



UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
ENGENHARIA DA COMPUTAÇÃO

Selmo Eduardo Rodrigues Júnior

**MODELO BASEADO EM DADOS METEOROLÓGICOS PARA ESTIMAR O VOLUME DE
ÁGUA APLICADO EM UMA PLANTAÇÃO DE DENDÊ**

São Luís - MA

2014

Selmo Eduardo Rodrigues Júnior

**MODELO BASEADO EM DADOS METEOROLÓGICOS PARA ESTIMAR O VOLUME DE
ÁGUA APLICADO EM UMA PLANTAÇÃO DE DENDÊ**

Trabalho de Conclusão de Curso apresentado
à Universidade Estadual do Maranhão para a
obtenção do título de Bacharel em Engenharia
da Computação.

Orientador: Msc. José Pinheiro de Moura
Coorientador: Dr. Hamilton Jesus Santos
Almeida

São Luís - MA

2014

Rodrigues Júnior, Selmo Eduardo

Modelo baseado em dados meteorológicos para estimar o volume de água aplicado em uma plantação de dendê / Selmo Eduardo Rodrigues Júnior. – São Luís - MA, 2014.

89 p. : il. (algumas color.) ; 30 cm.

Monografia (Graduação) - Curso de Engenharia da Computação, Universidade Estadual do Maranhão, 2014.

Orientador: Msc. José Pinheiro de Moura

1. Irrigação. 2. Dendê. 3. Volume de água. 4. KDD. 5. Regressão linear múltipla. I.
Título

CDU: 519.6:628.1.034.3

Selmo Eduardo Rodrigues Júnior

**MODELO BASEADO EM DADOS METEOROLÓGICOS PARA ESTIMAR O VOLUME DE
ÁGUA APLICADO EM UMA PLANTAÇÃO DE DENDÊ**

Trabalho de Conclusão de Curso apresentado
à Universidade Estadual do Maranhão para a
obtenção do título de Bacharel em Engenharia
da Computação.

Trabalho aprovado em São Luís - MA, 6 de agosto de 2014.

Msc. José Pinheiro de Moura - UEMA
Orientador

Msc. Mauro Sérgio Silva Pinto
Primeiro Membro da Banca

Msc. Sidney Nascimento Cerqueira Júnior
Segundo Membro da Banca

AGRADECIMENTOS

Agradeço a Deus por ter me iluminado e por ter me oferecido todas as condições, saúde e força para superar as dificuldades.

Ao professor Msc. José Pinheiro de Moura pela orientação, confiança e apoio na elaboração deste trabalho. Ao meu coorientador Dr. Hamilton Jesus Santos Almeida pelo suporte, pelos incentivos e pela cooperação na pesquisa.

A Universidade Estadual do Maranhão, em especial, ao Curso de Engenharia da Computação e a todas as pessoas com que convivi ao longo desses 5 anos, onde a comunhão com os amigos e o aprendizado em grupo foram as melhores experiências na minha formação acadêmica.

Ao professor Mário Chaves Lima, por ter me concedido alguns conselhos essenciais na elaboração da pesquisa. Agradeço também a Valdir, que me ajudou na coleta de informações da plantação de dendê.

Às pessoas do Núcleo Geoambiental da Universidade Estadual do Maranhão, por terem auxiliado essa pesquisa cedendo dados que foram essenciais em sua produção.

Aos meus pais e a todos os meus familiares que me ensinaram os princípios e valores que levarei comigo para sempre. Agradecimento especial ao meu pai, que me fortaleceu e que foi muito importante para mim nessa jornada.

A minha namorada Emilly Nascimento Mota, pelo companheirismo, pelo carinho e por sua ajuda nessa pesquisa.

Agradeço a todos que, diretamente ou indiretamente, fizeram parte da minha formação acadêmica. O meu muito obrigado a vocês.

*"Quem tem medo de subir as montanhas,
não enxerga o horizonte."
(Alex Husfor)*

RESUMO

Um dos maiores desafios da agricultura é estimar o volume de água adequado a ser aplicado em uma plantação. Irrigar de forma eficiente, além de economizar recursos hídricos, eleva a produtividade das plantas e as deixa mais saudáveis. No entanto, determinar essa quantidade de água depende de diversos fatores climáticos, de particularidades do solo, do tipo de sistema de irrigação adotado e da cultura a ser irrigada. Nessa pesquisa, é apresentado um modelo de conhecimento para estimação do volume de água para irrigar uma plantação de *Elaeis guineensis*, também conhecida como dendê. Essa plantação está localizada na Fazenda Escola de São Luís, Universidade Estadual do Maranhão. Esse modelo foi desenvolvido seguindo o fluxo de etapas do processo de KDD (*Knowledge Discovery in Databases*), utilizando a técnica de regressão linear múltipla em uma base de dados fornecida pelo Núcleo Geoambiental da Universidade Estadual do Maranhão. O resultado desse procedimento foi uma equação linear que calcula o volume de água para irrigar essa plantação e que busca auxiliar o irrigante na tomada de decisão. Essa equação apresentou bons resultados analíticos, indicando uma forte correlação positiva entre as variáveis e baixos índices de erros.

Palavras-chaves: irrigação, dendê, volume de água, KDD, regressão linear múltipla.

ABSTRACT

One of the biggest challenges of the agriculture is to estimate the water volume need to be applied on a plantation. Irrigate efficiently, besides saving water, increases plants productivity and makes them healthier. However, determining the quantity of water depends on various weather conditions, characteristics of the soil, the type of irrigation system used and the plant to be irrigated. In this research, is presented a knowledge model for estimating the volume of water to irrigate a plantation *Elaeis guineensis*, also known as oil palm. This plantation is located at the Farm School in São Luís, State University of Maranhão. This model was developed following the flow of process steps of KDD (Knowledge Discovery in Databases), using the technique of multiple linear regression on a database provided by the Geoenvironmental Center from the State University of Maranhão. The result of this procedure was a linear equation that calculates the volume of water to irrigate that plantation, helping the irrigator in the decision making. This equation showed good analytical results, indicating a strong positive correlation between the variables and low error rates.

Key-words: irrigation, oil palm, water volume, KDD, multiple linear regression.

LISTA DE ILUSTRAÇÕES

Figura 1 – Consumo de Água Doce.	25
Figura 2 – Distribuição Global da Água.	26
Figura 3 – Distribuição dos Recursos Hídricos, da Superfície e da População - em % do Total no Brasil.	27
Figura 4 – Ciclo Hidrológico.	28
Figura 5 – Métodos de Irrigação e seus Principais Sistemas.	32
Figura 6 – Exemplo de Microaspersor.	33
Figura 7 – Tanque Classe A.	35
Figura 8 – Evolução do Coeficiente de Cultura K_c ao Longo do Ciclo Fenológico de uma Planta.	38
Figura 9 – Plantação de Dendê na Fazenda Escola de São Luís, Universidade Estadual do Maranhão (UEMA).	41
Figura 10 – Etapas do Processo de KDD.	45
Figura 11 – Gráfico das Notas de Pós-graduação em Relação as Notas de Graduação.	48
Figura 12 – Agrupamentos de Registros Similares.	49
Figura 13 – Tipos de Regressão.	52
Figura 14 – Exemplos de Coeficientes de Correlação.	55
Figura 15 – Relação entre Duas Variáveis.	56
Figura 16 – Tela <i>Editor</i> do <i>MATLAB</i>	59
Figura 17 – Tela Inicial do <i>WEKA</i>	61
Figura 18 – Metodologia de Manejo de Irrigação da Pesquisa.	65
Figura 19 – Guia <i>Preprocess</i> do <i>WEKA Explorer</i>	68
Figura 20 – Guia <i>Classify</i> do <i>WEKA Explorer</i>	69
Figura 21 – Resultados da Validação Cruzada.	72
Figura 22 – Comparação entre o Volume Fornecido e o Volume Estimado.	73
Figura 23 – Esquema da Plantação de Dendê da Fazenda Escola de São Luís, UVT.	83
Figura 24 – Histogramas dos Dados Meteorológicos no <i>WEKA</i>	89

LISTA DE TABELAS

Tabela 1 – Valores Estimados para a Eficiência de Irrigação (E_a).	39
Tabela 2 – Classificação por Renda de Indivíduos.	47
Tabela 3 – Resumo de Valores Obtidos do Universo de Estudo.	64
Tabela 4 – Amostragem dos Dados Preparados Para Mineração.	67
Tabela 5 – Parâmetros Estatísticos do Modelo.	72
Tabela 6 – Resultados dos Modelos para uma Amostra dos Dados.	74

LISTA DE ABREVIATURAS E SIGLAS

FAO	<i>Food and Agriculture Organization of the United Nations</i> - Organização das Nações Unidas para Agricultura e Alimentação.
GNU	<i>General Public License</i> - Licença Pública Geral.
IDC	<i>International Data Corporation</i> - Corporação Internacional sobre Dados.
KDD	<i>Knowledge Discovery in Database</i> - Descoberta de Conhecimento em Bases de Dados.
MATLAB	<i>Matrix Laboratory</i> - Laboratório de Matrizes
ONU	Organização das Nações Unidas.
UEMA	Universidade Estadual do Maranhão.
UVT	Unidade de Validação Tecnológica.
WEKA	<i>Waikato Environment for Knowledge Analysis</i> - Ambiente para Análise de Conhecimento de Waikato.

SUMÁRIO

1	INTRODUÇÃO	21
2	FUNDAMENTOS DA AGRICULTURA IRRIGADA	25
2.1	O Cenário da Água e sua Relação com a Agricultura	25
2.2	Aspectos Gerais sobre Irrigação	29
2.3	Irrigação Localizada	31
2.3.1	Sistema de Irrigação por Microaspersão	32
2.4	Procedimentos para o Manejo de Irrigação	34
2.4.1	Evapotranspiração de Referência	34
2.4.1.1	<i>Método do Tanque de Classe A</i>	35
2.4.1.2	<i>Método de Blaney-Criddle</i>	36
2.4.1.3	<i>Método de Penman-Monteith</i>	36
2.4.1.4	<i>Método da Radiação Solar</i>	37
2.4.2	Evapotranspiração da Cultura	38
2.4.3	Lâmina Bruta de Irrigação	38
2.4.4	Volume de Água Aplicado na Irrigação	39
2.4.5	Tempo de Irrigação	40
2.5	Dendezeiro	40
3	PROCESSO DE AQUISIÇÃO DE CONHECIMENTO	43
3.1	Aspectos Básicos sobre a Descoberta de Conhecimento	43
3.2	Etapas do Processo de Descoberta de Conhecimento	44
3.3	Etapa de Preparação dos Dados	46
3.4	Etapa de Mineração de Dados	46
3.4.1	Tarefas de Mineração de Dados	47
3.4.1.1	<i>Classificação</i>	47
3.4.1.2	<i>Estimação</i>	48
3.4.1.3	<i>Clustering ou Agrupamento</i>	49
3.4.1.4	<i>Associação</i>	50
3.4.2	Técnicas de <i>Data Mining</i>	50
3.4.3	Regressão e Correlação	51
3.4.3.1	<i>Métodos dos Mínimos Quadrados</i>	53
3.4.3.2	<i>Coeficiente de Correlação</i>	54
3.5	Avaliação do Modelo	54
3.6	Apresentação do Conhecimento	57

4	SOFTWARES USADOS NA PESQUISA	59
4.1	Funcionalidades Básicas do <i>MATLAB</i>	59
4.2	O <i>Software WEKA</i>	60
5	METODOLOGIA DE DESENVOLVIMENTO DO MODELO PARA ESTIMAÇÃO DO VOLUME	63
5.1	Características do Universo de Estudo	63
5.2	Coleta dos Dados Brutos	64
5.3	Metodologia de Irrigação da Pesquisa	65
5.4	Preparação dos Dados Brutos	66
5.5	Mineração dos Dados	67
6	RESULTADOS DA PESQUISA	71
6.1	Apresentação dos Resultados	71
6.2	Discussão dos Resultados	73
7	CONSIDERAÇÕES FINAIS	75
	Referências	77
	APÊNDICE A – PLANTAÇÃO DE DENDÊ	83
	APÊNDICE B – CÓDIGO <i>MATLAB</i>	85
	APÊNDICE C – HISTOGRAMAS DOS DADOS	89

1 INTRODUÇÃO

Uma das questões mais críticas na agricultura é determinar a quantidade de água para irrigar uma plantação. Segundo Testezlaf (2011), atender as necessidades hídricas das culturas melhora o rendimento delas, proporcionando uma maior produtividade e melhor qualidade nessa produção. Porém, se a irrigação for exagerada, o excesso de água pode ocasionar danos às plantas e prejudicar o solo. Caso contrário, quando o volume de água não é suficiente para satisfazer as exigências das culturas, a produtividade e o metabolismo delas são afetados.

Devido a esses fatores, busca-se aplicar água de maneira eficiente nas plantações. Todavia, realizar essa tarefa é um procedimento complexo, pois envolve diversas variáveis meteorológicas e características da localidade da plantação, como por exemplo, o tipo de solo, o sistema de irrigação considerado e a cultura específica (USA, 1993). Assim, há muitas técnicas para calcular o volume de água para irrigação, cada uma requisitando um conjunto de variáveis diferente.

Entretanto, nem sempre os dados que essas técnicas exigem estão disponíveis ou são facilmente obtidos. À vista disso, o tema desse estudo envolve alternativas para estabelecer o volume de água que irá irrigar uma plantação de dendê (*Elaeis guineensis*) na Fazenda Escola de São Luís, localizada na Universidade Estadual do Maranhão. Sendo assim, busca-se responder o seguinte questionamento: qual a quantidade de água que irá atender as exigências hídricas diárias dessa plantação de dendê?

Tendo como referência esse questionamento, essa pesquisa se baseia na hipótese de que é possível estimar esse volume de água com base em dados meteorológicos da região, em características da plantação citada e nas particularidades do sistema de irrigação local.

Desse modo, a finalidade principal desse trabalho é desenvolver um modelo que calcule o volume de água diário para irrigar a plantação de dendê na Fazenda Escola de São Luís. Esse modelo deve ser útil ao irrigante como apoio na tomada de decisão para essa tarefa. Busca-se desenvolvê-lo com base em dados meteorológicos dessa localidade, coletados de um período de, aproximadamente, 2 anos. Para extrair informações sobre o volume de água para irrigação a partir desses dados, usou-se o processo de descoberta de conhecimento em base de dados ou, simplesmente, KDD (*Knowledge Discovery in Database*)

O processo de KDD é aplicado em uma base de dados com o objetivo de se obter padrões ou conhecimentos convenientes a partir dela. Esses padrões e conhecimentos gerados buscam apoiar os especialistas em suas tomadas de decisão. O KDD é constituído por uma sequência de etapas, que são o pré-processamento e a preparação dos dados, a mineração, a avaliação dos padrões encontrados e a exibição do conhecimento. Na etapa de mineração de dados, existem algumas tarefas e numerosas técnicas que podem ser utilizadas pelos pesquisadores. As tarefas estão relacionadas aos objetivos da aplicação com KDD, enquanto as técnicas indicam a forma

ou como que se irá atingir a finalidade de determinada tarefa. O investigador seleciona a tarefa e a técnica que serão adotadas conforme o formato dos dados e o tipo de conhecimento que ele deseja alcançar (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Dessa maneira, esta pesquisa seguiu esse fluxo de etapas do processo de descoberta de conhecimento para determinar o volume de água para irrigação da plantação de dendê citada. Os dados meteorológicos brutos foram cedidos pelo Núcleo Geoambiental da Universidade Estadual do Maranhão e posteriormente analisados. Após a realização de um estudo sobre os fundamentos da irrigação, esses dados foram preparados e minerados usando a tarefa de estimação aplicada com a técnica de regressão linear múltipla. Realizou-se também um aprendizado prévio sobre o universo de estudo considerado, isto é, sobre as particularidades da plantação de dendê, sobre o tipo sistema de irrigação usado nela e sobre características do local. Por fim, foi realizada uma avaliação do modelo de conhecimento desenvolvido, discutindo seus resultados.

Metodologicamente, esse trabalho adotou o tipo de pesquisa aplicada quanto a sua finalidade, visto que ele tenta resolver um problema concreto e apresenta um objetivo prático. Quanto aos meios de investigação, essa pesquisa se caracteriza por ser um estudo de caso, já que é limitada a um universo circunscrito, que é a plantação de dendê (VERGARA, 2003).

A importância deste trabalho se reflete na sociedade e no meio ambiente, pois uma alternativa foi investigada para economizar recursos hídricos e para elevar a produtividade das culturas, através da aplicação de água de forma eficiente em uma plantação. É necessário pesquisar meios para que as plantas produzam mais alimentos, para sustentar um número crescente de pessoas, e bens de consumo, principalmente para países que dependem economicamente de produtos agrícolas. O dendê, por exemplo, manifesta um grande potencial bioenergético, além de apresentar taxas consideráveis de sequestro de carbono. Para essa cultura ter boa produtividade, suas necessidades hídricas devem ser atendidas.

Além desses fatores, essa pesquisa apresentou uma contribuição na área científica e acadêmica, comentando sobre as características do processo de KDD, destacando a sua interdisciplinaridade e realizando um enfoque para resolver um problema prático, seguindo as suas etapas para obter um conhecimento útil a partir da extração de padrões de dados brutos.

A estrutura do trabalho está organizada como se segue. O Capítulo 2 descreve os principais conceitos das ciências agrárias que serão necessários para essa pesquisa, apresentando sobre a relação da água com a agricultura, a importância da irrigação e algumas de suas técnicas e metodologias para o cálculo do volume de água aplicado. Este capítulo também exhibe algumas das principais características da cultura de dendê.

O Capítulo 3 fornece uma fundamentação teórica sobre o processo de descoberta de conhecimento, explicitando suas particularidades fundamentais e as suas etapas, assim como algumas tarefas e técnicas da etapa de mineração de dados. Já o Capítulo 4 comenta sobre as funcionalidades básicas dos *softwares* usados no trabalho. O Capítulo 5 mostra como foi

desenvolvido o modelo de conhecimento para estimação do volume de água para irrigar a plantação de dendê, isto é, os procedimentos metodológicos da pesquisa são esclarecidos. O Capítulo 6 mostra os resultados e as principais discussões do trabalho. O Capítulo 7 refere-se às conclusões e às considerações finais da pesquisa. Finalmente, o texto apresenta as Referências e os Apêndices, contendo o código elaborado e algumas figuras produzidas.

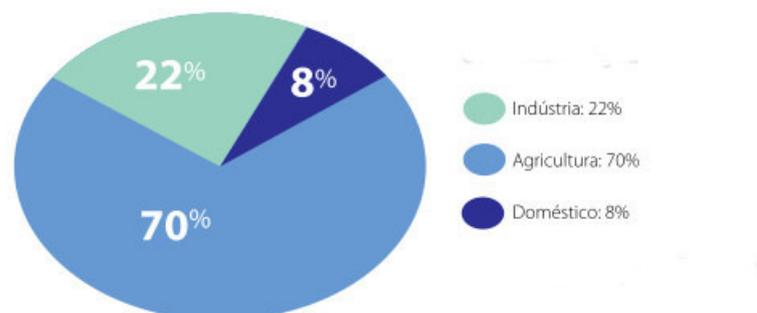
2 FUNDAMENTOS DA AGRICULTURA IRRIGADA

Este capítulo trata da importância da água e a sua distribuição no globo, além de comentar como o ser humano vem utilizando esse valioso recurso natural. Também é citada a relação da água com a agricultura e como a aplicação dela ou não nas plantações impacta no metabolismo e na produtividade das culturas. Em seguida, caracteriza-se a irrigação, destacando a importância dela ser administrada de forma eficiente e seus conceitos fundamentais. Por conseguinte, algumas variáveis para o manejo e planejamento da irrigação são definidas, assim como as equações para obtenção do volume de água para irrigação. Para finalizar, as principais características do dendezeiro, a planta de referência desse trabalho, serão comentadas.

2.1 O Cenário da Água e sua Relação com a Agricultura

A água é um recurso natural indispensável à vida, além de ser fundamental para manter o equilíbrio dos ecossistemas, da biodiversidade e do clima em nosso planeta. No organismo dos seres humanos, a água atua como meio para a troca de substâncias e para a manutenção da temperatura, compondo cerca de 70% da massa corporal humana (BRASIL, 2005).

Figura 1 – Consumo de Água Doce.



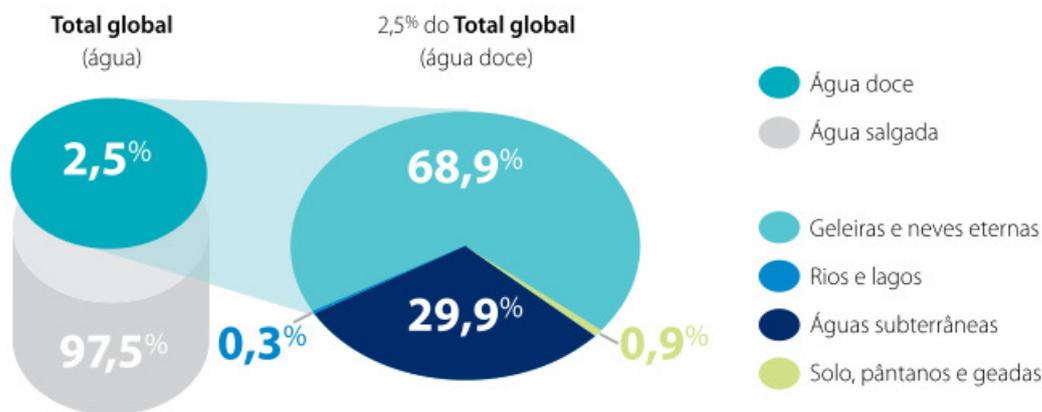
Fonte: (BRASIL, 2005).

Ademais, a água é um fator de produção essencial de vários bens de consumo oriundos da agropecuária, como por exemplo, a produção de alimentos. Utiliza-se também este recurso natural para higiene pessoal, para lavar roupas ou utensílios e para a manutenção da limpeza de nossas habitações. Ela pode ser aproveitada para a produção de energia elétrica, na limpeza das cidades, na construção de obras, no combate a incêndios, entre outras muitas aplicações. As indústrias utilizam grandes quantidades de água, quer seja como matéria-prima, quer seja na remoção de impurezas, na geração de vapor e na refrigeração. Contudo, é a agricultura aquela que mais consome esse líquido, onde cerca de 70% de toda a água derivada de rios, lagos e

aquíferos subterrâneos consumida no planeta é utilizada pela irrigação (Figura 1) (BRASIL, 2005).

A água ocupa cerca de 70% da superfície do planeta Terra, sendo que 97,5% é salgada. A porcentagem restante é referente à água doce que é apropriada para consumo. Dessa parcela de água doce, 68,9% encontra-se nas geleiras, nas calotas polares ou em regiões montanhosas, 29,9% em águas subterrâneas, 0,9% compõe a umidade do solo e dos pântanos e apenas 0,3% constitui a porção superficial de água doce presente em rios e lagos (BRASIL, 2005). A Figura 2 exibe o panorama da distribuição da água no mundo.

Figura 2 – Distribuição Global da Água.

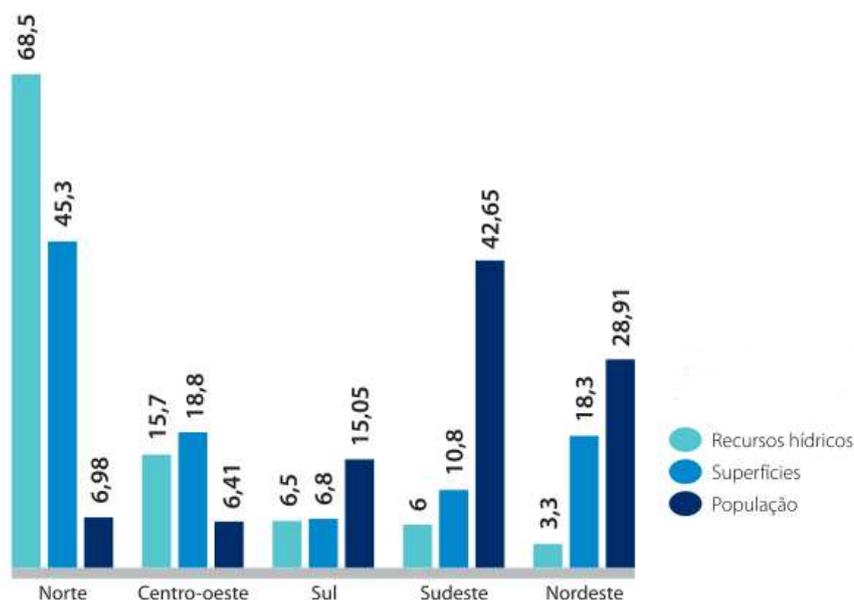


Fonte: (BRASIL, 2005).

O Brasil é um país privilegiado por possuir a maior reserva de água doce do planeta. Isso corresponde, em números, a 12% do total mundial. Todavia, esse recurso natural não é distribuído uniformemente em sua extensão territorial. A Região Amazônica é detentora da maior bacia fluvial do mundo e o volume de água do rio Amazonas é considerado o maior do globo. Ao mesmo tempo, é uma das regiões menos habitadas do Brasil. Em contrapartida, as regiões com maior concentração populacional e que exigem mais recursos hídricos estão distantes dos grandes rios. O maior problema de escassez hídrica é no Nordeste brasileiro, onde algumas de suas regiões ficam sem água por longos períodos (GOMES, 2011). Esse cenário pode ser visualizado através da Figura 3.

Um fator preocupante no cenário brasileiro é que um grande número de rios e lagos vem sendo comprometido pela queda de qualidade da água disponível para captação e tratamento, devido às ações impactantes dos seres humanos. Essa queda de qualidade é causada principalmente por despejos de esgotos domésticos e industriais, uso de agrotóxicos na agricultura e uso dos rios como transportadores de lixo (GOMES, 2011).

Figura 3 – Distribuição dos Recursos Hídricos, da Superfície e da População - em % do Total no Brasil.



Fonte: (BRASIL, 2005).

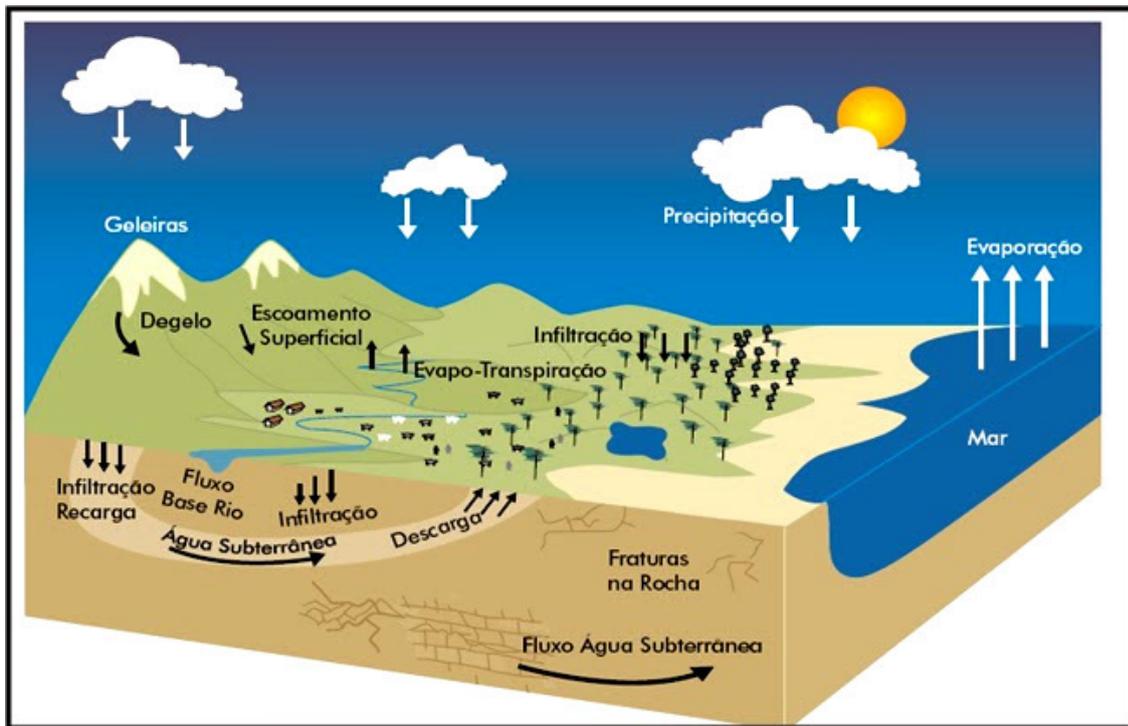
Conforme dados da FAO¹ que podem ser observados na Figura 1, a agricultura, além de ser a atividade humana que mais consome recursos hídricos, é também a que mais desperdiça água. Em números, ela desperdiça cerca de 2.500 km^3 de água por ano, valor muito superior ao que a indústria perde (117 km^3) e ao que o uso doméstico também perde ($64,5 \text{ km}^3$). Isso ocorre principalmente devido a métodos ultrapassados e ineficientes de irrigação, que quando realizada de forma inadequada, pode afetar drasticamente a qualidade dos solos e ocasionar grandes desperdícios de água. Outro motivo que contribui para a baixa eficiência da irrigação é a falta de informação de uma parcela dos irrigantes, que não assimila os princípios fundamentais da agricultura irrigada, dificultando o próprio entendimento da eficiência de irrigação e suas vantagens (COELHO; FILHO; OLIVEIRA, 2005).

Apesar disso, a água é de suma importância para o metabolismo das plantas, principalmente para o processo de transpiração delas. Elas absorvem água do solo através de suas raízes, sendo que uma pequena parcela desse líquido é incorporada à matéria vegetal e a outra grande parte é perdida pelas folhas através dos estômatos na forma de vapor de água. Essa água entra no ciclo hidrológico juntamente com a água evaporada, como mostra a Figura 4. Em seguida, pode ser reutilizada pelas plantas através de precipitações (BRASIL, 2005).

Entretanto, quando a quantidade de água das precipitações não é suficiente para atender as exigências hídricas das plantas, a aplicação da irrigação é necessária. Caso as necessidades

¹ Food and Agriculture Organization of the United Nations ou Organização das Nações Unidas para Agricultura e Alimentação.

Figura 4 – Ciclo Hidrológico.



Fonte: (CICLO. . . , 2011).

hídricas de certa cultura sejam satisfeitas, haverá um ganho de qualidade e uma elevação na produtividade da cultura. Porém, na ocasião em que há uma irrigação insuficiente, a produtividade é reduzida devido a falta de água nos processos metabólicos das plantas, conforme cita Testezlaf (2011).

Por outro lado, não se pode aplicar uma quantidade excessiva de água, pois as consequências podem ser muito mais destrutivas. Se a irrigação for realizada de forma exagerada constantemente, ocorrerá a saturação do solo, a lixiviação de nutrientes, maior salinização e a elevação do lençol freático, de forma que poderá sufocar a planta em certo prazo (ALBUQUERQUE, 2010).

A agricultura irrigada é responsável, em média, por 42% da produção mundial, sendo que ela corresponde a uma pequena parcela do total cultivado em nosso planeta: apenas 18% (275 milhões de hectares) (COELHO; FILHO; OLIVEIRA, 2005).

Esse aumento de produtividade proporcionado pela agricultura irrigada é essencial para que se possa alimentar um número crescente de pessoas. Segundo a ONU (Organização das Nações Unidas), a produção de alimentos terá de aumentar entre 70% e 100% até o ano de 2050 para que seja possível alimentar uma população mundial com um número estimado de 9 bilhões de habitantes nesse ano (SPITZCOVSKY, 2013).

Além disso, o ganho em produtividade no setor primário também é importante economi-

camente, principalmente aos países em desenvolvimento, produzindo alimentos e matéria-prima para exportação. O Brasil tem um grande e ascendente potencial agrícola. De acordo com dados do Ministério da Agricultura, as exportações do agronegócio do Brasil alcançaram o valor recorde de 99,97 bilhões de dólares em 2013, com alta de 4,3% em relação ao ano anterior (AZEVEDO, 2014).

Tendo em vista o cenário atual da água e sua importância na agricultura irrigada, identifica-se a necessidade de ferramentas eficientes de gestão e de manejo dos recursos hídricos, uma vez que o uso racional dela está relacionado à sobrevivência, à qualidade de vida da população mundial e ao equilíbrio do ecossistema do planeta. Com a incorporação de novas tecnologias de gestão de irrigação, busca-se melhorar a produtividade das culturas e proporcionar o uso eficiente da água.

2.2 Aspectos Gerais sobre Irrigação

Há vários conceitos feitos por diferentes autores sobre a irrigação. Do ponto de vista de Testezlaf (2011), pode-se conceituar irrigação como "as técnicas, formas ou meios utilizados para aplicar água artificialmente às plantas, procurando satisfazer suas necessidades e visando a produção ideal para o seu usuário".

De acordo com Mello e Silva (2007), a irrigação é definida conforme a seguir:

A técnica da irrigação pode ser definida como sendo a aplicação artificial de água ao solo, em quantidades adequadas, visando proporcionar a umidade adequada ao desenvolvimento normal das plantas nele cultivadas, a fim de suprir a falta ou a má distribuição das chuvas (MELLO; SILVA, 2007).

Ela não deve ser vista como uma simples ferramenta, mas como uma ciência que busca a aplicação eficiente de água para atender as exigências hídricas das plantas e aumentar a produtividade delas, além de evitar desperdícios e não alterar a fertilidade do solo (TESTEZLAF, 2011).

A irrigação foi uma das primeiras modificações que o homem fez no ambiente. As tentativas de irrigação iniciais eram bem rudimentares, contudo, já se começava a perceber a importância da água. Dados históricos comprovam essa importância, pois as primeiras civilizações se formaram a beira de grandes rios como o rio Nilo no Egito, o Tigre e o Eufrates na Mesopotâmia e o Rio Amarelo na China. A agricultura foi um dos fatores que fizeram com que o homem deixasse de ser nômade para se estabelecer em territórios que lhe sustentasse (MELLO; SILVA, 2007).

Nas civilizações antigas, a irrigação era praticada fazendo-se represamentos de água cercados por diques. No Egito Antigo, por exemplo, foram construídos diques que faziam com

que a água fosse represada em grandes reservatórios e que escoava esse líquido até os campos através de canais e comportas, controlando-se a quantidade de água aplicada (MELLO; SILVA, 2007).

O desenvolvimento de várias civilizações antigas está relacionado com o sucesso das técnicas de irrigação, porque era necessário elevar a produção de alimentos para lidar com o aumento populacional. Em outras palavras, através da irrigação foi possível alcançar uma estabilidade na produção de alimentos e suportar populações mais densas (MELLO; SILVA, 2007).

Atualmente, os cientistas buscam o desenvolvimento de técnicas de irrigação para o uso eficiente de água e aumento da produção das culturas. Como referência em tecnologia e em pesquisa na área de agricultura irrigada está Israel. O país localizado no Oriente Médio possui uma tecnologia de ponta que transformou seus desertos em pomares e lavouras de alta produtividade. Israel é líder mundial em pesquisa de recursos hídricos e hoje reutiliza 85% da água que produz, contando com tecnologias como o bombardeio de nuvens, a reciclagem de água de esgoto e o uso de bactérias luminosas, que verificam a qualidade da água em poucos segundos (BLECHER, 2012?).

Para Testezlaf (2011), existem 6 unidades básicas que podem compor um sistema de irrigação. Porém, há sistemas que não possuem todas elas, pois isso depende das condições locais, como a topografia, as características do solo, a quantidade e a qualidade da água, o clima e a cultura. As unidades citadas por esse autor são:

- **Unidade de Bombeamento:** responsável por elevar a água da fonte superficial ou subterrânea até inseri-la no sistema de irrigação por meio do bombeamento. Ela fornece a pressão necessária para que esse líquido possa se movimentar até a área da plantação;
- **Unidade de Condução:** compreende o conjunto de estruturas hidráulicas responsável pelo transporte da água bombeada entre as fontes de água e os locais a serem irrigados. O dimensionamento dos canais de transporte da água precisam atender os requisitos econômicos e técnicos, permitindo que se tenha a vazão necessária para que o sistema de irrigação funcione adequadamente;
- **Unidade de Tratamento:** busca a filtragem da água proveniente da fonte, já que os emissores com pequeno diâmetro requerem que ela esteja isenta de partículas que possam obstruí-los parcialmente ou entupi-los. Essa unidade é constituída de filtros de areia, de tela ou disco;
- **Unidade de Controle:** unidade responsável por funções de controle do sistema, determinando o momento de se iniciar a irrigação e a quantidade de água aplicada, permitindo a automação de práticas agrícolas, como a aplicação de diferentes produtos químicos pela unidade de condução, por exemplo, fertilizantes;

- **Unidade de Aplicação:** tem o objetivo de aplicar a água eficientemente, sem perdas excessivas e de forma uniforme sobre a área a ser irrigada;
- **Unidade de Drenagem:** responsável por captar o volume de água em excesso que não foi utilizado pela cultura, direcionando-a para sistemas de tratamento que irão prepará-la para reutilização.

Ainda na visão de Testezlaf (2011), há 4 métodos ou modos principais e mais utilizados para irrigar uma plantação:

- **Aspersão:** aplicação de água sobre a folhagem e o solo sob a forma de uma chuva mais ou menos intensa e uniforme sobre a superfície, usando dispositivos mecânicos chamados aspersores;
- **Superfície:** aplicação de água somente sobre a superfície do terreno, aproveitando-se do efeito da gravidade para que a água se infiltre no solo;
- **Localizada:** aplicação de água em uma área limitada sobre a superfície do solo, preferencialmente dentro da área sombreada pela copa das plantas, buscando abranger a área radicular delas. Esse método se caracteriza pela baixa vazão e alta frequência de irrigação;
- **Subsuperficial ou subterrânea:** aplicação de água abaixo da superfície do solo, dentro da região de alcance das raízes das plantas. Aproveita-se o efeito do fenômeno de ascensão capilar, em que a água sobe ao longo do perfil do solo por diferença de potencial.

É importante destacar a diferença entre métodos e sistemas. Os métodos referem-se à maneira que se realiza a irrigação de uma plantação, enquanto os sistemas podem ser definidos como um conjunto de equipamentos, acessórios, formas de operação e manejo que de forma organizada realizam a aplicação de água nas culturas (TESTEZLAF, 2011). A Figura 5 exibe os principais métodos e sistemas de irrigação.

2.3 Irrigação Localizada

A irrigação localizada tem como princípio o emprego de água somente próxima a região onde estão as raízes das plantas, de forma que esse líquido seja utilizado com melhor aproveitamento. Esse método de irrigação utiliza um pequeno volume de água e apresenta uma alta frequência de aplicação. Dessa forma, a umidade do solo pode ser mantida em um nível de acordo com as necessidades hídricas das plantas (SILVA; SILVA, 2005).

No dizer de Testezlaf (2011), a irrigação localizada é recomendada quando se têm recursos hídricos caros e escassos, quando o solo é salino, pedregoso ou com topografia muito acidentada e quando se produz culturas com alto valor comercial.

Figura 5 – Métodos de Irrigação e seus Principais Sistemas.



Fonte: (TESTEZLAF, 2011).

Além de permitir o melhor aproveitamento dos recursos hídricos, o método de irrigação localizada propicia um aumento da produtividade, devido a umidade permanecer razoavelmente constante. Esse método possibilita ainda a prática da técnica de quimigação, que consiste na aplicação de produtos químicos como fertilizantes e inseticidas via água de irrigação (SILVA; SILVA, 2005).

Entretanto, a irrigação localizada exige um alto custo inicial, o que faz com que ela seja recomendada para culturas de elevado valor econômico. Por exemplo, ela necessita de um investimento em filtragem da água porque os pequenos emissores podem apresentar problemas de entupimento, causado principalmente por partículas de areia, fertilizantes, algas, bactérias, óxido de ferro e precipitados químicos (TESTEZLAF, 2011).

Os principais sistemas de irrigação localizada são por gotejamento, em que os emissores empregam água em forma de gotas, e por microaspersão, onde são utilizados os microaspersores que aplicam água de forma circular ou em faixa contínua (TESTEZLAF, 2011).

2.3.1 Sistema de Irrigação por Microaspersão

Na irrigação por microaspersão, os emissores trabalham lançando pequenos jatos, cobrindo uma distância limitada até atingir o solo. Comparando-se com a irrigação por gotejamento, os microaspersores contêm vazões superiores aos gotejadores e molham áreas maiores do solo na forma de pequenos círculos ao longo do seu raio (SILVA; SILVA, 2005).

Figura 6 – Exemplo de Microaspersor.



Fonte: (TESTEZLAF, 2011).

Nesse sistema de irrigação, as dimensões do bulbo molhado dependem dos parâmetros de alcance e de intensidade de aplicação ao longo do raio do microaspersor e do volume de água utilizado por irrigação. Assim, a forma que se umedece o solo é determinado pelo modelo de microaspersor utilizado. Como descrito por Testezlaf (2011), os microaspersores podem ser classificados em:

- **Microaspersor rotativo:** possuem partes móveis que rotacionam e ajudam na distribuição do jato na área de aplicação;
- **Microaspersor estacionário ou spray:** o emissor é composto por partes fixas, isto é, somente placas defletoras do jato. Apresentam a vantagem de permitir projetos para áreas irregulares.

A microaspersão proporciona a facilidade de visualizar a distribuição de água na superfície do solo. Dos sistemas que utilizam o método de irrigação localizada é o que tem menos risco de ocorrer entupimentos, devido aos diâmetros dos microaspersores serem maiores. Porém, pode favorecer o aparecimento de doenças. Isso porque parte do caule da planta geralmente é umedecida. Este tipo de sistema pode ser vulnerável a ventos fortes e altas taxas de evaporação, mas requer uma frequência de manutenção menor que outros tipos de sistemas de irrigação (TESTEZLAF, 2011).

2.4 Procedimentos para o Manejo de Irrigação

Como foi mencionado anteriormente, quando a precipitação pluvial é insuficiente para atender à demanda das culturas, usa-se a irrigação para suprir as necessidades hídricas das plantas. No entanto, o uso desta técnica deve utilizar a água de forma eficiente. Para isso, estratégias de manejo de irrigação podem ser adotadas. Essas estratégias buscam definir o momento e a quantidade de água a ser aplicada (ALBUQUERQUE, 2010).

Segundo Albuquerque (2010), o manejo de irrigação visa também:

- elevar a produtividade da cultura;
- aprimorar a qualidade da produção;
- reduzir o custo da água e da energia;
- aumentar a eficiência de fertilizantes;
- diminuir a incidência de doenças e pragas;
- manter ou melhorar as condições químicas e físicas do solo.

Apesar desses benefícios, muitos irrigantes não utilizam os procedimentos de manejo de irrigação de forma apropriada, devido principalmente ao seu alto custo de investimento e a falta de difusão desse conhecimento para uma parcela dos agricultores (ALBUQUERQUE, 2010).

Para realizar o manejo de irrigação, é necessário obter informações sobre diversos parâmetros e variáveis ligados ao clima, ao solo, ao sistema de irrigação adotado e à cultura, por exemplo. Os parâmetros exigidos dependem da metodologia seguida para o manejo de irrigação (ALBUQUERQUE, 2010). Nessa pesquisa, será comentada a metodologia de microirrigação, que faz a aplicação de água diariamente e em um pequeno intervalo de tempo.

2.4.1 Evapotranspiração de Referência

Borges e Mendiondo (2007) definem a evapotranspiração de referência (ET_o) como "o processo de perda de água para a atmosfera por meio de uma superfície padrão gramada, cobrindo a superfície do solo e sem restrição de umidade". Em outras palavras, essa variável representa a evaporação de água da superfície somada ao efeito da transpiração das plantas.

Os métodos que estimam o valor da ET_o são classificados em métodos diretos, que fazem um balanço de água no solo, e em métodos indiretos, que usam dados meteorológicos. Apesar dos métodos diretos serem mais exatos, eles são muito caros e complexos, exigindo equipamentos e instalações especiais. Por essas dificuldades, os métodos indiretos são os mais empregados, produzindo bons resultados para estimar a evapotranspiração de referência (SOARES; COSTA; NASCIMENTO, 2006).

Uma das formas para obter o valor da evapotranspiração de referência é através de estações agrometeorológicas situadas em um raio, no máximo, de 40 km em relação à área que se deseja analisar (SOARES; COSTA; NASCIMENTO, 2006). Alguns dos principais métodos indiretos para determinação da ET_o serão citados nas subseções a seguir.

2.4.1.1 Método do Tanque de Classe A

O método do tanque de evaporação classe A é um dos mais utilizados para determinação da evapotranspiração de referência. O tanque tem um formato circular, com 120 cm de diâmetro e com altura de 25 cm, e é geralmente constituído por uma chapa galvanizada ou por aço inoxidável. Ele deve ser posto sobre um estrado de madeira com cerca de 15 cm de altura, conforme a Figura 7. A água contida no tanque deve ser regularmente renovada para evitar a turbidez (USA, 1993).

Figura 7 – Tanque Classe A.



Fonte: (TANQUE... , 2011).

O método do tanque classe A busca obter uma medida da evaporação, relacionando um conjunto de variáveis como a radiação, o vento, a temperatura e a umidade relativa do ar sobre ela. Essas variáveis influenciam no comportamento das plantas em relação às suas necessidades hídricas (MEDEIROS et al., 2013).

As informações da evaporação do tanque podem ser usadas para estimar a ET_o utilizando a seguinte equação (USA, 1993):

$$ET_o = K_p E_t \quad (2.1)$$

onde ET_o é a evapotranspiração de referência (mm/dia), K_p é o coeficiente do tanque classe A (adimensional) e E_t é a evaporação do tanque (mm/dia). Snyder (1992) desenvolveu uma equação para fazer uma estimativa do coeficiente do tanque classe A (K_p):

$$K_p = 0,482 + 0,024 \ln(R) - 0,02356u_{10} + 0,0045UR_{min} \quad (2.2)$$

sendo que R é o raio de bordadura vegetada ao redor do tanque (m), u_{10} é a velocidade do vento média diária a 10 m de altura (m/s) e UR_{min} é a umidade relativa mínima diária (%).

2.4.1.2 Método de Blaney-Criddle

O método de Blaney-Criddle calcula a ET_o utilizando a temperatura e a porcentagem de horas diurnas como variáveis climáticas. Ele é muito adotado no oeste dos Estados Unidos para quantificar a irrigação, pois apresenta bons resultados para determinar o valor de ET_o , de acordo com Borges e Mendiondo (2007). Segundo esse autor, a equação que descreve esse método é:

$$ET_o = kp(0,46T_a + 8,13) \quad (2.3)$$

em que k (adimensional) é o coeficiente mensal, sendo que seu valor depende da vegetação, localidade e estação do ano e varia de 0,5 a 1,2. A variável p (%) é referente a porcentagem de insolação máxima diária em relação ao horário de insolação teórico do ano, que é 4380 horas. A temperatura média diária do ar é representada por T_a ($^{\circ}C$).

2.4.1.3 Método de Penman-Monteith

O método de Penman-Monteith é o recomendado pela FAO para o cálculo da evapotranspiração de referência. A equação usa variáveis climáticas como radiação solar, temperatura do ar, umidade e velocidade do vento. Para assegurar a integridade dos cálculos, as medições devem ser feitas a 2 metros acima de uma superfície de grama verde, protegendo o solo e sem escassez de água (ALLEN et al., 1998). A equação que representa esse método é:

$$ET_o = \frac{0,408\Delta(R_n - G) + \gamma \frac{900}{T_a + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0,34u_2)} \quad (2.4)$$

onde Δ é o gradiente da curva pressão vapor por temperatura ($kP_a^{\circ}C^{-1}$). A variável R_n refere-se a radiação solar líquida disponível ($MJm^{-2}d^{-1}$). O fluxo de calor no solo é representado por G ($MJm^{-2}d^{-1}$), γ é a constante psicrométrica ($kP_a^{\circ}C^{-1}$), u_2 corresponde a velocidade do vento a 2 metros (m/s), e_s é a pressão de saturação do vapor de água atmosférico (kPa) e e_a equivale a pressão atual do vapor de água atmosférico (kPa). Por último, T_a representa a temperatura média diária do ar em $^{\circ}C$ (ALLEN et al., 1998).

2.4.1.4 Método da Radiação Solar

Foi comentado que o método de Penman-Monteith é recomendado pela FAO para determinar a evapotranspiração de referência por apresentar bons resultados. Contudo, nem sempre todos os dados requisitados por ele estão disponíveis. O método da radiação solar é o que apresenta melhores resultados quando se têm informações sobre a radiação solar e a temperatura do ar. Esse método foi adotado por este trabalho, pois as informações de que ele necessita estão acessíveis e porque é um dos métodos indicados pelo Departamento de Agricultura dos Estados Unidos (USA, 1993). A equação deste método é:

$$ET_o = -0,012 + \frac{\Delta}{\Delta + \gamma} b_r + \frac{R_s}{\lambda} \quad (2.5)$$

O método da radiação solar utiliza o sistema inglês de unidades. A variável Δ corresponde a inclinação da curva de pressão de vapor ($mb/^{\circ}F$ - milibar por grau *Fahrenheit*). Ela é calculada por:

$$\Delta = 0,051 \left(\frac{164,8 + T}{157} \right)^7 \quad (2.6)$$

onde T é a temperatura do ar ($^{\circ}F$). A variável γ da Equação 2.5 é a constante psicrométrica ($mb/^{\circ}F$). Ela é determinada por:

$$\gamma = c_p \frac{BP}{0,622\lambda} \quad (2.7)$$

sendo que c_p é o calor específico do ar seco ($lang/polegadas/^{\circ}F$) e λ é o calor de vaporização ($langs/polegada$), que depende da temperatura do ar ($^{\circ}F$). A equação que representa λ é:

$$\lambda = 1543 - 0,796T \quad (2.8)$$

A variável b_r na Equação 2.5 é um fator de ajuste que depende da umidade relativa do ar média diária dada em % (RH_a) e da velocidade do vento média diária em milhas por dia (U_d). A equação a seguir mostra como essa variável é calculada:

$$b_r = 1,06 - 0,0013RH_a + 8,38 \cdot 10^{-4}U_d - 3,73 \cdot 10^{-6}RH_aU_d - 0,315 \cdot 10^{-4}RH_a^2 - 3,82 \cdot 10^{-7}U_d^2 \quad (2.9)$$

Por último, tem-se a variável R_s , que é referente a radiação solar global diária (cal/cm^2). Neste método, a evapotranspiração de referência é dada em polegadas por dia.

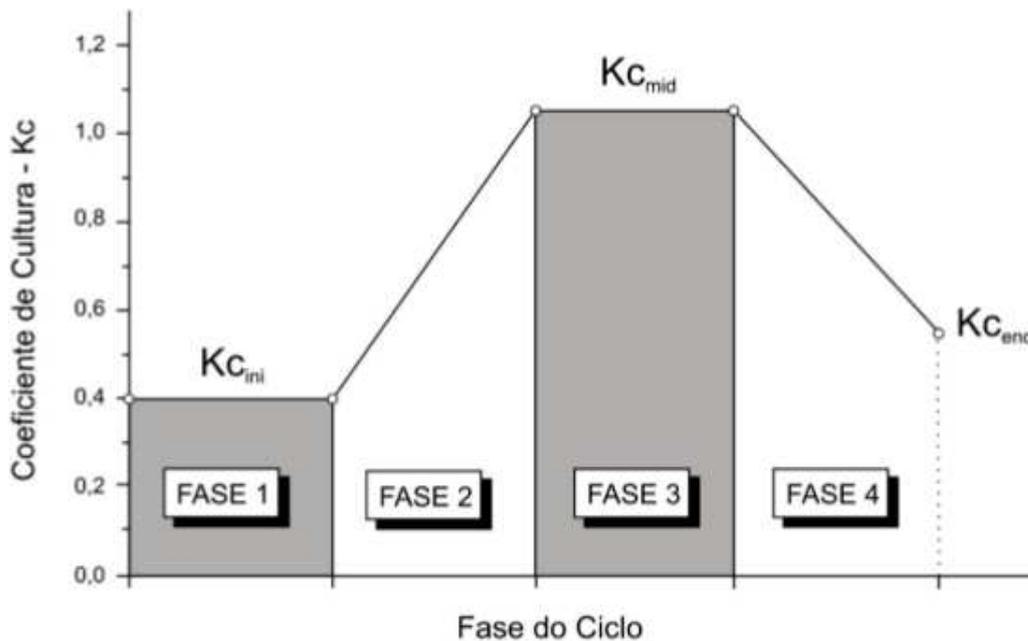
2.4.2 Evapotranspiração da Cultura

A evapotranspiração da cultura (ET_c) varia de acordo com as características de cada planta, como a sua arquitetura foliar, seu estágio de desenvolvimento e a época de plantio. A quantidade de água evapotranspirada depende também do solo e do clima onde se encontra determinada cultura (MELLO; SILVA, 2007). O valor diário da ET_c pode ser obtido pela equação a seguir (USA, 1993):

$$ET_c = K_c ET_o \quad (2.10)$$

sendo que ET_c é a evapotranspiração da cultura (mm/dia), K_c é o coeficiente da cultura (adimensional), cujo valor depende da espécie, variedade, idade da planta, práticas culturais, entre outros. A variável ET_o é a evapotranspiração de referência comentada na subseção 2.4.1.

Figura 8 – Evolução do Coeficiente de Cultura K_c ao Longo do Ciclo Fenológico de uma Planta.



Fonte: (ALBUQUERQUE, 2010).

De acordo com Mello e Silva (2007), o valor do coeficiente da cultura (K_c) varia conforme a planta vai se desenvolvendo. A Figura 8 exhibe como é comportamento geral de K_c .

2.4.3 Lâmina Bruta de Irrigação

A lâmina bruta de irrigação (L_b) equivale a quantidade de água em mm que deverá ser aplicada em uma plantação, prevendo-se perdas e buscando a uniformidade de distribuição. O cálculo de L_b pode ser realizado por meio da seguinte equação (SOARES; COSTA; NASCIMENTO, 2006):

Tabela 1 – Valores Estimados para a Eficiência de Irrigação (E_a).

Sistema de Irrigação	Eficiência da Irrigação - E_a
Aspersão	70%
Microaspersão	85%
Gotejamento	95%

Fonte: (SOARES; COSTA; NASCIMENTO, 2006).

$$L_b = \frac{ET_c - P_e}{E_a} \quad (2.11)$$

onde P_e é a precipitação efetiva (mm), que é a quantidade de água das chuvas que efetivamente é aproveitada pelas plantas, e E_a corresponde a eficiência de irrigação (%).

A precipitação efetiva pode ser determinada efetuando-se alguns cálculos com as equações a seguir, citadas por Jensen, Burman e Allen (1990):

$$P_e = f(A_r)[1,25P^{0,824} - 2,93][10^{0,0000955E_i}] \quad (2.12)$$

$$f(A_r) = 0,53 + 0,0116A_r - 8,94 \times 10^{-5}A_r^2 + 2,32 \times 10^{-7}A_r^3 \quad (2.13)$$

$$A_r = 10(\theta_{cc} - \theta)D_gP_r \quad (2.14)$$

em que $f(A_r)$ é o fator de correção (adimensional), P é a precipitação pluviométrica (mm), A_r equivale a capacidade de armazenamento do solo na profundidade efetiva da raiz (mm), θ_c e θ são referentes aos conteúdos de água na capacidade de campo e no momento de irrigação ($cm^3 \cdot cm^{-3}$). A variável D_g representa a densidade do solo (g/m^3) e P_r corresponde a profundidade efetiva da raiz (m) (SOARES; COSTA; NASCIMENTO, 2006).

Contudo, para Azevedo et al. (2008), é possível estabelecer uma aproximação para a precipitação efetiva (P_e). Em seu trabalho, foram considerados efetivos 75% dos valores de precipitações acima de 10 mm . Chuvas abaixo de 10 mm foram declaradas 100% aproveitadas.

Em relação a eficiência da irrigação, pode-se estimar valores aceitáveis, conforme cita Soares, Costa e Nascimento (2006). A Tabela 1 exibe o valor de E_a para alguns tipos de sistemas de irrigação.

2.4.4 Volume de Água Aplicado na Irrigação

O volume de água bruto aplicado por planta em cada evento de irrigação pode ser calculado pela seguinte fórmula (SOARES; COSTA; NASCIMENTO, 2006):

$$V_b = L_b F_c A_p \quad (2.15)$$

em que V_b é o volume total bruto de água aplicado por planta (L), L_b é a lâmina bruta de irrigação (mm), F_c equivale ao fator de cobertura do solo (%) e A_p caracteriza a área da planta (m^2).

A variável F_c é referente a porcentagem que cobre o solo, ou seja, a razão entre a área coberta pela cultura e a área ocupada por seu espaçamento. O valor de F_c sempre está entre compreendido entre 0 e 1 (SOARES; COSTA; NASCIMENTO, 2006).

2.4.5 Tempo de Irrigação

O cálculo do tempo de irrigação para sistemas de irrigação localizada por microaspersão é baseado em V_b , como mostra a equação a seguir (SOARES; COSTA; NASCIMENTO, 2006):

$$t = N \frac{V_b}{q_a} \quad (2.16)$$

sendo que N é a razão entre o número de plantas (N_p) e o número de microemissores (N_a) que são utilizados por planta. A variável V_b se refere ao volume total bruto de água em litros aplicado por planta e q_a é vazão do microaspersor em litros por hora (L/h).

2.5 Dendzeiro

O dendzeiro (*Elaeais guineenses*) ou óleo de palma é uma planta originária da costa ocidental africana encontrada desde o Senegal até Angola. Atualmente, essa planta é cultivada em várias partes do mundo, destacando-se a Malásia, a Nigéria e a Indonésia como os maiores produtores de dendê do planeta. Na América, ela foi introduzida no século XVI, coincidindo com o tráfico de escravos da África para esse continente. No território brasileiro, o dendzeiro chegou a diversas regiões, porém foi na Bahia que se adaptou melhor, posteriormente chegando ao Pará e ao Amapá e, recentemente, introduzido no Maranhão em 14 de julho de 2011 (ALMEIDA, 2013).

O fruto é oval, liso, duro e brilhante, do tamanho de uma azeitona pequena. Ele nasce em cachos, na cor preta e quando está maduro alcança uma cor que varia entre o amarelo forte, o alaranjado e o vermelho rosado. Sua polpa fibrosa não passa de 2 mm de espessura, sendo possível que, em algumas variedades, ela esteja ausente (ALMEIDA, 2013).

O dendê é uma espécie perene, tropical que se comporta bem em ambientes com temperatura média entre 24°C e 28°C, com máxima 33°C e mínima não inferior a 18°C, luminosidade de radiação solar superior a 1.800 horas/ano e com uma precipitação pluviométrica bem distribuída superior a 1.800 mm no decorrer do ano. Além disso, essa cultura se destaca por ter uma elevada produtividade em óleo/ha (4 t a 6 t de óleo), uma menor quantidade de área para plantio e uma boa adaptação a áreas degradadas de floresta, reduzindo os impactos do desmatamento (MACIEL et al., 2013).

Figura 9 – Plantação de Dendê na Fazenda Escola de São Luís, Universidade Estadual do Maranhão (UEMA).



Fonte: Autoria Própria.

O dendê é muito usado para a produção de 2 tipos de óleo. O primeiro, que é extraído da polpa ou do mesocarpo, é chamado de azeite de dendê ou óleo de palma. Segundo estatísticas da FAO, o dendê ocupa hoje o segundo lugar em produção mundial de óleos e ácidos graxos, devendo ultrapassar a soja. Isso devido a seu baixo custo de produção, boa qualidade e ampla utilização. Hoje, 80% da produção mundial desse óleo é destinada à culinária. O segundo é extraído da semente, chamado de óleo de palmiste, usado na confeitaria, na produção de margarinas, sorvetes, sabões, detergentes e graxas (ALMEIDA, 2013).

Ademais, o dendzeiro tem grande potencial agroenergético, sendo a espécie oleaginosa de maior potencial produtivo e com alta eficiência na conversão energética para a produção de biodiesel. Isso porque, além de utilizar uma baixa quantidade de combustível fóssil, seu óleo figura como substituto direto do óleo diesel em motores multicompostíveis (MACIEL et al., 2013).

3 PROCESSO DE AQUISIÇÃO DE CONHECIMENTO

Este capítulo apresenta a metodologia do processo de descoberta de conhecimento. Esse método busca extrair conhecimento e padrões úteis a partir de um conjunto massivo de dados. Ele é constituído de várias etapas, como a preparação dos dados, a mineração e a avaliação do conhecimento encontrado. Na etapa de mineração dos dados, o pesquisador precisa escolher que tarefas e técnicas devem ser adotadas de acordo com a aplicação. As tarefas são referentes aos objetivos que o investigador deseja alcançar, já as técnicas estão relacionadas a como ele irá atingir a finalidade da tarefa selecionada. Algumas tarefas serão comentadas nesse capítulo. A regressão será destacada, pois foi a técnica adotada nessa pesquisa.

3.1 Aspectos Básicos sobre a Descoberta de Conhecimento

Em diversas áreas de conhecimento e atividades dos seres humanos, um número grandioso de dados é produzido em um ritmo cada vez mais acelerado. Eles são provenientes de empresas, do meio científico e acadêmico ou do governo, por exemplo. De acordo com um estudo feito pela IDC¹, a quantidade de dados gerados dobra a cada dois anos (MATSSURA, 2012).

Há também uma grande evolução na tecnologia para o armazenamento desses dados, com isso, a capacidade de analisá-los e compreendê-los precisa acompanhar esse avanço. Geralmente, quanto mais se conhece os dados, melhor para se tomar decisões e definir estratégias no domínio ou área de conhecimento deles. Dessa forma, há muitos dados disponíveis, mas nem sempre é fácil extrair informações e conhecimentos úteis a partir deles, pois estes podem estar desorganizados e contendo ruídos (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

A metodologia tradicional para extração de conhecimentos a partir de dados brutos depende da análise e interpretação manual. Ou seja, especialistas e analistas que estão familiarizados com os dados levam o conhecimento adquirido até os usuários que irão utilizá-lo. Contudo, esse método tradicional pode-se tornar lento, caro e altamente subjetivo para volumes grandiosos de elementos (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Devido a essas condições, surge a necessidade do desenvolvimento de teorias e ferramentas inteligentes que auxiliem os pesquisadores na descoberta de conhecimentos aproveitáveis a partir de um conjunto de dados brutos. Essas teorias e ferramentas estão inseridas no domínio do KDD² (RAMOS; LOBO, 2003).

No ponto de vista de Fayyad, Piatetsky-Shapiro e Smyth (1996), a definição de KDD é: "o processo não trivial de identificação válida, original, potencialmente útil, de padrões

¹ *International Data Corporation* ou Corporação Internacional sobre Dados.

² *Knowledge Discovery in Database* ou Descoberta de Conhecimento em Base de Dados.

compreensíveis existentes nos dados". Em outras palavras, o KDD está relacionado com a aplicação de métodos e técnicas que buscam dar sentido aos dados, utilizando algoritmos de mineração (*Data Mining*) para descobrir padrões neles.

Fayyad, Piatetsky-Shapiro e Smyth (1996) destacam ainda que é fundamental não confundir a mineração de dados com o processo de KDD. Enquanto o KDD refere-se a todo o procedimento de descoberta de conhecimento, a mineração é apenas uma das etapas desse processo.

Quando se trabalha com o KDD, é muito importante saber diferenciar dados de informação e de conhecimento. Os dados podem ser considerados como itens primários coletados de eventos, observações ou atividades, não apresentando um significado específico. Esses dados quando estão agrupados de forma que tenham sentido e significado são nomeados de informação. Por último, o conhecimento consiste em informações organizadas e processadas para transmissão de compreensão, aprendizado e experiência, auxiliando os seus usuários em tomadas de decisão (TURBAN; MCLEAN; WETHERBE, 2004).

Outras questões que os pesquisadores devem considerar quando o KDD é utilizado são os problemas de eficiência e de desempenho, o nível de privacidade das informações, a confiabilidade das fontes de dados, a escolha correta da metodologia de extração de conhecimento, entre outras preocupações (RAMOS; LOBO, 2003).

Fayyad, Piatetsky-Shapiro e Smyth (1996) consideram que há duas categorias de critérios para selecionar potenciais aplicações de KDD: os critérios práticos e os técnicos. Os primeiros se preocupam com o impacto potencial de uma aplicação, levando em consideração a existência ou não de boas soluções e a dificuldade de obtê-las. Já os critérios técnicos levam em conta a disponibilidade de dados, baixa incidência de ruídos e dados errôneos, além da existência de atributos relevantes.

Geralmente, quanto mais atributos existem e quanto maior for a dificuldade de extrair padrões, serão exigidos mais dados e tempo de processamento para desenvolver um modelo de conhecimento. No entanto, um estudo prévio do domínio da aplicação e o pré-processamento dos dados são muito importantes, pois podem reduzir a complexidade da aplicação e melhorar bastante os resultados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

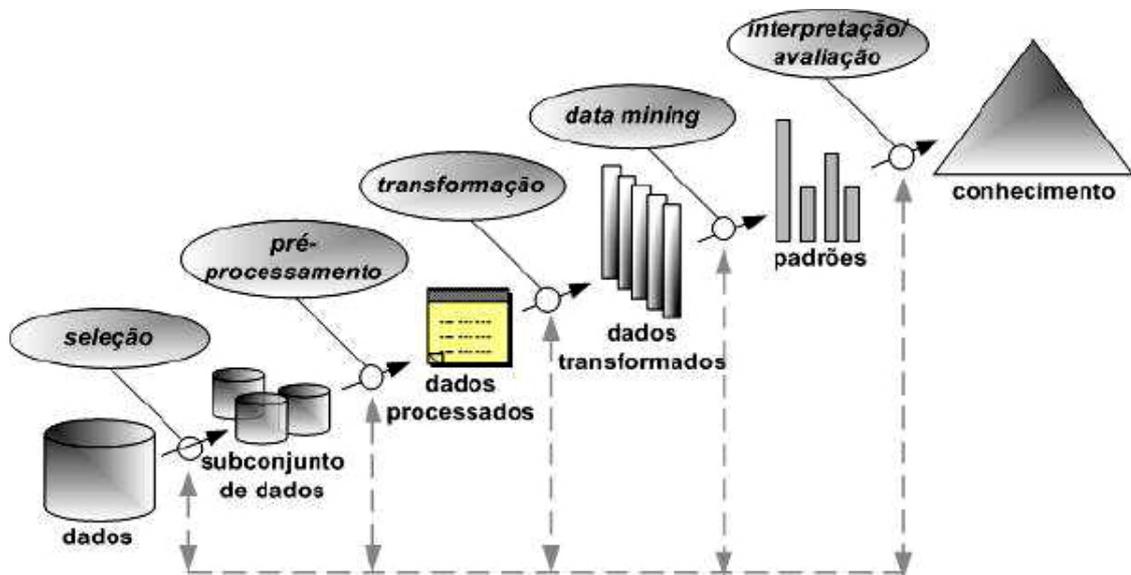
3.2 Etapas do Processo de Descoberta de Conhecimento

O processo de KDD tem a característica de ser interativo e iterativo, porque seus estágios são influenciados por tomadas de decisão de usuários, sendo que é possível eles se repetirem. Logo, o KDD é composto por um conjunto de etapas, conforme a Figura 10. Em primeiro lugar, tem-se a etapa de seleção de dados, o pré-processamento, a subamostragem e a transformação dos dados. Nessa primeira fase, eliminam-se as inconsistências dos elementos, realiza-se a integração entre as diversas fontes de dados caso haja mais de uma e obtém-se o conjunto de registros

relevantes para a aplicação, preparando-os para serem minerados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Por conseguinte, os métodos de mineração de dados são implementados através de algoritmos ou técnicas que são selecionados dependendo da tarefa ou do objetivo do processo de KDD. A etapa de *data mining* busca padrões sobre os dados por meio dessas técnicas. Por último, os padrões encontrados são avaliados e interpretados, verificando se os conhecimentos adquiridos são úteis e apresentando os resultados dessa análise ao usuário final. Nem sempre todas essas etapas são necessárias, dependendo da aplicação específica onde está se empregando o processo de KDD (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Figura 10 – Etapas do Processo de KDD.



Fonte: (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Como alternativa às etapas descritas acima, Fayyad, Piatetsky-Shapiro e Smyth (1996) enumeram uma sequência de passos básicos para a implementação do processo de KDD:

1. Adquirir um conhecimento relevante sobre o domínio da aplicação e identificar o objetivo do processo de KDD no ponto de vista do cliente;
2. Selecionar um conjunto de dados ou amostras, dependendo do objetivo do processo de KDD;
3. Realizar a limpeza de dados (removendo ruídos e lidando com registros vazios) e seu pré-processamento;
4. Promover uma redução da quantidade de variáveis e buscar recursos úteis para representar os dados, dependendo do objetivo;

5. Escolher a tarefa de mineração de dados que irá ser adotada, como por exemplo, classificação, estimação, associação e *clustering*;
6. Selecionar os algoritmos ou as técnicas de mineração de dados que serão usadas para buscar padrões sobre eles, por exemplo, árvores de decisão, redes neurais e regressão;
7. Minerar os dados, de acordo com as tarefas e técnicas escolhidas nas etapas anteriores;
8. Interpretar os padrões obtidos, avaliando se eles são úteis para serem considerados como conhecimento;
9. Agir sobre o conhecimento adquirido, usando-o diretamente, incorporando-o em outro sistema com novas funcionalidades ou apenas fazendo sua documentação.

3.3 Etapa de Preparação dos Dados

A maior parte dos dados brutos coletados precisam de um tratamento para ser usado nas etapas posteriores. Os dados brutos podem estar com vários campos nulos ou redundantes, em formatos inadequados ou com valores errôneos e discrepantes. Caso isso não seja feito, há uma grande probabilidade que os resultados não sejam satisfatórios (LAROSE, 2005).

É comum nessa etapa quatro tipos de procedimentos, a limpeza de dados (*data cleaning*), a integração de dados (*data integration*), a seleção de dados (*data selection*) e a transformação de dados (*data transformation*). Na primeira, os elementos com campos nulos, ausentes e com ruídos são removidos ou modificados com valores coerentes. A união de diversas fontes de dados é realizada pela integração. O *data selection* escolhe apenas o conjunto relevante de acordo com a ocasião. Por último, faz-se a formatação desses dados através do estágio de transformação, preparando-os para serem minerados (HAN; KAMBER, 2006).

Esses elementos que foram tratados nessas etapas podem ser divididos em duas classes, os dados de treinamento, que serão minerados para o desenvolvimento do modelo de conhecimento, e o conjunto de dados de avaliação, cuja função é ser utilizado como avaliador desse modelo que foi gerado. Quando não há disponibilidade de dados para essa divisão, é comum usar outros meios de avaliação que serão comentados na seção 3.5 (WITTEN; FRANK; HALL, 2011).

3.4 Etapa de Mineração de Dados

A mineração de dados pode ser definida simplesmente como a etapa do processo de KDD que busca a descoberta de padrões de dados. Esses padrões descobertos devem ser considerados proveitosos e significativos, fornecendo vantagens econômicas e realizando estimções de novos dados, por exemplo. Para efetuar a mineração, deve-se primeiramente conhecer o objetivo que se deseja alcançar, ou seja, escolher a tarefa de *data mining* que será usada. Em seguida, é preciso analisar que técnicas ou algoritmos são mais adequados para atingir o objetivo da tarefa

selecionada e para satisfazer os requisitos do usuário final (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

3.4.1 Tarefas de Mineração de Dados

De acordo com Larose (2005), algumas das tarefas que usualmente são adotadas para a realização da etapa de mineração de dados são listadas a seguir:

- Classificação;
- Estimação;
- *Clustering* ou Agrupamento;
- Associação.

3.4.1.1 Classificação

A tarefa de classificação visa desenvolver um modelo que organiza os dados, distribuindo-os em uma das classes pré-definidas de uma variável. Exemplos de métodos de classificação utilizados como parte de aplicações de descoberta de conhecimento incluem a identificação de tendências nos mercados financeiros e a categorização automática de objetos de interesse em grandes bancos de dados de imagens (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

A mineração de dados com a tarefa de classificação examina um grande número de registros, onde cada um deles contém um conjunto de variáveis de entrada. Em seguida, eles são classificados de acordo com um rótulo ou uma variável alvo. Larose (2005) utiliza um exemplo para mostrar como a tarefa de classificação se caracteriza, por meio da Tabela 2.

Tabela 2 – Classificação por Renda de Indivíduos.

Registro	Idade	Gênero	Cargo	Renda
001	47	F	Engenheiro de Software	Alta
002	28	M	Consultor de <i>Marketing</i>	Média
003	35	M	Desempregado	Baixa

Fonte: (LAROSE, 2005).

Neste exemplo, o objetivo é desenvolver um modelo de conhecimento para classificar pessoas com base em características, como idade, sexo e cargo de outros indivíduos, contidas em registros armazenados em um banco de dados. Em primeiro lugar, todos os registros devem ser agrupados segundo um rótulo de referência, que no caso é a renda. O aglomerado de dados formado é chamado de conjunto de treinamento. Dessa forma, aplica-se uma técnica de mineração selecionada para desenvolver um modelo classificador, que aprende sobre a relação entre as

variáveis categóricas dos registros. Com o modelo disponível, é possível atribuir as categorias alta, média ou baixa renda em novos registros de entrada. Por exemplo, uma professora de 63 anos de idade, pode ser classificada na faixa de alta renda (LAROSE, 2005).

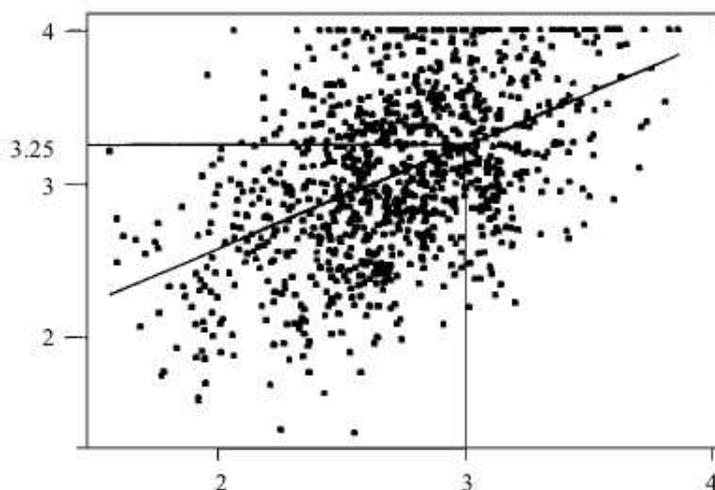
As técnicas mais usadas na tarefa de classificação são os k-vizinhos mais próximos, as árvores de decisão e as redes neurais (LAROSE, 2005).

3.4.1.2 Estimação

A tarefa de estimação é bem semelhante à de classificação, exceto que ela trabalha com variáveis numéricas em vez de categóricas. Os modelos são construídos usando registros completos que fornecem o valor de uma determinada variável. Em outras palavras, é fornecido ao modelo um conjunto de variáveis independentes como entrada para calcular o valor da variável dependente ou alvo.

Como exemplo, deseja-se estimar a pressão arterial de um paciente, com base na idade dele, seu sexo, seu índice de massa corporal e os níveis de sódio no sangue. A relação entre a pressão arterial e as variáveis de previsão no conjunto de treinamento fornecerá um modelo de estimativa. Então, pode-se posteriormente aplicar esse modelo obtido para novos casos (LAROSE, 2005).

Figura 11 – Gráfico das Notas de Pós-graduação em Relação as Notas de Graduação.



Fonte: (LAROSE, 2005).

Larose (2005) cita outros exemplos onde se pode utilizar essa tarefa, como estimar a quantidade de dinheiro que uma família gastará com compras ou determinar a nota média de um estudante de pós-graduação, tendo como base suas notas na graduação.

A Figura 11 mostra um gráfico deste último exemplo, onde se deseja estimar as notas na pós-graduação (eixo y) em relação as notas de aluno no curso de graduação (eixo x). Foram

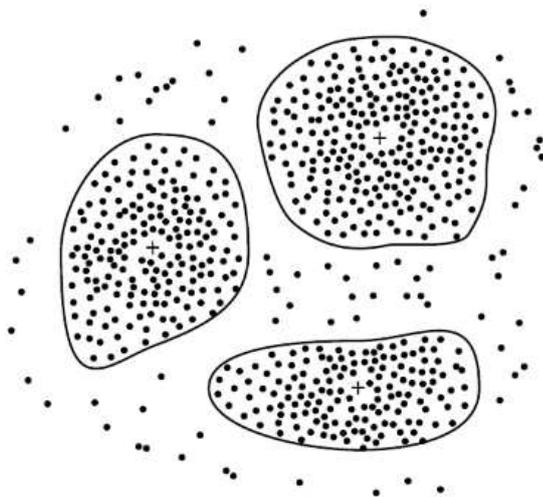
considerados mil estudantes. Neste caso, a técnica de regressão linear simples foi utilizada, onde se obteve uma função (representada pela reta linear na Figura 11) que melhor aproxima a relação entre estas duas variáveis (LAROSE, 2005).

A função linear obtida foi $y = 1.24 + 0.67x$, onde y representa a nota na pós-graduação e x é a nota na graduação. Portanto, essa função linear é o modelo de estimativa para este caso. Além da regressão, é comum utilizar outras técnicas na tarefa de estimação como as redes neurais (LAROSE, 2005).

3.4.1.3 Clustering ou Agrupamento

A tarefa de agrupamento tenta identificar e aproximar dados similares, criando grupos ou *clusters* de registros com características semelhantes. As técnicas que implementam essa tarefa buscam criar agrupamentos homogêneos de dados, onde a similaridade entre os registros do *cluster* deve ser maximizada e a semelhança com outros registros fora deste mesmo grupo seja reduzida (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Figura 12 – Agrupamentos de Registros Similares.



Fonte: (HAN; KAMBER, 2006).

Larose (2005) aponta alguns exemplos de aplicações para a tarefa de *clustering*, como a identificação de transações financeiras autênticas e suspeitas, redução da dimensão dos dados quando estes estiverem em enormes quantidades, taxonomia de plantas e animais, reconhecimento de padrões e processamento de imagens.

As técnicas mais utilizadas para realizar o *clustering* são as redes neurais, os algoritmos de k-médias e as redes de Kohonen (LAROSE, 2005).

3.4.1.4 Associação

A tarefa de associação consiste em descobrir as regras que relacionam os atributos. Essas regras são da forma "SE atributo X então atributo Y", sendo que há uma medida de confiança relacionada a elas. Um exemplo dado por Larose (2005) é de um supermercado, onde, de um número total de 1000 clientes, 200 compram fraldas e 50 desses compram cerveja. Assim, a regra de associação que pode ser formulada é "SE comprar fralda, então comprar cerveja", sendo que essa regra possui um fator de 25% (50/200) de confiança para 20% (200/1000) dos clientes em questão (LAROSE, 2005).

Outras aplicações para essa tarefa são: determinar os casos onde um medicamento pode apresentar efeitos colaterais e identificar se os usuários respondem bem a certa oferta de serviços (LAROSE, 2005).

3.4.2 Técnicas de *Data Mining*

Enquanto as tarefas dizem respeito aos objetivos que os pesquisadores possuem ao minerar dados, as técnicas se referem a que maneira esse objetivo será alcançado. Os métodos ou técnicas de *Data Mining* podem ser classificados em aprendizado supervisionado e não-supervisionado. A maioria dos métodos de mineração de dados é da classe de aprendizado supervisionado, onde há uma variável de referência pré-definida, logo, os dados são categorizados em relação a ela. Geralmente, as técnicas que implementam a classificação e a estimação são do tipo supervisionada (LAROSE, 2005).

O aprendizado não-supervisionado não necessita dessa variável de referência. Usualmente, os métodos que usam esse tipo de aprendizado visam medir a similaridade entre os atributos e encontrar padrões entre os dados. As técnicas que são usadas para aplicação do *clustering* são tipicamente não supervisionadas (LAROSE, 2005).

Embora exista essa divisão, não há uma fronteira rígida entre esses dois tipos de aprendizado, sendo que alguns de seus conceitos podem se misturar (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Os métodos listados abaixo são alguns dos mais utilizados na etapa de mineração de dados:

- Árvores de decisão;
- K-vizinhos mais próximos;
- Redes neurais;
- Redes de Kohonen;
- Regressão.

3.4.3 Regressão e Correlação

Quando pesquisas, trabalhos científicos, teses entre outros trabalhos têm como objetivo estabelecer relações entre variáveis para estimar o comportamento de outra, pode-se utilizar a técnica de regressão. Ela permite calcular o valor médio de uma variável em termos dos valores conhecidos de outras. Dessa maneira, busca-se expressar ou aproximar as relações entre as grandezas conhecidas e as variáveis que devem ser determinadas por meio de equações matemáticas com o uso dessa técnica (FREUND, 2006).

Toda vez que se usam dados observados para se obter uma equação matemática que descreve a relação entre as variáveis, o processo de ajuste de curvas precisa ser realizado. Esse processo tenta solucionar três questões principais (FREUND, 2006):

1. Selecionar o tipo de curva e de equação para previsão;
2. Encontrar a equação mais adequada para a aplicação;
3. Analisar os méritos da equação escolhida.

No primeiro problema, escolhe-se qual o tipo de curva que melhor descreve os padrões de dados, por exemplo, uma parábola, uma reta ou uma exponencial. Quando a representação é feita por uma reta, diz-se que a regressão é linear. A regressão linear é caracterizada por uma função do primeiro grau. Esse tipo de regressão pode ser classificado em dois tipos de acordo com o número de variáveis independentes envolvidas: em simples ou múltipla. A denominada regressão linear simples trata-se de um modelo linear que possui apenas uma variável independente que está presente na relação para o cálculo da variável dependente. Esse tipo de regressão tem o formato exibido na equação da reta a seguir, onde y é o valor estimado para um dado valor de x (LAUBER; FARBER, 2010).

$$y = mx + b \quad (3.1)$$

A inclinação m e a intersecção b com o eixo y são dadas pelas equações abaixo, sendo que n corresponde à quantidade de dados do problema (LAUBER; FARBER, 2010).

$$m = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} \quad (3.2)$$

$$b = \frac{\sum y}{n} - m \frac{\sum x}{n} \quad (3.3)$$

A regressão linear múltipla trabalha com mais de uma variável independente. Ela pode ser classificada como total quando considera o efeito de todas as variáveis simultaneamente ou parcial quando algumas variáveis puderem ser definidas como constantes. O modelo de

regressão linear múltipla trata de eventos que somente são razoavelmente descritos por mais de uma variável independente. Esse tipo de regressão também ele é descrito por uma equação linear conforme a seguir (FREUND, 2006):

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (3.4)$$

De forma semelhante a equação da regressão linear simples, a Equação 3.4 contém k variáveis conhecidas (x_1, x_2, \dots, x_k) e $k + 1$ constantes numéricas a serem determinadas ($b_0, b_1, b_2, \dots, b_k$). A relação linear entre elas determina o valor de y que é a variável dependente que se deseja estimar. O problema para determinação de uma equação linear consiste em encontrar as constantes numéricas e, geralmente, usa-se o método dos mínimos quadrados (subseção 3.4.3.1) para determiná-las (FREUND, 2006).

Caso não seja viável modelar de forma linear, é preciso considerar algum outro tipo de função matemática, como uma parábola, curva ou exponencial. O tipo de regressão que trabalha com funções desse formato é chamado de não-linear. Da mesma forma que a regressão linear, a não-linear pode ser simples, quando contém uma variável independente, ou múltipla, apresentando mais de uma variável independente (FREUND, 2006). Os tipos de regressão podem ser visualizados pelo diagrama na Figura 13.

Figura 13 – Tipos de Regressão.



Fonte: Autoria Própria.

As equações que descrevem uma regressão do tipo não-linear apresentam diversos formatos e alguns exemplos estão a seguir:

$$y = ab^x \quad (3.5)$$

$$y = ax^2 + bx + c \quad (3.6)$$

$$y = e^x + b \quad (3.7)$$

$$y = x_0^2 - b \sin(x_1) \quad (3.8)$$

3.4.3.1 Métodos dos Mínimos Quadrados

O segundo tipo de problema que foi enumerado anteriormente consiste em encontrar a equação mais adequada para certa situação, isto é, determinar a equação particular que representa o melhor ajuste de acordo com dados fornecidos. Um método muito utilizado para tratar dessa situação é o método dos mínimos quadrados, que é um dos recursos estatísticos mais comuns nas ciências experimentais. Esse método é bastante aplicado na regressão, tanto na simples quanto na múltipla. Ele busca a curva de melhor ajuste, isto é, a soma dos quadrados das distâncias verticais aos pontos dessa curva deve ser mínima. Essa curva pode ser linear (uma reta) ou não-linear (FREUND, 2006).

Em outras palavras, esse método tenta minimizar o erro quadrático médio (S) das medidas. Considerando a Equação 3.1 para regressão linear simples com um conjunto de k medidas, pode-se calcular S como a seguir (LAUBER; FARBER, 2010):

$$S = \sum_{i=1}^k \Delta S_i = \sum_{i=1}^k (y - y_i)^2 \quad (3.9)$$

Como nesse caso a equação é linear simples e é representada por uma função linear, o objetivo é somar os ΔS_i de todas as k medidas e traçar essa reta para que a soma dos ΔS_i seja mínima. Isso quer dizer que a derivada parcial de S em relação a m e a derivada parcial de S em relação a b devem ser iguais a zero, ou seja, (LAUBER; FARBER, 2010):

$$\frac{\partial S}{\partial m} = -2 \sum_{i=1}^k (y_i - m - bx_i) = 0 \quad (3.10)$$

$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^k (y_i - m - bx_i) x_i = 0 \quad (3.11)$$

O método busca obter uma reta que passe por todos os k pontos considerados. Assim, os parâmetros m e b podem ser determinados pelas equações abaixo:

$$m = \frac{k \sum_{i=1}^k x_i y_i - \sum_{i=1}^k x_i \sum_{i=1}^k y_i}{k \sum_{i=1}^k x_i^2 - (\sum_{i=1}^k x_i)^2} \quad (3.12)$$

$$b = \frac{\sum_{i=1}^k x_i^2 \sum_{i=1}^k y_i - \sum_{i=1}^k x_i y_i \sum_{i=1}^k x_i}{k \sum_{i=1}^k x_i^2 - (\sum_{i=1}^k x_i)^2} \quad (3.13)$$

A aplicação desse método para regressão linear múltipla é semelhante ao procedimento acima, expandindo as equações de acordo com a quantidade de variáveis (COELHO-BARROS et al., 2008).

3.4.3.2 Coeficiente de Correlação

Para Lauber e Farber (2010), a correlação é definida como a seguir: "uma correlação é uma relação entre duas variáveis. Os dados podem ser representados por pares ordenados (x, y) , onde x é a variável independente (ou explanatória) e y é a variável dependente."

Pode-se interpretar a correlação usando os diagramas de dispersão ou o coeficiente de correlação de Pearson. Os diagramas de dispersão exibem a relação entre duas variáveis, com a finalidade de determinar se existe uma correlação entre elas. Já o coeficiente de correlação de Pearson (r) é uma medida de força e direção de uma relação linear. Ele pode ser calculado pela fórmula a seguir, tendo como referência todo o conjunto de pares ordenados (x, y) (LAUBER; FARBER, 2010):

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}} \quad (3.14)$$

O coeficiente de correlação fica compreendido entre -1 e 1. Caso seu valor seja próximo de 1, existe uma correlação linear positiva e forte. Se o coeficiente for próximo de -1, a correlação entre as variáveis é dita negativa e forte. Quando o seu valor for próximo de 0, não há uma correlação linear ou existe uma correlação fraca entre as variáveis (LAUBER; FARBER, 2010). Alguns exemplos de correlações são mostrados na Figura 14.

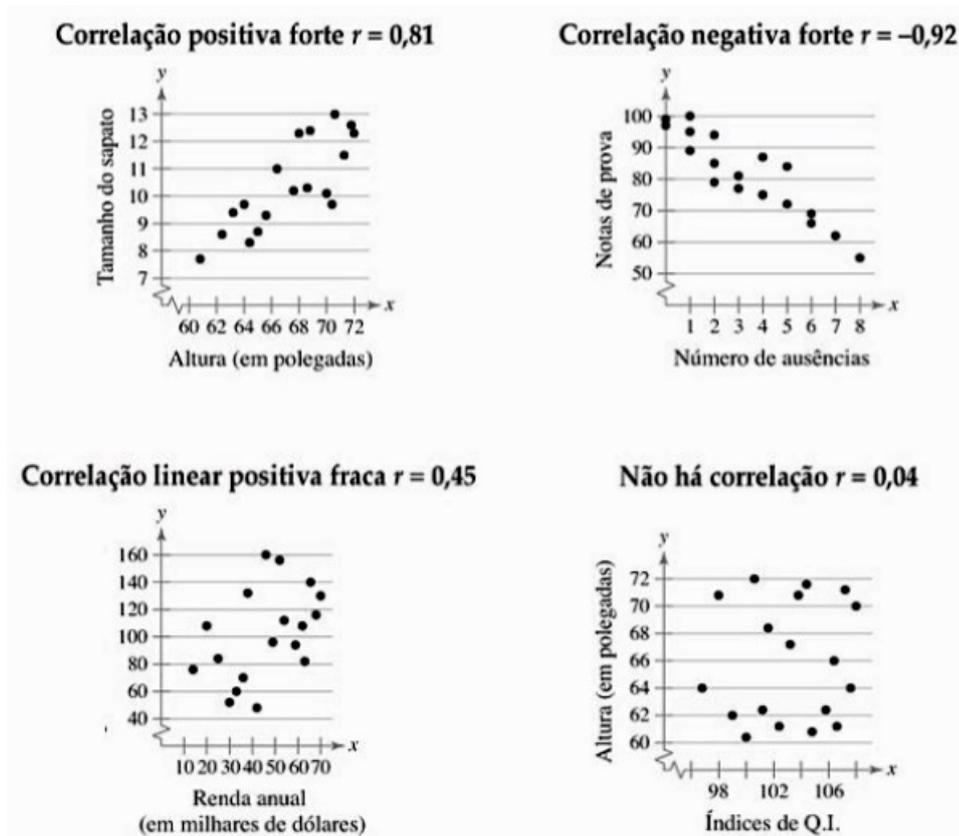
3.5 Avaliação do Modelo

Uma das etapas mais importantes no processo de descoberta de conhecimento é a avaliação do modelo desenvolvido por meio da mineração de dados, verificando se ele consegue alcançar os requisitos desejados. Pode-se ainda comparar diferentes métodos nesta fase (WITTEN; FRANK; HALL, 2011).

Todavia, esse estágio não é tão simples, pois os parâmetros e as taxas de erros apresentados ao gerar o modelo de treinamento não são bons indicadores de desempenho. Como esse modelo foi desenvolvido com os dados de treinamento, qualquer estimativa de desempenho para testá-lo utilizando esses mesmos dados será muito otimista. Assim, é preciso testar o desempenho na prática, com base em outro conjunto de dados que não foi incluído no treinamento do modelo. Quando há disponível uma grande oferta de dados, pode-se dividi-los em duas partes, uma para realizar o treinamento e outra para avaliar o modelo gerado (WITTEN; FRANK; HALL, 2011).

No entanto, se não houver muitos dados acessíveis, pode-se utilizar outras técnicas para avaliação de um modelo. Uma das mais conhecidas é a validação cruzada (*cross-validation*). Ela

Figura 14 – Exemplos de Coeficientes de Correlação.



Fonte: (LAUBER; FARBER, 2010).

divide a base de dados em uma quantidade n de partes ou *folds*. Em seguida, essa técnica usa $n - 1$ partes para o treinamento do modelo e uma apenas para realização do teste. Isso é repetido n vezes até que cada parte já tenha sido usada como teste. (HAN; KAMBER, 2006).

Em relação aos modelos de regressão, geralmente eles são avaliados por meio de taxas de erros, onde se compara o valor real de uma variável (y_i) com a estimativa feita pelo modelo de regressão (y'_i), e através do coeficiente de correlação (subseção 3.4.3.2). As funções de erros mais comuns são do erro absoluto (Equação 3.15) e do erro quadrático (Equação 3.16) (WITTEN; FRANK; HALL, 2011).

$$|y_i - y'_i| \quad (3.15)$$

$$(y_i - y'_i)^2 \quad (3.16)$$

Baseando-se nessas fórmulas, pode-se inferir o erro em média ao longo de uma quantidade de dados n (WITTEN; FRANK; HALL, 2011). A seguir, é mostrado como se calcula o erro absoluto médio e o erro quadrático médio, respectivamente:

$$\frac{\sum_{i=1}^n |y_i - y'_i|}{n} \quad (3.17)$$

$$\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n} \quad (3.18)$$

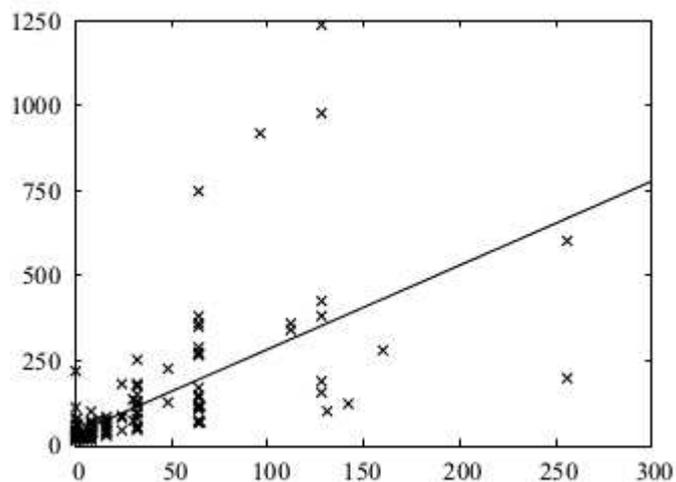
Quando se calcula a raiz quadrada do erro quadrático médio, o erro resultante é denominado de raiz do erro médio quadrado, que exhibe o erro na mesma ordem de grandeza da variável estimada. Existem ainda os erros relativos que relacionam a soma dos erros (absolutos ou quadráticos) de todos os dados com o somatório da diferença entre cada valor da variável e sua média (\bar{y}) (WITTEN; FRANK; HALL, 2011). As equações abaixo exibem o cálculo do erro absoluto relativo e do erro quadrático relativo nessa ordem:

$$\frac{\sum_{i=1}^n |y_i - y'_i|}{\sum_{i=1}^n |y_i - \bar{y}|} \quad (3.19)$$

$$\frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.20)$$

Além dessas funções de erros apresentadas, o coeficiente de correlação também é comumente utilizado para avaliação de um modelo de regressão. Esse parâmetro indica o grau de correlação entre duas variáveis, classificando essa relação em tipo forte ou fraco e positiva ou negativa (LAUBER; FARBER, 2010).

Figura 15 – Relação entre Duas Variáveis.



Fonte: (WITTEN; FRANK; HALL, 2011).

3.6 Apresentação do Conhecimento

A fase final do processo de descoberta de conhecimento é apresentar os padrões obtidos pela etapa de mineração de dados aos usuários finais. Há muitas maneiras diferentes para representar esse conhecimento de acordo com cada tipo de técnica e do tipo de dados que se tem em mãos (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Para dados numéricos, por exemplo, usualmente utiliza-se o método de regressão. Assim, a representação do conhecimento gerado no caso da regressão pode ser feita por meio de equações, como por exemplo, as equações apresentadas na subseção 3.4.3 (WITTEN; FRANK; HALL, 2011).

Outra forma de apresentação de dados numéricos quando se usa a tarefa de regressão é por meio de gráficos que representam a relação linear ou não-linear entre as variáveis consideradas (WITTEN; FRANK; HALL, 2011). A Figura 15 exibe uma relação com regressão linear simples entre duas variáveis quaisquer.

4 SOFTWARES USADOS NA PESQUISA

Este capítulo mostra os *softwares* que foram utilizados na pesquisa. Primeiramente, comentou-se sobre o *MATLAB* (*Matrix Laboratory*), que é um software poderoso com aplicações matemáticas e científicas diversas. Em seguida, são comentadas as principais funcionalidades do *WEKA* (*Waikato Environment for Knowledge Analysis*). Ele contém vários algoritmos oriundos de diferentes abordagens da inteligência artificial, trabalhando principalmente com a aprendizagem de máquinas.

4.1 Funcionalidades Básicas do *MATLAB*

O *MATLAB* é uma ferramenta computacional conhecida mundialmente. Ele pode ser utilizado como linguagem de programação ou como um ambiente de cálculo interativo para solucionar problemas matemáticos, científicos e tecnológicos. O nome *MATLAB* significa *MATrix LABoratory* e representa bem o funcionamento deste *software*, pois todas as variáveis definidas em seu ambiente estão na forma de matrizes. Por causa dessa característica, esse programa é adequado quando se necessita trabalhar com uma grande coleção de dados (GILAT, 2006).

Figura 16 – Tela *Editor* do *MATLAB*.

```

1  % Selmo Eduardo Rodrigues Junior 05/07/2014
2  % Engenharia da Computacao - UEMA
3  % Sao Luis, MA
4
5  % VARIAVEIS DE ENTRADA
6
7  % T = Temperatura do ambiente em graus C;
8  % RH = Umidade relativa media diaria em %;
9  % Pa = Precipitacao em mm;
10 % Ud = Velocidade do vento media diaria em km/h;
11 % RS = Radiacao solar global em cal/cm2;
12
13 % CONSTANTES
14
15 % E = Nivel acima do mar em ft (pes);
16 % Cp = Calor especifico do ar seco em lang/in/F de agua;

```

Fonte: Autoria Própria.

Esse *software* possui aplicações em análise numérica, no cálculo com matrizes, no desenvolvimento de algoritmos, na simulação, na manipulação de uma vasta quantidade de dados e na construção de gráficos científicos e de engenharia. Além das funções mais comuns a várias

outras ferramentas matemáticas, o *MATLAB* disponibiliza um conjunto de *toolboxes*, que são uma coleção de funcionalidades para resolver problemas específicos, por exemplo, processamento de sinais, sistemas de controle e lógica *fuzzy* (GILAT, 2006).

A principal janela desse programa é a Janela de Comando (*Command Window*), que é onde se inicializa as variáveis, executa comandos e programas e onde se gerencia o uso da ferramenta. Outra janela importante é a *Editor* (Figura 16), cuja função é servir como um ambiente onde o usuário pode escrever e editar programas no formato de arquivo *.m*, que é o padrão do *MATLAB*. Além disso, ele ainda fornece uma tela de ajuda, apenas digitando "*help*" na *Command Window* para pesquisar sobre diversos assuntos, exemplos e funcionalidades (GILAT, 2006).

Quando é necessário analisar um conjunto de dados, é preciso importá-lo ao ambiente *MATLAB*, assim como dados gerados por ele devem ser exportados para serem utilizados por outras aplicações. Essa ferramenta é capaz de manipular diversos tipos de informações, como valores numéricos, textos, áudio e imagens. A importação de dados pode ser realizada através de comandos ou por um Assistente de Importação disponível no *MATLAB*. Para importar arquivos no formato *.xls* do *Microsoft Excel*, por exemplo, basta usar o comando *xlsread("nome do arquivo")* e, para exportar dados para esse mesmo formato, usa-se o comando *xlswrite("nome do arquivo", nome da variavel)* (GILAT, 2006).

4.2 O Software WEKA

O *software WEKA* (*Waikato Environment for Knowledge Analysis* ou Ambiente para Análise de Conhecimento de Waikato) é composto por uma coleção de algoritmos de aprendizado de máquina no estado da arte e ferramentas para o pré-processamento de dados. Esse programa foi desenvolvido na Universidade de Waikato, que fica na Nova Zelândia. O *WEKA* disponibiliza um suporte para todo o processo de descoberta de conhecimento, fornecendo um ambiente para preparação de dados de entrada, para avaliação de sistemas de aprendizagem e para visualização dos resultados das aplicações das técnicas. Esse *software* possui uma *interface* que possibilita aos usuários compararem os resultados das diferentes técnicas disponíveis e identificarem qual delas é mais adequada para certa aplicação (MARKOV; RUSSEL, 2006).

Essa ferramenta é implementada em *Java*, é gratuita, distribuída sob a licença *GNU* (*General Public License* ou Licença Pública Geral) e é multiplataforma, ou seja, pode ser instalada em sistemas *Linux*, *Windows* e *Mac OS*. Para importar os dados ao *WEKA*, é preciso convertê-los para o formato de arquivo *.arff* (MARKOV; RUSSEL, 2006).

Existem três usos mais comuns para o *WEKA*. Pode-se usá-lo aplicando um método de aprendizagem sobre uma coleção de dados e analisar sua saída para obter conhecimento a partir desse conjunto. Outra alternativa é desenvolver modelos que possam gerar previsões sobre novas instâncias. E uma terceira aplicação é comparar o desempenho de técnicas distintas para

determinada situação (WITTEN; FRANK; HALL, 2011).

O usuário pode trabalhar com quatro tipos de interfaces diferentes: o *Explorer* (Explorador), o *Knowledge Flow* (Fluxo do Conhecimento), o *Experimenter* (Experimentador) ou por meio da *Command-Line* ou linhas de comandos (WITTEN; FRANK; HALL, 2011). Esses ambientes podem ser escolhidos a partir da tela principal do WEKA, como pode ser observado na Figura 17.

Figura 17 – Tela Inicial do WEKA.



Fonte: Autoria Própria.

O ambiente *Explorer* dá acesso a todas as aplicações do WEKA, orientando o usuário através de menus para que ele siga uma ordem apropriada para trabalhar com essa ferramenta. Nessa *interface*, o usuário envia os dados para o programa no formato *.arff*. Ele pode ainda utilizar filtros para selecionar os registros que devem ser considerados no problema e escolher uma técnica para mineração dos dados dentre muitas que estão disponíveis. Sem deixar de mencionar também sobre a possibilidade de ver os resultados graficamente (WITTEN; FRANK; HALL, 2011).

No entanto, a *interface* do *Explorer* tem a desvantagem de carregar todo o conjunto de dados na memória principal. Dessa forma, o *Explorer* pode ser aplicado a problemas de pequeno a médio porte, a não ser que se tenha a disposição uma máquina com grande capacidade de processamento (WITTEN; FRANK; HALL, 2011).

Uma alternativa para esse caso, é utilizar algoritmos incrementais que processam grandes conjuntos de dados. Para implementar esses algoritmos, o WEKA fornece o ambiente de *Knowledge Flow*, que permite o usuário projetar configurações para o fluxo de processamento de dados. Essa *interface* trabalha com ícones que representam algoritmos de aprendizagem, fontes de dados, ferramentas de pré-processamento, métodos de avaliação e módulos de visualização de conhecimento. Em seguida, os usuários organizam esses ícones na configuração desejada. Com o fluxo de conhecimento organizado, é possível que os dados sejam carregados e processados de forma incremental (WITTEN; FRANK; HALL, 2011).

O terceiro tipo de interface é o *Experimenter*. Ele é adequado para comparar as técnicas de aprendizagem e analisar qual funciona melhor para certa ocasião. Com o ambiente *Explorer* é possível fazer essa tarefa, contudo, a interface *Experimenter* automatiza o processo, facilitando o ajuste de parâmetros dos classificadores e filtros e coletando estatísticas de desempenho. Além desses três ambientes, o *WEKA* disponibiliza ainda uma interface de linha de comando para usuários avançados (WITTEN; FRANK; HALL, 2011).

No menu *Classify* do ambiente *Explorer*, está a disposição uma vasta quantidade de algoritmos que são subdivididos em categorias: classificadores bayesianos, árvores, regras, funções, *lazy classifiers* e os classificadores de várias instâncias (WITTEN; FRANK; HALL, 2011).

Em destaque para esse trabalho, tem-se o algoritmo *LinearRegression*, que está inserido na categoria de funções. As funções representam uma classe de técnicas que inclui uma variedade de classificadores que podem ser representados como equações matemáticas. O algoritmo *LinearRegression* realiza o método dos mínimos quadrados de regressão linear múltipla (WITTEN; FRANK; HALL, 2011).

5 METODOLOGIA DE DESENVOLVIMENTO DO MODELO PARA ESTIMAÇÃO DO VOLUME

Esse capítulo mostra como foi realizado o desenvolvimento do modelo de estimação do volume de água para irrigar a plantação de dendê na Fazenda Escola de São Luís. Primeiramente, comenta-se sobre as particularidades do universo de estudo dessa pesquisa, que é a região da plantação citada. Em seguida, são relatados que tipos de dados devem ser levados em conta nesse trabalho e como esses foram obtidos. Por conseguinte, é mencionada a metodologia de irrigação adotada. Por último, são exibidas todas as etapas do KDD para extração do conhecimento a partir dos dados brutos coletados.

Esse trabalho pode ser classificado quanto a sua finalidade e quanto aos meios de alcançar o seu objetivo. Quanto aos fins, essa pesquisa é do tipo aplicada, pois é motivada pela necessidade de resolver um problema concreto. No caso, deseja-se resolver o problema da decisão da quantidade de água para irrigar a plantação de dendê. Quanto aos meios de investigação, pode ser considerada como um estudo de caso, já que ela é circunscrita à localização e às culturas dessa plantação (VERGARA, 2003).

5.1 Características do Universo de Estudo

O universo de estudo observado por esse trabalho foi a Unidade de Validação Tecnológica (UVT) que está instalada na Fazenda Escola na Universidade Estadual do Maranhão. Ela está localizada na cidade de São Luís, Maranhão, a 35 metros acima do nível do mar, segundo o levantamento aerofotogramétrico de 2003 (AEROCONSULT, 2003), e com coordenadas geográficas de latitude $2^{\circ}31'47''$ Sul e longitude $44^{\circ}18'10''$ Oeste. Essa região se caracteriza por manifestar um clima quente e úmido, com temperatura mínima geralmente entre $20^{\circ}C$ e $23^{\circ}C$ e a máxima entre $29^{\circ}C$ e $31^{\circ}C$. Ela apresenta duas estações distintas: a estação seca (agosto a dezembro) e a estação chuvosa (janeiro a junho), cuja média pluviométrica anual é de 1.500 mm (ALMEIDA, 2013).

Dentre as espécies de plantas presentes na UVT, o dendezeiro foi a cultura de referência para este trabalho. A área da plantação de dendê na UVT é de 120 metros por 70 metros (8400 m^2), contabilizando 120 plantas presentes, com 24 espécies diferentes e 5 para cada variedade. O espaçamento entre as plantas é em formato de um triângulo equilátero, com 8,75 metros em cada lado. Para aplicação de água nessa plantação, usa-se um sistema de irrigação localizada por microaspersão, sendo que os microaspersores são do tipo estacionário. O esquema dessa plantação está presente no Apêndice A.

Algumas constantes apresentadas nas equações introduzidas na seção 2.1 foram medidas ou pesquisadas, baseando-se nas características do universo de estudo. Admitiu-se a constante K_c para o dendê igual a 1, segundo o trabalho de Antonini et al. (2013), que considera plantas

similares e no mesmo período de vida das espécies de dendê presentes na UVT. Em relação a área dessa planta, mediu-se em campo o raio de algumas espécies, sendo obtido o valor médio de 1,525 m, portanto, com área circular de 7,31 m². O fator de cobertura foi adquirido conforme cita a pesquisa de Amorim (2007), onde é apresentado o valor de F_c para um coqueiro, que é uma cultura com características similares ao dendê. Logo, o valor de F_c admitido foi de 0,7 para este caso. Cada planta possui apenas um microaspersor. Dessa forma, o valor de N é igual a 1.

As constantes para o sistema de irrigação localizada por microaspersão dessa plantação também foram definidas. A eficiência de irrigação para um sistema de irrigação deste tipo, de acordo com a Tabela 1, é 85% ($E_a = 0,85$). A vazão dos microaspersores do sistema de irrigação do dendezeiro é de 15 litros por hora. A plantação de dendê vem sendo irrigada somente na estação seca do ano, sendo que a cada três dias, a irrigação funciona por 3 horas.

Conforme o trabalho de Singh (2001), o calor específico do ar seco (c_p) pode ser considerado igual a $1013 \frac{J}{Kg.K}$. A Tabela 3 resume todos os valores das variáveis comentadas nesta seção.

Tabela 3 – Resumo de Valores Obtidos do Universo de Estudo.

Variável	Descrição	Valor
K_c	Constante do dendê	1
E_a	Eficiência do sistema de irrigação localizada por microaspersão	85%
F_c	Fator de cobertura	0,7
A_p	Área do dendê	7,31 m ²
N	Razão entre o número de plantas e o de emissores	1
Q_a	Vazão do microaspersor	15 L/h
E	Altitude média da UEMA, São Luís - MA	35 m
c_p	Calor específico do ar seco	$1013 \frac{J}{Kg.K}$

Fonte: Autoria Própria.

5.2 Coleta dos Dados Brutos

Os dados brutos utilizados para realização deste trabalho foram cedidos pelo Núcleo Geoambiental da Universidade Estadual do Maranhão. O período de tempo desses dados está compreendido entre o mês de Setembro de 2002 a Maio de 2014. Eles são quantitativos, isto é, são representados por valores numéricos. Esses dados contêm valores diários sobre atributos meteorológicos, que são:

- Temperatura máxima, mínima e média diária em °C;

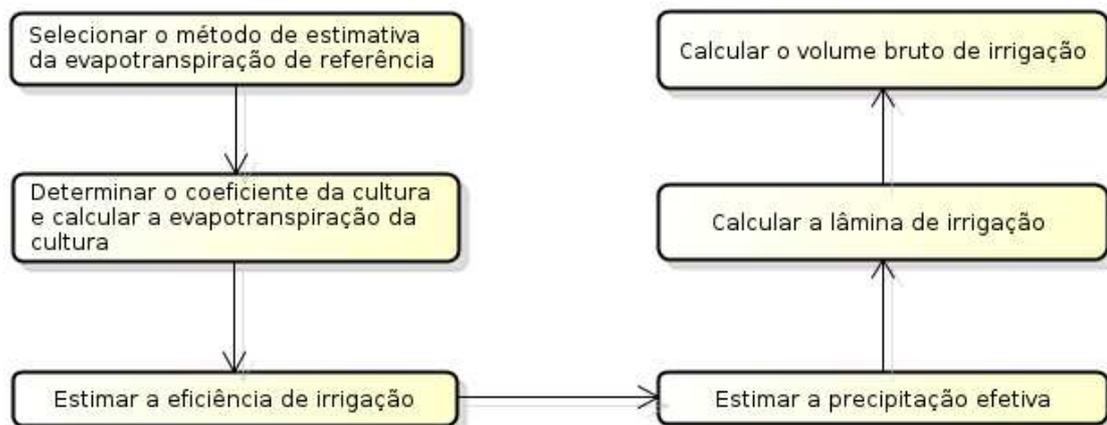
- Umidade relativa do ar máxima, mínima e média diária em %;
- Precipitação diária em *mm*;
- Velocidade do vento média e máxima diária em *km/h*;
- Radiação solar global em *cal/cm²*;
- Pressão atmosférica máxima, mínima e média diária em *atm*;

Há uma quantidade inicial de 4273 registros, contudo, existem campos nulos, ou seja, alguns registros não possuem valores para todos os atributos acima apresentados, além de existirem ainda ruídos e valores discrepantes.

5.3 Metodologia de Irrigação da Pesquisa

Ao observar esses dados meteorológicos e característicos do universo de estudo considerado, percebeu-se que não há informações sobre o volume de água usado na irrigação. A presença da variável volume é essencial para o desenvolvimento do modelo de conhecimento, pois este precisa conhecer como as outras variáveis influenciam na quantidade de água. Dessa forma, baseando-se nesses dados obtidos, utilizou-se o procedimento representado pela Figura 18, recomendado pelo Departamento de Agricultura dos Estados Unidos (USA, 1993), para determinar o volume de água a ser irrigado na plantação de dendê.

Figura 18 – Metodologia de Manejo de Irrigação da Pesquisa.



Fonte: Adaptado de (USA, 1993).

Sendo assim, foram pesquisados métodos em que as variáveis exigidas estão disponíveis e foram coletadas. A técnica da radiação solar (Equação 2.5 a Equação 2.9) e as expressões matemáticas propostas no trabalho de Soares, Costa e Nascimento (2006), que são a Equação 2.10 até a Equação 2.15, foram selecionadas, pois elas se mostraram mais adequadas a essa ocasião.

A técnica da radiação solar é um método indireto que calcula a evapotranspiração de referência (ET_o), enquanto as outras equações determinam o volume de água tendo como ponto de partida o valor de ET_o .

5.4 Preparação dos Dados Brutos

Para adquirir um modelo de conhecimento a partir desses dados coletados, a metodologia do processo de KDD apresentada no Capítulo 3 foi seguida. Em primeiro lugar, realizou-se uma exploração dos dados meteorológicos para aquisição de mais conhecimento sobre os mesmos. Depois disso, seguindo os passos propostos por Fayyad, Piatetsky-Shapiro e Smyth (1996), o objetivo do processo de KDD no ponto de vista do cliente foi destacado, que é determinar o volume de água para irrigar a plantação de dendê na UVT baseando-se nos dados meteorológicos coletados.

O passo seguinte foi localizar os registros que poderiam comprometer a qualidade dos resultados. Portanto, a limpeza dos dados foi iniciada, ou seja, eliminaram-se as inconsistências, assim como os registros com campos nulos.

Em seguida, vem a fase de integração, que é a união de diversas fontes de dados em apenas um repositório consistente. Contudo, essa fase não foi necessária para este caso, pois há apenas uma fonte de dados.

Posteriormente, a seleção dos dados ou *data selection* foi iniciada. Nesse estágio, é analisado que atributos dos dados serão úteis para a ocasião. Em relação aos dados meteorológicos coletados, apenas os atributos relevantes para o problema foram selecionados. Neste caso, os campos escolhidos foram: temperatura média, umidade relativa do ar média, precipitação, velocidade do vento e radiação solar global. Eles foram selecionados por serem necessários para que a metodologia e as equações comentadas na seção 5.3 sejam seguidas.

O próximo passo é começar o estágio de transformação dos dados, cujo objetivo é formatá-los adequadamente em conformidade com o objetivo do pesquisador. Nessa fase, adicionou-se ao conjunto de dados a variável volume, que foi calculada para cada registro, seguindo o procedimento apresentado na seção 5.3. Ela foi definida como a variável alvo, ou seja, aquela que o modelo desenvolvido deve exibir para usuário de acordo com os valores diários dos outros atributos escolhidos na fase de *data selection*.

Para realizar essa transformação, usou-se o *software MATLAB*. O código utilizado para essa fase está disponível no Apêndice B. Ele implementa o método da radiação (Equação 2.5 a Equação 2.9) para determinar a evapotranspiração de referência e as equações 2.10 a 2.15 para calcular o volume bruto de água que deve irrigar a plantação de dendê. Desse modo, armazenaram-se os dados que foram limpos anteriormente na matriz M , para que estes sejam manipulados pelo *MATLAB*. Todas as variáveis e constantes exigidas por esses métodos também foram declaradas no código.

Ademais, o código realiza todas as conversões de unidades para efetuar os cálculos exigidos por esses métodos. Para determinar a precipitação efetiva, o código *MATLAB* seguiu o critério proposto por Azevedo et al. (2008) (subseção 2.4.3), que considera 75% dos valores de precipitações acima de 10 mm e 100% para chuvas abaixo de 10 mm como efetivas.

Logo depois, atualizou-se a matriz *M*, adicionando a variável volume obtida de cada registro do conjunto de dados. Assim, os dados ficaram com os atributos temperatura média, umidade relativa do ar média, precipitação, velocidade do vento, radiação solar global e o volume bruto. A quantidade de dados após esse procedimento foi de 3070 registros.

Para finalizar, o código *MATLAB* exportou todos os registros contidos na matriz *M* para o formato *.arff*, a extensão de arquivo reconhecida pelo *software WEKA* que foi responsável pela etapa de *data mining*. A Tabela 4 apresenta uma amostragem dos dados, exibindo como eles estão formatados depois que a preparação dos dados para a fase de mineração foi finalizada.

Tabela 4 – Amostragem dos Dados Preparados Para Mineração.

Temp. (°C)	Umid. Relativa (%)	Precip. (mm)	V. vento (km/h)	R. Solar (cal/cm ²)	Vol. (L)
24,94	90,5	14,2	0,37	965,56	-18,86
28,18	78	0	6,9	501,9	40,72
26,06	87,5	0,2	1,81	444,54	30,26
26,81	85	0,4	2,5	406,3	27,46
27,37	77	0	2,62	325,04	24,85
23,87	87,5	86,25	2,18	11,95	-520,17

Fonte: Autoria Própria.

Percebe-se pela Tabela 4 que alguns valores para o volume estão negativos. O valor negativo indica que não há necessidade de irrigação, portanto a quantidade de água irrigada na plantação de dendê deve ser nula. Contudo, a mineração foi realizada com esses valores para que os resultados sejam melhores. Depois que o modelo foi gerado, realizou-se a gerência desses resultados para que o volume negativo fosse considerado igual a 0.

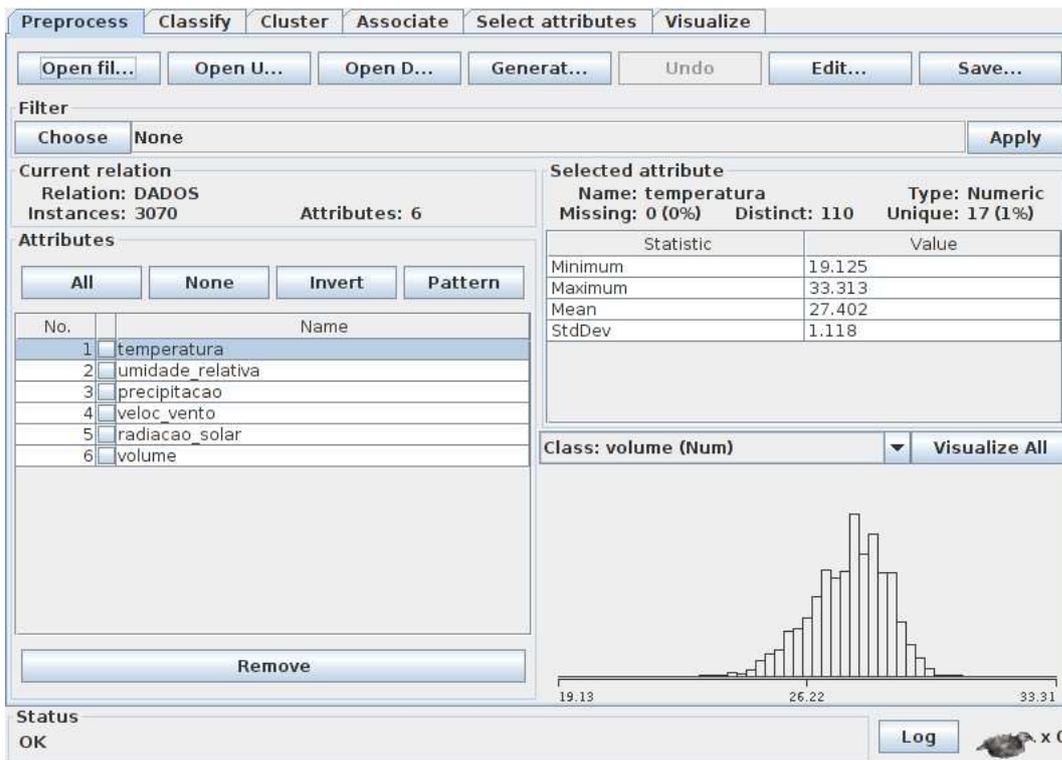
5.5 Mineração dos Dados

O próximo passo foi definir qual tarefa e qual técnica seriam adotados para promover a mineração dos dados. Logo, era preciso analisar os objetivos e as características da pesquisa, verificando quais métodos seriam mais adequados para buscar a solução do problema proposto.

Primeiramente, buscou-se identificar como os valores dos atributos do conjunto de dados estavam formatados, isto é, se eram categóricos ou numéricos. Verificou-se que eles estavam em formato numérico. Outro fator que deve ser analisado é o objetivo da mineração dos dados. Esse

trabalho busca determinar um valor para o volume de água para irrigar a plantação de dendê na UVT, baseando-se nos dados meteorológicos apresentados. Portanto, existe uma variável alvo ou dependente, representada pelo volume. Ambas as características apresentadas confirmam que a estimação é a tarefa mais adequada para essa ocasião (subseção 3.4.1.2).

Figura 19 – Guia *Preprocess* do *WEKA Explorer*.



Fonte: Autoria Própria.

A técnica escolhida foi a regressão linear múltipla, pois ela trabalha com valores numéricos e faz uma estimativa de um resultado de uma variável dependente, com base em valores de variáveis independentes. Em relação aos dados desse trabalho, todas as suas variáveis influenciam no volume de água aplicado na irrigação, logo, todas elas devem ser incluídas na mineração. Assim, a regressão múltipla se apresenta como adequada a ser utilizada nessa ocasião, já que ela se caracteriza por trabalhar com diversas variáveis independentes.

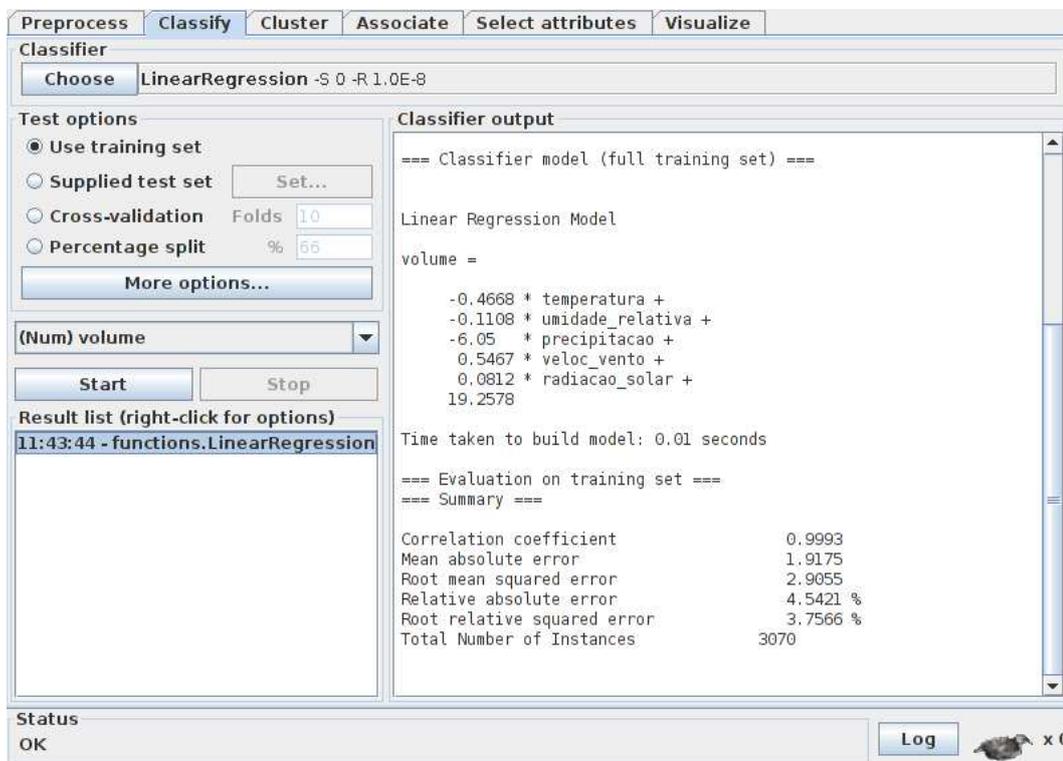
Apesar do problema de determinar a quantidade de água para irrigação ser tipicamente não-linear, escolheu-se a regressão do tipo linear para desenvolver um modelo que aproximasse linearmente o valor do volume de água, reduzindo a complexidade da análise dos resultados.

Com a tarefa e a técnica identificadas, o *software WEKA* foi utilizado para fazer a mineração dos dados. No ambiente *WEKA Explorer* e na aba *Preprocess*, pode-se visualizar algumas informações dos dados que foram importados para o software no formato *.arff*. A tela inicial do *WEKA Explorer* pode ser observada pela Figura 19.

Através dessa figura, é possível identificar os seis atributos, a quantidade de instâncias ou registros (3070), os valores mínimos, máximos, médios, o desvio padrão de cada variável e observar os histogramas de todas elas. No Apêndice C, esses histogramas são apresentados.

Por conseguinte, selecionou-se a guia *Classify*, onde foi realizada a mineração dos dados propriamente dita. A Figura 20 exibe o algoritmo escolhido para implementar a regressão múltipla, que foi o *LinearRegression*, e os resultados do treinamento utilizando esse algoritmo. A tela mostra também a equação gerada (Equação 6.1) para calcular o volume em relação aos outros atributos apresentados. Pode ser percebido ainda alguns parâmetros de avaliação do modelo desenvolvido.

Figura 20 – Guia *Classify* do *WEKA Explorer*.



Fonte: Autoria Própria.

6 RESULTADOS DA PESQUISA

Este capítulo apresenta os resultados da aplicação do processo de KDD sobre os dados meteorológicos cedidos pelo Núcleo Geoambiental da Universidade Estadual do Maranhão. Eles são exibidos por meio de gráficos, tabelas e medidas estatísticas. Finalizando, os resultados dessa pesquisa são discutidos.

6.1 Apresentação dos Resultados

A equação linear resultante da etapa de mineração dos dados promovida pelo *software WEKA* foi:

$$Vol = -0.4668T - 0.1108Ur - 6.05Pr + 0.5467Vv + 0.0812Rs + 19.2578 \quad (6.1)$$

onde Vol é o volume aplicado de água para a plantação de dendê em litros, T é a temperatura em $^{\circ}C$, Ur representa a umidade relativa do ar em %, Pr é a precipitação em mm , a velocidade do vento é Vv em km/h e Rs é a radiação solar global em cal/cm^2 .

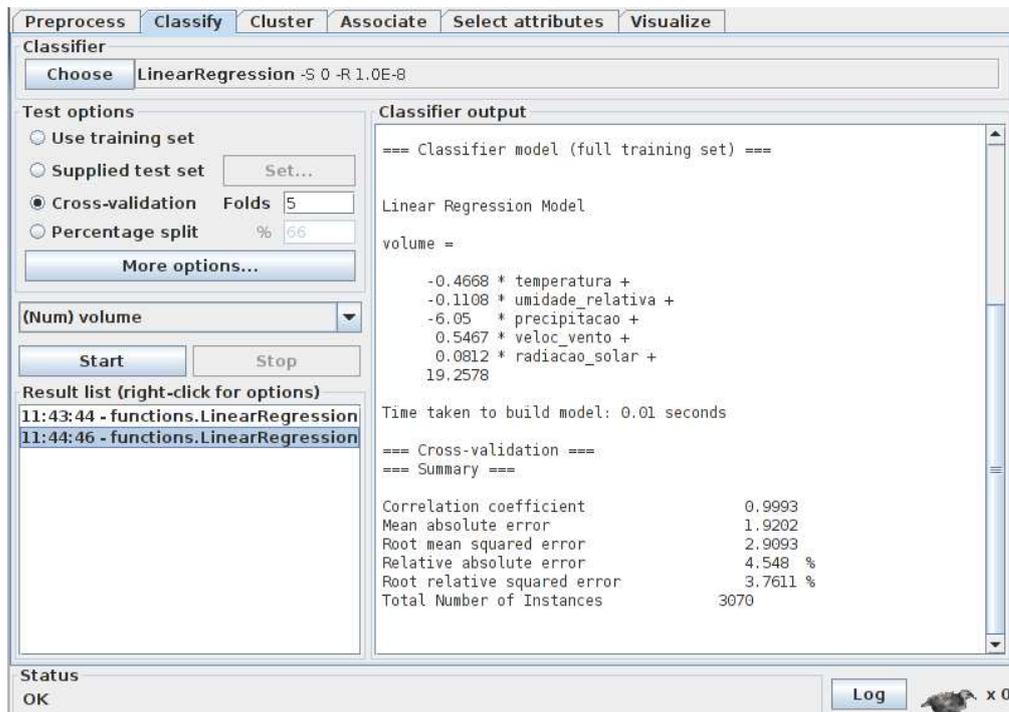
A Figura 20 exibe na janela de saída os resultados do treinamento desse modelo linear. No entanto, foi comentado na seção 3.5 que a avaliação do treinamento não é um bom indicador de desempenho do modelo, pois se utilizou os próprios dados para realização dos testes. Além do mais, não há uma disponibilidade de dados muito grande para segmentá-los em um conjunto para o treinamento e outro para avaliação. Assim, como alternativa, empregou-se a técnica de validação cruzada.

A ferramenta *WEKA* disponibiliza uma funcionalidade para promover essa técnica de avaliação. Essa função foi executada, determinando-se cinco *folds* ou partes de dados, isto é, o conjunto de dados foi dividido em 5 partes. Dessa maneira, o *WEKA* considerou 4 *folds* para realizar o treinamento e um deles para avaliar o modelo. Essa operação foi repetida até cada uma das partes terem sido usadas para efetuar os testes. Os resultados da validação cruzada são exibidos na Figura 21.

Os parâmetros estatísticos resultantes da avaliação do modelo pelo método da validação cruzada (*cross validation*) podem ser melhor visualizados na Tabela 5.

A Figura 22 exibe dois gráficos que comparam o valor do volume usado no treinamento (abscissa) com o volume estimado pela equação (ordenada). O gráfico mais acima mostra o comportamento original do modelo linear desenvolvido. Percebe-se que ele fornece vários valores negativos de volume. Na prática, isso significa que não há necessidade de irrigação para a plantação de dendê. Dessa forma, o modelo foi modificado e melhorado para considerar nulo qualquer valor negativo para a quantidade de água. A região que está abaixo da linha horizontal

Figura 21 – Resultados da Validação Cruzada.



Fonte: Autoria Própria.

Tabela 5 – Parâmetros Estatísticos do Modelo.

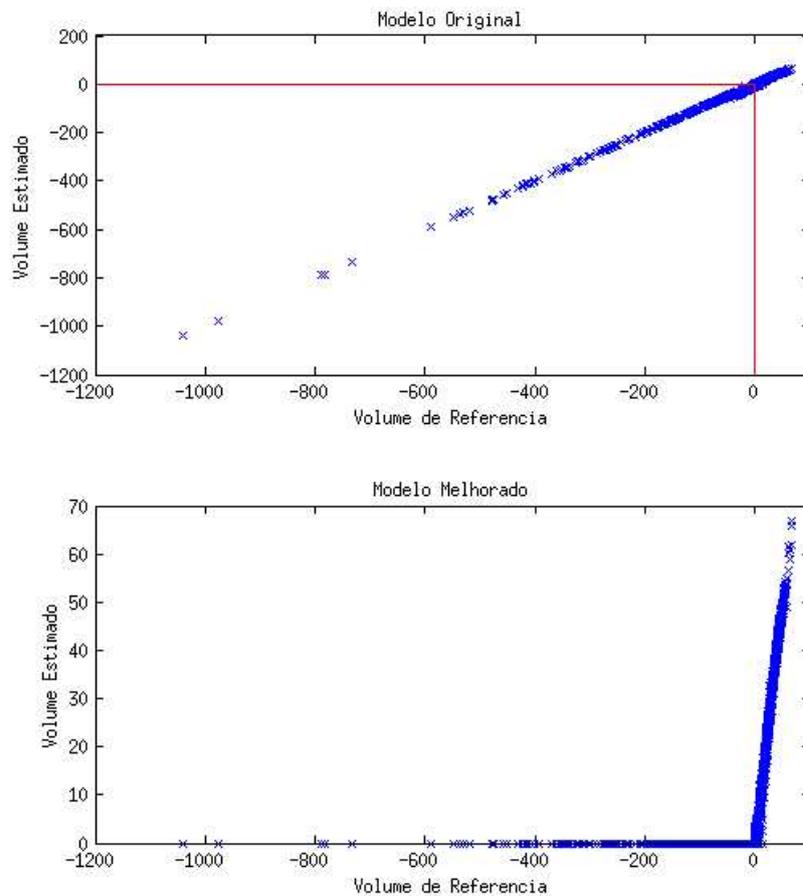
Descrição	Valor
Coefficiente de correlação	0,9993
Erro médio absoluto	1,9202
Erro relativo absoluto	4,548 %
Erro quadrático médio	2,9093
Erro quadrático relativo	3,7611 %

Fonte: Autoria Própria.

vermelha e a esquerda da linha vertical também vermelha compreende os pontos que o modelo indica que não há necessidade de irrigar essa plantação. Assim, o modelo resultante é o segundo gráfico mostrado pela Figura 22.

A Tabela 6 é uma extensão da Tabela 3, que considera os mesmos registros para exibição os resultados. Ela apresenta três medidas para o volume. A primeira variável de volume (ante-penúltima coluna) foi calculada pela metodologia apresentada na seção 5.3 e que foi fornecida como valor de referência para o treinamento do modelo. A penúltima coluna mostra o valor dessa variável que o modelo original desenvolvido representado pelo gráfico superior na Figura 22

Figura 22 – Comparação entre o Volume Fornecido e o Volume Estimado.



Fonte: Autoria Própria.

estimou. Pode-se indentificar que há valores negativos para alguns casos. Já a última coluna da tabela representa os resultados do modelo que foi melhorado, considerando todos os valores negativos como ausência do evento de irrigação para a plantação de dendê na UVT.

6.2 Discussão dos Resultados

Para o especialista utilizar o modelo de conhecimento desenvolvido por essa pesquisa, representado pela Equação 6.1, ele deve consultá-lo diariamente, pois os dados que foram usados para gerar essa fórmula são de condições diárias do universo de estudo. Ele indica a quantidade de água estimada para irrigar a plantação de dendê na UVT com base nas variáveis independentes representadas nessa equação.

Através da Tabela 5, observa-se que o coeficiente de correlação foi obtido com o valor de 0,9993. Pela classificação proposta por Freund (2006), o modelo apresenta uma correlação forte positiva, ou seja, as variáveis possuem uma forte associação linear positiva com o atributo volume. Além disso, os erros estatísticos podem ser visualizados. O erro médio absoluto foi

Tabela 6 – Resultados dos Modelos para uma Amostra dos Dados.

T (°C)	Ur (%)	P (mm)	Vv (km/h)	Rs (cal/cm ²)	Vol. Trein. (L)	Vol. Est. (L)	Vol. Melh. (L)
24,94	90,5	14,2	0,37	965,56	-18,86	-9,71	0
28,18	78	0	6,9	501,9	40,72	41,98	41,98
26,06	87,5	0,2	1,81	444,54	30,26	33,27	33,27
26,81	85	0,4	2,5	406,3	27,46	29,26	29,26
27,37	77	0	2,62	325,04	24,85	25,77	25,77
23,87	87,5	86,25	2,18	11,95	-520,17	-521,22	0

Fonte: Autoria Própria.

calculado com o valor de 1,9202. Esse valor representa o erro médio encontrado comparando-se o volume dos dados fornecidos no treinamento e o volume calculado pela equação linear. Ou seja, esse modelo possui uma diferença média de aproximadamente 2 litros de excesso ou falta de água a ser aplicada no sistema de irrigação. Já o erro relativo absoluto, que representa a razão entre o erro médio absoluto e o valor exato, possui um valor de 4,548 % em relação ao volume dos dados de treinamento. O método dos erros quadráticos também foi aplicado. O erro quadrático médio foi calculado com o valor de 2,9093 litros a mais ou a menos para o volume e o erro quadrático relativo foi 3,7611 % em relação ao volume de treinamento.

Por meio dos resultados, percebe-se que a equação gerada apresenta baixos índices de erros e uma forte correlação positiva entre as variáveis. Contudo, existem ainda outros erros acumulados antes de desenvolver o modelo. Eles são provenientes das medições das variáveis meteorológicas e dos métodos para determinação do volume de irrigação bruto. Esses erros são difíceis de estimar. Ao observar a Figura 22, pode ser analisado que a equação conseguiu representar linearmente de forma aproximada o comportamento da variável volume em relação as variáveis climáticas selecionadas. O modelo foi alterado e melhorado para considerar quantidades negativas de água iguais a 0.

Apesar desses fatores, não se pode considerar essa estimativa do volume de água para irrigação totalmente confiável, pois é necessário analisar seu comportamento na prática, as consequências de sua aplicação para a plantação e compará-lo com outras técnicas de cálculo do volume de água para avaliar o seu nível de confiabilidade.

7 CONSIDERAÇÕES FINAIS

Conhecer a quantidade de água adequada para aplicar em uma plantação é um dos principais objetivos de um projeto de irrigação. Pois, irrigar de forma eficiente, além de economizar recursos hídricos, eleva a produtividade das culturas, tornando-as mais resistentes e saudáveis. Caso a irrigação seja utilizada de forma descuidada, o excesso de água poderá sufocar as plantas e a falta dela possivelmente deixará as culturas mais suscetíveis a doenças. Existem diversas técnicas para determinar o volume de água, porém nem sempre todos os dados estão disponíveis para aplicá-las.

Sendo assim, essa pesquisa trabalhou com base na hipótese de que é possível desenvolver uma alternativa para estimar o volume de água para irrigar certa plantação baseando-se em dados meteorológicos históricos da região onde ela se encontra. Então, o objetivo desse trabalho foi desenvolver um modelo que calcule a quantidade de água de irrigação de uma plantação específica de dendê, localizada na Fazenda Escola de São Luís, Universidade Estadual do Maranhão. Esse modelo foi desenvolvido com base em dados meteorológicos da região, que foram cedidos pelo Núcleo Geoambiental da Universidade Estadual do Maranhão, variáveis medidas em campo, valores pesquisados em outros trabalhos referentes a essa cultura e ao tipo de sistema de irrigação dessa plantação.

Para gerar esse modelo, usou-se a metodologia do processo de descoberta de conhecimento (KDD). Os dados brutos coletados foram pré-processados e preparados para realização do treinamento desse modelo. Na etapa de preparação, foi necessário calcular a variável volume para cada registro do conjunto de dados. Isso porque o volume é a variável que se deseja estimar, portanto, o modelo deve conhecer a relação dela com os outros atributos. Para fazer esse cálculo, foram usados em conjunto o método da radiação solar, indicado pelo Departamento de Agricultura dos Estados Unidos (USA, 1993), e o método para o cálculo do volume citado no trabalho de Soares, Costa e Nascimento (2006) para se obter um valor de referência para o volume no treinamento. Esse treinamento foi efetuado utilizando tarefas e técnicas de mineração de dados.

Para desenvolver o modelo, analisou-se que a tarefa de estimação e a técnica de regressão linear múltipla eram mais adequadas para ocasião. A regressão foi implementada através do algoritmo *LinearRegression* presente no *software WEKA* para mineração dos dados. O resultado da etapa de mineração foi uma equação linear com múltiplas variáveis independentes e uma variável dependente, que é o volume bruto de água para irrigar a plantação de dendê.

Essa equação linear (Equação 6.1) apresentou uma forte correlação positiva entre as suas variáveis, ou seja, elas possuem uma associação linear forte e positiva. Além disso, os erros estatísticos encontrados foram considerados baixos. Entretanto, existem ainda os desvios das medições das variáveis meteorológicas, erros de previsões dessas variáveis e os erros dos

métodos usados nesse trabalho, que são alterações complicadas de se estimar e requerem novas pesquisas.

O modelo desenvolvido apresentou muitos valores negativos para o volume. Logo, ele foi modificado e melhorado para que considerasse qualquer valor negativo como ausência de irrigação. É importante lembrar que esse conhecimento resultante do processo de KDD pode ser aplicado somente para a plantação de dendê da fazenda Escola de São Luís, pois ele foi gerado de acordo com dados específicos da localidade dessa plantação. Porém, caso sejam coletados dados de outras localidades, pode-se utilizar a mesma metodologia dessa pesquisa para desenvolver modelos para outras plantações.

Apesar de essa equação ter apresentado bons resultados analiticamente, é preciso testá-lo em campo para observar as suas consequências para a plantação. É importante esse estudo pois se pode avaliar realmente a precisão do modelo. Portanto, não é aconselhável que esse modelo seja única referência para o especialista determinar a quantidade de água para irrigação. Contudo, ele pode ser utilizado como ferramenta de apoio às suas tomadas de decisão.

Como trabalhos futuros, pode-se fazer um estudo mais minucioso da plantação de dendê, coletando mais variáveis que podem influenciar no volume, desenvolvendo um novo conjunto de dados para treinamento. Pode-se também usar outras metodologias para determinar o volume de água para compor os dados de treinamento. É possível também abordar essa pesquisa de maneira intervencionista, isto é, modificar a realidade de forma participativa através da equação resultante. Esse modelo de conhecimento também pode ser utilizado em conjunto com outros sistemas que contenham atuadores e sensores que colem informações em tempo real para melhorar os resultados da irrigação.

REFERÊNCIAS

AEROCONSULT. *Levantamento Aerofotogramétrico*. São Luís: Elaborado para Prefeitura de São Luís, Cemar e Caema, 2003.

ALBUQUERQUE, P. E. P. de. *Estratégias de Manejo de Irrigação: Exemplos de cálculo*. Sete Lagoas: Circular Técnica 136. Embrapa, 2010. Disponível em: <http://www.cnpms.embrapa.br/publicacoes/publica/2010/circular/Circ_136.pdf>. Acesso em: 10 abr. 2014.

ALLEN, R. G. et al. *Crop evapotranspiration: Guidelines for computing crop water requirements*. Rome: Food and Agriculture Organization of the United Nations, 1998. Disponível em: <http://www.engr.scu.edu/~emaurer/classes/ceng140_watres/handouts/FAO_56_Evapotranspiration.pdf>. Acesso em: 30 mai. 2014.

ALMEIDA, H. J. S. de. *Implantação da Unidade de Validação Tecnológica - UVT de Dendê, na Fazenda Escola de São Luís - FESL - CCA/UEMA*. São Luís: Universidade Estadual do Maranhão, 2013.

AMORIM, J. R. A. de. *A Cultura do Coqueiro*. Aracaju: Embrapa, 2007. Disponível em: <<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Coco/ACulturadoCoqueiro/irrigacao.htm>>. Acesso em: 16 jun. 2014.

ANTONINI, J. C. dos A. et al. *Consumo de água de irrigação e produtividade da palma de óleo cultivadas nas condições edafoclimáticas de savana tropical*. Ribeirão Preto: VII Workshop Agroenergia, 2013. Disponível em: <http://www.infobibos.com/agroenergia/cd/Resumos/ResumoAgroenergia_2013_042.pdf>. Acesso em: 16 jun. 2014.

AZEVEDO, J. A. de et al. *Parâmetros de Irrigação Durante o Período Seco em Plantas Jovens de Dendê Cultivadas no Cerrado*. Planatina, DF: Comunicado Técnico 146. Embrapa, 2008. Disponível em: <www.cpac.embrapa.br/download/1274/t>. Acesso em: 30 mai. 2014.

AZEVEDO, R. *Exportações do agronegócio somaram quase US\$ 100 bilhões em 2013*. 2014. Disponível em: <<http://migre.me/iJA3b>>. Acesso em: 12 abr. 2014.

BLECHER, B. *A tecnologia que domou os desertos*. Revista Globo Rural., 2012? Disponível em: <<http://revistagloborural.globo.com/Revista/Common/0,,ERT292798-18281,00.html>>. Acesso em: 16 abr. 2014.

BORGES, A. C.; MENDIONDO, E. M. *Comparação entre equações empíricas para estimativa da evapotranspiração de referência na Bacia do Rio Jacupiranga*. Campina Grande: Revista Brasileira de Engenharia Agrícola e Ambiental, 2007. Disponível em: <http://www.scielo.br/scielo.php?pid=S1415-43662007000300008&script=sci_arttext>. Acesso em: 25 abr. 2014.

BRASIL. *Água: Um recurso cada vez mais ameaçado*. Ministério do Meio Ambiente, 2005. Disponível em: <http://www.mma.gov.br/estruturas/sedr_proecotur/_publicacao/140_publicacao09062009025910.pdf>. Acesso em: 6 abr. 2014.

CICLO Hidrológico. il. color, 2011. Disponível em: <<http://4.bp.blogspot.com/-jIkI57J6zNM/TrHpC8Ko12I/AAAAAAAAABk/rEzPEQhGp0/s1600/Ciclo+Hidro\%25C3\%25B3gico.bmp>>. Acesso em: 10 abr. 2014.

COELHO-BARROS, E. A. et al. *Métodos de estimação em regressão linear múltipla: aplicação a dados clínicos*. Revista Colombiana de Estatística, 2008. Disponível em: <<http://www.kurims.kyoto-u.ac.jp/EMIS/journals/RCE/body/v31n1a07CoelhoEtAl.pdf>>. Acesso em: 6 jul. 2014.

COELHO, E. F.; FILHO, M. A. C.; OLIVEIRA, S. L. *Agricultura irrigada: eficiência de irrigação e de uso de água*. Cruz das Almas, 2005. v. 7, n. 1.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *From Data Mining to Knowledge Discovery in Databases*. Rhode Island: American Association for Artificial Intelligence, 1996. Disponível em: <<http://128.148.32.110/courses/cs227/archives/2001/groups/custom/papers/1996-Fayyad.pdf>>. Acesso em: 22 jun. 2014.

FREUND, J. E. *Estatística Aplicada: Economia, administração e contabilidade*. 11 ed. Porto Alegre: Bookman, 2006.

GILAT, A. *MATLAB: com aplicações em engenharia*. 2. ed. Porto Alegre: Bookman, 2006.

GOMES, M. A. F. *Água: sem ela seremos o planeta marte de amanhã*. Embrapa Meio Ambiente, 2011. Disponível em: <http://www.cnpma.embrapa.br/down_hp/464.pdf>. Acesso em: 8 abr. 2014.

HAN, J.; KAMBER, M. *Data Mining: Concepts and techniques*. 2. ed. San Francisco: Elsevier, 2006.

JENSEN, M. E.; BURMAN, R. D.; ALLEN, R. G. *Evapotranspiration and irrigation water requirements*. New York: ASCE, 1990.

LAROSE, D. T. *Discovering Knowledge in Data: An introduction to data mining*. New Jersey: John Wiley & Sons, Inc. Hoboken, 2005.

LAUBER, R.; FARBER, B. *Estatística Aplicada: 2. ed.* São Paulo: Pearson Prentice Hall, 2010.

MACIEL, F. C. da S. et al. *Desenvolvimento vegetativo de cultivares de palma de óleo dos 14 aos 34 meses de idade em ecossistemas de Roraima*. Boa Vista: Revista Agroambiente. Centro de Ciências Agrárias, Universidade Federal de Roraima, 2013. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/96000/1/1261-6861-1-PB.pdf>>. Acesso em: 1 jun. 2014.

MARKOV, Z.; RUSSEL, I. *Introduction to the WEKA Data Mining System*. Bologna, Italy: Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education, 2006.

MATSSURA, S. *Produção de dados dobra a cada dois anos, diz consultoria do IDC*. O Globo, 2012. Disponível em: <<http://migre.me/jXl6C>>. Acesso em: 21 jun. 2014.

MEDEIROS, S. de S. et al. *Manejo de Irrigação Utilizando o Tanque de Classe A*. Campina Grande: Ministério da Ciência, Tecnologia e Inovação. Instituto Nacional do Semiárido, 2013. Disponível em: <<http://www.insa.gov.br/wp-content/uploads/2013/07/Documento-técnico-Tanque.pdf>>. Acesso em: 28 abr. 2014.

MELLO, J. L. P.; SILVA, L. D. B. da. *Irrigação*. Rio de Janeiro: Instituto de Tecnologia. Departamento de Engenharia. Universidade Federal Rural do Rio de Janeiro, 2007.

RAMOS, C.; LOBO, F. *Descoberta de Conhecimento em Bases de Dados*. Faro, Portugal: Universidade do Algarve, 2003. Disponível em: <http://www.academia.edu/1501519/Descoberta_de_Conhecimentos_em_Base_de_Dados>. Acesso em: 22 jun. 2014.

SILVA, C. A. da; SILVA, C. J. da. *Avaliação de Uniformidade em Sistemas de Irrigação Localizada*. Uberlândia: Revista Científica Eletrônica de Agronomia, 2005. Disponível em: <http://faef.revista.inf.br/imagens_arquivos/arquivos_destaque/Tm9d5yhlcpezey1x_2013-4-29-15-39-59.pdf>. Acesso em: 17 jun. 2014.

SINGH, R. P. *Introduction to food engineering*. San Diego: Academic Press, 2001. Disponível em: <<http://books.google.com.br/books?id=xXZh9qugyGgC>>. Acesso em: 17 jun. 2014.

SNYDER, R. L. *Equation for evaporation pan to evapotranspiration conversions*. New York: Journal of Irrigation and Drainage Engineering, 1992.

SOARES, J. M.; COSTA, F. F. da; NASCIMENTO, T. *Recomendações básicas para o manejo de água em fruteiras*. Petrolina: Circular Técnica 82. Embrapa, 2006. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/CPATSA/34153/1/CTE82.pdf>>. Acesso em: 26 abr. 2014.

SPITZCOVSKY, D. *Produção de alimentos precisa crescer 70% até 2050*. São Paulo: Revista Planeta Sustentável. Abril, 2013. Disponível em: <<http://migre.me/jHN93>>. Acesso em: 11 abr. 2014.

TANQUE Classe A. il. color, 2011. Disponível em: <<http://jornalagricola.files.wordpress.com/2011/08/tanque-classe-a.jpg>>. Acesso em: 20 abr. 2014.

TESTEZLAF, R. *Irrigação: Métodos, sistemas e aplicações*. Campinas: Universidade Estadual de Campinas. Faculdade de Engenharia Agrícola, 2011.

TURBAN, E.; MCLEAN, E.; WETHERBE, J. *Tecnologia da informação para gestão: transformando os negócios da economia digital*. Porto Alegre: Bookman, 2004.

USA. *Irrigation Water Requirements*. United States Department of Agriculture. National Engineering Handbook. Part 623, 1993. Disponível em: <<ftp://ftp.wcc.nrcs.usda.gov/wntsc/waterMgt/irrigation/NEH15/ch2.pdf>>. Acesso em: 2 jun. 2014.

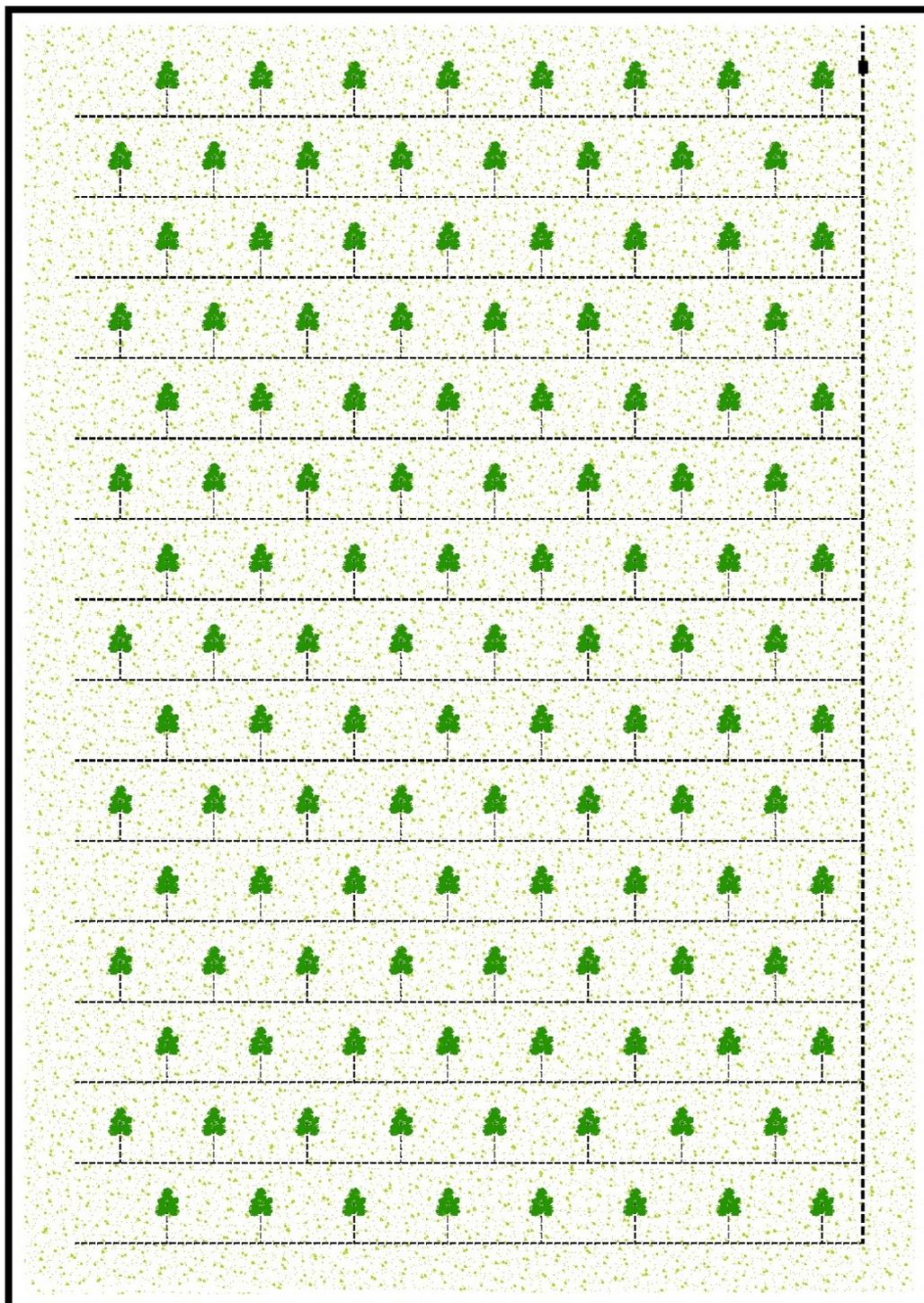
VERGARA, S. C. *Projetos e relatórios em administração*. São Paulo: Atlas, 2003.

WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: practical machine learning tools and techniques*. 3. ed. Burlington: Elsevier, 2011.

Apêndices

APÊNDICE A – PLANTAÇÃO DE DENDÊ

Figura 23 – Esquema da Plantação de Dendê da Fazenda Escola de São Luís, UVT.



Fonte: Autoria Própria.

APÊNDICE B – CÓDIGO MATLAB

```

% Selmo Eduardo Rodrigues Junior 05/07/2014
% Engenharia da Computacao – UEMA
% Sao Luis , MA

% VARIAVEIS DE ENTRADA

% T = Temperatura do ambiente em graus C;
% RH = Umidade relativa media diaria em %;
% Pa = Precipitacao em mm;
% Ud = Velocidade do vento media diaria em km/h;
% RS = Radiacao solar global em cal/cm2;

% CONSTANTES

% E = Nivel acima do mar em ft (pes);
% Cp = Calor especifico do ar seco em lang/in/F de agua;
% (in = polegadas)

% VARIAVEIS METODO RADIACAO

% delta = Inclinação da curva de pressão de vapor em mb/F;
% lambda = Calor de vaporização da água em lang/in;
% BP = Pressão barométrica em mb;
% gamma = Constante psicrométrica em mb/F;
% ETou = Evapotranspiração de referência em in/d;

% CONSTANTES DA SITUACAO PROBLEMA

% Kc = Constante da cultura;
% Perc = Percolação profunda em mm/dia;
% Ea = Eficiência de irrigação em %;
% Fc = Fator de cobertura do solo em %;
% Ap = Área da planta em metro quadrado;
% N = Razão entre o número de plantas e o número de
% microaspersores;
% Qa = Vazão do microaspersor em L/h;

% VARIAVEIS PARA DETERMINACAO DO VOLUME

% ETc = Evapotranspiração da cultura em mm/dia
% Lb = Lamina bruta de irrigação em mm
% Pe = Precipitação efetiva em mm
% Vb = Volume de água bruto aplicado em L

```

% LEITURA DOS DADOS

clear all

format bank

dir = '/home/selmo/dadosPreprocessados.xls';

M = xlsread (**dir**);

% DEFINICAO CONSTANTES DO METODO RADIACAO

Eu = 35; *% nivel acima do mar para UEMA, 35 metros*

E = Eu*3.2808399; *% Conversao para pes (ft)*

cpu = 1013; *% calor especifico do ar seco*

cp = cpu/5527.62; *% conversao para lang/in/F*

% DEFINICAO CONSTANTES PARA IRRIGACAO

Kc = 1; *% constante do dende*

Ea = 0.85; *% eficiencia do sistema de microirrigacao*

Fc = 0.7; *% fator de cobertura*

Ap = 7.31; *% area da planta*

N = 1; *% razao entre plantas e emissores*

Qa = 15; *% vazao em L/h do microaspersor*

% CALCULO DA ETO PELO METODO RADIACAO

for L = 1:length(M)

% Obter temperatura e converter para Fahrenheit

T = M(L,1)*1.8 + 32;

% Obter Umidade Relativa

RH = M(L,2);

% Obter Precipitacao

Pa = M(L,3);

% Obter velocidade do vento e converter para mi/d

Ud = M(L,4)*14.9129086;

% Obter radiacao solar

Rs = M(L,5);

% Metodo pela Radiacao Solar

```

delta = 0.051*((164.8 + T)/157)^7;
lambda = 1543 - 0.796*T;
BP = 1.013*(1 - E/145350)^(5.26);
gamma = cp*(BP/(0.622*lambda));
br = 1.06 - 0.0013*RH + 8.38*10^(-4)*Ud ...
      - 3.73*10^(-6)*RH*Ud - 0.315*10^(-4)*RH^2 ...
      - 3.82*10^(-7)*Ud^2;
ETou = -0.012 + (delta/(delta + gamma))*br*(Rs/lambda);

ETo = ETou*25.4; % Converter para mm/dia

% Calcular Evapotranspiracao da cultura
ETc = Kc*ETo;

% Estimativa para precipitacao efetiva (Secao 2.4.3)
if (Pa > 10)
    Pe = Pa;
else
    Pe = 0.75*Pa;
end

Lb = (ETc - Pe)/Ea; % Lamina Irrigacao

Vb = Lb*Fc*Ap; % Calcular volume de agua

M(L,6) = Vb; % inserir na matriz de dados
end

% EXPORTAR PARA ARQUIVO .ARFF

arquivo = fopen('fonte.arff','wt');

fprintf(arquivo, '@relation_ 'DADOS' _\n\n');

fprintf(arquivo, '@attribute_temperatura_numeric_\n');
fprintf(arquivo, '@attribute_umidade_relativa_numeric_\n');
fprintf(arquivo, '@attribute_precipitacao_numeric_\n');
fprintf(arquivo, '@attribute_veloc_vento_numeric_\n');
fprintf(arquivo, '@attribute_radiacao_solar_numeric_\n');
fprintf(arquivo, '@attribute_volume_numeric_\n\n');

fprintf(arquivo, '@data\n');

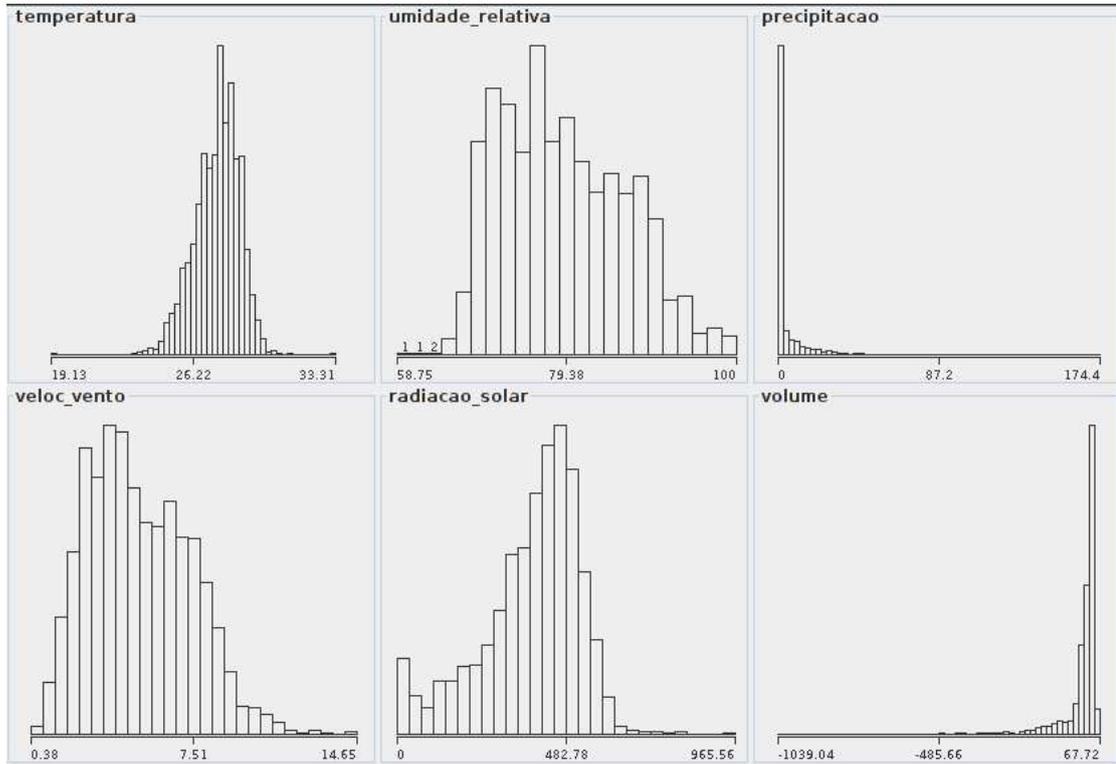
for L = 1:length(M)
    fprintf(arquivo, '%f, _%f, _%f, _%f, _%f, _%f\n', M(L,1), ...
            M(L,2), M(L,3), M(L,4), M(L,5), M(L,6));
end

fclose(arquivo);

```


APÊNDICE C – HISTOGRAMAS DOS DADOS

Figura 24 – Histogramas dos Dados Meteorológicos no WEKA



Fonte: Autoria Própria.