



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA AGRÍCOLA



Nádia Vieira Ribeiro

Desenvolvimento de modelos preditivos utilizando árvores de regressão para análise de infestação de broca (*Diatraea saccharalis*) na cultura da cana-de-açúcar.

CAMPINAS

2016



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA AGRÍCOLA



Nádia Vieira Ribeiro

Desenvolvimento de modelos preditivos utilizando árvores de regressão para análise de infestação de broca (*Diatraea saccharalis*) na cultura da cana-de-açúcar.

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de **Engenheiro Agrícola** à Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas.

CAMPINAS

2016

Agência(s) de fomento e nº(s) de processo(s): Não se aplica.

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

R354d Ribeiro, Nádía Vieira
Desenvolvimento de modelos preditivos utilizando árvores de regressão para análise de infestação de broca (*Diatraea saccharalis*) na cultura da cana-de-açúcar / Nádía Vieira Ribeiro. – Campinas, SP : [s.n.], 2016.

Orientador: Luiz Henrique Antunes Rodrigues.
Trabalho de Conclusão de Curso (graduação) – Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola.

1. Árvores - Doenças e pragas. 2. Mineração de dados. 3. Árvores de decisão. 4. Broca-da-cana-de-açúcar. I. Rodrigues, Luiz Henrique Antunes, 1959-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Agrícola. III. Título.

Informações adicionais, complementares

Título em outro idioma: Development of predictive models using regression trees for analysis of borer infestation (*Diatraea saccharalis*) in sugarcane culture

Palavras-chave em inglês:

Trees - Diseases and pests

Data mining

Decision trees

Sugarcane drill

Titulação: Engenheira Agrícola

Banca examinadora:

Luiz Henrique Antunes Rodrigues [Orientador]

Paulo Rodrigues Peloia

Matheus Uzelotto Lopes

Data de entrega do trabalho definitivo: 01-12-2016

Desenvolvimento de modelos preditivos utilizando árvores de regressão para análise de infestação de broca (*Diatraea saccharalis*) na cultura da cana-de-açúcar.

Nádia Vieira Ribeiro

BANCA EXAMINADORA

.....
Prof. Dr. Luiz Henrique Antunes Rodrigues
Orientador

.....
Dr. Paulo Rodrigues Peloia

.....
Matheus Uzelotto Lopes

AGRADECIMENTOS

Agradeço à minha mãe, Teresinha, que fez uso da sua formação em psicologia, escolhendo as melhores palavras para acalmar os altos - e, principalmente - os baixos da minha graduação, que fez incontáveis marmitas ao longo destes 5 anos e que trabalhou muito para garantir o estudo dos filhos. Ao meu pai, Antonio, por sair de sua zona de conforto para ganhar a vida, por sempre buscar o seu melhor e por ser tão carinhoso e atencioso com sua família. Meus pais são os protagonistas desta conquista, grandes exemplos de amor, honestidade e perseverança.

Ao meu irmão, Jáder, por desbravar o desconhecido antes de mim, facilitando as minhas escolhas e por ser meu amigo para toda vida. Ao meu namorado, Marcos, pelo companheirismo em todas as situações, sendo (muito!) paciente, carinhoso e altruísta. Este é o maior presente que a Unicamp me deu. À minha família, por sempre permanecer unida e torcer pelo meu sucesso.

A todos os amigos da FEAGRI, por me acompanharem na jornada da Engenharia Agrícola; em especial ao Prof. Lique, por todas as oportunidades que me concedeu e por me orientar construtivamente e de forma íntegra. Aos amigos Bocca, Matheus e Monique, que colaboraram tecnicamente com a realização deste trabalho e que foram essenciais por me motivarem nas horas que precisei. Aos avaliadores, Paulo e Matheus, por aceitarem prontamente o convite e por se disporem a contribuir.

A todos que participaram direta e indiretamente da minha formação pessoal e profissional e aos que utilizarem esta obra como fonte de estudo.

Todos vocês me trouxeram até aqui!

RESUMO

A cultura da cana-de-açúcar no Brasil é alvo de infestação pela *Diatraea saccharalis*, mais conhecida como broca. A broca, em sua fase larval, penetra no colmo, levando à abertura de inúmeras galerias. Tais galerias reduzem a massa do colmo, propiciam o acamamento e podem até mesmo provocar a morte das gemas, prejudicando a produção e reduzindo a produtividade em até 20%. Além disso, os orifícios realizados pela broca são convenientes para pragas secundárias, pois facilitam a entrada destes patógenos que acentuam os prejuízos do processamento da cana com a contaminação microbiológica. Para conter a infestação, práticas como controles químico e/ou biológico são utilizadas, aumentando ainda mais o custo produtivo, reduzindo as margens do produtor e da usina e nem sempre sendo sustentável o controle econômico da praga. O controle poderia ser mais efetivo se fosse possível antecipar quais as áreas em que as condições são propícias ao aparecimento da praga. Técnicas de mineração de dados podem contribuir com a solução do problema citado, visto que possibilitam extrair informações de conjuntos de dados. Assim, o objetivo deste trabalho foi desenvolver modelos preditivos dos índices de infestação de broca (*Diatraea saccharalis*) na cultura da cana-de-açúcar utilizando árvores de regressão. Para isso, foram realizados procedimentos de descoberta de conhecimento em um conjunto de dados de uma usina do estado de São Paulo utilizando árvores de regressão. Os atributos iniciais disponíveis no conjunto de dados da usina foram retrabalhados a fim de facilitar as análises. Resultados preliminares indicaram que condições meteorológicas e propriedades como variedade e número de cortes são os fatores mais relacionados ao índice de infestação. Constatou-se que as situações críticas para infestação de broca ocorrem na variedade SP83-2847, nos cultivares RB946015 e RB976931 e em canaviais de primeiro ciclo (cana-planta). Pelos dados de manejo, conclui-se que o controle biológico só é efetivo quando praticado em infestação inicial. Uma vez extrapolado o índice de infestação, aplicações adicionais aparentam não ser efetivas na redução da infestação pela broca.

Palavras chave: Modelagem de pragas, mineração de dados, árvores de decisão, índice de infestação, tempo térmico.

ABSTRACT

Sugar cane is Brazil a potential target to *Diatraea saccharalis*, commonly known as sugarcane borer. While in the larvae phase, the borer penetrates the sugarcane's stalk, feeding on it. This reduces the stalk's mass, increases the chance of lodging and may even cause the death of the buds, reducing productivity up to 20%. In addition, the holes made by the borer are convenient for secondary pests, since they facilitate the entry of these pathogens that accentuate the losses of the cane processing with the microbiological contamination. Chemical and/or biological control is used to manage infestation, which increases costs of the product, reduces producer's margins, and is not always sustainable. These control form's effectiveness could be increased if it was possible to preemptively identify areas which are more likely to be infested. Data mining techniques may help with a solution to this problem, since it enables to extract information from data sets. From this, the objective of this work was to develop predictive models of borer infestation rates (*Diatraea saccharalis*) in the sugarcane crop using regression trees. Knowledge discovery procedures were performed in a data set from São Paulo state mill using regression trees. The initial attributes available in the data set were reworked to facilitate analysis. Preliminary results indicated that meteorological conditions and properties such as variety and number of cuts are the factors most related to the infestation index. It was verified that the critical situations for borer infestation occur in the variety SP83-2847, in cultivars RB946015 and RB976931 and for sugarcane in its first cycle (cane-plant). Based on management data, it is concluded that biological control is only effective when practiced at levels of borer infestation that are still in the beginning. Once the infestation index is extrapolated, additional applications appear to be ineffective in reducing infestation by the drill.

Key words: Pest modeling, data mining, decision trees, infestation index, thermal time.

SUMÁRIO

1. INTRODUÇÃO	8
1.1. Justificativa	8
1.2. Objetivos	9
1.2.1. Objetivo Geral	9
1.2.2. Objetivos Específicos	9
2. REVISÃO DE LITERATURA	9
2.1. Tempo térmico	9
2.2. Árvores de regressão	12
2.3. Broca (<i>Diatraea saccharalis</i>)	12
2.4. Análise exploratória	14
3. METODOLOGIA	14
3.1. Caracterização dos dados da usina	15
3.2. Caracterização dos dados meteorológicos	16
3.3. Tratamento dos dados	17
3.4. Elaboração dos atributos	17
3.5. Agrupamento espacial	19
3.6. Análise exploratória	19
3.7. Divisão dos conjuntos de treino e teste	19
3.8. Ajuste de hiperparâmetros dos algoritmos	20
3.9. Validação do modelo	20
4. RESULTADOS E DISCUSSÃO	20
4.1. Análise exploratória	20
4.2. Modelagem preliminar	25
4.3. Modelagem	27
5. CONCLUSÕES	31
6. REFERÊNCIAS	32
7. ANEXO	34

1. INTRODUÇÃO

A cana-de-açúcar é uma das principais culturas para a economia do país. A produção desta cultura no Brasil em 2015/2016 foi de 658,7 milhões de toneladas (CONAB, 2016), sendo a região sudeste a maior contribuidora. Além de ser o maior produtor de cana, o país é o primeiro na produção de açúcar e etanol e conquista o mercado externo com o uso do biocombustível como alternativa energética (MAPA, 2016).

Embora o potencial de produtividade seja elevado, existem fatores que reduzem a produção do canavial como, por exemplo, o ataque de pragas. Uma das principais pragas causadoras de danos fisiológicos à cultura e, conseqüentemente, prejuízos econômicos, é a broca (*Diatraea saccharalis*). Os danos ocorrem quando a praga encontra-se na fase larval, pois, após a eclosão dos ovos que foram depositados nas folhagens da cana, as lagartas perfuram a casca do colmo e promovem a abertura de galerias.

A abertura de galerias compromete o desenvolvimento da cana-de-açúcar. Em canaviais jovens, algumas conseqüências podem ser a morte da gema apical e a secagem das folhas mais novas, originando o sintoma conhecido como "coração morto", o que acarreta na mortalidade de grande número de perfilhos (DINARDO-MIRANDA; VASCONCELOS; LANDELL, 2010). Em canaviais desenvolvidos, ocorre a secagem dos colmos, levando à morte dos colmos ou quebra devido ao vento por estarem mais fragilizados. Assim, a queda de produtividade devido à perda de massa é inevitável, podendo alcançar reduções de 10,5% (TERÁN; SÁNCHEZ; PRECETTI, 1985) a 19,80% (ROSSATO et al., 2013). O problema relatado acima foi confirmado pelos representantes da Usina Santa Fé durante uma reunião com o grupo de pesquisa, o que foi determinante para a escolha do tema da monografia. Além disso, os orifícios causados pela broca também favorecem a contaminação microbiológica, pois fragilizam a estrutura do colmo e servem de entrada para patógenos que prejudicam o processamento da cana (DINARDO-MIRANDA; FRACASSO; PERECIN, 2011).

A mineração de dados é conveniente para este tipo de situação em que se faz necessário o entendimento e a descoberta de conhecimentos em extenso banco de dados. Este trabalho associa a demanda por informações da usina à ferramenta preditiva (árvore de regressão) da mineração de dados.

1.1. Justificativa

Muitas empresas possuem bancos de dados com um enorme potencial, que normalmente é subutilizado. Uma forma de aproveitar melhor estes bancos de dados é utilizando-os para a descoberta de conhecimento, particularmente de forma a mitigar impactos econômicos nos custos de produção e auxiliar nas tomadas de decisão. Este trabalho é um exemplo de que informações relevantes podem ser extraídas quando o conjunto de dados é submetido às técnicas adequadas.

1.2. Objetivos

Este trabalho teve como objetivo desenvolver modelos preditivos dos índices de infestação de broca (*Diatraea saccharalis*) na cultura da cana-de-açúcar utilizando árvores de regressão em um conjunto de dados concedido pela Usina Santa Fé, localizada no município de Nova Europa – SP.

1.2.1. Objetivo Geral

Analisar o comportamento da infestação de broca na cana-de-açúcar, relacionando-a às demais variáveis disponíveis no banco de dados da usina.

1.2.2. Objetivos Específicos

- Adequar o conjunto de dados disponibilizado para o uso das técnicas de árvore;
- Realizar análise exploratória do conjunto de dados para identificar atributos mais correlacionados com a infestação da praga;
- Elaborar modelos de predição utilizando árvores de regressão;
- Identificar quais talhões estão mais propensos a infestação para direcionar o levantamento de campo.

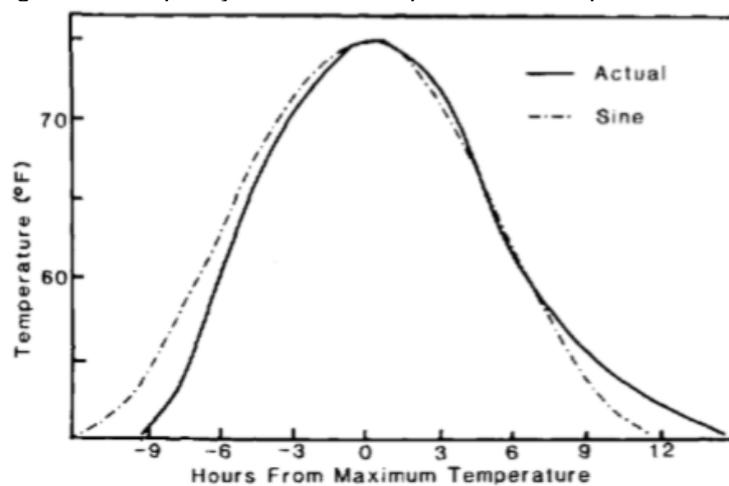
2. REVISÃO DE LITERATURA

2.1. Tempo térmico

Um dos conceitos chave nos estudos de infestação de pragas são os índices bioclimáticos. O conceito de tempo térmico originou-se do fato de que plantas e insetos não são capazes de manter suas temperaturas constantes, por isso, seu desenvolvimento é influenciado pela temperatura do ambiente em que se encontram. Assim, na tentativa de estimar a taxa de crescimento de pragas e viabilizar tomadas de decisão, desenvolveu-se o conceito de tempo térmico que, segundo Bonhomme (2000), mede os “graus-dia” de crescimento dos organismos combinando o tempo e a temperatura ambiente. Para o cálculo de tempo térmico, são utilizadas duas temperaturas: a mínima e a máxima para o desenvolvimento da praga, denominadas temperatura base inferior de crescimento (T_{base}) e temperatura base superior de crescimento (T_{sup}).

Pruess et al. (1983) revisaram e compararam 5 métodos de cálculo de tempo térmico acumulado para manejo de pragas, incluindo o Método do Seno (ou *Sine Wave Method*). Segundo eles, "a principal vantagem deste método é para organismos que respondem à temperatura durante a primavera em climas temperados. Esse é o caso de muitos insetos e, conseqüentemente, um argumento lógico para a utilização do Método do Seno pelos entomologistas. Além disso, tais métodos podem ser associados aos métodos de recorte ("*cutoff*") apresentados por Allen et al. (1976). A Figura 1 ilustra a proximidade entre a curva de temperaturas real e a curva do Método do Seno.

Figura 1: Comparação entre o comportamento do perfil de temperaturas real e senoidal.

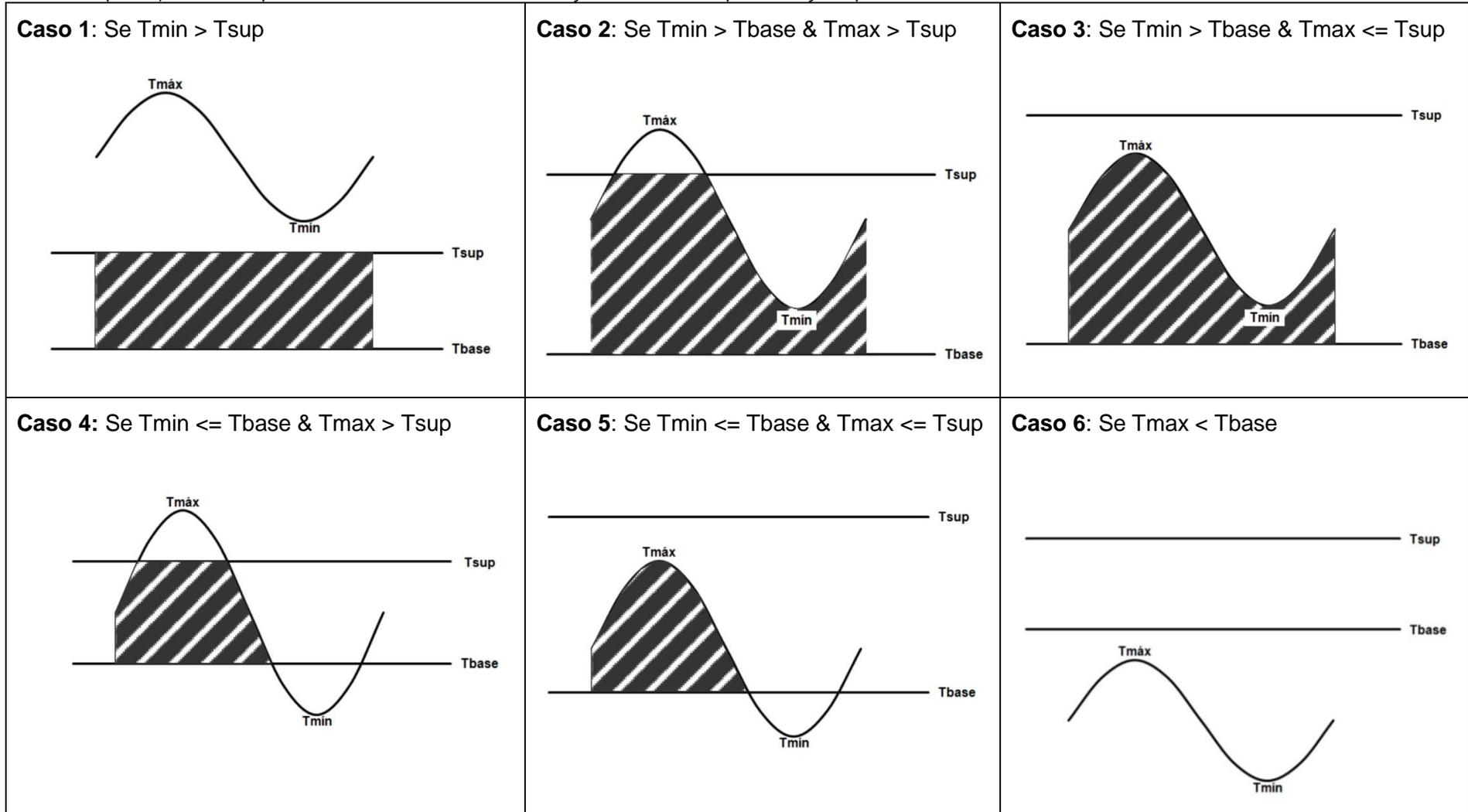


FONTE: PRUESS, 1983.

A determinação de tempo térmico acumulado é realizada através do cálculo da área do perfil de temperatura do dia em relação aos limites de temperaturas de desenvolvimento da praga.

Desta forma, tornam-se possíveis 6 diferentes cenários de estudo, ilustrados na Tabela 1.

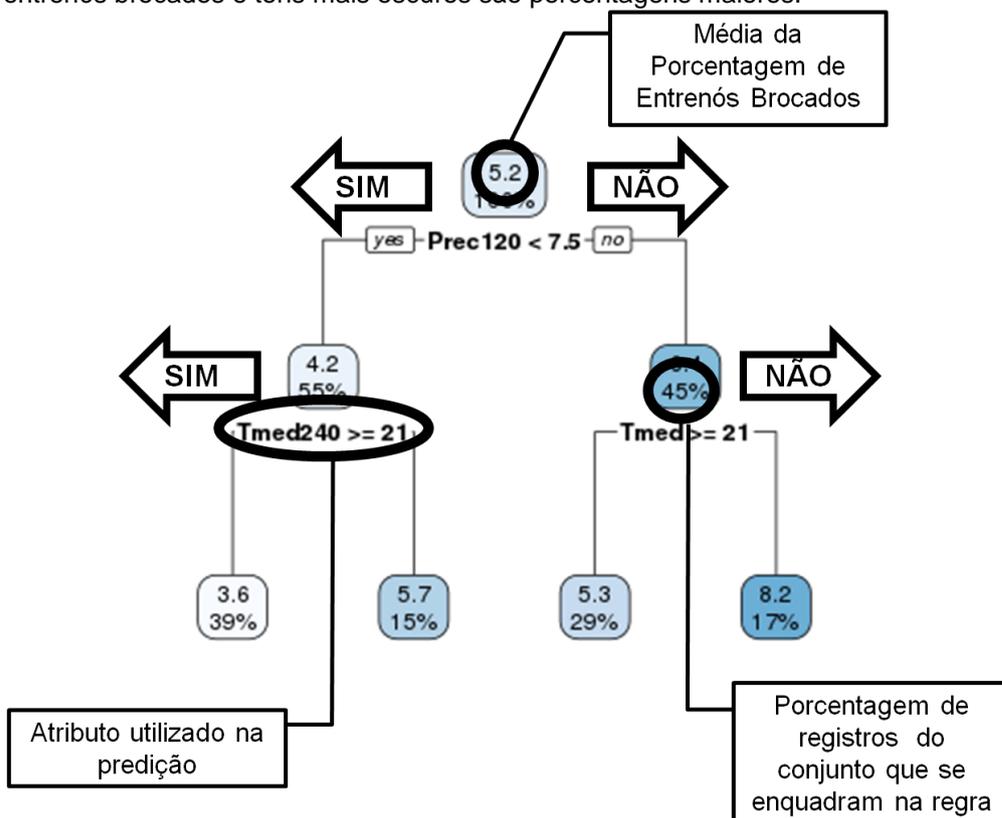
Tabela 1: Acúmulo de graus-dias (área hachurada) calculado a partir dos limites de temperatura ideais para o desenvolvimento da broca (temperaturas de base e superior) e das temperaturas máxima e mínima alcançadas no dia. Representações para o recorte horizontal a ser descrito.



2.2. Árvores de regressão

Os dois principais tipos de tarefas para predição em Mineração de Dados são a classificação e a regressão. A classificação baseia-se na predição de um atributo categórico, o que, para este trabalho, significaria dizer que a infestação da broca seria dada em "alta/média/baixa infestação" ou "sim/não", por exemplo. Já na regressão, o atributo a ser predito possui um valor numérico. Neste caso, a infestação seria representada por um número, que por sua vez, seria relacionado ao grau de infestação. Uma das formas de se elaborar modelos de regressão é utilizando o algoritmo rpart (THERNEAU; ATKINSON; RIPLEY, 2010), implementado no pacote de mesmo nome no software R (BREIMAN et al., 1993). A Figura 2, abaixo, exemplifica a representação dos resultados das árvores provenientes deste algoritmo:

Figura 2: Exemplo de árvore resultante do algoritmo rpart. O primeiro nível é a chamada raiz da árvore, os últimos níveis são denominados folhas. Tons mais claros são menores porcentagens de entrenós brocados e tons mais escuros são porcentagens maiores.



FONTE: A autora (2016).

2.3. Broca (*Diatraea saccharalis*)

A broca é uma mariposa de cor amarela-palha cuja larva penetra nos colmos ocasionando a abertura de galerias como a apresentada na Figura 3. Além dos prejuízos físicos provocados por estes orifícios, como propiciar o acamamento, as galerias favorecem a entrada de fungos que aumentam ainda mais os danos da produção.

Figura 3: Exemplificação da abertura de galerias nos colmos da cana-de-açúcar causada pela broca.



FONTE: DINARDO-MIRANDA; VASCONCELOS; LANDELL, (2010)

Segundo Dinardo-Miranda et al.(2010), a duração dos ciclos de desenvolvimento da broca é bastante variável, tendo ciclos inteiros com duração média de 40 a 70 dias nas condições ambientais do Brasil, podendo ocorrer até quatro gerações anuais. Geremias (2013), em seu estudo sobre as exigências térmicas da praga para a cultura do milho, constatou diferentes durações de ciclos para diferentes temperaturas em laboratório (Tabela 2).

Tabela 2: Resultados encontrados por Geremias (2013) para duração média dos ciclos de desenvolvimento da broca para diferentes temperaturas em laboratório (UR 70 ± 5% e 14h de fotofase) em dieta artificial.

Temperaturas (° C)	Estágios de desenvolvimento						
	Ovo	n	Lagarta	n	Pupa	n	Ovo-adulto*
18	24,01 ± 1,38a	59	86,41 ± 1,85a	66	23,83 ± 0,45a	53	134,25
20	13,50 ± 0,06b	648	45,15 ± 0,73b	85	18,92 ± 0,22b	74	77,58
22	9,06 ± 0,19c	597	37,70 ± 0,69c	97	13,04 ± 0,16c	84	59,79
25	6,53 ± 0,02d	734	27,16 ± 0,43d	97	10,17 ± 0,10d	92	43,87
28	6,13 ± 0,09de	644	24,48 ± 0,36e	94	8,92 ± 0,12e	89	39,53
30	4,58 ± 0,04f	615	20,47 ± 0,18f	100	7,81 ± 0,12e	94	32,86
32	4,44 ± 0,05f	670	21,68 ± 0,48f	93	8,23 ± 0,6e	55	34,34
34	4,59 ± 0,02f	795	37,50 ± 3,56c	8	-	-	-
36	5,26 ± 0,30ef	528	-	-	-	-	-
38	-	578	-	-	-	-	-
40	-	622	-	-	-	-	-
F	122,75		808,72		183,47		
gl	8		7		6		
P	<0,0001		<0,0001		<0,0001		

Médias seguidas pela mesma letra, nas colunas não são significativamente diferentes (Tukey, P<0,05). O número de repetições é dado pelo valor de n. Ovo-adulto* é o somatório das médias das fases de ovo, lagarta e pupa.

FONTE: GEREMIAS (2013)

Um terceiro estudo realizado com ambos gêneros da broca para 9 dietas diferentes (ROE; HAMMOND; SPARKS, 1982), apresenta outras durações de ciclos para as fases de desenvolvimento da praga (Tabela 3).

Tabela 3: Resultados encontrados por Roe, Hammond e Sparks (1982) para duração dos ciclos de desenvolvimento da broca, considerando gêneros e dietas diferentes.

Diet a	Larval		% mortalidade real ^d	Pupal		% mortalidade real ^d	% de contaminação por fungos
	Desenvolvimento (dias)			Peso (mg) ^e			
	♂	♀		♂	♀		
1	35,1a	38,2a	6,2a	75,4a	112,3a	23,4a	0,0a
2	39,0b	42,9b	19,8b	82,8b	132,1b	1,3b	0,0a
3	24,8c	26,7c	3,3a	94,1c	157,7c	3,3b	0,0a
4	23,2cd	25,1cde	5,2a	101,6d	193,6d	4,5b	0,0a
5	21,1e	23,2ef	3,3a	100,9d	166,2ce	2,7b	1,3a
6	23,8cd	25,9cd	2,8a	103,1d	188,0df	0,7b	16,2b
7	29,6f	30,3g	5,2a	102,1d	177,9ef	0,0b	0,6a
8	20,5e	22,3f	1,3a	103,9d	180,9df	1,3b	8,0ab
9	22,0de	23,9def	0,7a	97,7cd	157,6c	0,7b	57,2c

^a Código para diferentes dietas segundo Roe, Hammond e Sparks (1982).

^b Médias derivadas de, no mínimo, 150 brocas examinadas por dieta.

^c Médias onde cada coluna seguida pela mesma letra não é significativamente diferente de $\alpha = 0,05$, pelo procedimento de Tukey.

^d Número de indivíduos mortos na fase larval e pupal, respectivamente, como porcentagem do número incluindo a geração observada.

^e Peso úmido das pupas com um dia de idade.

FONTE: ROE; HAMMOND; SPARKS, (1982).

A comparação dos estudos apresentados acima nos permite concluir que a duração dos ciclos de desenvolvimento da broca é variável e depende principalmente das condições - ambientais e nutricionais - a que estes indivíduos estão submetidos.

2.4. Análise exploratória

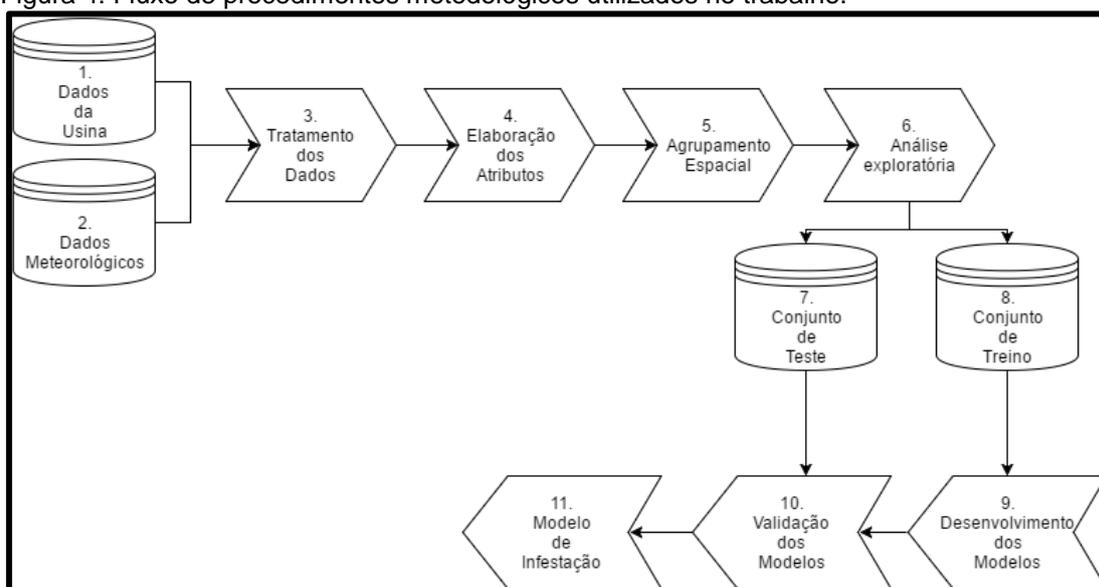
A análise exploratória do conjunto é um processo que auxilia no entendimento dos dados e possivelmente traz melhorias significativas no resultado final. Trata-se de diversas abordagens que visam detalhar e encontrar relações entre os dados para que estes possam ser manuseados e adaptados de maneira otimizada, facilitando a extração de informações úteis.

Steiner et al. (2006) discorrem sobre a influência da análise exploratória em um estudo de caso e concluem que dos 5 métodos utilizados, apenas 1 não apresentou melhoria significativa em seu desempenho com a análise exploratória.

3. METODOLOGIA

Os procedimentos a seguir foram baseados na metodologia CRISP-DM (**C**Ross-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining) (CHAPMAN et al., 2000). O fluxograma da Figura 4 ilustra a sequência das etapas, que será detalhada posteriormente. De modo geral, os dados fornecidos pela Usina e os dados meteorológicos foram tratados, foram criados atributos e foi avaliada a correlação espacial entre eles. O conjunto de dados foi então separado em dois grupos: um de treino, onde foram aplicadas técnicas para criação de modelos, e um de teste, que foi utilizado para validar o modelo gerado mediante métricas pré-estabelecidas - Erro Médio Absoluto (MAE, *Mean Absolute Error*), Erro Médio (ME, *Mean Error*) e Coeficiente de correlação de postos de Spearman (ρ).

Figura 4: Fluxo de procedimentos metodológicos utilizados no trabalho.



FONTE: A autora (2016).

3.1. Caracterização dos dados da usina

O conjunto de dados utilizado foi concedido pela Usina Santa Fé (etapa 1 do fluxograma da Figura 4), localizada no município de Nova Europa – SP, e contém informações das safras da cana-de-açúcar dos períodos 2013/2014, 2014/2015 e 2015/2016, com aproximadamente 5240 registros por safra. Destes, 1764 registros compuseram o conjunto de dados final da modelagem; os demais foram excluídos por motivos apresentados a seguir. No conjunto de dados, constam as seguintes informações:

1. Número de cortes;
2. Dados de insumos, com suas respectivas descrições, quantidades, local e data da aplicação. Registros que não constavam no conjunto de dados de manejo foram excluídos do conjunto de dados de modelagem. Esses talhões são compostos predominantemente por áreas de terceiros que constam no conjunto de dados de colheita, mas sobre os quais não foi disponibilizada informação sobre manejo aplicado na área;
3. Variedades dos cultivares de canas-de-açúcar: Como o conjunto de dados possuía algumas variedades com poucas observações, as variedades com menos de 10 instâncias foram agrupadas na categoria "outras". Ao todo, o conjunto possuía as 46 variedades, listadas no anexo 1;
4. Dados de infestações:
 - a. Identificação da praga: O conjunto de infestação possuía informações de outras pragas além da praga de interesse. Assim, foi trabalhada apenas a planilha da broca, selecionando apenas os atributos relevantes para o projeto;
 - b. Intensidade da infestação (por meio de cana brocada e % de entrenós brocados);

5. Dados de colheita, colheita anterior e plantio das safras 2013/2014, 2014/2015 e 2015/2016. Registros que possuíam informações incompletas ou não consistentes sobre estes três atributos foram removidos. Registros em que o ciclo teve duração de mais de 20 meses e menos de 8 meses também foram removidos.

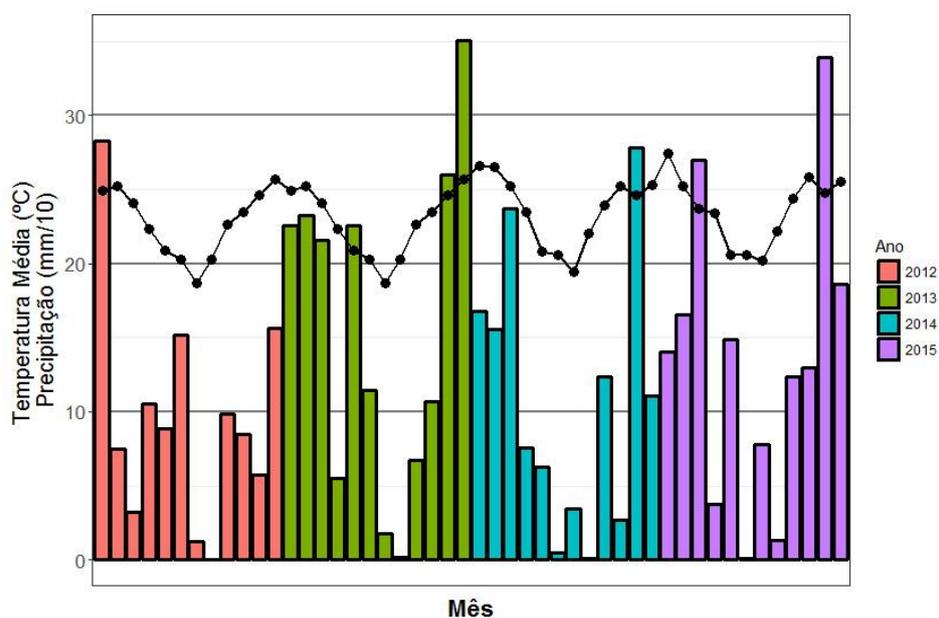
3.2. Caracterização dos dados meteorológicos

Os dados meteorológicos (etapa 2 do fluxograma da Figura 4) foram obtidos através de estações meteorológicas situadas nos arredores da Usina Santa Fé. Tal conjunto de dados é composto pelas seguintes informações diárias - dos anos de 2012, 2013, 2014 e 2015:

1. Temperatura:
 - a. Máxima;
 - b. Média;
 - c. Mínima;
2. Precipitação
 - a. Estação de Nova Europa
 - b. Estação de Ibitinga.

A princípio, os dados meteorológicos foram disponibilizados em um único arquivo, com as informações separadas em abas por ano e por atributos. O conjunto de dados foi unificado para representar os atributos meteorológicos como colunas para entradas diárias por linhas. As médias mensais de temperatura e precipitação do período de estudo estão ilustradas na Figura 5.

Figura 5: Médias mensais de temperatura e precipitação do período entre 01/01/2012 e 31/12/2015.



FONTE: A autora (2016).

3.3. Tratamento dos dados

O tratamento dos dados (etapa 3 do fluxograma da Figura 4) visa maximizar a extração de informações pelas técnicas que serão utilizadas. Este procedimento envolveu a inclusão de valores padrões, retirada de valores irregulares e preenchimento de valores faltantes.

Inicialmente, a modelagem do problema foi realizada de forma a estimar diretamente o nível de infestação. Entretanto, fez-se necessário também a modelagem de graus de infestação. Para isso, discretizou-se o atributo meta das seguintes formas:

- 1ª metodologia:
 - Porcentagem de Entrenós Brocados entre $0 < 2\%$ → categoria "baixo";
 - Porcentagem de Entrenós Brocados entre $2 \leq 5\%$ → categoria "médio";
 - Porcentagem de Entrenós Brocados entre $5 \leq \text{Inf}$ → categoria "alto";
- 2ª metodologia - discretização por quantil:
 - Porcentagem de Entrenós Brocados entre $0 < 1/3$ → categoria "baixo";
 - Porcentagem de Entrenós Brocados entre $1/3 \leq 2/3$ → categoria "médio";
 - Porcentagem de Entrenós Brocados entre $2/3 \leq 1$ → categoria "alto";

Após a discretização, a categoria "médio" foi removida, com o intuito de fazer com que o modelo se especializasse em situações extremas.

3.4. Elaboração dos atributos

Dispondo destas informações referenciadas por datas, foi possível dar início aos cálculos de tempo térmico acumulado e aplicação dos métodos de recorte explicitados a seguir (etapa 4 do fluxograma da Figura 4). Para o cálculo de tempo térmico acumulado da broca, as temperaturas mínima e máxima para o desenvolvimento - T_{base} e T_{sup} - foram de 9,4 e 32,0 °C (DE MELO; PARRA, 1988), respectivamente. Pelo motivo explicitado na seção 2.1, utilizou-se o Método do Seno para realizar estimativas de tempo térmico acumulado Método do Seno Simples (tradução própria de *Single Sine Method*) (PRUESS, 1983), que supõe que as variações de temperatura ao longo do dia se comportam como uma senóide e que cada metade do dia possui a mesma duração.

Todos os casos descritos na seção 2.1 são passíveis de recortes. Os métodos de recorte referem-se à variação da área de cálculo de tempo térmico acumulado em relação ao limite de temperatura superior T_{sup} . Nesta monografia, foram abordados 3 recortes diferentes:

- Sem recorte: Assume que o desenvolvimento da praga é sempre proporcional ao aumento da temperatura;
- Recorte horizontal: Assume que o desenvolvimento continua a uma taxa constante em temperaturas acima da temperatura máxima de desenvolvimento T_{sup} da praga;
- Recorte intermediário: Assume que, para temperaturas acima da T_{sup} , o desenvolvimento da praga é desacelerado ao invés de constante.

A Tabela 4 sintetiza as 18 combinações possíveis entre o cálculo de graus-dia e os métodos de recorte.

Tabela 4: Equações utilizadas para os cálculos dos diferentes tipos de recortes em função dos 6 diferentes cenários.

	Sem Recorte	Recorte Horizontal	Recorte Intermediário
Caso 1	$DD = \frac{T_{sup} - T_{base}}{2} + \frac{T_{máx} - T_{sup}}{2}$	$DD = \frac{T_{sup} - T_{base}}{2}$	$DD = \frac{T_{sup} - T_{base}}{2} - \frac{T_{máx} - T_{sup}}{2}$
Caso 2	$DD = \frac{T_{mín} - T_{base}}{2} + \frac{T_{máx} - T_{mín}}{2}$	$DD = \frac{T_{mín} - T_{base}}{2} + \frac{T_{máx} - T_{mín}}{2} - \frac{T_{máx} - T_{sup}}{2}$	$DD = \frac{T_{mín} - T_{base}}{2} + \frac{T_{máx} - T_{mín}}{2} - 2 \times \frac{T_{máx} - T_{sup}}{2}$
Caso 3	$DD = \frac{T_{mín} - T_{base}}{2} + \frac{T_{máx} - T_{mín}}{2}$	$DD = \frac{T_{mín} - T_{base}}{2} + \frac{T_{máx} - T_{mín}}{2}$	$DD = \frac{T_{mín} - T_{base}}{2} + \frac{T_{máx} - T_{mín}}{2}$
Caso 4	$DD = \frac{T_{máx} - T_{base}}{2}$	$DD = \frac{T_{máx} - T_{base}}{2} - \frac{T_{máx} - T_{sup}}{2}$	$DD = \frac{T_{máx} - T_{base}}{2} - 2 \times \frac{T_{máx} - T_{sup}}{2}$
Caso 5	$DD = \frac{T_{máx} - T_{base}}{2}$	$DD = \frac{T_{máx} - T_{base}}{2}$	$DD = \frac{T_{máx} - T_{base}}{2}$
Caso 6	DD = 0	DD = 0	DD = 0

Para cada tipo de recorte, foram criados atributos que continham os cálculos de tempo térmico acumulado nos 120, 240 e 360 dias anteriores à colheita. O mesmo raciocínio foi utilizado para a precipitação. Para as temperaturas, também foram criados atributos de dias anteriores à colheita, mas não mais do acúmulo e sim da média. As precipitações das estações de Ibitinga e Nova Europa foram unificadas da seguinte forma: para registros mais próximos à estação de Ibitinga, foram utilizados os dados de precipitação de Ibitinga. Para os mais próximos à Nova Europa, foram utilizados os dados de precipitação de Nova Europa. Para registros sem informações de coordenadas do talhão, foi utilizada a média das precipitações.

Para os insumos, foram criados atributos que contemplavam:

1. Separação dos produtos por tipo de alvo (produtos aplicados visando: broca, pragas de solo, cigarrinha e nematoides).;
2. Normalização por produto, para cada produto (divisão pelo valor máximo aplicado para cada produto);
3. Segmentação: Foram somadas as aplicações feitas por bloco de 120 dias;

3.5. Agrupamento espacial

Na metodologia proposta inicialmente para o projeto, o agrupamento espacial (etapa 5 do fluxograma da Figura 4) se valeria das distâncias entre os talhões. Isso não foi realizado dado que o contorno georreferenciado da área não contemplava todos os registros do conjunto de dados. Sendo assim, os talhões que foram colhidos no mesmo bloco foram identificados, utilizando informações do conjunto de dados.

Esta informação foi utilizada para estratificar os conjuntos de treino/teste e os 10 *folds* da validação cruzada, ou seja, registros que pertenciam a um bloco eram atribuídos exclusivamente ao conjunto de treino ou de teste.

3.6. Análise exploratória

A análise exploratória (etapa 6 do fluxograma da Figura 4) foi realizada com o conjunto de dados antes que este fosse separado em conjunto de treino e teste. Utilizando histogramas, boxplots e gráficos de dispersão observou-se a distribuição das observações em função do atributo meta, dos dias da colheita e do número de cortes, além da análise da duração do ciclo de desenvolvimento de cada fase.

3.7. Divisão dos conjuntos de treino e teste

O conjunto de teste (etapa 7 do fluxograma da Figura 4) foi composto por 31% das observações do conjunto de dados, totalizando 550 registros. O conjunto de treino (etapa 8 do fluxograma da Figura 4) foi composto por 69% das observações do conjunto de dados, totalizando 1214 registros.

3.8. Ajuste de hiperparâmetros dos algoritmos

A configuração utilizada para o desenvolvimento dos modelos (etapa 9 do fluxograma da Figura 4) foi a seguinte:

1. Número de folds: 10. Valor recomendado por Friedman; Hastie; Tibshirani (2001);
2. CP (*complexity parameter*): 10^{-4} . Optou-se por um valor mínimo neste hiperparâmetro para garantir maior "grau de liberdade" para as árvores geradas;
3. Profundidade máxima: 2 e 3. Restringiu-se a profundidade da árvore para que fosse mantida a simplicidade do modelo, visto que quanto maior o número de regras na árvore, maior a dificuldade de interpretação do modelo;
4. Número mínimo de observações por folha: 60, 120, 180 e 240. Definiram-se estes valores (aproximadamente 5, 10, 15 e 20 % do conjunto de dados) para avaliar diferentes cenários entre casos mais e menos comuns.

3.9. Validação do modelo

Para a validação do modelo (etapa 10 do fluxograma da Figura 4), foram utilizadas as seguintes medidas de avaliação de desempenho:

- Erro Absoluto Médio (MAE - *Mean Absolute Error*);

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

- Erro Médio (ME - *Mean Error*)

$$ME = \frac{\sum_{i=1}^n y_i - \hat{y}_i}{n}$$

- Coeficiente de correlação de postos de Spearman (ρ);

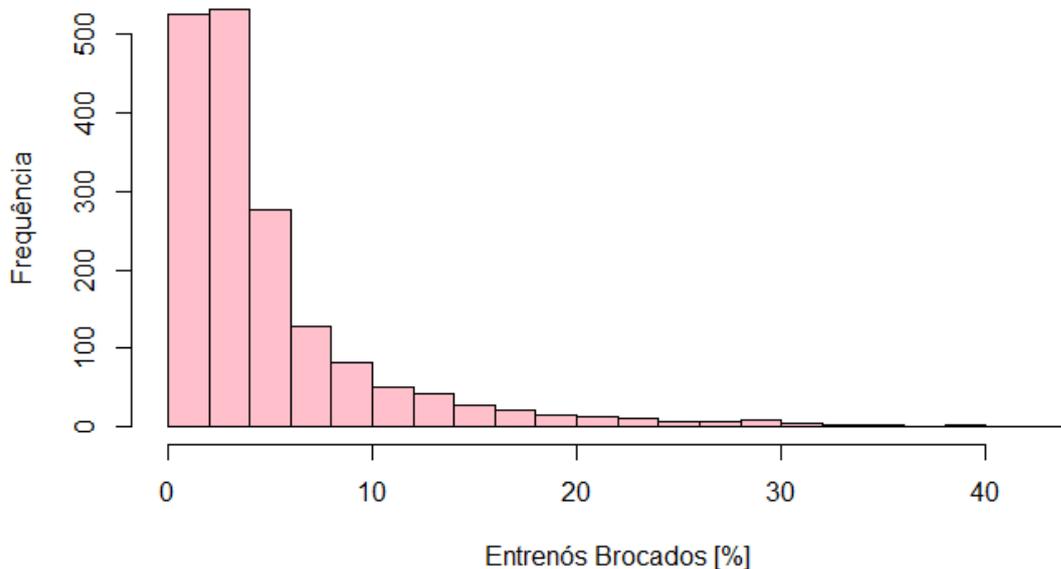
$$\rho = 1 - \frac{6 \sum d_i^2}{(n^3 - n)}$$

4. RESULTADOS E DISCUSSÃO

4.1. Análise exploratória

Inicialmente, buscou-se mais informações sobre o atributo meta - Entrenós Brocados (%) - e verificou-se que aproximadamente 88% do conjunto de dados enquadravam-se em até 10% de entrenós brocados, 70% dos dados enquadravam-se em até 5% de infestação e 47% do conjunto estudado apresentou infestação abaixo do nível de dano econômico (3%) (MACEDO; MACEDO, 2004). Estes níveis de infestação estão apresentados na Figura 6.

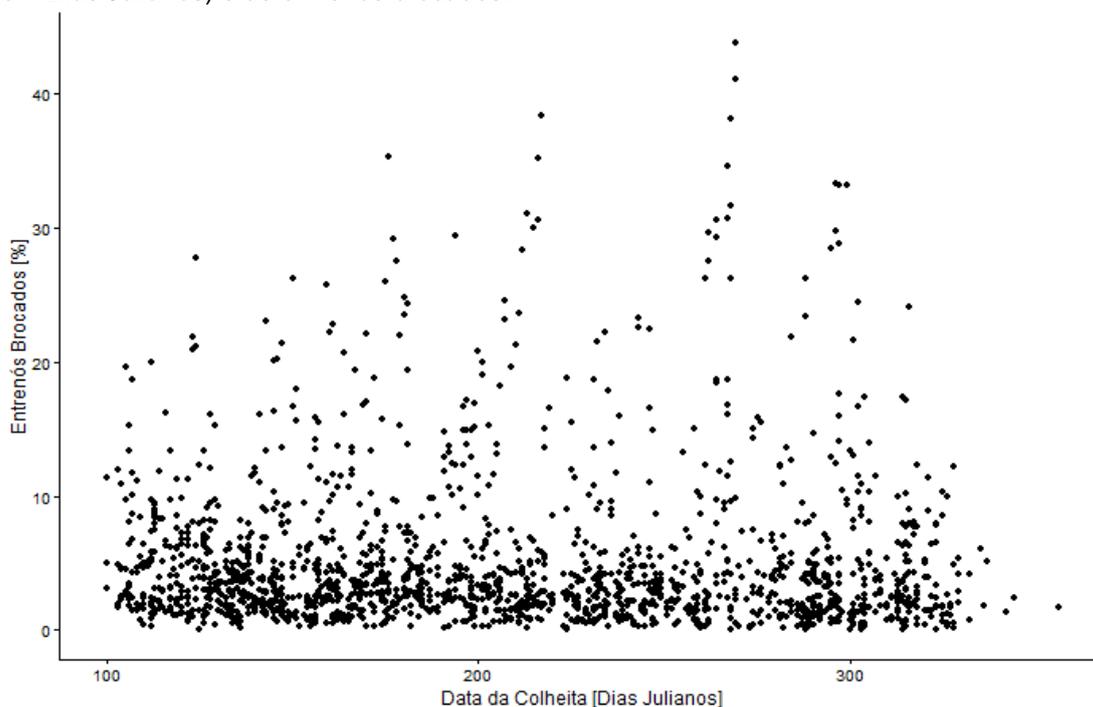
Figura 6: Intensidade de infestação (em % de entrenós brocados) obtida a partir do conjunto de dados utilizado na modelagem.



FONTE: A autora (2016).

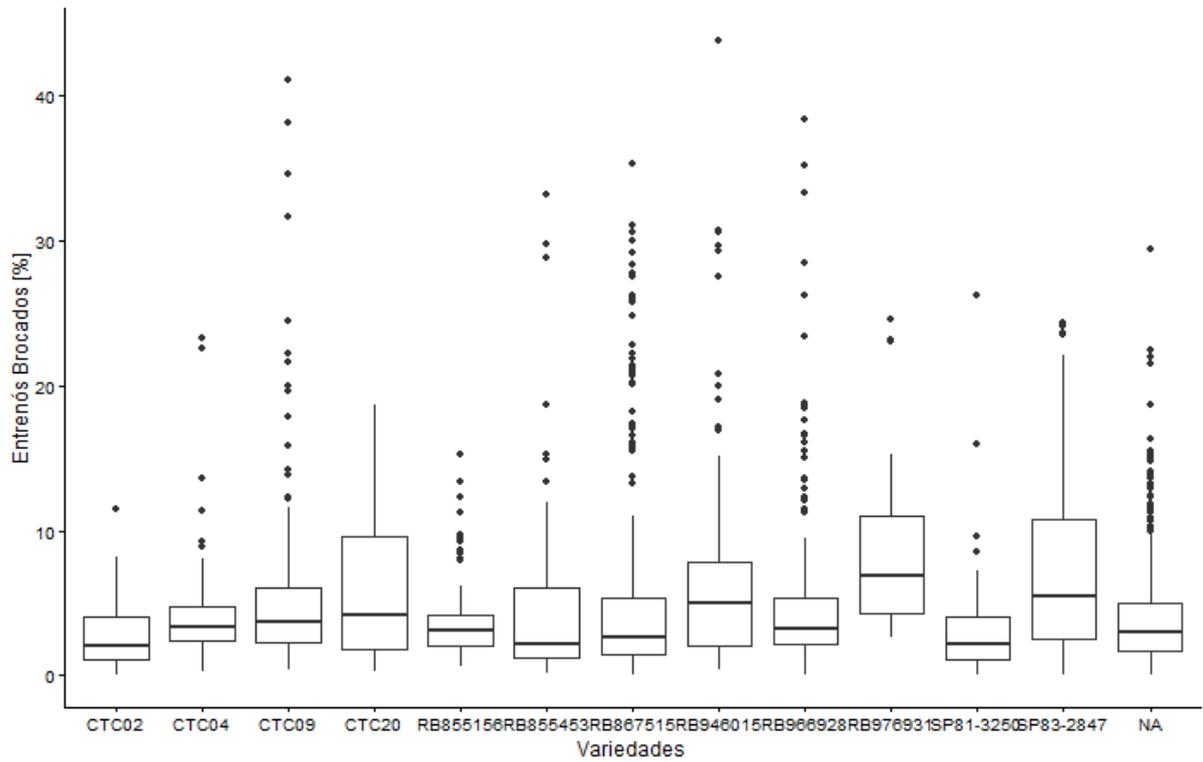
Posteriormente, foi avaliada em um subconjunto a variação da população de broca, estimada por Entrenós Brocados (%), em relação a outras variáveis. Destas primeiras análises, não foi possível observar nenhum padrão que indicasse relação direta entre o atributo meta e a data de corte (Figura 7). Para variedade (Figura 8), verificou-se que algumas eram mais resistentes à infestação que outras, constatação já obtida em trabalhos anteriores como o de Dinardo-Miranda et al.(2012).

Figura 7: Distribuição das observações contidas no conjunto de dados em função da data de corte (em Dias Julianos) e de entrenós brocados.



FONTE: A autora (2016).

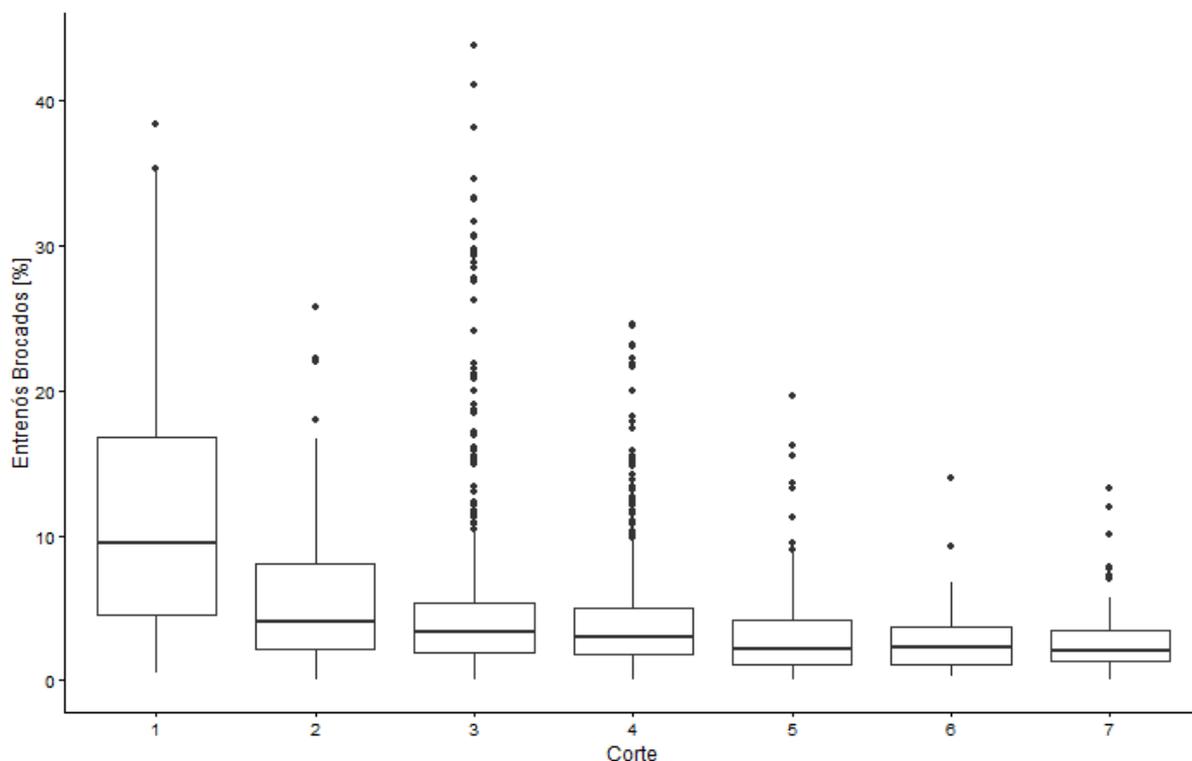
Figura 8: Distribuição das observações contidas no conjunto de dados em função da variedade e de entrenós brocados.



FONTE: A autora (2016).

Em relação aos cortes (Figura 9), observou-se que canas mais jovens (ou seja, com menor número de cortes) apresentavam maiores índices de infestação, o que pode ser explicado pelo maior vigor vegetativo se comparada com as canas-soca (DINARDO-MIRANDA; VASCONCELOS; LANDELL, 2010).

Figura 9: Distribuição das observações contidas no conjunto de dados em função do número de cortes e de entrenós brocados.



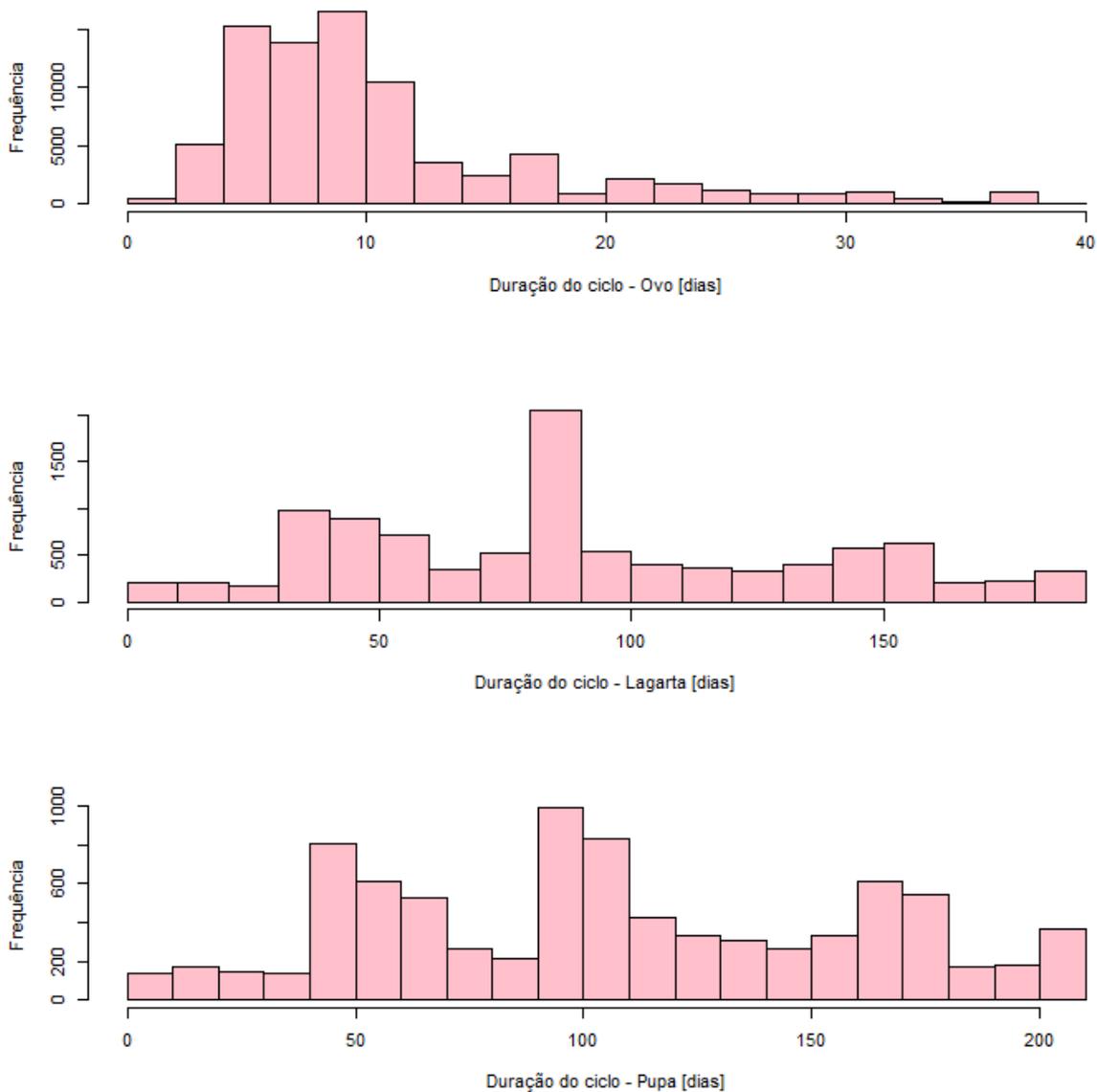
FONTE: A autora (2016).

Baseado nas exigências térmicas relatadas por Melo e Parra (1988), foi analisada a duração do ciclo - em dias - de cada fase de desenvolvimento da broca. As durações dos ciclos da fase ovo, lagarta e pupa tiveram diferentes durações para os diferentes recortes. Os resultados encontram-se na Tabela 5 e Figura 10:

Tabela 5: Duração dos ciclos de cada fase de desenvolvimento da broca em relação ao tipo de recorte aplicado.

	Sem recorte	Recorte horizontal	Recorte intermediário
Ciclo - Ovo	11	10	11
Ciclo - Lagarta	90	83	85
Ciclo - Pupa	108	99	102

Figura 10: Distribuição das durações dos ciclos de desenvolvimento (ovo, lagarta, pupa) da broca, em dias, sem recorte.



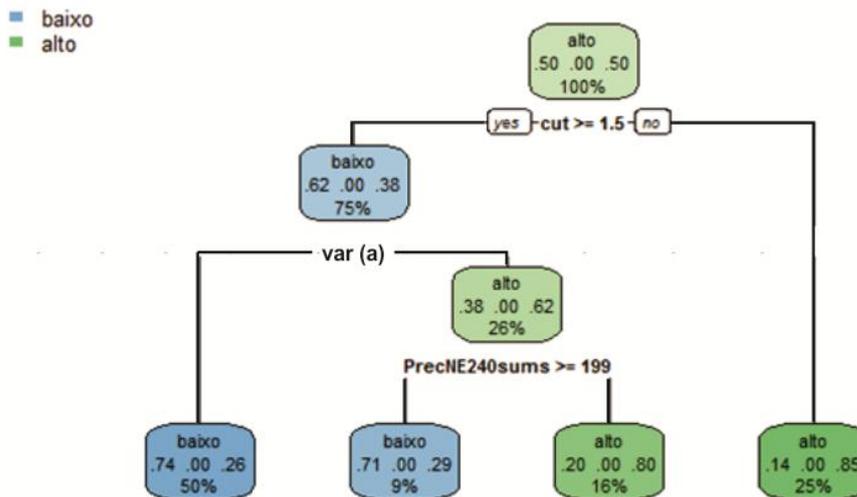
FONTE: A autora (2016).

Embora a duração dos ciclos possua alta variabilidade dependendo das condições em que o desenvolvimento da praga esteja inserido, os resultados obtidos a partir do conjunto de dados são semelhantes aos resultados encontrados na literatura (seção 2.3), comprovando a coerência do conjunto de dados.

4.2. Modelagem preliminar

Uma modelagem preliminar utilizando o conjunto de treino chamou a atenção para os RMSEs (*Root Mean Squared Error*) insatisfatórios. Por este motivo, verificou-se se a modelagem utilizando árvores de classificação obteriam melhores resultados. Entretanto, como a hipótese não foi constatada, a metodologia com a validação cruzada foi realizada apenas com árvores de regressão. Algumas das principais árvores de classificação e regressão encontradas estão apresentadas na Figura 11 e 12. A Figura 2 da seção 2.2 exemplifica como interpretar as árvores apresentadas a seguir, tanto para regressão quanto para classificação.

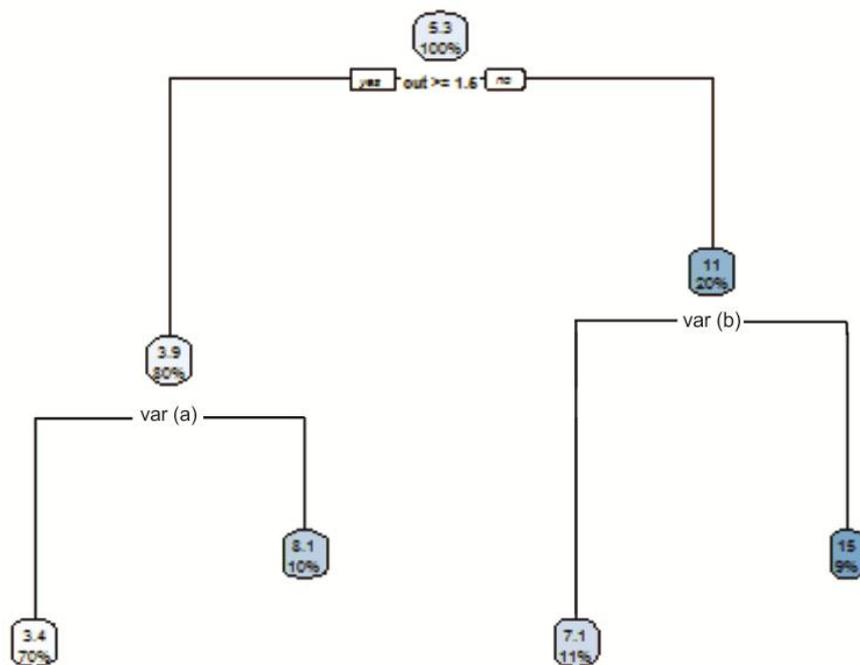
Figura 11: Exemplo de árvore de classificação gerada pelo algoritmo rpart nas modelagens preliminares utilizando a 1ª metodologia.



var(a): Refere-se às variedades CTC02, CTC15, CTC20, CTC24, RB855156, RB855453, RB935744, RB965902 e SP81-3250.

FONTE: A autora (2016).

Figura 12: Exemplo de árvore de regressão gerada pelo algoritmo rpart nas modelagens preliminares.



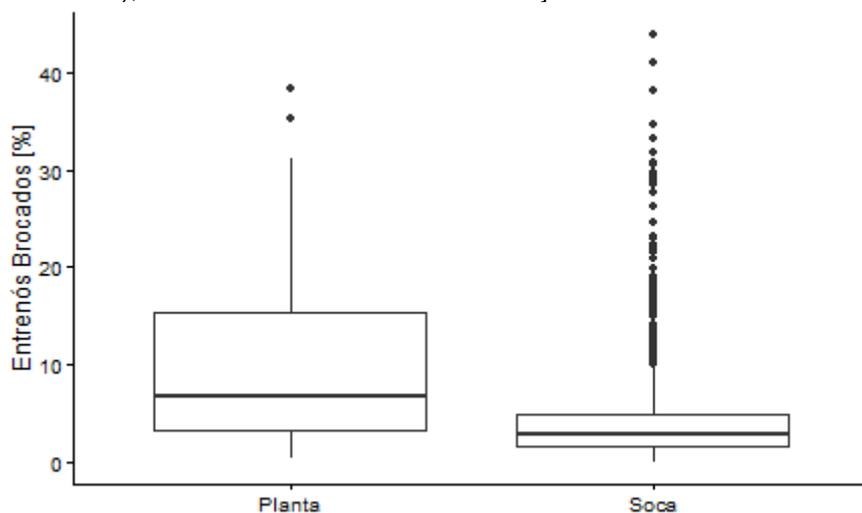
var (a): Refere-se às variedades CTC02, CTC04, CTC09, CTC20, CTC24, IACSP95-5000, outras, RB855156, RB855453, RB887515, RB935744, RB965902, RB855453 e SP81-3250.

var (b): Refere-se às variedades CTC02, CTC20, IACSP95-5000, outras, RB855156, RB965902, RB966928 e RB975952.

FONTE: A autora (2016).

Após análise das árvores resultantes da análise exploratória, observou-se que o corte prevalecia como principal atributo preditivo e, após buscar justificativa, constatou-se que cortes mais baixos predominam no conjunto de dados (Figura 9) e por isso o atributo de corte se destacava nas árvores. Visando minimizar este efeito, o corte foi retirado no conjunto de dados para a modelagem final.

Figura 13: Proporções entre o número de registros com um corte (cana-planta) e mais que dois cortes (cana-soca), relacionados ao índice de infestação - em % de entrenós brocados.

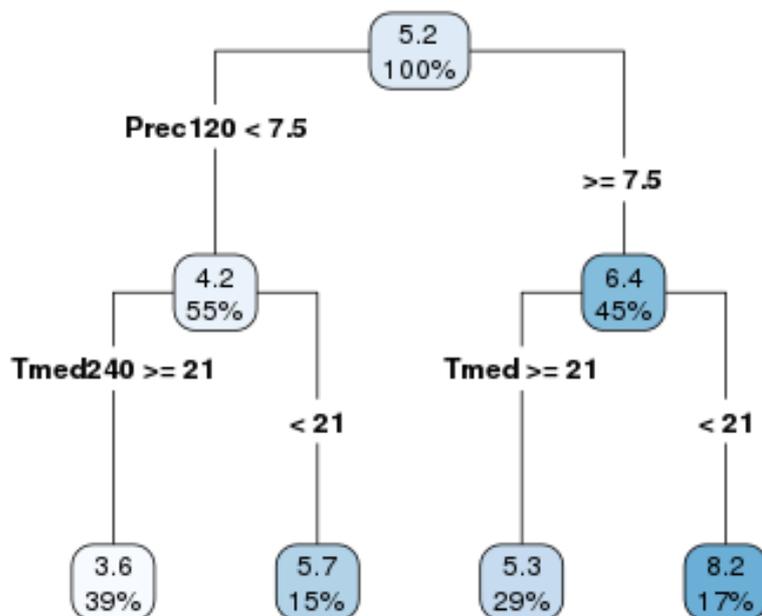


FONTE: A autora (2016).

4.3. Modelagem

Das 60 árvores geradas, 3 principais foram selecionadas baseando-se nos desempenhos - aferidos pelos indicadores descritos anteriormente - e nas considerações qualitativas, como simplicidade do modelo e utilidade em tomada de decisões. Tais árvores encontram-se ilustradas a seguir, com suas respectivas métricas:

Figura 14: Primeira árvore selecionada dentre os 60 modelos finais obtidos.



FONTE: A autora (2016).

Tabela 6: Erro Médio Absoluto (MAE), Erro Médio (ME) e Coeficiente de correlação de postos de Sperman (ρ) para os conjuntos de treino e teste da árvore representada na Figura 14.

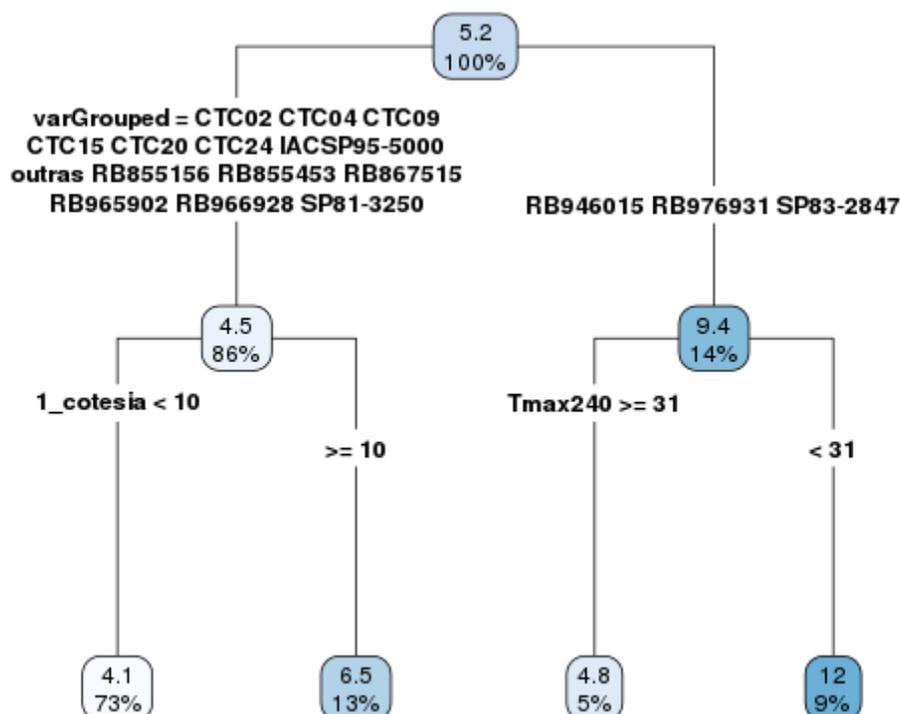
	MAE	ME	ρ
Treino	3,21	0,00	0,23
Teste	3,26	-0,17	0,23

Das 60 árvores finais, a apresentada na Figura 14 foi a que possuiu maior ρ . Entretanto, a diferença entre os erros desta e das demais árvores não é expressivo. Observa-se que os atributos preditivos deste modelo são estritamente meteorológicos e que a temperatura de corte para ambos os ramos foi igual, com valor de 21 °C, indicando que temperaturas médias diárias superiores a 21°C podem reduzir o grau de infestação, possivelmente por estarem associadas a altas temperaturas diurnas. Ressalta-se que os atributos meteorológicos compuseram todas as 3 árvores selecionadas.

Os atributos meteorológicos elaborados de períodos anteriores estiveram presentes em 2 das 3 modelagens e, analisando as 60 árvores, constatou-se que em todos os modelos que possuíam atributos meteorológicos do período corrente e de períodos anteriores, os atributos de períodos anteriores apareceram em níveis mais altos (mais próximos à raiz da árvore) que os atributos do período corrente. Para exemplificar tal padrão, vide a Figura 14, em que a Precipitação de 120 dias anteriores aparece em nível superior se comparado à Temperatura Média. Isso pode implicar que a ocorrência de níveis de temperaturas em períodos bem anteriores à colheita leva a um grau de infestação que se intensifica ou não dependendo das condições mais próximas à colheita.

Temperaturas elevadas apareceram frequentemente nos modelos, indicando que temperaturas acima de determinados valores implicam em baixos níveis de entrenós brocados. Ressalta-se também a proximidade entre os valores de corte das temperaturas máximas (31 e 28 °C nas Figuras 15 e 16, respectivamente) e a temperatura superior de desenvolvimento da broca (32 °C). Isso reforça a hipótese de que temperaturas acima da T_{sup} são desfavoráveis ao desenvolvimento da praga e respalda o fato de que se temperaturas elevadas são relevantes, os atributos referentes aos tempos térmicos acumulados também deveriam ter entrado nas modelagens.

Figura 15: Segunda árvore selecionada dentre os 60 modelos finais obtidos.



FONTE: A autora (2016).

Tabela 7: Erro Médio Absoluto (MAE), Erro Médio (ME) e Coeficiente de correlação de postos de Sperman (ρ) para os conjuntos de treino e teste da árvore representada na Figura 15.

	MAE	ME	ρ
Treino	3,00	0,00	0,35
Teste	3,51	-0,06	0,13

Já a segunda árvore selecionada (Figura 15), embora possua menor coeficiente de correlação de postos de Spearman, utiliza atributos mais interessantes para predição, como variedade e indivíduos de cotésia, que são parâmetros passíveis de modificação e controle por parte da Usina.

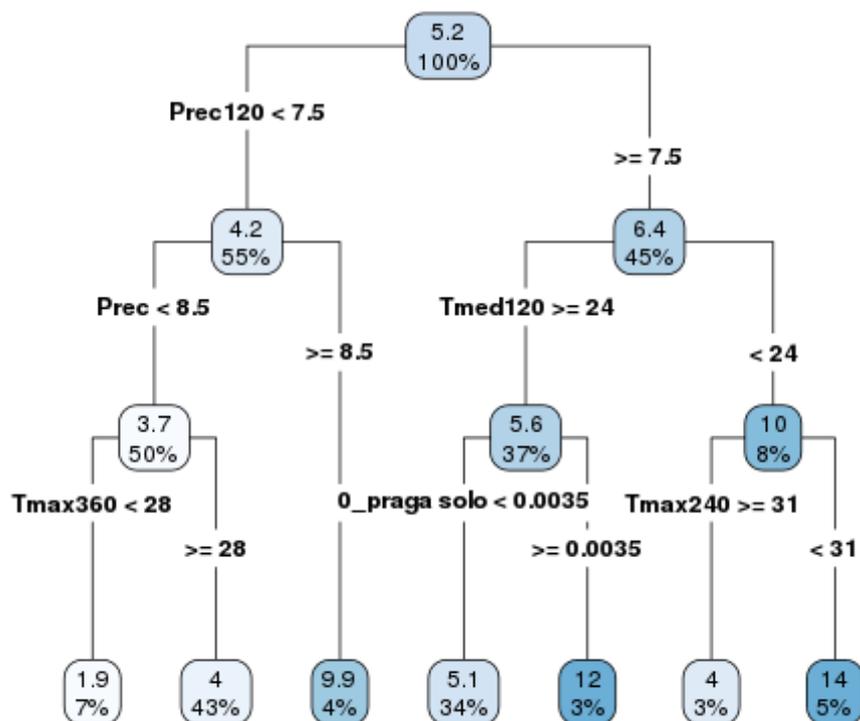
Em relação às variedades, observou-se que das 46 iniciais, 3 eram mais susceptíveis à infestação, sendo elas a variedade SP83-2847 e os cultivares RB946015 e RB976931, também ilustrado na Figura 8 anteriormente. A característica comum entre os 3 mais susceptíveis é a alta produtividade.

A cotésia foi utilizada no segundo modelo principal, com um valor de corte de 10 copos. É válido ressaltar que, segundo o ramo esquerdo do modelo da Figura 15, se o número de indivíduos de cotésia for menor que 10 copos, a porcentagem de entrenós brocados é menor; se o número de indivíduos é maior ou igual a 10 copos, a porcentagem de entrenós brocados também é maior.

Uma justificativa plausível para este resultado é que quando a infestação é relativamente baixa, menos indivíduos de cotésia são capazes de controlá-la. Por outro lado, quando a infestação já está em níveis mais elevados, por mais que sejam aplicados mais indivíduos de cotésia, estes não são capazes de conter e/ou reduzir a população da broca. A partir da análise da Figura 15, constata-se que o valor de corte em que o controle com cotésia deixa de ser eficaz é a partir de 10 copos.

Em um estudo sobre o impacto de *Cotesia flavipes* como um agente de biocontrole para a broca da cana-de-açúcar em arroz, Lv et al. (2011) constataram que a incapacidade da cotesia para proporcionar controle econômico em áreas com temperaturas subtropicais é devido ao seu alto custo de criação, a baixa taxa de pesquisa efetiva, o número máximo de hospedeiros parasitados por fêmeas e a falha de parasitoides emergentes na primavera para encontrar hospedeiros.

Figura 16: Terceira árvore selecionada dentre os 60 modelos finais obtidos.



FONTE: A autora (2016).

Tabela 8: Erro Médio Absoluto (MAE), Erro Médio (ME) e Coeficiente de correlação de postos de Sperman (ρ) para os conjuntos de treino e teste da árvore representada na Figura 16.

	MAE	ME	ρ
Treino	2,93	0,00	0,39
Teste	3,40	-0,13	0,17

Em linhas gerais, as métricas das árvores apresentadas acima, por possuírem valores próximos entre treino e teste, permitem concluir que os modelos não sofreram *overfitting*, ou seja, eram modelos genéricos o suficiente para representar o fenômeno e não estavam demasiadamente ajustados para o conjunto. Esta informação também pode ser verificada na Tabela 9, que contém a média das métricas das 60 árvores de regressão.

Tabela 9: Médias das métricas Erro Médio Absoluto (MAE), Erro Médio (ME) e Coeficiente de correlação de postos de Sperman (ρ) para os 60 modelos preditivos finais obtidos.

	MAE	ME	ρ
Treino	3,10	0,00	0,29
Teste	3,36	-0,01	0,20

Todavia, os valores obtidos não são satisfatórios para modelos preditivos, pois, conforme ilustrado no início do trabalho (Figura 6), 70% das observações do conjunto de dados encontram-se abaixo de 5% de entrenós brocados; logo, errar em 3,36% (MAE do conjunto de teste) implica praticamente dobrar a porcentagem de entrenós brocados. Analisando o erro médio, é possível concluir que os modelos subestimavam o conjunto de teste (ME <0). Nenhum dos 60 modelos finais utilizou os atributos de graus-dia para predição e durante o trabalho, observou-se que não havia distinção relevante entre os três métodos de recorte avaliados. Analisando a Tabela 10 a seguir, nota-se que os valores de correlação de Spearman entre o atributo meta e os atributos de tempo térmico possuem baixo valor absoluto. Isso indica que o relacionamento entre tais variáveis é pequeno, o que justifica a ausência destes atributos no modelo preditivo. Para os casos de valores negativos, a relação é inversamente proporcional. É possível que o conjunto de dados contendo somente os valores finais, ao invés de níveis de infestação que caracterizem o ciclo todo, não favoreça a exploração deste atributo baseado na biologia da praga.

Tabela 10: Correlação entre os atributos de tempo térmico (do período atual e de períodos anteriores) e o índice de infestação (em % de entrenós brocados).

	0 dias	120 dias	240 dias	360 dias
Sem recorte	-0.01	0.00	-0.08	-0.08
Recorte horizontal	0.00	-0.01	-0.09	0.01
Recorte intermediário	0.01	-0.01	-0.10	0.04

O período de 4 meses (120 dias) foi o mais recorrente nos modelos de regressão, além de ser o maior valor de correlação se comparado aos demais períodos. Embora o período de 240 dias também tenha composto alguns modelos, sempre que atributos de ambos períodos compunham a árvore, o atributo referente a 120 dias aparecia em um nível superior (Figura 16), mostrando-se mais relevante para a modelagem.

5. CONCLUSÕES

Conclui-se que, embora os atributos de tempo térmico não estejam inseridos nos modelos finais, a frequente presença das temperaturas elevadas e das precipitações indica que os atributos meteorológicos são determinantes para o índice de infestação. Os atributos meteorológicos de períodos de 4 meses anteriores à colheita mostraram-se relevantes.

Constata-se também que as situações críticas para infestação de broca ocorrem na variedade SP83-2847, nos cultivares RB946015 e RB976931 e para canaviais em seu primeiro ciclo (canas-planta), devendo assim priorizar os levantamentos de infestação e atentar o manejo em talhões com estas características. É recomendável comparar os manejos de tais talhões aos demais com o objetivo de detectar se há alguma diferença que possa causar índices de infestação diferentes.

Pela análise dos indivíduos de cotésia, observou-se que o controle biológico só é efetivo quando praticado em níveis de infestação de broca ainda passíveis de controle. Uma vez extrapolado o índice de infestação, por mais que sejam aplicados mais copos de cotésia, não será possível reduzir a infestação pela broca. O estudo constatou que a partir de 10 copos de cotésia o controle da infestação já não é mais eficaz.

6. REFERÊNCIAS

- ALLEN, J. C. A modified sine wave method for calculating degree days. **Environmental Entomology**, v. 5, n. 3, p. 388–396, 1976.
- BONHOMME, R. Bases and limits to using “degree. day”units. **European journal of agronomy**, v. 13, n. 1, p. 1–10, 2000.
- BREIMAN, L. et al. Classification and Regression Trees. Wadsworth, 1984. **Intelligence**, p. 1002–1007, 1993.
- CHAPMAN, P. et al. **CRISP-DM 1.0 - Step-by-step data mining guide**. [s.l.] SPSS Inc., 2000.
- CONAB, C. N. DE A. **Monitoramento agrícola – Cana-de-açúcar – Safra 2015/16**, 26 fev. 2016. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_02_23_17_34_53_boletim_cana_portugues_-_3o_lev_-_15-16.pdf>. Acesso em: 26 fev. 2016
- DE MELO, A. B.; PARRA, J. R. Exigências térmicas e estimativa do número de gerações anuais de broca-da-cana-de-açúcar em quatro localidades canavieiras de São Paulo. **Pesquisa Agropecuária Brasileira**, v. 23, n. 7, p. 691–696, 1988.
- DE S. ROSSATO, J. A. et al. Characterization and Impact of the Sugarcane Borer on Sugarcane Yield and Quality. **Agronomy Journal**, v. 105, n. 3, p. 643, 2013.
- DINARDO-MIRANDA, L. L. et al. Resistance of sugarcane cultivars to *Diatraea saccharalis*. **Pesquisa Agropecuária Brasileira**, v. 47, n. 1, p. 1–7, jan. 2012.
- DINARDO-MIRANDA, L. L.; FRACASSO, J. V.; PERECIN, D. Spatial variability of *Diatraea saccharalis* populations in sugarcane fields and suggestion of sampling method. **Bragantia**, v. 70, n. 3, p. 577–585, 2011.
- DINARDO-MIRANDA, L. L.; VASCONCELOS, A. C. M. DE; LANDELL, M. G. DE A. **Cana-de-açúcar**. 1ª ed. Campinas: IAC: [s.n.].
- FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The elements of statistical learning**. [s.l.] Springer series in statistics Springer, Berlin, 2001. v. 1
- GEREMIAS, L. D. **Exigências térmicas e danos de *Diatraea saccharalis* (Fabricius, 1794) (Lepidoptera: Crambidae) em milho e base para seu controle com *Trichogramma galloi* Zucchi, 1988 (Hymenoptera: Trichogrammatidae)**.
- LV, J. et al. Impact of *Cotesia flavipes* (Hymenoptera: Braconidae) as an augmentative biocontrol agent for the sugarcane borer (Lepidoptera: Crambidae) on rice. **Biological Control**, v. 56, n. 2, p. 159–169, 2011.
- MACEDO, N.; MACEDO, D. As pragas de maior incidência nos canaviais e seus controles. 2004.
- MAPA. **Cana-de-açúcar**. Disponível em: <<http://www.agricultura.gov.br/portal/page/portal/Internet-MAPA/pagina-inicial/vegetal/culturas/cana-de-acucar>>. Acesso em: 3 jun. 2016.

PRUESS, K. P. Day-degree methods for pest management. **Environmental Entomology**, v. 12, n. 3, p. 613–619, 1983.

ROE, R. M.; HAMMOND, A. M.; SPARKS, T. C. Growth of larval *Diatraea saccharalis* (Lepidoptera: Pyralidae) on an artificial diet and synchronization of the last larval stadium. **Annals of the Entomological Society of America**, v. 75, n. 4, p. 421–429, 1982.

STEINER, M. T. A. et al. Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados. **Gest Prod**, v. 13, n. 2, p. 325–37, 2006.

TERÁN, F. O.; SÁNCHEZ, A. G.; PRECETTI, A. A. C. M. Estudos sobre resistencia da cana a broca em telado: primeiros resultados. **Boletim Técnico COPERSUCAR**, 1985.

THERNEAU, T. M.; ATKINSON, B.; RIPLEY, B. rpart: Recursive Partitioning. R package version 3.1–42. **Computer software program retrieved from <http://CRAN.R-project.org/package=rpart>**, 2010.

7. ANEXO

Anexo 1: Número de registros por variedade utilizados para modelagem.

Variedade	Número de registros	Variedade	Número de registros
RB966928	225	CTC16	4
RB867515	165	RB72454	4
CTC09	98	RB855536	4
CTC02	92	RB92579	4
SP83-2847	83	CTC07	3
RB855156	81	IACSP94-2101	3
SP81-3250	77	PAV94-09	3
RB946015	62	RB835486	3
RB855453	51	CTC11	2
CTC20	49	RB925211	2
CTC04	32	RB945961	2
RB976931	25	SP80-1842	2
RB965902	24	SP80-3280	2
CTC15	21	CT95-1445	1
IACSP95-5000	17	CTC14	1
CTC24	16	CTC17	1
IAC87-3396	8	CTC19	1
RB935744	8	DIVERSAS	1
VIVEIRO	8	RB975952	1
CTC22	7	SP80-1816	1
IAC91-1099	6	SP87-365	1
RB965911	6	SP89-1115	1
CT99-1906	5	SP91-1049	1