFERRA; CORRÊA, 2003).

A Mineração de Dados é uma metodologia que objetiva melhorar a qualidade e a eficiência das decisões, através do descobrimento de conhecimento útil, baseada nos dados históricos armazenados (LEMOS; STEINER; NIEVOLA, 2005).

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Esse cenário envolve também o setor de Agronegócios, em que a tecnologia e a qualidade da informação são fundamentais para que se atinja os resultados esperados. Assim, não basta apenas as organizações deterem a tecnologia e os dados se não souberem transformá-los em informações úteis, de forma rápida.

Para Toloí e Reiner (2011), é relevante destacar que o cenário do agronegócio está cada vez mais competitivo e complexo fazendo com que as organizações elaborem estratégias em busca de vantagem competitiva em relação aos concorrentes e também pelas condições adversas proporcionadas pela economia.

Pesquisas sobre como otimizar os resultados no agronegócio são de grande relevância para o setor, assim como estudos sobre a Mineração de Dados, ferramenta a qual está sendo útil em diversos setores organizacionais. Porém, verifica-se que o tema ainda é recente. Sendo assim, esse estudo busca responder o seguinte questionamento: Qual a produção científica brasileira sobre aplicação de mineração de dados para o agronegócio?

O estudo caracteriza-se como documental através da análise de artigos publicados sobre Mineração de Dados no Agronegócio, nos periódicos classificados como “Administração, Ciências Contábeis e Turismo”, classificação essa realizada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes).

O estudo apresenta no segundo capítulo o Referencial Teórico, seguido pela metodologia de pesquisa, resultados e por fim as considerações finais nos capítulos 3, 4 e 5 respectivamente.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

2 REFERENCIAL TEÓRICO

2.1 Mineração de Dados

A busca por padrões úteis em dados pode ser representada por diversos termos: extração de conhecimento, descoberta de informações, colheita de informações, arqueologia de dados e padrão de processamento de dados. Porém, “mineração de dados” é o termo mais usado, principalmente por estatísticos, pesquisadores de banco de dados, e mais recentemente no mundo dos negócios (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a).

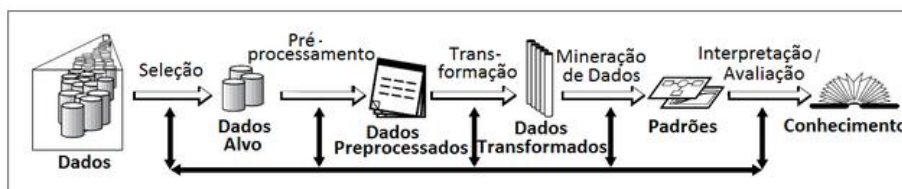
A Mineração de dados, ou *Data Mining*, é uma tecnologia que envolve métodos e padrões de análise de dados com algoritmos sofisticados para processar grande volume de dados, sendo que pode ser aplicada em diversas ocasiões. As técnicas são organizadas para agir sobre grande banco de dados com o intuito de descobrir padrões úteis que poderiam permanecer ignorados (TAN; STEINBACH; KUMAR, 2009).

A mineração de dados é uma etapa especial no processo denominado Descoberta de Conhecimentos em Bases de Dados, ou KDD (*Knowledge Discovery in Databases*), o qual refere-se a todo processo de descoberta de conhecimento útil de dados. A Mineração de Dados refere-se a uma etapa desse processo, onde envolve aplicação de algoritmos para extrair modelos dos dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996b).

O processo de KDD é um conjunto de atividades contínuas, composto por cinco etapas, que compartilham o conhecimento descoberto a partir de bases de dados. As cinco etapas do processo são: seleção dos dados; pré-processamento e limpeza dos dados; transformação dos dados; mineração de dados e por fim interpretação e avaliação dos resultados, representadas pela Figura 1.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Figura 1 - Processo de Descoberta de Conhecimentos em Base de Dados (KDD)



Fonte: Adaptado de Fayyad *et al.* (1996).

Para Dantas, Júnior, Lima e Azevedo (2008), o início se dá na análise da aplicação e dos objetivos a serem realizados. Após, a seleção de dados a ser minerada é realizada, podendo ser um conjunto de dados ou um subconjunto de variáveis onde a extração será feita. A próxima etapa denominada *Data Cleaning* e Pré-Processamento busca assegurar a qualidade dos dados envolvidos no KDD realizando operações básicas como a remoção de ruídos, que podem ser, por exemplo, atributos nulos. A fase seguinte consiste na Seleção e Transformação dos dados em que serão escolhidos os atributos realmente interessantes ao usuário, além de transformados, utilizando o padrão ideal para aplicar algoritmos de mineração. Após, a Mineração de Dados (*Data Mining*) é iniciada. Esta fase é a mais relevante do processo KDD, com a escolha do método e do algoritmo mais compatível com o objetivo da extração, para resultar padrões nos dados que sirva de subsídios para descobrir conhecimentos ocultos. A etapa Avaliação ou Pós-Processamento tem a função de identificar, entre os padrões extraídos na etapa de *Data Mining*, os padrões interessantes ao critério estabelecido pelo usuário, podendo voltar à fase inicial para novas iterações. Ao término da avaliação, o conhecimento descoberto deverá ser implantado e incorporado ao sistema, sempre documentando e publicando os métodos, a fim de apresentar o conhecimento descoberto ao usuário.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Para Lemos, Steiner e Nievola (2005), a técnica de Mineração de Dados apresenta soluções para diversas áreas como vendas, finanças, seguros, transporte, medicina, telecomunicações e mercado financeiro, demonstradas no Quadro 1.

Quadro 1 - Utilização Da Mineração De Dados

ÁREA	UTILIZAÇÃO DE MINERAÇÃO DE DADOS
Vendas	Reter clientes, através da identificação dos passíveis de migração para o concorrente objetivando retê-los; detectar associações entre produtos; identificar padrões de comportamento de consumidores; encontrar características dos consumidores de acordo com a região demográfica; prever quais consumidores serão atingidos nas campanhas de marketing e, nesses casos, enviar mala-direta anunciando o produto apenas para aqueles prováveis compradores, tarefa denominada de mala-direta direcionada.
Finanças	Detectar padrões de fraudes no uso dos cartões de crédito; identificar os consumidores que estão tendendo a mudar de companhia de cartão de crédito; identificar regras a partir dos dados do mercado; encontrar correlações escondidas nas bases de dados .
Seguros	Determinar os procedimentos médicos requisitados ao mesmo tempo; prever quais consumidores têm tendência a comprar novas apólices; identificar comportamentos fraudulentos.
Transporte	Determinar a distribuição dos horários entre os vários caminhos; analisar padrões de sobrecarga.
Medicina	Caracterizar o comportamento dos pacientes para prever novas consultas; identificar terapias de sucesso para diferentes doenças; prever quais pacientes têm maior probabilidade de contrair uma certa doença, em função de dados históricos de pacientes e doenças.
Telecomunicações	Identificar fraudes em ligações telefônicas dentre um número de ligações efetuadas pelos clientes.
Mercado Financeiro	Prever as ações que irão subir ou descer na bolsa de valores, em função de dados históricos com preços de ações e valores de índices financeiros.

Fonte: Adaptado de Lemos, Steiner e Nievola (2005).

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Segundo Castro (2012), na década de 90, diversos setores passaram a concorrer com *players* externos, o que causou a busca pelo desempenho diferencial, focando na qualidade, eficiência e competitividade, enfatizando a relevância da gestão nos setores. Este choque de mercado e competitividade veio acompanhado da valorização da inovação, como fator de crescimento para as organizações ligadas ao agronegócio. Inovações tecnológicas e gerenciais tornaram-se fundamentais para a competitividade das organizações envolvidas na produção agrícola.

O setor agronegócio apresenta um papel estratégico para o crescimento da economia nacional, considerando o cenário dos próximos anos, em que a economia mundial e nacional tende a ser retraída, os aumentos de produtividade e participação relativa do setor agrícola do país colocarão em relevância o agronegócio. Sendo assim deve ser conduzido como uma política de Estado visando a garantir o crescimento econômico para o país (BARROS, 2013 apud JÚNIOR et al, 2014).

2.2 Mineração de Dados no Agronegócios

A agricultura se baseava na geração de capital e de mão-de obra, mas conforme o setor foi evoluindo, as propriedades rurais adotaram novas tecnologias de produção e de gestão administrativa, que permitiram uma nova postura diante das exigências do mercado, principalmente nos países mais desenvolvidos. No Brasil, essa situação é característica em regiões mais desenvolvidas e com propriedades de grande porte (UECKER; UECKER; BRAUN, 2005).

A necessidade de uma maior eficiência nos processos produtivos, devido a competitividade global, demonstra a relevância de um sistema que forneça informações essenciais para o entendimento e o aperfeiçoamento das atividades operacionais das

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

empresas inseridas no setor Agronegócios (CALLADO; CALLADO; ALMEIDA, 2008).

A utilização dos procedimentos relacionados a agricultura de precisão é um desafio atual no País. Diversas dúvidas semelhantes apontadas por outros países mais desenvolvidos vão surgindo, principalmente as que estão relacionadas a como interpretar a grande quantidade de dados e transformá-la em ferramentas que auxiliem na tomada de decisões. Essa variabilidade em termos de diferenças de produtividade ao longo da lavoura, tem causas das mais variadas e de maneira alguma pode ser atribuída a um fator isoladamente (SOUZA et al., 2010).

No Agronegócio são produzidos produtos agropecuários e agroindustriais, mas o processo de produção e comercialização fornece uma diversidade de dados sobre insumos, como rações, adubos, capital, terra e trabalho, bem como sobre os produtos da atividade agropecuária, como preços, quantidades produzidas e vendidas etc. Essas informações devem receber atenção (MOLLER, 2011).

O planejamento e o gerenciamento de áreas agrícolas abrangem atividades de manejo das culturas tendo como pré requisito um conhecimento simultâneo de atributos do solo, do relevo e do rendimento das áreas cultivadas. O armazenamento, o tratamento, a sobreposição, a análise e a visualização em conjunto dessas informações espaciais sob a forma de mapas digitais, torna-se possível com técnicas da geoestatística, da estatística clássica e pela indução por árvore de decisão permitindo tomadas de decisões eficientes, favorecendo a minimização dos custos, a otimização dos recursos e das atividades produtivas, ganho de produtividade e maximização dos lucros (SOUZA et al., 2010).

Pela grande quantidade de fontes de dados e de diferentes tipos de informação, realizar análises sobre o agronegócio é sempre um desafio. Mas, se os conteúdos

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

estiverem tratados para gerar informação consolidada, é possível agilizar a obtenção de variáveis para apoiarem as pesquisas e análises para o agronegócio. O uso de *data mining* para o agronegócio deve receber uma atenção maior, tanto por pesquisas como por aplicações estatísticas, o que possibilitará melhorar a qualidade dos dados disponíveis (CORREA, 2010).

3 METODOLOGIA

A atual pesquisa é de caráter exploratório, onde ocorre a familiarização com o tema, sem pretender a descrição em profundidade. Para Gil (1999) abordagem exploratória tem como finalidade desenvolver, esclarecer e modificar conceitos e ideias, observando a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores. Objetiva proporcionar visão geral de determinado fato.

Quanto aos procedimentos de coleta de dados, utilizou-se a pesquisa bibliográfica e documental, com o levantamento de informações previamente elaboradas e publicadas. Os dados foram coletados no período de março e abril de 2015.

A pesquisa bibliográfica vem sendo aplicada em estudos exploratórios ou descritivos, casos em que o objeto de estudo proposto é pouco estudado. É desenvolvida a partir de material já elaborado, formado principalmente de livros e artigos científicos. Apesar da maioria dos estudos exigirem algum tipo de trabalho desta natureza, há pesquisas desenvolvidas exclusivamente a partir de fontes bibliográficas. A sua indicação para esses estudos deve-se pela aproximação com o objeto ser dada a partir de fontes bibliográficas. Portanto, a pesquisa bibliográfica possibilita um amplo alcance de informações, além de permitir a utilização de dados dispersos em inúmeras publicações, auxiliando também na construção, ou na melhor definição do quadro conceitual que envolve o objeto de estudo proposto (GIL, 1999).

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

A classificação da amostra foi não-probabilística, sendo escolhidos os periódicos através dos seguintes critérios:

- a) Referentes a área Administração, Ciências Contábeis e Turismo, classificados pela CAPES,
- b) Publicações entre janeiro de 2005 a abril de 2015 (Período em que a pesquisa foi concluída);
- c) Periódicos classificados entre os estratos A1, A2 e B1;
- d) Os critérios de busca foram os artigos que continham como palavra-chave, resumo, título ou assunto o termo Mineração de Dados. As Palavras *Data Mining* e *DataMining* também foram objetos de pesquisa, pelo fato de muitos autores utilizarem o termo em Inglês.

Após a identificação dos periódicos e dos artigos, os mesmos foram analisados, a fim de classificar artigos que envolvem o tema Mineração de Dados no Agronegócio.

Os artigos resultantes foram analisados de forma qualitativa, com a interpretação dos temas e conteúdo, e quantitativa com os seguintes critérios:

- a) Volume de artigos publicados no período;
- b) Volume de artigos por periódicos;
- c) Volume de artigos por estrato de avaliação;
- d) Evolução anual das publicações;
- e) Evolução anual das publicações por estrato de avaliação;
- f) Caracterização de autoria:
 - Instituições de origem dos autores;
 - Parcerias com instituições internacionais;
 - Número de autores por artigo;
 - Interação entre autores de diferentes instituições;

No capítulo seguinte são expostos os resultados obtidos.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

4 RESULTADOS

Através dos critérios adotados, inicialmente foram identificados 191 periódicos nacionais referentes a área Administração, Ciências Contábeis e Turismo, sendo excluídos 56 por apresentarem temáticas divergentes da pesquisa. Após essa seleção, 135 periódicos foram consultados, em busca de artigos com o tema da pesquisa. Desses periódicos pesquisados, apenas 15 resultaram algum artigo com os termos definidos (Mineração de Dados, *Datamining e Data Mining*), sendo no total 30 artigos, distribuídos em 15 periódicos, conforme demonstrado na Tabela 1.

Tabela 1 - Resultado de pesquisa a periódicos

RESULTADOS DE PERIÓDICOS	Número de artigos
Base (UNISINOS)	1
Ciência e Agrotecnologia (UFLA)	1
Ciência Rural	6
Encontros Bibli	1
Engenharia Agrícola	3
Gestão & Produção	2
Informação & Sociedade (UFPB)	2
RAC. Revista de Administração Contemporânea	1
RAI: Revista de Administração e Inovação	1
RAM. Revista de Administração Mackenzie	3
Revista Ambiente & Água	2
Revista de Administração da UFSM	1
Revista de Ciências da Administração (CAD/UFSC)	1

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

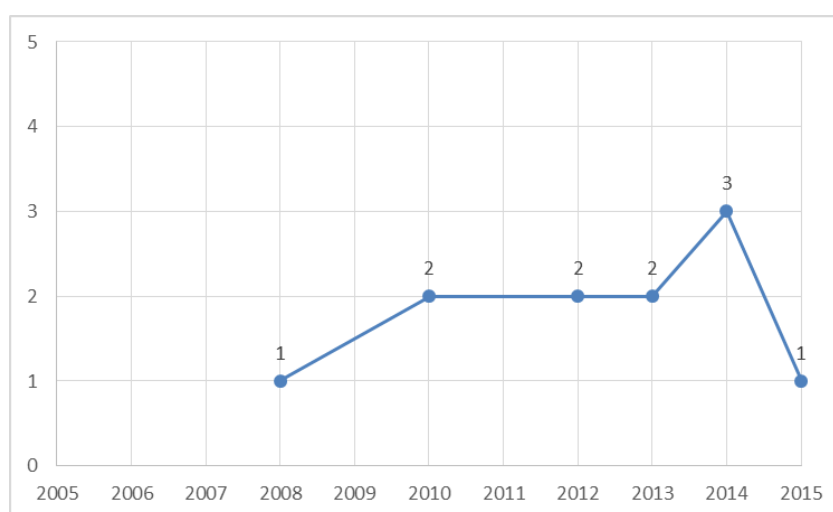
Revista de Gestão da Tecnologia e Sistemas de Informação	4
Transinformação	1
Total Geral	30

Fonte: Dados da pesquisa

Após a pesquisa, os artigos identificados foram analisados e classificados através de suas características como período da publicação, periódico publicado, número e afiliação dos autores, as palavras chaves utilizadas, o tema central do estudo, a metodologia aplicada e os respectivos resultados gerados. A análise resultou em 11 artigos que envolvem o assunto abordado na pesquisa, ou seja, dos 30 artigos resultantes sobre Mineração de dados, 11 referem-se ao tema Mineração de Dados no Agronegócio.

Objetivando identificar a produção por ano, são apresentados os dados de publicações, na Figura 2, demonstrando uma elevação no ano de 2014.

Figura 2 - Evolução anual de publicações



Fonte: Dados da pesquisa.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Destaca-se entre os periódicos o estrato Qualis A2 e B1, com 10 e 1 artigos respectivamente. Entre os periódicos, onde foi verificado a maior frequência de artigos do tema foi a revista Ciência Rural, demonstrados no Tabela 2.

Tabela 2- Resultado de pesquisas a periódicos – Tema Mineração de Dados

Periódicos	Qualis	2008	2010	2012	2013	2014	2015	Total
Ciência e Agrotecnologia (UFLA)	A2		1					1
Ciência Rural	A2	1	1	1	1	1	1	6
Engenharia Agrícola	A2			1	1	1		3
Revista de Gestão da Tecnologia e Sistemas de Informação	B1					1		1

Fonte: Dados da pesquisa

Dos 11 artigos analisados, 8 foram com interações entre distintas universidades e/ou instituições, sendo uma internacional. O número de autores por artigo variou entre 2 a 6. Foram identificados 33 autores, com destaque para 14 afiliados a UNICAMP e 6 a UFRGS, conforme Tabela 3.

Tabela 3 - Autores por instituição

Instituição	Número de Autores
Universidade Estadual de Campinas - UNICAMP	14
Universidade Federal do Rio Grande do Sul - UFRGS	6
Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA	2
Universidade de São Paulo - USP	2
Universidade Estadual do Oeste do Paraná – UNIOESTE	1

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Agroinfo Tecnologia da Informação Ltda	1
Cooperativa Regional de Cafeicultores em Guaxupé Ltda -Cooxupé	1
Instituto Federal de Goiás - IF	1
Universidade Federal de Goiás - UFG	1
Universidade Federal de Mato Grosso - UFMT	1
Universidade Federal de Santa Catarina - UFSC	1
Universidade Federal de Santa Maria - UFSM	1
University of Wisconsin-Madison	1
Total Geral	33

Fonte: Dados da Pesquisa

As palavras-chaves utilizadas assim como a frequência nos artigos são demonstradas no Tabela 4.

Tabela 4 - Frequência de Palavras-Chaves

Palavra-Chave	Frequência	Palavra-Chave	Frequência
Mineração de dados	7	Mapa de solos	1
<i>Data mining</i>	2	Maxver	1
<i>Spectral behavior</i>	1	Modelagem preditiva	1
Agricultura de precisão	1	Modis	1
<i>Annual crop</i>	1	Monitor de produtividade	1
Aprendizado de máquina	1	Monitoramento ambiental	1
<i>Artificial intelligence</i>	1	<i>Orbital remote sensing</i>	1
<i>Artificial neural network</i>	1	Pedologia	1

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Classificação supervisionada de imagem	1	Phakopsora pachyrhizi	1
<i>Credit unionism</i>	1	Poultry science	1
<i>Crop monitoring</i>	1	Previsão de doenças de plantas	1
<i>Decision tree</i>	1	Rede de sensores sem fio	1
Enxurradas	1	Sensoriamento remoto	1
Epidemiologia	1	Sig	1
Florestamento	1	Sistemas de suporte à decisão	1
Kdd	1	Solos	1
K-means	1	<i>Spot vegetation</i>	1
Lógica Fuzzy	1	Uso da terra	1
<i>Management</i>	1	Variabilidade espacial	1
Total de palavras chaves			38

Fonte: Dados da pesquisa

As pesquisas foram analisadas de forma qualitativa, sendo observado aspectos como os objetivos, metodologia, resultados e sugestões para pesquisas futuras. Também foi possível identificar as principais ferramentas de Mineração de Dados utilizadas.

Nääs, Queiroz, Moura e Brunassi (2008), realizaram uma pesquisa no setor pecuário, onde duas análises foram realizadas: a primeira foi a aplicação da Lógica Fuzzy, elaborada em uma base de regras, com registros retirados do banco de dados de informações históricas do rebanho com medidas de movimentação, produção de leite, dias em lactação e presença do comportamento de monta, para analisar os atributos mais significativos na determinação do estro de vacas leiteiras. A segunda analisou os mesmos parametros com os dados do subconjunto de 15 vacas do conjunto total das 100

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

vacas do experimento, para análise na previsão de cio. Os dados obtidos foram submetidos a uma mineração de dados, sendo fundamentada na construção de uma árvore de decisão. Identificou-se que essas técnicas quantitativas preditivas permitiram associar todos os dados disponíveis na produção, de maneira a tornar o processo de determinação do estro mais preciso, fundamentando melhor a tomada de decisão do momento ideal para inseminação artificial e reduzindo as perdas, conseqüentemente elevando a eficiência na produção leiteira.

Lima e Rodrigues (2010) realizaram um trabalho, com o objetivo de explorar a técnica Árvore de Decisão em banco de dados de uma organização do setor avícola, visando a elaboração de padrões de incubação. Foram disponibilizados pela empresa, dados de incubação entre os anos de 2002 e 2006. Dois experimentos foram realizados, onde foi possível observar que com a técnica estudada, os dados utilizados no gerenciamento de produção são suficientes para identificar conhecimentos novos, úteis e aplicáveis a fim de melhorar a produtividade das empresas incubadoras, atendendo à demanda com diminuição do desperdício.

Souza et al. (2010), avaliaram os atributos químicos do solo e a produtividade da cultura de cana-de-açúcar por meio da geoestatística e mineração de dados pela indução da árvore de decisão. A indução de árvores de decisão permitiu verificar que a altitude é a variável com maior potencial para interpretar os mapas de produtividade de cana de açúcar, auxiliando na agricultura de precisão e mostrando-se uma ferramenta adequada para o estudo.

Lamparelli, et al. (2012), utilizaram a mineração de dados a fim de identificar condições diferentes de plantas de café após o período de colheita. As imagens foram pré-processadas usando a técnica do MNF (*Minimum Noise Fraction Transform*), que tornou possível a eliminação de ruídos. Depois, realizou-se a correção atmosférica da imagem com a aplicação do algoritmo Flaash. Para executar a tarefa, foi construído um

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

banco de dados com informações analisadas, como cultivo, idade, área, inclinação e espaçamento. O banco de dados foi estruturado em um sistema de informação geográfica (SIG). Com o objetivo de diferenciar as condições do café após a colheita, um perfil espectral para cada etapa foi gerado. Teste *t Student* com nível de significância de 5% foi aplicado para verificar a igualdade dos meios de comprimentos de onda em todos os grupos realizados. Os resultados mostraram cinco clusters segundo a condição pós-colheita, possibilitando melhorar ações futuras de intervenção.

Teske e Giasson (2012), utilizaram a técnica de mineração de dados para gerar mapas preditores de áreas aptas ao plantio de eucalipto. Foi realizada uma amostragem aleatória constando 3.500 pontos, sendo que em cada ponto de amostragem, foram coletadas informações sobre as variáveis preditoras e se era ou não uma área plantada com eucalipto. Esses dados foram analisados utilizando o software Weka, onde a equação obtida possibilitou a estimativa da probabilidade de ser uma área apta ao plantio, a fim de gerar um mapa binário, mostrando as áreas aptas e não aptas ao plantio de eucalipto. A regressão logística demonstrou viabilidade de seu uso para a estimativa de áreas aptas para o plantio de eucalipto, possibilitando seu uso na avaliação de terras para análises iniciais de aquisição de terras, quantificação e qualificação de áreas para florestamento com eucalipto.

Giasson, Hartemink, Tornquist, Teske e Bagatini (2013), realizaram um estudo testando cinco algoritmos de árvores de decisão. Para a geração das árvores de decisão, as informações amostradas foram tabeladas e analisadas estatisticamente com a ferramenta computacional Weka. Assim foi possível identificar os modelos de árvore de decisão capazes de produzir mapas de solos com maior semelhança com o mapa de referência. A avaliação dos resultados e metodologia do estudo indicou que, em estudos futuros, deve-se utilizar mapas de solos tradicionais de referência, com qualidade conhecida. Também foi possível identificar que os mapas gerados por Mapeamento

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Digital de Solo (MDS) devem ser avaliados através da verdade de campo, o que poderá dar indicações mais úteis sobre a qualidade dos resultados.

Johann, et al., (2013), trabalharam com dados obtidos através dos perfis temporais do produto NDVI decendial. O NDVI (Índice de Vegetação por Diferença Normalizada) é um indicador numérico que demonstra graficamente o status da planta. Utilizou-se a abordagem de geração dos clusters pelos algoritmos K-Means, MAXVER e DBSCAN no software Weka. Foram gerados clusters com base no perfil temporal médio do índice NDVI dos 277 municípios produtores de soja do Estado do Paraná, na safra ano de 2005/2006, resultando como melhor método o algoritmo K-Means, agrupando os municípios em seis clusters demonstrando os bons resultados encontrados com a metodologia proposta, em termos de identificação de áreas homogêneas. O estudo é útil para estimar a área plantada.

Megeto, Oliveira, Ponte e Meira (2014), analisaram a influência de variáveis meteorológicas analisando uma grande quantidade de dados de ocorrência de ferrugem referentes a quatro safras (2007/2008 a 2010/2011), através da árvore de decisão. Os dados foram preparados no software R. As árvores de decisão foram induzidas no software Weka 3.6. O conjunto de treinamento para a modelagem foi composto de 45 variáveis de chuva e temperatura e 12.591 registros. O modelo preditivo escolhido resultou em uma árvore de decisão com aproximadamente, 78% de taxa de acerto e 108 regras, determinadas por validação cruzada. O modelo interpretado, com 28 regras, considerou variáveis de temperatura como mais importantes, sendo que temperaturas abaixo de 15 °C e acima de 30 °C foram relacionadas com eventos de não ocorrência, enquanto temperaturas dentro da faixa favorável foram associadas com eventos de ocorrência.

Para Ruiz, Caten e Dalmolin (2014), a cobertura da terra é uma informação espacial de extrema relevância para uma série de modelos, sendo utilizada para estimar

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

a produção de sedimentos e para mensurar por exemplo a potencialidade da paisagem em sequestrar carbono. Os autores realizaram um estudo em um modelo por Árvore de Decisão, através da ferramenta Weka, a fim de discriminar as classes de cobertura da terra e avaliar o tamanho da Arvore de Decisão gerada quanto ao seu número de folhas. Preocupou-se também em identificar as classes da cobertura da terra de mais difícil classificação. As classes de cobertura da terra estabelecidas foram: campo, lavoura de soja, lavoura de arroz, floresta de Mata Atlântica, solo exposto e água. A aplicação da Arvore de Decisão propiciou identificar os usos da terra de difícil classificação, que, neste trabalho, foram as florestas e as lavouras de arroz.

Sousa e Figueiredo (2014), através de suas pesquisas desenvolveram modelos para analisar a capacidade dos membros de uma cooperativa de crédito para liquidar seus compromissos, com a mineração de dados. A base utilizada é formada por dados históricos dos clientes, do período de 2003 a 2007. As simulações realizadas utilizando a ferramenta experimentador do Weka indicam que o processo de descoberta de conhecimento e o uso de modelos baseados em mineração de dados desenvolvido pode fornecer uma melhor classificação e previsão o comportamento dos membros da cooperativa.

Dota, Cugnasca e Barbosa (2015) propõem uma abordagem usando técnicas com o objetivo de classificar a qualidade da água a ser usada na agricultura e na sua classificação ambiental. Para a análise aplicou-se a Árvore de Decisão (AD) com o uso da ferramenta Weka. Os dados coletados durante o experimento foram submetidos a diferentes algoritmos de classificação para verificar se a classificação proposta é coerente e se os dados de cada classe têm forte relação estatística entre si. Além disso, foram utilizados os algoritmos de classificação para construir um modelo para identificar a qualidade da água. Os resultados mostram que a classificação proposta para a qualidade da água neste cenário é coerente, sendo que diferentes algoritmos indicaram

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

uma forte relação estatística entre as classes e suas instâncias. Como estudo futuro, os autores pretendem usar os dados dos sensores para monitoramento ambiental em tempo real.

Foi possível identificar as ferramentas utilizadas pelos autores, para a aplicação das técnicas de mineração de dados, com destaque para o software Weka (Waikato Environment for Knowledge Analysis), desenvolvido pela *University of Waikato* da Nova Zelândia, conforme tabela 5.

Tabela 5 - Softwares utilizados para aplicação de Mineração de Dados

Ferramentas utilizadas	
Software Weka	8
SAS Enterprise Miner	2
Software ArcGis	1
Software R	1

Fonte: Dados da pesquisa

O software Weka apresenta ferramentas de processamento de dados, contendo praticamente todos os algoritmos, possibilitando experimentar métodos em novos conjuntos de dados de forma flexível. Ele fornece uma ampla gama de ferramentas de pré-processamento. Essa diversificada e abrangente ferramenta é acessada através de uma interface comum, de modo que seus usuários possam comparar diferentes métodos e identificar aqueles que são mais adequadas para o problema (WITTEN; FRANK; HALL, 2005).

O Weka é reconhecido como um marco no sistema mineração de dados. Alcançou ampla aceitação nos meios acadêmicos e de negócios, e tornou-se uma

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

ferramenta amplamente utilizada para pesquisa de mineração de dados (HALL et al, 2009).

5 CONSIDERAÇÕES FINAIS

Com o estudo foi possível identificar uma diversificada utilização da técnica de mineração de dados para o Agronegócio. O estudo possibilitou relacionar o que vem sendo pesquisado sobre o tema abordado e quais são as instituições que mais publicam sobre o tema.

As pesquisas foram desenvolvidas por autores de diversas instituições, porém destaca-se os autores da UNICAMP e da UFRGS. Destaca-se também o uso por diversos autores do software Weka, como ferramenta para aplicação do processo KDD e Mineração de Dados.

Os estudos demonstram que a técnica é bem útil ao agronegócio, sendo muitas vezes possível a sua replicação entre as organizações. Por ser um tema de nova abordagem, sugere-se outros estudos como forma de identificar formas inovadoras da utilização do processo KDD e da Mineração de dados, não só no Agronegócios como em outras áreas.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

THE BRAZILIAN SCIENTIFIC PRODUCTION ABOUT DATA MINING APPLICATION FOR AGRIBUSINESS

ABSTRACT

The study identified the Brazilian scientific literature on Data Mining application for Agribusiness. The research is justified because it is the data mining a valuable tool for various industries. Agribusiness has significant participation in the economy, generating development and innovation. The study was exploratory, through bibliographic research, which analyzed journals related area Administration, Accounting and Tourism, classified by the CAPES (Coordination for the Improvement of Higher Education Personnel) in the strata A1, A2 and B1 published between January 2005 to April 2015, during which the survey was completed. The search identified 30 articles on the theme Data Mining, which were analyzed individually. Of these, 11 studies involving the topic Data Mining in Agribusiness. It concludes that the matter is still little explored but with a high possibility of application in various areas of the agribusiness sector.

Key words: Data Mining. Agribusiness. KDD. Knowledge Discovery in Databases.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

REFERÊNCIAS

ARAUJO, J. B. De; ZILBER, S. N. Adoção de E-Business e mudanças no modelo de negócio: Inovação organizacional em pequenas empresas dos setores de comércio e serviços. **Revista Gestão e Produção**, v. 20, n. 1, p. 147–161, 2013.

CALLADO, A. L. C.; CALLADO, A. A. C.; ALMEIDA, M. A. A utilização de indicadores de desempenho não-financeiros em organizações agroindustriais: Um estudo exploratório. **Organizações Rurais & Agroindustriais**, p. 35–48, 2008.

CASTRO, A. De. Prospecção de cadeias produtivas e gestão da informação. **Transinformação**, p. 1–27, 2012.

CORREA, F. E. **Representação de comercialização agropecuária através de modelo de data warehouse**. 2010. Escola Politécnica da USP, 2010.

DANTAS, E. R. G.; JÚNIOR, J. C. A. P.; LIMA, D. S. De; AZEVEDO, R. R. De. O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões. **V Simpósio de Excelencia em Gestão e Tecnologia**, p. 1–10, 2008.

DOTA, M. a; CUGNASCA, C. E.; BARBOSA, D. S. Comparative analysis of decision tree algorithms on quality of water contaminated with soil. **Ciencia Rural**, v. 45, n. 2, p. 267–273, 2015.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, n. 11, p. 27–34, 1996a.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, p. 37–54, 1996b.

GIASSON, E.; HARTEMINK, A. E.; TORNQUIST, C. G.; TESKE, R.; BAGATINI, T. Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na Bacia do Lageado.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Ciencia Rural, v. 43, n. 11, p. 1967–1973, 2013.

GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 1999.

GOMES, C. M.; KRUGLIANSKAS, I. Indicadores e características da gestão de fontes externas de informação tecnológica e do desempenho inovador de empresas brasileiras. **Rac**, v. 13, n. 2, p. 172–188, 2009.

HALL, Mark et al. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, v. 11, n. 1, p. 10-18, 2009.

JOHANN, J. A.; ROCHA, J. V.; OLIVEIRA, S. R. de M.; RODRIGUES, L. H. A.; LAMPARELLI, R. A. C. Data mining techniques for identification of spectrally homogeneous areas using NDVI temporal profiles of soybean crop. **Engenharia Agrícola**, v. 33, n. 3, p. 511–524, jun. 2013.

JÚNIOR, E. G.; COSTA, A. D.; GONÇALVES, F. de O.; DUENHAS, R. A. Sistema de Inovação do Agronegócio Brasileiro? Dualismo estrutural-tecnológico e desafios para o desenvolvimento do país. **Desenvolvimento em questão**, p. 279–317, 2014.

LAMPARELLI, R. A. C.; JOHANN, J. A.; SANTOS, E. R. Dos; ESQUERDO, J. C. D. M.; ROCHA, J. V. Use of data mining and spectral profiles to differentiate condition after harvest of coffee plants. **Engenharia Agrícola**, v. 32, n. 1, p. 184–196, 2012.

LEMOS, E. P.; STEINER, M. T. A.; NIEVOLA, J. C. Análise de crédito bancário por meio de redes neurais e árvores de decisão : uma aplicação simples de data mining. **Rausp**, p. 225–234, 2005.

LIMA, M. G. F.; RODRIGUES, L. H. A. Arvore de decisao aplicada em dados de incubacao de matrizes de postura Hy-Line W36. **Ciencia e Agrotecnologia**, v. 34, n. 6, p. 1550–1556, 2010.

MEGETO, G. a. S.; OLIVEIRA, S. R. M.; DEL PONTE, E. M.; MEIRA, C. A. A.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

Modelagem Da Temperatura E Da Precipitação Em Ocorrências Da Ferrugem Asiática Da Soja Por Meio Da Técnica De Árvore De Decisão. **Engenharia Agrícola**, v. 34, n. 3, p. 590–599, 2014.

MOLLER, H. D. Agronegócio. In: CALLADO, A. A. C. (Ed.). **Agronegócio . Métodos quantitativos aplicados ao Agronegócio**. p. 183–203.

NÄÄS, A.; DE, I.; QUEIROZ, G.; PAULO, M.; DE, D. J.; BRUNASSI, A.; MARIA, S. Dairy cows estrus estimation using predictive and quantitative methods. **Ciência Rural**, p. 2383–2387, 2008.

RUIZ, L. F. C.; CATEN, A. Ten; DALMOLIN, R. S. D. Árvore De Decisão E a Densidade Mínima De Amostras No Mapeamento Da Cobertura Da Terra. **Ciência Rural**, v. 44, n. 6, p. 1001–1007, 2014.

SFERRA, H. H.; CORRÊA, Â. M. C. J. Conceitos e Aplicações de Data Mining. **Revista de Ciência & Tecnologia**, p. 19–34, 2003.

SOUSA, M. D. M.; FIGUEIREDO, R. S. Credit Analysis Using Data Mining: Application in the Case of a Credit Union. **Journal of Information Systems and Technology Management**, v. 11, n. 2, p. 379–396, 2014.

SOUZA, Z. M. De; CERRI, D. G. P.; COLET, M. J.; RODRIGUES, L. H. A.; MAGALHÃES, P. S. G.; MANDONI, R. J. A. Análise dos atributos do solo e da produtividade da cultura de cana-de-açúcar com o uso da geoestatística e árvore de decisão. **Ciência Rural**, v. 40, n. 4, p. 840–847, 2010.

TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introdução ao data mining mineração de dados**. Rio de Janeiro: Moderna, 2009.

TESKE, R.; GIASSON, E. Delimitação de áreas para plantio de eucalipto utilizando regressões logísticas. **Ciência Rural**, v. 42, n. 7, p. 1219–1222, 2012.

José Vinícius Santos Barboza, Sandra Mara Stocker Lago, Jerry Adriani Johann

TOLOI, R. C.; REINERT, J. N. Contribuição do programa de pós-graduação em agronegócio da Universidade Federal de Mato Grosso do Sul no desenvolvimento do agronegócio do Estado de Mato Grosso do Sul, Brasil. **Acta Scientiarum. Human and Social Sciences**, v. 33, n. 1, p. 55–65, 2011.

UECKER, G. L.; UECKER, A. D.; BRAUN, M. B. S. A gestão dos pequenos empreendimentos rurais num ambiente competitivo global e de grandes estratégias. **Congresso da SOBER**, 2005.

WITTEN, I. H.; FRANK, E.; HALL, M. a. **Data mining: practical machine learning tools and techniques**. San Francisco: Elsevier, 2005.