

MESTRADO

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
GOIANO
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
MESTRADO PROFISSIONAL EM OLERICULTURA

USO DA TÉCNICA DE MINERAÇÃO DE DADOS PARA
ESTUDO DA INFLUÊNCIA DE FATORES METEOROLÓGICOS
SOBRE A PRODUÇÃO DE TOMATE PARA PROCESSAMENTO
INDUSTRIAL

Autor: Hiury Luiz dos Santos

Orientador: Prof. Dr. Nadson de Carvalho Pontes

Coorientador(a): Prof^ª. Dra. Leila Roling Scariot da Silva

HIURY LUIZ DOS SANTOS

2019

MORRINHOS-GO
2019

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
GOIANO – CAMPUS MORRINHOS
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
MESTRADO PROFISSIONAL EM OLERICULTURA

USO DA TÉCNICA DE MINERAÇÃO DE DADOS PARA
ESTUDO DA INFLUÊNCIA DE FATORES
METEOROLÓGICOS SOBRE A PRODUÇÃO DE TOMATE
PARA PROCESSAMENTO INDUSTRIAL

Autor: Hiury Luiz dos Santos
Orientador: Prof. Dr. Nadson de Carvalho Pontes
Coorientador(a): Prof^a. Dra. Leila Roling Scariot da Silva

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
GOIANO – CAMPUS MORRINHOS
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
MESTRADO PROFISSIONAL EM OLERICULTURA

USO DA TÉCNICA DE MINERAÇÃO DE DADOS PARA
ESTUDO DA INFLUÊNCIA DE FATORES
METEOROLÓGICOS SOBRE A PRODUÇÃO DE TOMATE
PARA PROCESSAMENTO INDUSTRIAL

Autor: Hiury Luiz dos Santos

Orientador: Prof. Dr. Nadson de Carvalho Pontes

Coorientador(a): Prof^a. Dra. Leila Roling Scariot da Silva

Dissertação apresentada, como parte das exigências para obtenção do título de MESTRE EM OLERICULTURA, ao Programa de Pós-Graduação em Olericultura do Instituto Federal de Educação, Ciência e Tecnologia Goiano – Campus Morrinhos - Área de Concentração Olericultura.

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas – SIBI/IF Goiano Campus Morrinhos

S237u Santos, Hiury Luiz dos.

Uso da Técnica de mineração de dados para estudo da influência de fatores meteorológicos sobre a produção de tomate para processamento industrial. / Hiury Luiz dos Santos. – Morrinhos, GO: IF Goiano, 2018.
36 f. : il.

Orientador: Dr. Nadson de Carvalho Pontes.

Coorientadora: Dra. Leila Roling Scariot da Silva.

Dissertação (mestrado) – Instituto Federal Goiano Campus Morrinhos, Programa de Pós-Graduação Mestrado Profissional em Olericultura, 2018.

1. Tomate. 2. Mineração de dados (Computação). 3. Produtividade agrícola. I. Pontes, Nadson de Carvalho. II. Silva, Leila Roling Scariot da. III. Instituto Federal Goiano. IV. Título.

CDU 635.64

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM OLERICULTURA

USO DA TÉCNICA DE MINERAÇÃO DE DADOS PARA
ESTUDO DA INFLUÊNCIA DE FATORES
METEOROLÓGICOS SOBRE A PRODUÇÃO DE TOMATE
PARA PROCESSAMENTO INDUSTRIAL

Autor: Hiury Luiz dos Santos
Orientador: Nadson de Carvalho Pontes

TITULAÇÃO: Mestre em Olericultura-Área de Concentração em Manejo
Fitossanitário em Olerícolas.

APROVADO em 13 de dezembro de 2018



Prof. Dr. Nadson de Carvalho Pontes
Presidente da Banca



Prof. Dr. Cícero José da Silva
Avaliador Externo
Instituto Federal Goiano - Campus Morrinhos



Prof^a. Dr^a. Leila Roling Scariot da Silva
Avaliadora Externa
Instituto Federal Goiano - Campus Morrinhos

AGRADECIMENTOS

A Deus pelo dom da vida e por me dar saúde para encarar todos os desafios.

À minha família, pelo apoio incondicional durante toda minha vida e nesta trajetória da pós-graduação.

Ao Instituto Federal Goiano Campus Morrinhos, em especial à Coordenação do Programa de Pós-graduação em Olericultura, pela oportunidade de ingresso e de conclusão do curso.

Aos professores Nadson de Carvalho Pontes e Leila Roling Scariot da Silva, pela orientação e apoio no desenvolvimento deste trabalho.

Ao Prof. Cícero José da Silva, pelo auxílio no processamento dos dados climatológicos.

À Empresa Conservas Olé, por ter disponibilizado todos os dados de áreas de produção de tomate para processamento.

A todos os professores do Programa de Pós-graduação em Olericultura, pelos ensinamentos repassados ao longo do curso.

Aos meus colegas de curso, pelo companheirismo e por todo o auxílio durante o curso de mestrado.

A todos que, diretamente ou indiretamente, contribuíram para a execução deste trabalho, dedico.

BIOGRAFIA DO AUTOR

Hiury Luiz dos Santos, filho de José Dias dos Santos e Uilda Luiza dos Santos, nasceu em 25 de outubro de 1976, na cidade de Morrinhos-GO.

Em 2001, graduou-se em Análise de Sistemas pela Universidade Salgado de Oliveira.

Em 2004, especializou em Administração de Sistema de Informação pela Universidade Federal de Lavras - UFLA.

Em setembro de 2016, iniciou o curso de Mestrado Profissional em Olericultura no Instituto Federal de Educação, Ciência e Tecnologia Goiano – Campus Morrinhos.

ÍNDICE

	Página
RESUMO.....	iv
ABSTRACT.....	v
1. INTRODUÇÃO GERAL.....	1
2. REVISÃO DE LITERATURA.....	2
2.1 A cultura do tomateiro	2
2.2 Produção de tomate para processamento industrial	3
2.3 Influência de fatores climáticos sobre a produção de tomate industrial	3
2.4 Mineração de dados como ferramenta de estudo	4
2.5 Referências bibliográficas	6
3. CAPÍTULO I	8
RESUMO.....	8
ABSTRACT.....	9
3.1 Introdução	10
3.2 Material e Métodos	12
3.2.1 Descrição dos algoritmos utilizados.....	15
3.3 Resultados e Discussão	15
3.4 Conclusão.....	24
3.5 Referências bibliográficas	25
4. CONCLUSÃO GERAL.....	27

RESUMO

SANTOS, HIURY LUIZ. Instituto Federal Goiano Campus Morrinhos, dezembro 2018.
Uso da técnica de mineração de dados para estudo da influência de fatores meteorológicos sobre a produção de tomate para processamento industrial.
Orientador(a): Prof. Dr. Nadson de Carvalho Pontes. Coorientador(a): Prof^a. Dra. Leila Roling Scariot da Silva

O Tomate industrial é uma das hortaliças de maior importância econômica no estado de Goiás e no Mundo. O objetivo deste trabalho é comprovar, pelo método de mineração de dados, quais variáveis que influenciam na classificação do ciclo do cultivo, produtividade, qualidade e teor de sólidos solúveis (°Brix) do tomate industrial. Foram analisadas as médias por ciclo do cultivo das variáveis: temperatura, precipitação pluviométrica e umidade relativa do ar. Outra variável analisada foi a Altitude do cultivo.

Com a classificação através da mineração de dados, foi possível concluir que as variáveis temperatura, precipitação e umidade, durante o ciclo do cultivo, impactam na produtividade do tomate industrial no município de Morrinhos, estado de Goiás, conforme os outros estudos já consolidados nesta área.

PALAVRAS-CHAVE: crescimento, mineração de dados, produtividade, qualidade, temperatura

ABSTRACT

SANTOS, HIURY LUIZ. Instituto Federal Goiano (Goiano Federal Institute) Morrinhos Campus, December 2018. **Data mining technique use to study the influence of meteorological factors on tomato production for industrial processing.** Advisor: Prof. Dr. Pontes, Nadson de Carvalho. Co-advisor: Prof Dr. Silva, Leila Roling Scariot da.

Industrial tomato is one of the most important vegetables economically in the state of Goiás, Brazil, and worldwide. This study aims to demonstrate which variables influence on the crop cycle classification, productivity, quality, and soluble solids content (°Brix) of the industrial tomato by data mining method. Temperature, rainfall, and relative humidity average per crop cycle was analyzed, as well as. the cultivation altitude. On the basis of data mining classification, it was possible to conclude that temperature, precipitation, and humidity during the crop cycle impact on the industrial tomato productivity in the municipality of Morrinhos, Goiás State, Brazil, is according to other studies already consolidated in this field.

Keywords: data mining, growth, productivity, quality, temperature

1. INTRODUÇÃO GERAL

O tomateiro (*Solanum lycopersicum*) é uma planta nativa da região andina, desde o Chile, Colômbia, Equador, Bolívia e Peru, sendo esta região considerada seu centro de origem do tomate, porém sua domesticação teve início no México (ALVARENGA., 2013). Pertence à família Solanaceae, sendo atualmente a segunda hortaliça com maior importância econômica entre as hortaliças no Brasil, (LUZ et al., 2010). É uma das culturas mais importantes, apresentando, sua produção, tendência de crescimento (JUNIOR et al., 2015).

Classificada como um dos mais importantes produtos do agronegócio em nível nacional e mundial (VILELA et al., 2012), o tomate é uma hortaliça presente diariamente na dieta alimentar da maioria da população brasileira. Em 2016, foi a segunda hortaliça mais produzida no Brasil. O último levantamento sobre a área cultivada no país mostrou que Goiás produziu mais de 978 mil toneladas em uma área superior a 12 mil hectares, correspondendo a 26,17% da produção nacional, com produtividade superior à 81t/ha, com isso Goiás lidera o ranking nacional como maior produtor desta olerícola no país (IBGE, 2017).

Boa parte da produção goiano de tomate é destinada ao processamento industrial, sendo o estado responsável por mais de 70% da produção nacional neste segmento. Quando a produção de tomate é para fim industrial, além da produtividade, outros fatores são importantes, como o teor de sólidos solúveis, o ciclo da cultura e a coloração dos frutos. Fatores climáticos como temperatura, umidade relativa e pluviosidade podem interferir na produção e na qualidade do tomate. Deste modo, o presente trabalho visa, pela técnica de mineração de dados, a avaliar o efeito de fatores climáticos sobre os componentes de produção do tomate industrial.

2. REVISÃO DE LITERATURA

2.1 A cultura do tomateiro

Nativo da região andina, o tomateiro (*Solanum lycopersicum*) é uma planta pertencente à família Solanaceae, sendo atualmente a segunda hortaliça com maior importância econômica entre as hortaliças no Brasil, (LUZ et al., 2010). Apesar do seu centro de origem compreender uma região que abrange Chile, Colômbia, Equador, Bolívia e Peru, sua domesticação teve início no México (ALVARENGA, 2013).

O tomateiro é uma das culturas mais importantes e vinha apresentando tendência ao crescimento da produção (JUNIOR et al., 2015). Entretanto, segundo o IBGE (2018), a produção brasileira de tomate deve ficar em 4,2 milhões de toneladas, 2,7% menor. Esta produção se concentra nos estados de Goiás (32,2% de participação no total nacional), São Paulo (19,1%), Minas Gerais (17,6%) e Espírito Santo (4,1%). Em setembro, os maiores declínios nas estimativas de produção foram verificados em São Paulo (12,4%), Alagoas (22,6%), Mato Grosso (4,4%) e Ceará (1,5%). Em relação a 2017, a produção de tomate apresentou redução de 3,0%, reflexo da retração de 3,5% na área a ser colhida, apesar de ligeiro aumento de 0,6% no rendimento médio.

O tomate pode ser destinado ao consumo *in natura*, ou para o processamento industrial, utilizado como matéria-prima para as indústrias processadoras de derivados. A finalidade a que se destina é de fundamental importância na escolha das variedades e do sistema de produção, haja vista que cada segmento tem demandas distintas. No caso do estado de Goiás, maior produtor nacional, apesar da sua forte contribuição para a produção de tomate *in natura*, é no segmento para processamento que o estado se destaca, sendo responsável por mais de 70% da produção nacional.

2.2 Produção de tomate para processamento industrial

Classificado como um dos mais importantes produtos do agronegócio, em nível nacional e mundial, o tomate para processamento corresponde à principal atividade geradora de renda para um grande número de produtores, tornando-se significativa fonte de renda regional (VILELA et al., 2012). Além disso, a cultura do tomate para processamento industrial movimentou indústrias paralelas de insumos, compreendendo embalagens, máquinas agrícolas e equipamentos de irrigação.

Além de níveis adequados de produtividade, alguns fatores são importantes para a qualidade da matéria-prima, como, por exemplo, o teor de sólidos solúveis, integridade e coloração de frutos. Quanto maior o teor de sólidos solúveis (°Brix), maior será o rendimento industrial e menor o gasto de energia no processo de concentração da polpa. Em termos práticos, para cada aumento do °Brix na matéria-prima, há um incremento de 20% no rendimento industrial. Obtém-se também a estimativa do rendimento de polpa por meio da seguinte fórmula: $P(\text{t/ha de polpa}) = [(\text{produção}(\text{t/ha}) \times 0,95) \times \text{°Brix do suco}] / 28$.

O teor de sólidos solúveis no fruto, além de ser uma característica genética da cultivar, é influenciado pela adubação, temperatura e irrigação. Os valores médios de °Brix na matéria-prima recebida pelas indústrias no Brasil têm sido bastante baixos (4,5 °Brix). Entretanto, existem cultivares que têm maior potencial genético, apresentando, em determinadas condições, valores próximos de 6,0 °Brix.

A cor é um parâmetro essencial para classificar o produto industrializado. O fruto deve apresentar cor vermelha intensa e uniforme, externa e internamente. Tomates com boa coloração apresentam teores de licopeno (pigmento responsável pela coloração vermelha) na faixa de 5 a 8 mg/100 gramas de polpa. Algumas cultivares apresentam "ombro verde" em razão da maturação tardia da região superior do fruto. Os tecidos nessa região podem ficar endurecidos e amarelados. Para melhor observar a cor e avaliar a uniformidade de maturação, devem ser feitos cortes transversais na região do ombro e na base do fruto.

2.3 Influência de fatores climáticos sobre a produção de tomate industrial

É sabido que fatores climáticos como temperatura, umidade relativa do ar, índice pluviométrico e o fotoperíodo podem interferir na produção e qualidade do tomate.

Segundo Alvarenga, temperaturas elevadas (acima de 32°C) favorecem a síntese de caroteno em detrimento do licopeno, causando manchas amareladas e maturação desuniforme. Tal fato, por si só, já seria indesejável para o tomate destinado à indústria, por interferir na coloração da polpa, reduzindo sua qualidade.

A temperatura ideal para a síntese do licopeno está na faixa de 20 a 24°C.

De maneira geral, a cultura do tomateiro se desenvolve bem em uma faixa de temperatura que vai de 15 a 25°C. Temperaturas acima desse máximo são desfavoráveis, resultando em abortamento, alteração de cor e problemas fisiológicos.

A umidade relativa é outro ponto de extrema importância para a cultura do tomateiro. Sabe-se que, em regiões com alta umidade relativa, há maior multiplicação de fungos e bactérias fitopatogênicos. Nestas regiões, a formação do orvalho e sua duração podem predispor as plantas à ocorrência de doenças. O tomateiro deve ser cultivado em ambiente com umidade variando entre 50 a 70% da umidade relativa.

O excesso de chuvas é um outro problema. Em geral, o cultivo de tomate para processamento industrial no Estado de Goiás está limitado ao período de fevereiro a outubro de cada ano agrícola. O ideal seria o cultivo entre os meses de maio a setembro, evitando a exposição das lavouras às chuvas. Entretanto, como a capacidade de processamento das indústrias é limitada, os plantios precisam ser distribuídos ao longo de todo esse período, havendo cultivos que sofrem com as chuvas de fevereiro e março em seu início, ou aqueles cuja colheita é feita no início das chuvas, em outubro. Quando da ocorrência de altos níveis de precipitação, as ocorrências de doenças aumentam e, com elas, as perdas.

2.4 Mineração de dados como ferramenta de estudo

A crescente geração de dados possibilitada pelos avanços tecnológicos na era da informação vem criando o desafio da obtenção de conhecimento tendo como referência grandes quantidades de dados. Com o aumento do volume de dados na era da informação e a facilidade de armazenamento, tornaram-se imprescindíveis a análise e a recuperação de informação, útil em meio à saturação de dados sem importância. Contudo, procedimentos convencionais de classificação digital se deparam com um grande volume de dados para avaliar e encontrar uma relação entre eles.

Nesse sentido, a mineração de dados surge como ferramenta capaz de elucidar os padrões implícitos nessas grandes massas de dados. Mineração de Dados, também

conhecida como “Descoberta de Conhecimento a partir de Dados” (“Knowledge Discovery from Data”, KDD), pode ser definida como a extração de padrões que representam os conhecimentos armazenados em grande banco de dados (HAN et al., 2011). A área da mineração de dados surgiu no final da década de 80 e vem se tornando uma das principais práticas para análise de dados. Para tanto, são utilizados diversos algoritmos, cada qual objetivando um modo de descobrir conhecimento.

A descoberta de conhecimento em bancos de dados é um processo não trivial de identificar padrões válidos, não conhecidos, potencialmente úteis e interpretáveis, consistindo, basicamente, em descobrir conhecimento nos dados armazenados, pela aplicação de técnicas de mineração de dados, da aplicação dos padrões obtidos e da interpretação dos resultados (FAYYAD et al., 1996). De acordo com Chinchuluun et al. (2010), o uso de técnicas de mineração de dados é algo recente nas áreas de agrícola e ambiental.

Para a cultura de milho, a água é um fator limitante, principalmente na fase de germinação e nos períodos de floração e enchimento de grãos. Desta forma, conhecer os elementos climáticos é uma estratégia para definir manejos que possibilitem minimizar riscos climáticos, contribuindo para melhorar a política de alocação de recursos para crédito agrícola e para o aumento de produção (SANS et al. 2001).

A aplicação de técnicas de mineração de dados em uma base de dados da produção agrícola tem como finalidade de criar recursos para obter informações que auxiliem no planejamento agrícola (TAKAMOTO et al. 2011).

As técnicas de mineração de dados são apropriadas para descobrir conhecimento em grandes conjuntos de dados, e sua aplicação nos dados referentes ao processo de produção da cana-de-açúcar pode ser eficaz para aumentar a produtividade dessa cultura. (TOMAZELA et al. 2015).

2.5 Referências bibliográficas

ALVARENGA, M.A.R. Tomate: Produção em Campo, Casa de Vegetação e Hidroponia. 2º ed. Lavras: Ed. Lavras, 2013. p. 13-14.

ALVARENGA, M.A.R. Tomate: Produção em Campo, Casa de Vegetação e Hidroponia. 2º ed. Lavras: Ed. Lavras, 2013. p. 33-37.

CHINCHULUUN, A.; XANTHOPOULOS, P.; TOMAINO, V.; PARDALOS, P. M. Data Mining techniques in agricultural and environmental sciences. International Journal of Agricultural and Environmental Information System. Hershey, v.1, n.1, p. 26-40, Jan./June 2010. Disponível em: <http://www.academia.edu/1769745/Data_Mining_Techniques_in_Agricultural_and_Environmental_Sciences>. Acesso em: 23 de novembro de 2018.

ELMASRI, RAMEZ. Conceito de Data Mining. IN: Sistema de Banco de Dados. Revisor técnico Luis Ricardo de Figueiredo. São Paulo: Pearson Addison Wesley, 2005. p. 624-645.

FAYYAD, U.; PIATESKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. Advances in knowledge discovery and data mining. Cambridge: MIT Press, 1996. 560 p.

GOOGLE EARTH. Google Earth PRO. Version 7.3.2, 2018. Morrinhos-Goiás. Disponível em: https://www.google.com/intl/pt-BR_ALL/earth/versions/ Acesso em: 07 de dezembro de 2018.

HAN, J.; KAMBER, M. Data mining: concepts and techniques. 3º ed. Amsterdam: Elsevier, 2011. 744 p.

IBGE 2018. IBGE prevê safra 6% menor que a de 2017. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/22780-em-setembro-ibge-preve-safra-6-menor-que-a-de-2017>> Acesso em 23 de novembro de 2018.

EMBRAPA HORTALIÇAS SISTEMA DE PRODUÇÃO 2003. Cultivo de tomate para industrialização. Disponível em <<https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Tomate/TomateIndustrial/cultivares.htm>> Acesso em 23 de novembro de 2018

JUNIOR ARS; RIBEIRO WM; NASCIMENTO AR; SOUZA CB. 2015. Cultivo do tomate industrial no estado de Goiás: Evolução das áreas de plantio e produção. *Conjuntura econômica goiana* 34: 107.

LUZ JMQ; BITTAR CA; QUEIROZ AA; CARREON R. 2010. Produtividade de tomate 'Débora Pto' sob adubação organomineral via foliar e gotejamento. *Horticultura Brasileira* 28: 489-494.

MELO, Paulo César T. de; VILELA, Nirlene J.. Desempenho da cadeia agroindustrial brasileira do tomate na década de 90. *Hortic. Bras.* vol.22 no.1 Brasília Jan./Mar. 2004.

VILELA NJ; MELO PCT; BOITEUX LS; CLEMENTE FMVT. 2012. Perfil socioeconômico da cadeia agroindustrial no Brasil. In: CLEMENTE FMVT; BOITEUX LS. (eds). Produção de tomate para processamento industrial. Brasília: Embrapa Hortaliças. p. 17.

SANS Luiz Marcelo Aguiar, ASSAD Eduardo Delgado, GUIMARÃES Daniel Pereira, AVELLAR Gisela. 2001. Zoneamento de riscos climáticos para a cultura de milho na Região Centro-Oeste do Brasil e para o estado de Minas Gerais. Disponível em: <<http://www.cnpt.embrapa.br/pesquisa/agromet/pdf/revista/cap17.pdf>>. Acesso em 08 de dezembro de 2018.

TAKAMOTO Miriam, TOMAZELA Maria das Graças J. M., 2011. Aplicação de técnicas de mineração de dados para planejamento agrícola no estado de São Paulo. Disponível em: <<http://reverte.fatecid.com.br/index.php/revista/article/view/52>>. Acesso em 08 de dezembro de 2018.

TOMAZELA Maria das Graças J. M., CAMPOS Fernando Celso de, DANIEL Luiz Antônio. 2015. Mineração de Dados Aplicada à Produtividade de Cana de açúcar. Disponível em: <http://www.abepro.org.br/biblioteca/TN_STO_206_226_27367.pdf>. Acesso em 8 de dezembro de 2018.

WEKA. Machine Learning Group. Disponível em: <http://www.cs.waikato.ac.nz/ml/index.html>. Acesso em 23 de novembro de 2018.

3. CAPÍTULO I

USO DA TÉCNICA DE MINERAÇÃO DE DADOS PARA ESTUDO DA INFLUÊNCIA DE FATORES METEOROLÓGICOS SOBRE A PRODUÇÃO DE TOMATE PARA PROCESSAMENTO INDUSTRIAL

RESUMO

O Tomate industrial é uma das hortaliças de maior importância econômica no estado de Goiás e no Mundo. O objetivo deste trabalho é comprovar, pelo método de mineração de dados, quais variáveis influenciam na classificação do ciclo do cultivo, produtividade, qualidade e teor de sólidos solúveis (°Brix) do tomate industrial. Foram analisadas as médias por ciclo do cultivo das seguintes variáveis: temperatura, precipitação pluviométrica e umidade relativa do ar. Outra variável analisada foi a altitude do cultivo. Com a classificação através da mineração de dados, foi possível concluir que as variáveis temperatura, precipitação e umidade, durante o ciclo do cultivo, impactam na produtividade do tomate industrial no município de Morrinhos, estado de Goiás, conforme os outros estudos já consolidados nesta área.

PALAVRAS-CHAVE: crescimento, mineração de dados, produtividade, qualidade, temperatura

ABSTRACT

Industrial tomato is one of the most important vegetables economically in the state of Goiás, Brazil, and worldwide. This study aims to demonstrate which variables influence on the crop cycle classification, productivity, quality, and soluble solids content (°Brix) of the industrial tomato by data mining method. Temperature, rainfall, and relative humidity average per crop cycle was analyzed, as well as. the cultivation altitude. On the basis of data mining classification, it was possible to conclude that temperature, precipitation, and humidity during the crop cycle impact on the industrial tomato productivity in the municipality of Morrinhos, Goiás State, Brazil, is according to other studies already consolidated in this field.

Keywords: data mining, growth, productivity, quality, temperature

3.1 Introdução

O tomateiro é uma das culturas mais importantes, apresentando tendência ao crescimento da produção, seja pelo seu valor nutricional, seja pela produção de subprodutos cada vez mais comuns em nosso dia a dia (JUNIOR et al., 2015). É considerada a segunda hortaliça com maior importância econômica no Brasil e no mundo, atrás apenas da batata (LUZ et al., 2010). Seu consumo pode ser na forma *in natura* ou como produto processado.

O último levantamento sobre a área cultivada de tomate no Brasil mostrou que Goiás foi o maior produtor, com mais de 978 mil toneladas produzidas em uma área superior a 12 mil hectares, correspondendo a 26,17% da produção nacional (IBGE, 2017). Com isso, o estado lidera o ranking nacional como maior produtor desta olerícola no país. Quando falamos apenas no segmento destinado ao processamento industrial, a liderança de Goiás é ainda maior, sendo responsável por mais de 70% da produção nacional (VILELA et al., 2012).

O segmento de tomate destinado ao processamento industrial tem grande importância socioeconômica. Além dos empregos gerados diretamente nos campos de produção e nas indústrias, esta cadeia agroindustrial movimenta indústrias paralelas de insumos, compreendendo embalagens, máquinas agrícolas e equipamentos de irrigação (VILELA et al., 2012). A produção de tomate para indústria tem algumas peculiaridades, e o produto utilizado como matéria-prima deve atender à alguns padrões de qualidade, como elevado teor de sólidos solúveis, boa coloração e a integridade dos frutos.

Quanto maior o teor de sólidos solúveis (°Brix), maior será o rendimento industrial e menor o gasto de energia no processo de concentração da polpa. O teor de sólidos solúveis no fruto, além de ser uma característica genética da cultivar, é influenciado pela adubação, irrigação e temperatura. A temperatura, por sinal, influencia em outro parâmetro de qualidade de frutos para indústria, que é a cor. Temperaturas elevadas (acima de 32°C) favorecem a síntese de caroteno em detrimento do licopeno, causando manchas amareladas e maturação desuniforme dos frutos do tomateiro (ALVARENGA, 2013).

Fatores climáticos, assim como a temperatura, tem grande influência sobre a agricultura, não sendo diferente na produção de tomate para processamento. Em épocas de alta umidade relativa e precipitação pluviométrica, são comuns os problemas com doenças na cultura. Épocas de temperaturas mais baixas costumam prolongar o ciclo da

cultura. Desta forma, para que haja um planejamento quanto à disponibilidade e qualidade da matéria-prima, os produtores devem monitorar o clima e entender seu efeito sobre a produção agrícola. (ALVARENGA, 2013).

A agricultura é considerada historicamente uma atividade de risco, mas o avanço tecnológico tem auxiliado na sua mitigação, de modo que o produtor tenha cada vez mais confiança no retorno do seu investimento. Uma estratégia para aumentar a previsibilidade dos sistemas de produção é a prática da agricultura de precisão, cuja utilização demanda coleta de enorme quantidade de informações e sua análise. Entre os fatores mais difíceis de controle, têm destaque as variações do clima, em função das mudanças climáticas que vêm acontecendo a nível global. O entendimento de sua influência sobre os componentes de produção tem grande importância para que o produtor possa tomar decisões para mitigar perdas ou estimar seus rendimentos.

Atualmente, com a disponibilidade de estações meteorológicas, seja em áreas particulares ou em instituições públicas, os técnicos dispõem de grande quantidade de dados meteorológicos que podem auxiliá-los no entendimento dos ciclos produtivos. Entretanto, com aumento do volume de dados na era da informação e a facilidade de armazenamento, tornaram-se imprescindíveis a análise e recuperação de informação, úteis em meio à saturação de dados sem importância. Nesse sentido, a mineração de dados surge como ferramenta capaz de elucidar os padrões implícitos nessas grandes massas de dados.

A mineração de dados, também conhecida como “Descoberta de Conhecimento a partir de Dados” (“Knowledge Discovery from Data”, KDD), pode ser definida como a extração de padrões que representam o conhecimento armazenado em um grande banco de dados (HAN et al., 2011). Nas técnicas de mineração, são utilizados diversos algoritmos, cada qual objetivando um modo de descobrir conhecimento. Estes algoritmos buscam identificar padrões válidos, não conhecidos, potencialmente úteis e interpretáveis, consistindo, basicamente, em descobrir conhecimento nos dados armazenados (FAYYAD et al., 1996).

De acordo com Chinchuluun et al. (2010), o uso de técnicas de mineração de dados é algo recente nas áreas agrícola e ambiental. Dessa forma, o presente trabalho apresenta um estudo da influência de algumas variáveis climáticas sobre os componentes de produção de tomate para processamento industrial por meio da aplicação da técnica de mineração de dados, especificamente, dos algoritmos KStar, J48 e OneR, baseando-se no software WEKA 3.9.0.

3.2 Material e Métodos

Foram coletados dados relativos a cultivos de tomate para processamento industrial no período entre 2007 e 2014, no município de Morrinhos, localizado na região Sul do estado de Goiás.

Os dados foram fornecidos por uma agroindústria da região. No total, foram avaliados 84 cultivos de tomate, todos destinados ao segmento industrial, sendo listados de 8 a 18 campos por ano. Estes dados foram organizados de modo a obter a duração do ciclo, produtividade média, teor de sólidos solúveis e rendimento industrial para cada cultivo.

Considerando o período de cada cultivo, foi feito o levantamento dos dados meteorológicos de modo que, para cada cultivo, fossem obtidos os valores diários de temperatura média (°C), umidade relativa do ar (%) e precipitação pluviométrica acumulada (mm). Os dados foram obtidos na Estação Meteorológica da Universidade Estadual de Goiás, Campus Morrinhos, sendo corrigidos pelo software DSSAT (WeatherMan), com base nos dados de localização (altitude, longitude e latitude) obtidos utilizando o programa Google Earth Pro 2018.

Para início do processo de avaliação e mineração de dados, foram seguidas as etapas do KDD, Figura 1, de acordo com Fayyad *et al.* (1996): (I) seleção de dados, (II) pré-processamento, (III) limpeza, (IV) transformação ou codificação (enriquecimento), (V) Mineração de Dados e construção de relatórios e (VI) Avaliação e apresentação da informação descoberta.

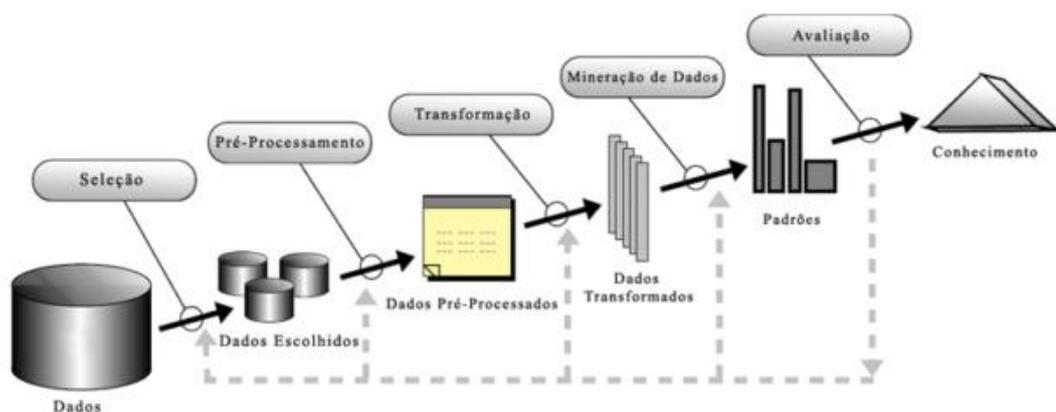


Figura 1: Representando o processo de KDD (Fayyad *et al.*, 1996)

Na fase de seleção dos dados, foram considerados para análise os dados meteorológicos correspondentes a cada ciclo de cultivos avaliados, relacionando-os com os respectivos valores de componentes de produção (ciclo, produtividade, teor de sólido solúveis e rendimento industrial).

Na fase de pré-processamento e limpeza, foi retirada a coluna dos nomes dos produtores.

A fase de transformação e enriquecimento dos dados para cada cultivo, foram inseridos mais atributos (colunas): tomados os valores médios diários das variáveis meteorológicas para os primeiros 30 dias após o transplante (DAT, 1º quarto do ciclo), de 30 a 60 DAT (2º quarto do ciclo), 60 a 90 DAT (3º quarto do ciclo), dos 90 DAT até a colheita (final do ciclo) e do ciclo inteiro. Tal delimitação teve por objetivo avaliar o efeito das variáveis meteorológicas em cada fase de desenvolvimento da cultura e sua influência nos componentes de produção. Esse procedimento gerou no final 15 colunas de atributos, relacionando cada ciclo de plantio com a temperatura, a umidade relativa e a precipitação pluviométrica.

Os valores dos componentes de produção (duração do ciclo, produtividade, teor de sólidos solúveis e rendimento industrial) foram transformados para classes, de modo que cada componente passou a ser um classificador.

Para a classificação do ciclo, os valores de duração do ciclo de cultivo (do plantio à colheita), originalmente descritos em dias, foram classificados em três classes de acordo com Alvarenga (2013): curto (<120 dias), médio (<130) e longo (≥ 130).

Em relação aos estudos de Alvarenga concernentes à produtividade, considerando a média nacional na casa das 70 toneladas por hectare, os valores obtidos foram classificados em cinco classes: ótimo (>110 t/ha), bom (>90 t/ha), regular (>70 t/ha), ruim (>50 t/ha) e péssimo (≤ 50 t/ha).

Os valores médios de teor de sólidos solúveis, com base nos estudos de Alvarenga, foram classificados em quatro classes: ótimo ($\geq 5,5^\circ\text{Brix}$), bom ($\geq 5^\circ\text{Brix}$), regular ($\geq 4^\circ\text{Brix}$) e ruim (<4 $^\circ\text{Brix}$). Por fim, tendo por base as classes determinadas para produtividade e teor de sólidos solúveis, os valores de rendimento industrial foram classificados em cinco classes: ótimo (≥ 20 t/ha), bom (≥ 15 t/ha), regular (≥ 10 t/ha), ruim (≥ 5 t/ha) e péssimo (<5 t/ha).

Após as fases de seleção de dados, limpeza, enriquecimento e transformação, foi feita a construção de relatórios. Para tal, os dados foram processados pela ferramenta

WEKA (software WEKA 3.9.0), utilizando algoritmos de classificação e de seleção de atributos.

Com o arquivo completo, os dados foram separados, cada componente de produção, sendo estes um atributo classificador, e avaliados separadamente.

Foram aplicados vários testes com algoritmos em cada arquivo, com seus respectivos classificadores para proceder à classificação, tendo sido escolhidos aqueles algoritmos que obtiveram a melhor matriz de confusão.

O classificador OneR avalia o nível de importância de um atributo medindo o ganho de informação em relação à classe. Com base nestes valores, foram selecionados os atributos mais importantes para cada classificador e montada uma árvore de decisão utilizando o Classificador J48, presente na ferramenta WEKA.

Na tabela 1, seguem as descrições das informações contidas nos gráficos.

Tabela 1: Legenda de identificação das unidades

Alt(m)	Altitude
Temp1	Média da Temperatura no 1º ciclo
Temp2	Média da Temperatura no 2º ciclo
Temp3	Média da Temperatura no 3º ciclo
Temp4	Média da Temperatura no 4º ciclo
UR1	Média da Umidade relativa ao 1º ciclo
UR2	Média da Umidade relativa ao 2º ciclo
UR3	Média da Umidade relativa ao 3º ciclo
UR4	Média da Umidade relativa ao 4º ciclo
PP1	Média da Precipitação no 1º ciclo
PP2	Média da Precipitação no 2º ciclo
PP3	Média da Precipitação no 3º ciclo
PP4	Média da Precipitação no 4º ciclo
TempTotal	Média da Temperatura Total dos Ciclos
UR Total	Média da Umidade Relativa total dos Ciclos
PP Total	Média da Precipitação Total dos Ciclos

3.2.1 Descrição dos algoritmos utilizados

O algoritmo KStar é um classificador baseado em instância, ou seja, a classe de uma instância de teste é baseada na classe dessas instâncias de treinamento semelhantes a ela, conforme determinado por alguma função de similaridade. Ele difere de outros aprendizes baseados em instâncias porque usa uma função de distância baseada em entropia.

O algoritmo OneR cria uma regra para cada atributo dos dados de treino e seleciona a regra com menor percentagem de erro como regra única. Para criar uma regra para um atributo, é necessário determinar a classe mais frequente para cada atributo. Como classe mais frequente, entende-se a classe que aparece mais vezes para um dado atributo.

O algoritmo J48 permite a criação de modelos de decisão em árvore. Utiliza uma tecnologia greedy para induzir árvores de decisão para posterior classificação. O modelo de árvore de decisão é construído pela análise dos dados de treino, sendo utilizado para classificar dados ainda não classificados. O J48 gera árvores de decisão, em que cada nó da árvore avalia a existência ou significância de cada atributo individual. As árvores de decisão são construídas do topo para a base pela escolha do atributo mais apropriado para cada situação.

3.3 Resultados e Discussão

Após feita a avaliação da influência dos atributos (médias de cada um dos parâmetros meteorológicos para cada etapa do ciclo e ciclo inteiro) para cada classificador (componente de produção), foi possível observar algumas diferenças entre os classificadores quanto aos atributos de maior importância. Para o ciclo de cultivo, por exemplo, o atributo de maior importância é a média de precipitação pluviométrica no quarto período. O atributo que teve maior importância para o classificador teor de sólidos solúveis (°Brix) foi a média da temperatura no primeiro período. Já para produtividade, o atributo que teve maior importância para o classificador foi a média da precipitação pluviométrica total; para o rendimento industrial, o atributo de maior importância foi a média de precipitação pluviométrica no segundo período da cultura.

Trabalho desenvolvido por Pontes et al. (2016) confirma os dados, sendo possível observar que quando as plantas apresentam maior severidade de doenças foliares, no caso a mancha bacteriana, há uma aceleração na maturação dos frutos, evidenciado

por aumento no número de frutos podres quando a colheita foi realizada apenas aos 120 dias. Do ponto de vista de mecanismos de sobrevivência das plantas, sabe-se que quando elas estão sob estresse, seja biótico ou abiótico, elas entram na fase reprodutiva mais rapidamente com o intuito fechar o seu ciclo e garantir a sobrevivência da espécie por meio de seus descendentes.

Em relação à produtividade de frutos, o fato da média de precipitação diária de todo o ciclo ter apresentado maior influência na determinação das classes também pode ser explicado sob o ponto de vista sanitário. Maior precipitação pluviométrica ao longo do cultivo aumenta a incidência de doenças, assim como índices elevados de chuvas no final do ciclo prejudicam a colheita e levam ao apodrecimento dos frutos. Quando avaliada a série histórica utilizada neste estudo, a menor média de produtividade foi observada no ano de 2009 (120,78 t/ha). Este também foi o ano em que os cultivos tiveram maior média de precipitação pluviométrica diária ao longo do ciclo (1,85mm).

Em revisão feita por Melo *et al.* (2011), quando avaliadas as médias de produtividade de tomate para processamento industrial na primeira década dos anos 2000, período em que a utilização de cultivares híbridas mais produtivas estava amplamente difundida, o ano de 2009 foi o que apresentou a menor produtividade. Segundo estes autores, a queda da produtividade neste ano ocorreu pelo excesso de chuvas, resultando em problemas fitossanitários. Tal afirmação se torna coerente quando observamos a árvore de decisão para o classificador 'produtividade', Figura 10, segundo a qual quando a média de precipitação pluviométrica diária ao longo do ciclo é superior a 0,63mm, na maioria das combinações as produtividades estimadas seriam classificadas como ruins ou péssimas (inferiores à 50 t/ha).

É de se esperar que com alta umidade relativa do ar haja maiores problemas fitossanitários, levando a planta à uma redução no seu desenvolvimento de modo a comprometer a translocação de fotoassimilados para os frutos. Em acordo com esta análise, se a média de temperatura no primeiro ciclo for maior que 24,22°C, Figura 04, o teor de sólidos solúveis estimado seria classificado como ruim (<4°Brix). Isso tanto poderia ser uma questão relacionada a problemas fitossanitários, bem como ao efeito deletério de altas temperaturas no desenvolvimento da planta.

Em relação ao rendimento industrial, os maiores efeitos sobre este classificador seria a média de precipitação pluviométrica no segundo período de cultivo. Em acordo com a árvore de tomada de decisão, Figura 13, valores elevados de precipitação neste período, aliados à alta umidade no quarto período e à precipitação no primeiro período e

a cultivos em baixa altitude, levam a estimativa do rendimento industrial para as piores classes (ruim e péssimo). A associação entre alta precipitação, altos teores de umidade e cultivos em baixas altitudes, onde é mais comum a ocorrência de temperaturas elevadas, leva a crer que esta perspectiva negativa seja prevista em função de estas condições serem propícias a problemas fitossanitários, que reduzem a produtividade e qualidade dos frutos.

Em relação ao ciclo, Figura 07, os maiores efeitos são a média da precipitação no quarto período acima de 1,22mm, média da umidade relativa no primeiro período e média da precipitação total, que levam a períodos curtos.

Na Figura 2, é utilizado o Classificador KStar e escolhendo a opção “use training set” para gerar a matriz de confusão para confirmar o resultado do classificador J48 nos classificadores de instâncias do Brix.

```

=== Summary ===

Correctly Classified Instances      84          100 %
Incorrectly Classified Instances    0           0 %
Kappa statistic                     1
Mean absolute error                 0.0008
Root mean squared error             0.0047
Relative absolute error              0.2596 %
Root relative squared error         1.1761 %
Total Number of Instances          84

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          1,000   0,000   1,000     1,000   1,000     1,000   1,000    1,000    Bom
          1,000   0,000   1,000     1,000   1,000     1,000   1,000    1,000    Regular
          1,000   0,000   1,000     1,000   1,000     1,000   1,000    1,000    Ruim
Weighted Avg.  1,000   0,000   1,000     1,000   1,000     1,000   1,000    1,000

=== Confusion Matrix ===

 a  b  c  <-- classified as
24  0  0  |  a = Bom
 0 56  0  |  b = Regular
 0  0  4  |  c = Ruim

```

Figura 2: Matriz de Confusão do KStar para Brix

Na Figura 3, utilizando o Classificador J48 e escolhendo a opção “use training set”, temos os seguintes dados para o Brix.

```

=== Summary ===

Correctly Classified Instances      80          95.2381 %
Incorrectly Classified Instances    4           4.7619 %
Kappa statistic                    0.9009
Mean absolute error                0.0565
Root mean squared error            0.1681
Relative absolute error            17.731 %
Root relative squared error        42.3969 %
Total Number of Instances         84

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
0,958  0,050  0,885  0,958  0,920  0,888  0,971  0,881  Bom
0,946  0,036  0,981  0,946  0,964  0,896  0,973  0,980  Regular
1,000  0,000  1,000  1,000  1,000  1,000  1,000  1,000  Ruim
Weighted Avg.  0,952  0,038  0,955  0,952  0,953  0,899  0,974  0,953

=== Confusion Matrix ===

 a  b  c  <-- classified as
23  1  0 | a = Bom
 3 53  0 | b = Regular
 0  0  4 | c = Ruim

```

Figura 3: Matriz de Confusão do J48 para Brix

Na Figura 4, utilizando o Classificador J48, temos a visualização da árvore para o Brix, cujo fator média da temperatura no primeiro ciclo maior que 24,22° celsius é condição ruim para o Brix.

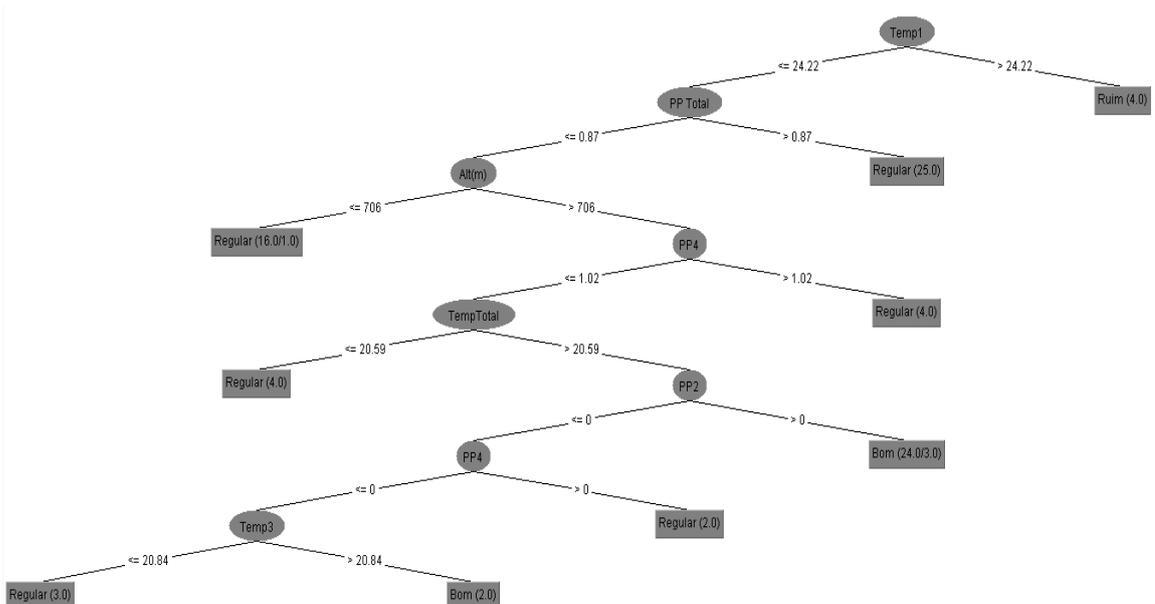


Figura 4: Árvore do J48 para Brix

Na Figura 5, é utilizado o Classificador KStar e escolhida a opção “use training set” para gerar a matriz de confusão para confirmar o classificador J48 nos classificadores de instâncias do Ciclo.

```

=== Summary ===

Correctly Classified Instances      84          100    %
Incorrectly Classified Instances    0           0    %
Kappa statistic                    1
Mean absolute error                 0
Root mean squared error             0.0003
Relative absolute error             0.0089 %
Root relative squared error        0.0626 %
Total Number of Instances          84

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          1,000   0,000   1,000     1,000   1,000     1,000   1,000    1,000    Medio
          1,000   0,000   1,000     1,000   1,000     1,000   1,000    1,000    Longo
          1,000   0,000   1,000     1,000   1,000     1,000   1,000    1,000    Curto
Weighted Avg.   1,000   0,000   1,000     1,000   1,000     1,000   1,000    1,000

=== Confusion Matrix ===

 a  b  c  <-- classified as
45  0  0 | a = Medio
 0 27  0 | b = Longo
 0  0 12 | c = Curto

```

Figura 5: Matriz de Confusão do KStar para Ciclo.

Na Figura 6, utilizando o Classificador J48 e escolhendo a opção “use training set”, temos os seguintes dados para o Ciclo.

```

=== Summary ===

Correctly Classified Instances      76          90.4762 %
Incorrectly Classified Instances    8           9.5238 %
Kappa statistic                    0.8343
Mean absolute error                 0.1011
Root mean squared error             0.2249
Relative absolute error             25.6255 %
Root relative squared error        50.7322 %
Total Number of Instances          84

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0,978   0,128   0,898     0,978   0,936     0,859   0,964    0,946    Medio
          0,852   0,035   0,920     0,852   0,885     0,834   0,959    0,930    Longo
          0,750   0,014   0,900     0,750   0,818     0,795   0,966    0,843    Curto
Weighted Avg.   0,905   0,082   0,905     0,905   0,903     0,842   0,963    0,926

=== Confusion Matrix ===

 a  b  c  <-- classified as
44  1  0 | a = Medio
 3 23  1 | b = Longo
 2  1  9 | c = Curto

```

Figura 6: Matriz de Confusão do J48 para Ciclo

Na Figura 7, utilizando o Classificador J48, temos a visualização da árvore para o Ciclo, em que, com o fator precipitação no quarto ciclo maior que 1,22 mm e média da umidade relativa do ar menor ou igual a 69,84, poderemos ter condições curtas para o ciclo do plantio.

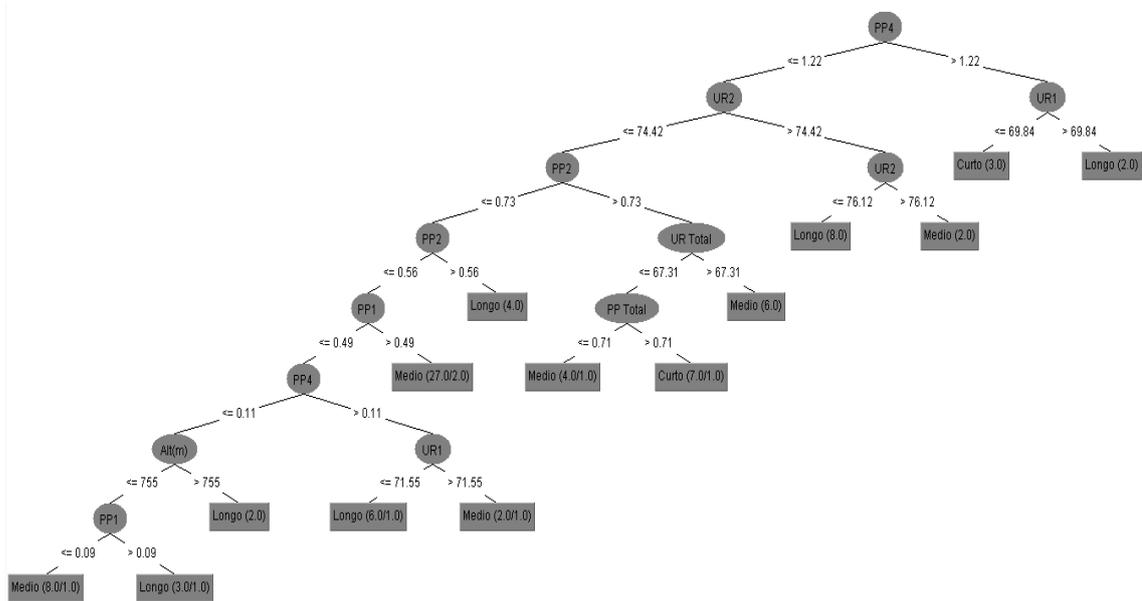


Figura 7: Árvore do J48 para o Ciclo

Na Figura 8, é utilizado o Classificador KStar e escolhida a opção “use training set” para gerar a matriz de confusão para confirmar o classificador J48 nos classificadores de instâncias da Produtividade.

=== Summary ===

Correctly Classified Instances	82	97.619 %
Incorrectly Classified Instances	2	2.381 %
Kappa statistic	0.9687	
Mean absolute error	0.0097	
Root mean squared error	0.069	
Relative absolute error	3.1757 %	
Root relative squared error	17.6663 %	
Total Number of Instances	84	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Otimo
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Bom
	1,000	0,032	0,917	1,000	0,957	0,942	0,999	0,994	Regular
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Pessimo
	0,857	0,000	1,000	0,857	0,923	0,913	0,998	0,986	Ruim
Weighted Avg.	0,976	0,008	0,978	0,976	0,976	0,970	0,999	0,996	

=== Confusion Matrix ===

```

a b c d e <-- classified as
15 0 0 0 0 | a = Otimo
0 27 0 0 0 | b = Bom
0 0 22 0 0 | c = Regular
0 0 0 6 0 | d = Pessimo
0 0 2 0 12 | e = Ruim

```

Figura 8: Matriz de Confusão do KStar para Produtividade

Na Figura 9, utilizando o Classificador J48 e escolhendo a opção “use training set”, temos os seguintes dados para a Produtividade.

```

=== Summary ===

Correctly Classified Instances      77          91.6667 %
Incorrectly Classified Instances    7           8.3333 %
Kappa statistic                    0.8912
Mean absolute error                 0.0478
Root mean squared error             0.1547
Relative absolute error             15.6281 %
Root relative squared error         39.5827 %
Total Number of Instances          84

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0,867   0,000   1,000     0,867   0,929     0,918   0,996    0,972    Otimo
                0,963   0,018   0,963     0,963   0,963     0,945   0,998    0,992    Bom
                0,864   0,032   0,905     0,864   0,884     0,844   0,983    0,948    Regular
                1,000   0,013   0,857     1,000   0,923     0,920   0,998    0,952    Pessimo
                0,929   0,043   0,813     0,929   0,867     0,841   0,983    0,874    Ruim
Weighted Avg.   0,917   0,022   0,922     0,917   0,917     0,895   0,991    0,955

=== Confusion Matrix ===

  a  b  c  d  e  <-- classified as
13  1  1  0  0  | a = Otimo
 0 26  1  0  0  | b = Bom
 0  0 19  0  3  | c = Regular
 0  0  0  6  0  | d = Pessimo
 0  0  0  1 13  | e = Ruim
    
```

Figura 9: Matriz de Confusão do J48 para a Produtividade

Na Figura 10, utilizando o Classificador J48, temos a visualização da árvore para a Produtividade. Temos condições péssimas para produtividade quando temos média de precipitação total maior que 0,63 mm e média de precipitação no primeiro ciclo maior que 8,42 mm.

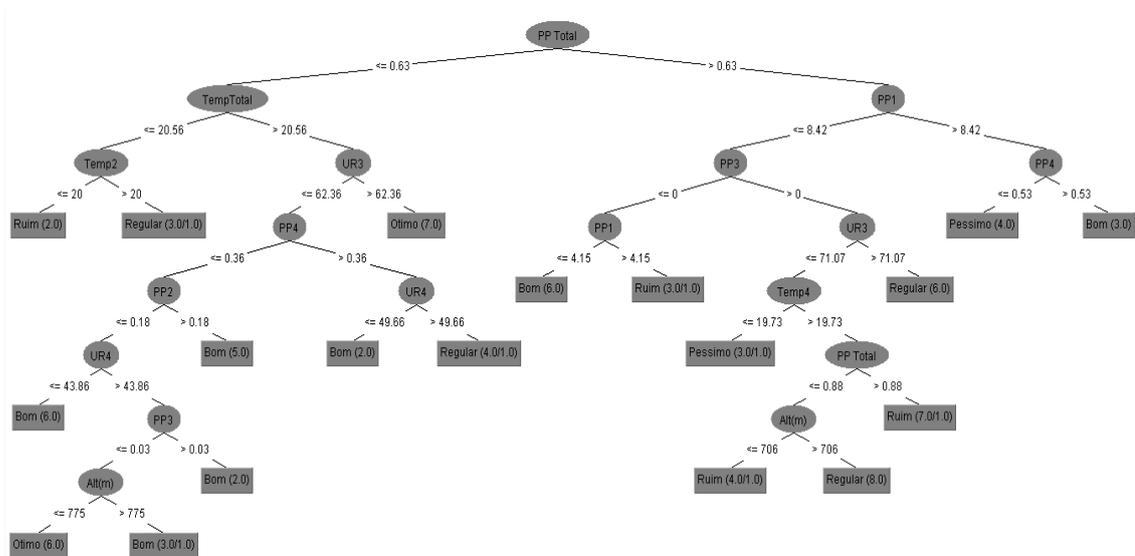


Figura 10: Árvore do J48 para Produtividade

Na Figura 11, é utilizado o Classificador KStar e escolhida a opção “use training set” para gerar a matriz de confusão para confirmar o classificador J48 nos classificadores de instâncias do Rendimento Industrial.

```

=== Summary ===

Correctly Classified Instances      81          96.4286 %
Incorrectly Classified Instances    3           3.5714 %
Kappa statistic                    0.9511
Mean absolute error                 0.0164
Root mean squared error             0.0868
Relative absolute error             5.5461 %
Root relative squared error        22.6534 %
Total Number of Instances          84

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      1,000    0,000    1,000     1,000    1,000     1,000    1,000    1,000    Bom
      1,000    0,000    1,000     1,000    1,000     1,000    1,000    1,000    Otimo
      1,000    0,050    0,889     1,000    0,941     0,919    0,997    0,990    Regular
      0,800    0,000    1,000     0,800    0,889     0,876    0,996    0,976    Ruim
      1,000    0,000    1,000     1,000    1,000     1,000    1,000    1,000    Pessimo
Weighted Avg.    0,964    0,014    0,968     0,964    0,963     0,955    0,998    0,993

=== Confusion Matrix ===

 a  b  c  d  e  <-- classified as
30  0  0  0  0  |  a = Bom
 0 13  0  0  0  |  b = Otimo
 0  0 24  0  0  |  c = Regular
 0  0  3 12  0  |  d = Ruim
 0  0  0  0  2  |  e = Pessimo

```

Figura 11: Matriz de Confusão do KStar para Rendimento Industrial

Na Figura 12, utilizando o Classificador J48 e escolhendo a opção “use training set”, temos os seguintes dados para o Rendimento Industrial.

```

=== Summary ===

Correctly Classified Instances      69          82.1429 %
Incorrectly Classified Instances    15          17.8571 %
Kappa statistic                    0.7518
Mean absolute error                 0.0961
Root mean squared error             0.2192
Relative absolute error            32.5605 %
Root relative squared error        57.1967 %
Total Number of Instances          84

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0,967    0,037    0,935     0,967    0,951     0,923    0,985    0,970    Bom
      1,000    0,000    1,000     1,000    1,000     1,000    1,000    1,000    Otimo
      0,917    0,200    0,647     0,917    0,759     0,660    0,896    0,670    Regular
      0,200    0,014    0,750     0,200    0,316     0,334    0,854    0,460    Ruim
      1,000    0,000    1,000     1,000    1,000     1,000    1,000    1,000    Pessimo
Weighted Avg.    0,821    0,073    0,831     0,821    0,791     0,756    0,939    0,799

=== Confusion Matrix ===

 a  b  c  d  e  <-- classified as
29  0  1  0  0  |  a = Bom
 0 13  0  0  0  |  b = Otimo
 1  0 22  1  0  |  c = Regular
 1  0 11  3  0  |  d = Ruim
 0  0  0  0  2  |  e = Pessimo

```

Figura 12: Matriz de Confusão do J48 para Rendimento Industrial

Na Figura 13, utilizando o Classificador J48, temos a visualização da árvore para o Rendimento Industrial. Temos precipitação no segundo ciclo acima de 0,21 mm e a altitude abaixo ou igual a 751 metros como fatores decisivos para um rendimento industrial tido como péssimo.

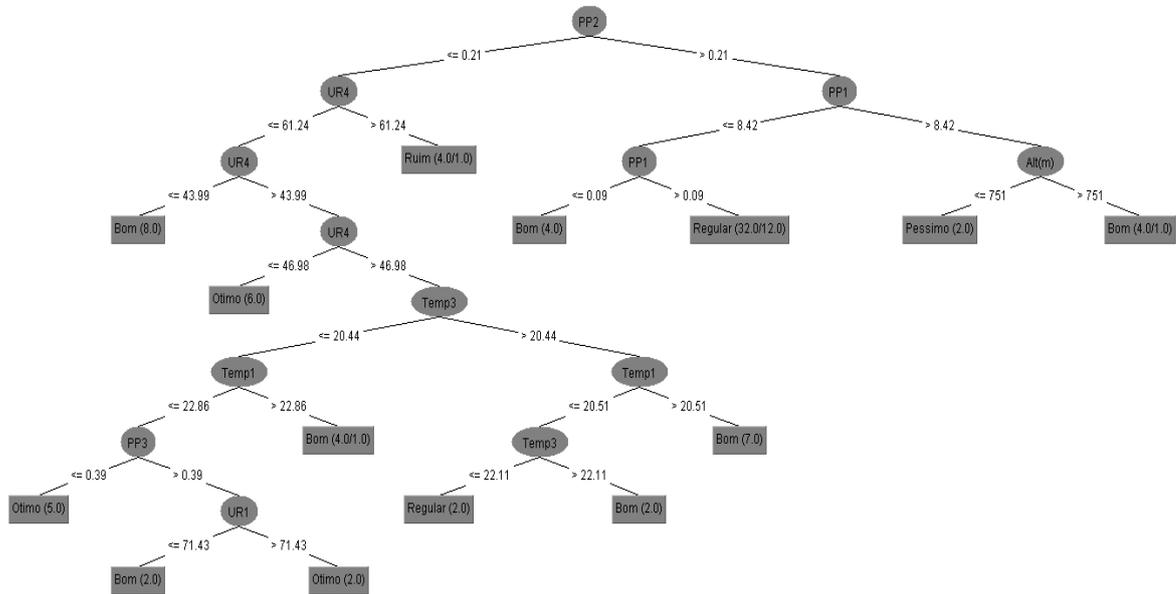


Figura 13: Árvore do J48 para Rendimento Industrial

3.4 Conclusão

No presente estudo, foi possível avaliar as influências de fatores meteorológicos sobre os componentes de produção por meio da técnica de mineração de dados, com resultados que se sustentam com as informações descritas na literatura. A alimentação desta base de dados com mais atributos, como radiação e fotoperíodo, ou outros classificadores, como sanidade da parte aérea, maturação e integridade de frutos, pode auxiliar o segmento da cadeia agroindustrial do tomate a entender melhor as influências dos fatores climáticos sobre os componentes de produção.

3.5 Referências bibliográficas

ALGORITMO ONER. Disponível em: <<http://www.saedsayad.com/oner.htm>>. Acesso em 23 de novembro de 2018.

ALGORITMO J45. Disponível em: < [https://books.google.com.br/books?hl=pt-BR&lr=&id=b3ujBQAAQBAJ&oi=fnd&pg=PP1&dq=Ross+Quinlan+\(1993\).+C4.5:+Programs+for+Machine+Learning.+Morgan+Kaufmann+Publishers,+San+Mateo,+CA.&ots=sQ6mRREoF4&sig=570rWhX1dGRlf0vDyq3ZvyluQvQ#v=onepage&q=Ross%20Quinlan%20\(1993\).%20C4.5%3A%20Programs%20for%20Machine%20Learning.%20Morgan%20Kaufmann%20Publishers%2C%20San%20Mateo%2C%20CA.&f=false](https://books.google.com.br/books?hl=pt-BR&lr=&id=b3ujBQAAQBAJ&oi=fnd&pg=PP1&dq=Ross+Quinlan+(1993).+C4.5:+Programs+for+Machine+Learning.+Morgan+Kaufmann+Publishers,+San+Mateo,+CA.&ots=sQ6mRREoF4&sig=570rWhX1dGRlf0vDyq3ZvyluQvQ#v=onepage&q=Ross%20Quinlan%20(1993).%20C4.5%3A%20Programs%20for%20Machine%20Learning.%20Morgan%20Kaufmann%20Publishers%2C%20San%20Mateo%2C%20CA.&f=false)> Acesso em 23 de novembro de 2018.

ALGORITMO K-STAR. Disponível em: <<http://weka.sourceforge.net/doc.dev/weka/classifiers/lazy/KStar.html>> Acesso em 23 de novembro de 2018.

ALVARENGA, M.A.R. Tomate: Produção em Campo, Casa de Vegetação e Hidroponia. 2º ed. Lavras: Ed. Lavras, 2013.

CHINCHULUUN, A.; XANTHOPOULOS, P.; TOMAINO, V.; PARDALOS, P. M. Data Mining techniques in agricultural and environmental sciences. International Journal of Agricultural and Environmental Information System. Hershey, v.1, n.1, p. 26-40, Jan./June 2010. Disponível em: <http://www.academia.edu/1769745/Data_Mining_Techniques_in_Agricultural_and_Environmental_Sciences>. Acesso em: 23 de novembro de 2018.

ELMASRI, RAMEZ. Conceito de Data Mining. IN: Sistema de Banco de Dados. Revisor técnico Luis Ricardo de Figueiredo. Sao Paulo: Pearson Addison Wesley, 2005. p. 624-645.

FAYYAD, U.; Piatetsky-shapiro, G.; Smyth, P.; Uthurusamy, R. Advances in knowledge discovery and data mining. Cambridge: MIT Press, 1996. 560 p.

GOOGLE EARTH. Google Earth PRO. Version 7.3.2, 2018. Morrinhos-Goiás. Disponível em: <https://www.google.com/intl/pt-BR_ALL/earth/versions/> Acesso em: 07 de dezembro de 2018.

HAN, J.; KAMBER, M. Data mining: concepts and techniques. 3º ed. Amsterdam: Elsevier, 2011. 744 p.

AGENCIA IBGE. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/22780-em-setembro-ibge-preve-safra-6-menor-que-a-de-2017>>. Acesso em: 07 de dezembro de 2018.

CULTIVO DO TOMATE PARA INDUSTRIALIZAÇÃO. Disponível em: <<https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Tomate/TomateIndustrial/cultivares.htm>>. Acesso em: 07 de dezembro de 2018.

JUNIOR ARS; RIBEIRO WM; NASCIMENTO AR; SOUZA CB. 2015. Cultivo do tomate industrial no estado de Goiás: Evolução das áreas de plantio e produção. *Conjuntura econômica goiana* 34: 107.

LUZ JMQ; BITTAR CA; QUEIROZ AA; CARREON R. 2010. Produtividade de tomate 'Débora Pto' sob adubação organomineral via foliar e gotejamento. *Horticultura Brasileira* 28: 489-494.

MELO, Paulo César T. de; VILELA, Nirlene J.. Desempenho da cadeia agroindustrial brasileira do tomate na década de 90. *Hortic. Bras.* vol.22 no.1 Brasília Jan./Mar. 2004.

MELO, P.C.T., FONTE, L.C. (2011) Brazil processing tomato season 2010: results and future perspectives. *Tomato News* 3:15-19.

PONTES, N. C., NASCIMENTO, A. R., GOLYNSKI, A., MAFFIA, L. A., OLIVEIRA, J. R., and QUEZADO-DUVAL, A. M. 2016. Intervals and number of applications of acibenzolar-S-methyl for the control of bacterial spot on processing tomato. *Plant Dis.* 100:2126-2133.

VILAS-BÔAS, G.L., MELO, P.E., CASTELO-BRANCO, M., GIORDANO, L.B., MELO, F.F., 2007. Desenvolvimento de um modelo de produção integrada de tomate indústria - PITI. In: Zambolim, L., Lopes, C.A., Picanço, M.C., Costa, H. (Eds.), *Manejo integrado de doenças e pragas - Hortaliças*. UFV/CNPH, Viçosa, pp. 349-362.

VILELA NJ; MELO PCT; BOITEUX LS; CLEMENTE FMVT. 2012. Perfil socioeconômico da cadeia agroindustrial no Brasil. In: CLEMENTE FMVT; BOITEUX LS. (eds). *Produção de tomate para processamento industrial*. Brasília: Embrapa Hortaliças. p. 17.

WEKA. Machine Learning Group. Disponível em: <<http://www.cs.waikato.ac.nz/ml/index.html>>. Acesso em 23 de novembro de 2018.

4. CONCLUSÃO GERAL

No presente estudo, foi possível avaliar a influência de fatores meteorológicos sobre os componentes de produção por meio da técnica de mineração de dados, com resultados que se sustentam com as informações descritas na literatura. A alimentação desta base de dados com mais atributos, como radiação e fotoperíodo, ou outros classificadores, como sanidade da parte aérea, maturação e integridade de frutos, pode auxiliar o segmento da cadeia agroindustrial do tomate a entender melhor as influências dos fatores climáticos sobre os componentes de produção.