

**FACULDADE DE TECNOLOGIA DE SÃO JOSÉ DOS CAMPOS  
FATEC PROFESSOR JESSEN VIDAL**

**ANA BEATRIZ SILVA DE ARAUJO LEITE**

RA 1460481821005 - 6º ADS

[ana.leite12@fatec.sp.gov.br](mailto:ana.leite12@fatec.sp.gov.br)

**CRISTIANO DONIZETE RIBEIRO**

RA 1461141521011 - 6º AMD

[cristiano.ribeiro@fatec.sp.gov.br](mailto:cristiano.ribeiro@fatec.sp.gov.br)

**GISELE BARBA DE LIMA LAPA**

RA 1460481811024 - 6º ADS

[gisele.lapa@fatec.sp.gov.br](mailto:gisele.lapa@fatec.sp.gov.br)

**THIAGO FRANCISCO**

RA 1460481811085 - 6º ADS

[thiagofrancisco3@fatec.sp.gov.br](mailto:thiagofrancisco3@fatec.sp.gov.br)



**PROJETO KERSYS**

**Orientador: M.e.Prof. José Walmir Gonçalves Duque  
Disciplina Inteligência Artificial**

São José dos Campos  
2020

## Sumário

<b>1. INTRODUÇÃO .....</b>	<b>3</b>
<b>2. KERSYS – SOLUÇÕES PARA GESTÃO FLORESTAL E AGRICULTURA .....</b>	<b>4</b>
• <b>A EMPRESA.....</b>	<b>4</b>
• <b>SOLUÇÕES .....</b>	<b>5</b>
• <b>KIA - KERSYS INTELIGÊNCIA ARTIFICIAL .....</b>	<b>6</b>
• <b>DESAFIO .....</b>	<b>6</b>
• <b>REQUISITOS .....</b>	<b>6</b>
<b>3. BASE DE DADOS .....</b>	<b>7</b>
• <b>DISPOSIÇÃO DO TALHÃO .....</b>	<b>7</b>
• <b>ANÁLISE DOS DADOS .....</b>	<b>8</b>
• <b>PREPARAÇÃO DOS DADOS .....</b>	<b>12</b>
<b>4. Modeling e Evaluation .....</b>	<b>15</b>
<b>6. GLOSSÁRIO .....</b>	<b>18</b>
<b>7. REFERÊNCIAS.....</b>	<b>19</b>

## 1. INTRODUÇÃO

A utilização da Tecnologia da Informação nos empreendimentos se faz cada vez mais necessária, principalmente ao almejar um crescimento no mercado de trabalho. Não seria diferente com as empresas de agronegócios: os produtores estão aumentando a inserção de soluções tecnológicas em suas produções, profissionalizando assim muitos procedimentos que são realizados. As empresas de tecnologia também têm investido em softwares de gestão e afins, que podem automatizar tarefas, economizando tempo, dinheiro, evitando erros humanos e auxiliando o produtor que deseja investir na tecnologia como melhoria do seu agronegócio.

É de extrema importância a inserção da tecnologia no setor agrícola, pois seus empresários enfrentam grandes desafios em suas gestões, surgindo assim a necessidade do investimento em ferramentas que os auxiliem a ter um planejamento estratégico mais assertivo, para se manterem atuando de forma satisfatória no mercado.

Tendo em vista este cenário, a **Kersys** - Soluções para Gestão Florestal e Agricultura – é uma empresa de tecnologia da informação que tem como missão *“Fornecer soluções de gestão para as empresas florestais e de agronegócio que possibilite aos clientes melhorar o retorno econômico dos investimentos realizados.”*. Com o intuito de *“simplificar e agilizar o processo de controle e gestão das fazendas”*, a Kersys auxilia os produtores do agronegócio a crescerem em seus empreendimentos utilizando de soluções tecnológicas, gerando resultados mais lucrativos.

Este projeto tem como objetivo principal, auxiliar a Kersys a desenvolver análises inteligentes extraídas de um banco de dados, utilizando das informações fornecidas para auxiliar em tomadas de decisões e avaliações do negócio florestal.

## 2. KERSYS – SOLUÇÕES PARA GESTÃO FLORESTAL E AGRICULTURA

### • A EMPRESA

A Kersys é uma empresa especialista em gestão de agronegócios, que ramifica suas vertentes em gestão florestal e agrícola, levando a tecnologia ao campo e permitindo que o agricultor, seja ele de pequeno ou grande plantio, acompanhe a produtividade e as variações impostas por fatores externos e ou internos ao negócio.

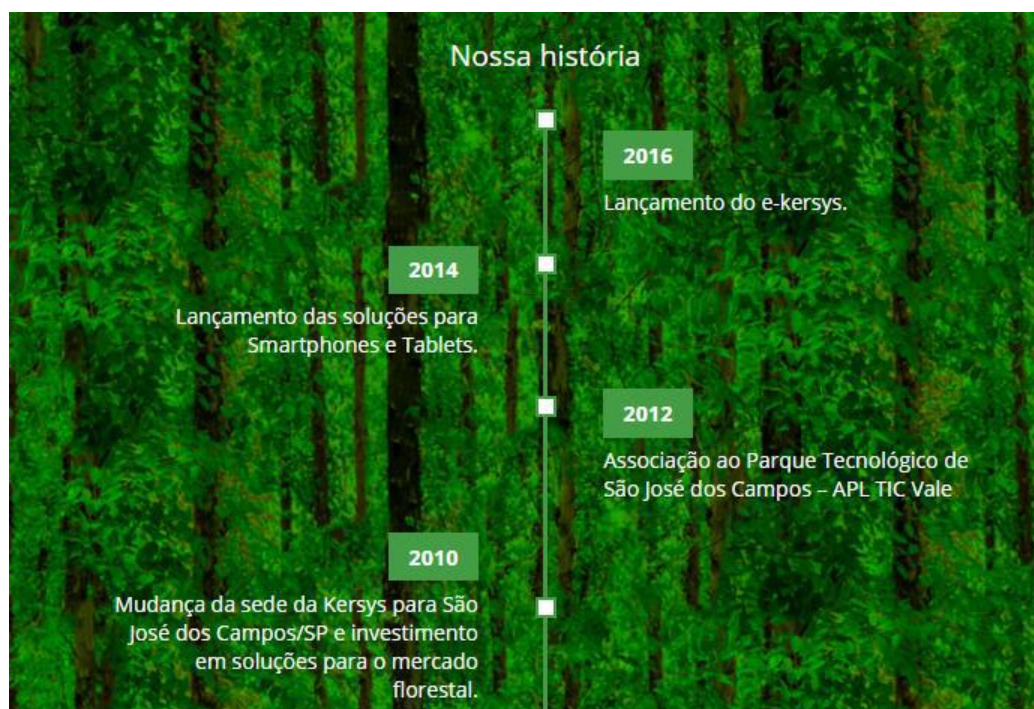
A partir do desenvolvimento de sistemas de fácil compreensão e manuseio, com possibilidade de acesso por variadas mídias e repositório em nuvem, a premissa da Kersys é levar de forma rápida e fácil a gestão informatizada ao homem do campo

Participa da gestão de 441 (quatrocentos e quarenta e uma) fazendas perfazendo um total de 225.121 (duzentos e vinte e cinco mil cento e vinte e um) hectares de solo em território abrangendo o Brasil todo.

“Administrar um projeto florestal utilizando um Sistema de Gestão rápido e eficiente e, especialmente, com informações disponibilizadas “em nuvem” são imprescindíveis para a tomada de decisão rápida e certa. Os sistemas de gestão da Kersys possibilitam isso através de interfaces simples e objetivas em qualquer tipo de mídia. Simples, precisos e seguros”.

Relata **Claudio Ramos**, CEO da BlackWood Forest Business & Consulting, parceiro da empresa que expõem seu depoimento no site desta<sup>1</sup>.

A Kersys ainda possui uma trajetória cheia de conquistas e evolução, incluindo a transferência da sede para São José dos Campos e uma parceira com o Parque da Tecnológico da cidade. Na figura a seguir, podemos visualizar um pouco mais de sua história:

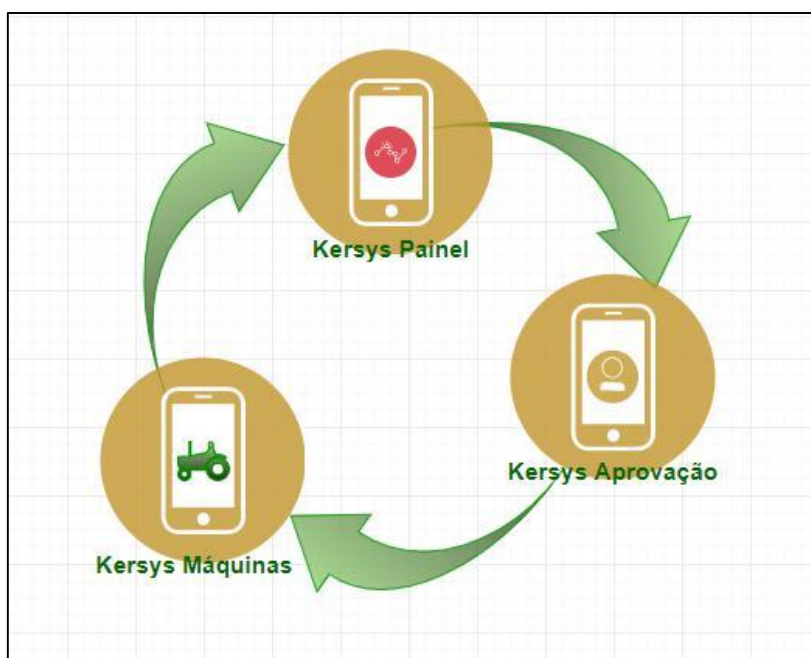


**Figura 1: Trajetória da empresa Kersys.**

## • SOLUÇÕES

Informar para crescer, é o que está intrínseco desde a raiz do nome até o propósito da empresa, fazendo uso das tecnologias mais modernas, informatizadas e automatizadas, em três sistemas de gestão desenvolvidos pela Kersys.

- **Kersys SAF** – Sistema de administração florestal, modular e customizável, para cadastro, planejamento da operação e análise e previsões climáticas, estoque, insumos entre outros fatores determinantes.
- **Kersys Agronegócios** – Solução WEB para gestão e fazendas, coloca de forma fácil, segura e econômica os indicadores de custo, colheita, sinistros e outros nas mãos do agricultor.
- **e-Kersys** - Sistema WEB desenvolvido com o objetivo de acompanhar o projeto florestal, agilizar as tomadas de decisões e maximizar o retorno. Colocando o negócio literalmente nas mãos do produtor, na gestão de equipamentos, nas autorizações e permissões e na análise dos indicadores, tudo da forma mais fácil e a distância se for necessário.



*Figura 2: Módulos independentes da Kersys.*

Temos ainda os *módulos independentes da Kersys*, utilizados para auxiliar o agricultor em necessidades específicas:

- **Kersys Aprovador:** aplicativo que tem como objetivo principal não permitir que os processos da empresa parem por falta de aprovação, caso os gestores estejam fora do escritório.
- **Kersys Painel:** permite a análise de dados do projeto de forma rápida e eficaz, emitindo relatórios e gráficos que auxiliam nas tomadas de decisões e reduções de custos.
- **Kersys Máquina:** permite o gerenciamento de todo maquinário na mesma ferramenta, possibilitando o controle total de todas as operações de tratores, colheitadeiras, caminhões, entre outros.

Tendo em vista todas essas soluções, temos ainda o projeto **KIA** (*Kersys Inteligência Artificial*), que trabalha para colocar nas mãos do gestor do negócio um quadro mais assertivo nas projeções e tomadas de decisões.

- **KIA - KERSYS INTELIGÊNCIA ARTIFICIAL**

O projeto tem o objetivo de otimizar e maximizar os lucros na produção planejada de eucaliptos, informação com base em ciência, materiais genéticos de qualidade, foco nas questões ambientais e uso responsável de insumos.

Em 2010 nasce a Kersys com o “sonho” de contagiar o mercado do agronegócio ao desenvolvimento contínuo e inovador com uso da tecnologia da informação, e hoje, seguindo esses passos, e fazendo uso da ciência de análise de dados com Inteligência Artificial, coloca a máquina e sua capacidade de processar essas informações em frações de segundos, analisando as ocorrências, os resultados, as condições, sejam elas climática, geográficas ou de manejo, para prever melhores decisões de plantios e projetos de plantios.

A pesquisa e análise está baseada na produção de eucaliptos e os dados disponibilizados espelham o comportamento de áreas a partir das operações realizadas e os resultados adquiridos.

Com uma base de dados já estruturada e que acompanha as características produtivas de projetos de plantio divididos por talhão de área, subdivisão feita para facilitar a mobilidade, as operações cotidianas e de sinistros entre os plantios, a expectativa é levantar na análise desta base direcionamentos e identificação de padrões que possam sugerir operações para maior produtividade.

- **DESAFIO**

É desafio do Projeto KIA disponibilizar ferramentas que visem uma projeção inteligente da produção e sua consequente maximização de lucros.

A análise assertiva da base de dados de forma eficaz e eficiente e fundamentalmente rápida para a condução embasada das decisões gerenciais e estratégicas do produtor, com responsabilidades e compromissos ambientais.

- **REQUISITOS**

Elaboração de algoritmos de identificação dos padrões que ocorrem entre as variáveis descritivas e de resultado do banco de dados.

Esse algoritmo deve calcular a projeção de produtividade a partir do padrão identificado.

Será feita o acompanhamento dos resultados a fim de validá-los com o conhecimento agroflorestal e identificar se os resultados estão de acordo com o que poderia ser observado no campo, o que poderá permitir a criação de mecanismos de feedback do próprio algoritmo.

### 3. BASE DE DADOS

A base de dados a ser analisada corresponde a planilha de extensão .xlsx composta pelos valores colhidos em campo correspondentes a 533 talhões, pertencentes a 20 fazendas e 36 projetos de plantio de eucalipto, com áreas correspondentes que variam de 0,44 a 85,29 há, todos no mesmo ciclo de plantio, sendo este ciclo um, espaçamento entre as árvores e as linhas em torno de 3,45 m e 3,5 m entre as linhas e 2,35, 3,20 e 2,15 m entre as árvores, 15 tipos de materiais genéticos diferentes, com datas de plantio de 2012 a 2016, meses diferentes de plantio ao longo do ano.

Com 37 atividades distintas registradas, podendo ser realizadas em 100% da área ou não, de forma manual, mecanizada, aérea ou de irrigação, executados de mais de um ano antes do plantio até quase 6,5 anos deste.

Idade do plantio, área inventariada, acréscimo anual de madeira em metros cúbicos, número de fustes por hectare, mortes, falhas, árvores quebradas ou tombadas, se houve chuva no dia e até 3 dias antes e 3 dias depois da atividade realizada, qual a quantidade de chuva por milímetros separadas em 5 faixas de registro, 1 mm, de 1 a 5 mm, de 5 a 15 milímetros, de 15 a 30 mm e acima de 30 mm de chuva.

A disponibilidade de água obtida pela relação da água armazenada e a capacidade de água disponível no solo, seccionadas entre até 20%, de 20 a 40%, de 40 a 60%, de 60 a 80% e de 80 a 100% de água disponível, sendo que o observado é que mais de 3/4 dos registros tem mais de 40% de água.

- **DISPOSIÇÃO DO TALHÃO**

Os talhões podem ter áreas distintas de forma a favorecerem o manuseio da plantação, são delimitados por passagens ou limites de propriedades.

Na figura 3 temos um exemplo de definição de talhão, que podem variar de acordo com necessidades de manejo. As áreas de passagem são planejadas para facilitar o acesso nas atividades diárias e nos repentes de sinistro. Torres de vigília podem ser implementadas em áreas de risco para acompanhamento constante e ações imediatas e aceiros são planejados e implementados.



*Figura 3: Exemplo de definição de talhão.*

- **ANÁLISE DOS DADOS**

Composição da base de dados dispostos em 45 colunas e 8,221 linhas com tipos relacionados abaixo, descrição das variáveis em documento anexo fornecido pela Kersys complementa as informações.

<b>COLUMNS</b>	<b>TYPE</b>
Z	int64
CODFAZENDA	int64
SIGLA_FAZENDA	object
SIGLA_PROJETO	object
CODTALHAO	int64
SIGLA_TALHAO	object
AREA_TALHAO	object
CICLO	int64
ROTACAO	int64
ESPACAMENTO	object
SIGLA_MAT_GEN	object
DATA_PLANTIO	object
MES_PLANTIO	object



10

ANO_PLANTIO	int64
DESC_GRP_ATIVIDADE	object
NOME_ATIVIDADE	object
COD_ATIVIDADE	float64
MODO_DE_OPERACAO	object
MODO_DE_APLICACAO	object
MODO_DE_ACAO	object
QTDE_REALIZADA_ATIVIDADE	object
%AREA	object
DATA_ATIVIDADE_REALIZ	object
MES_ATIVIDADE_REALIZ	object
ANO_ATIVIDADE_REALIZ	int64
DIAS_REF_PLANTIO_REALIZ	int64
IDADE_TALHAO_ATIVIDADE	object
IDADE_TALHAO_INVENTARIO	object
AREA_INVENTARIO	object
IMA_TALHAO	object
VOLUME_HA_TALHAO	object
FUSTES_HA	object
DIAM_MED_TALHAO	object
ALTURA_MED_TALHAO	object
MORTES_PERC	object
FALHA_PERC	object
QUEBRADAS_PERC	object
TOMBADAS_PERC	object
CHOVEU?	object
QUANTO?	int64
C.DIA_ANT?	object
C.DIA_POST?	object
C.3DIAS_A?	object
C.3DIAS_D?	object
ARM/CAD	int64
(8221, 45)	

### Análise de características por talhão:

### Influência da chuva na atividade de preparo do solo:

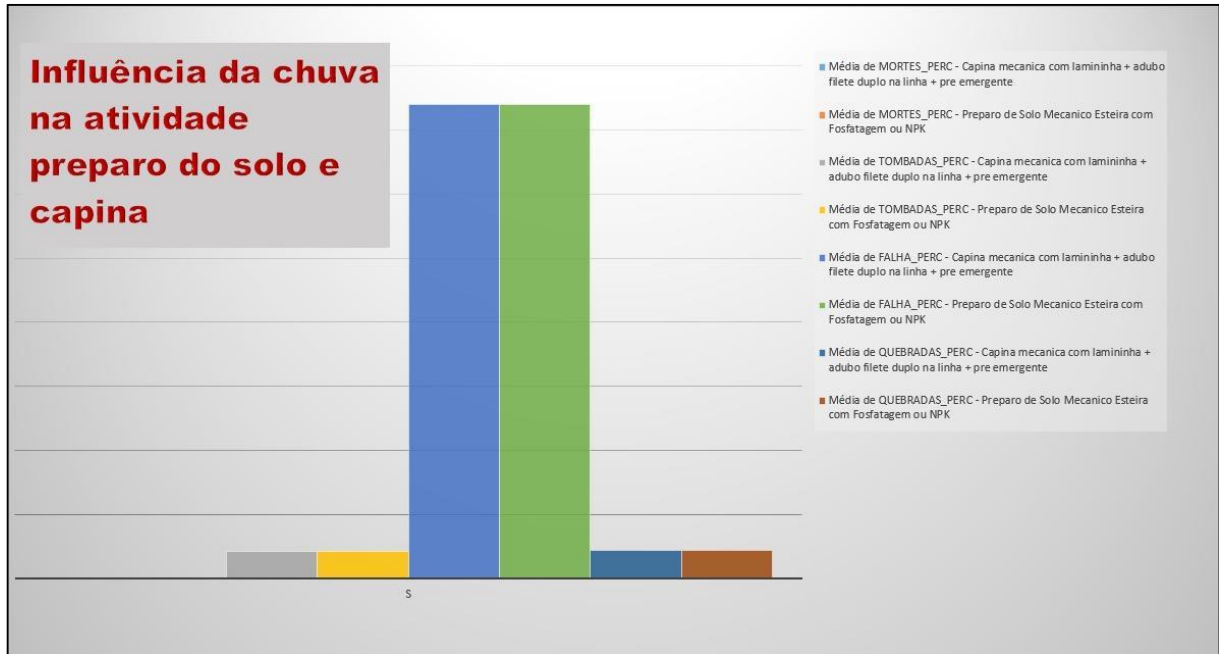


Figura 6: Gráfico contendo informações referentes a influência da chuva na atividade de preparo do solo.

### Influência da chuva na atividade de Combate as formigas:



Figura 7: Gráfico contendo informações referentes a influência da chuva na atividade de preparo do solo.

- PREPARAÇÃO DOS DADOS

Nesta etapa fizemos a preparação dos dados para análises e interpretações. Colunas foram renomeadas, alguns dados foram convertidos para numéricos, alguns dados são ordinais e outros nominais, classificados os números que tem pouca relevância e outros alta relevância, e os dados booleanos foram alterados para zero e um, ações tomadas para preparar os dados para interpretação pelo algoritmo de análise.

Abaixo a classificação dos dados fazendo uso da ferramenta Jupyter Notebook:

### Estrutura do Banco de dados a ser analisado:

```
In [4]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8221 entries, 0 to 8220
Data columns (total 45 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Z                                           8221 non-null   int64
1   CODFAZENDA                               8221 non-null   int64
2   SIGLA_FAZENDA                             8221 non-null   object
3   SIGLA_PROJETO                             8221 non-null   object
4   CODTALHAO                                 8221 non-null   int64
5   SIGLA_TALHAO                              8221 non-null   object
6   AREA_TALHAO                               8221 non-null   object
7   CICLO                                      8221 non-null   int64
8   ROTACAO                                    8221 non-null   int64
9   ESPACAMENTO                              8221 non-null   object
10  SIGLA_MAT_GEN                             8221 non-null   object
11  DATA_PLANTIO                             8221 non-null   object
12  MES_PLANTIO                               8221 non-null   object
13  ANO_PLANTIO                               8221 non-null   int64
14  DESC_GRP_ATIVIDADE                       8221 non-null   object
15  NOME_ATIVIDADE                           8221 non-null   object
16  COD_ATIVIDADE                             8221 non-null   float64
17  MODO_DE_OPERACAO                         8221 non-null   object
18  MODO_DE_APLICACAO                       8221 non-null   object
19  MODO_DE_ACAO                             8221 non-null   object
20  QTDE_REALIZADA_ATIVIDADE                 8221 non-null   object
21  % AREA                                    8221 non-null   object
22  DATA_ATIVIDADE_REALIZ                   8221 non-null   object
23  MES_ATIVIDADE_REALIZ                     8221 non-null   object
24  ANO_ATIVIDADE_REALIZ                     8221 non-null   int64
25  DIAS_REF_PLANTIO_REALIZ                  8221 non-null   int64

26  IDADE_TALHAO_ATIVIDADE                   8221 non-null   object
27  IDADE_TALHAO_INVENTARIO                  8221 non-null   object
28  AREA_INVENTARIO                          8221 non-null   object
29  IMA_TALHAO                               8221 non-null   object
30  VOLUME_HA_TALHAO                        8221 non-null   object
31  FUSTES_HA                                8221 non-null   object
32  DIAM_MED_TALHAO                         8221 non-null   object
33  ALTURA_MED_TALHAO                      8221 non-null   object
34  MORTES_PERC                              8221 non-null   object
35  FALHA_PERC                              8221 non-null   object
36  QUEBRADAS_PERC                          8221 non-null   object
37  TOMBADAS_PERC                           8221 non-null   object
38  CHOVEU?                                  8221 non-null   object
39  QUANTO?                                  8221 non-null   int64
40  C.DIA ANT?                               8221 non-null   object
41  C.DIA POST?                              8221 non-null   object
42  C.3 DIAS A?                              8221 non-null   object
43  C.3 DIAS D?                              8221 non-null   object
44  ARM/CAD                                  8221 non-null   int64
dtypes: float64(1), int64(10), object(34)
memory usage: 2.8+ MB
```

**Figura 8: Estrutura do banco de dados em análise.**



### Adequação dos dados para análise:

```
In [5]: df['SIGLA_MAT_GEN'], _ = pd.factorize(df['SIGLA_MAT_GEN'])
print(df['SIGLA_MAT_GEN'].unique())

[ 0  1  2  3  4  5  6  7  8  9 10 11]

In [6]: df['NOME_ATIVIDADE'], _ = pd.factorize(df['NOME_ATIVIDADE'])
print(df['NOME_ATIVIDADE'].unique())

[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64]

In [7]: df['CHOVEU?'], _ = pd.factorize(df['CHOVEU?'])
print(df['CHOVEU?'].unique())

[0 1]

In [8]: df['MORTES_PERC'], _ = pd.factorize(df['MORTES_PERC'])
print(df['MORTES_PERC'].unique())

[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
 72 73 74 75 76 77]
```

Figura 9: Adequação dos dados para análise.

```
In [9]: df['FALHA_PERC'], _ = pd.factorize(df['FALHA_PERC'])
print(df['FALHA_PERC'].unique())

[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53
 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89
 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107
 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125
 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143
 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161
 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179
 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197
 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215
 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233
 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251
 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269
 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287
 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305
 306 307 308 309 310 311 312 313 314]

In [10]: df['QUEBRADAS_PERC'], _ = pd.factorize(df['QUEBRADAS_PERC'])
print(df['QUEBRADAS_PERC'].unique())

[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92]
```

Figura 10: Continuação da adequação dos dados.

```
In [11]: feature_cols = ['SIGLA_MAT_GEN', 'SIGLA_MAT_GEN', 'CHOVEU?']
x_train = df[feature_cols]
y_train = df.MORTES_PERC

In [12]: df['QUEBRADAS_PERC'], class_names = pd.factorize(df['QUEBRADAS_PERC'])
print(class_names)
print(df['QUEBRADAS_PERC'].unique())

Int64Index([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67,
 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84,
 85, 86, 87, 88, 89, 90, 91, 92],
dtype='int64')
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92]
```

Figura 11: Continuação da adequação dos dados.

### Construção de uma análise de árvore de decisão:

```
In [13]: arvore = DecisionTreeClassifier(criterion='entropy')
          arvore

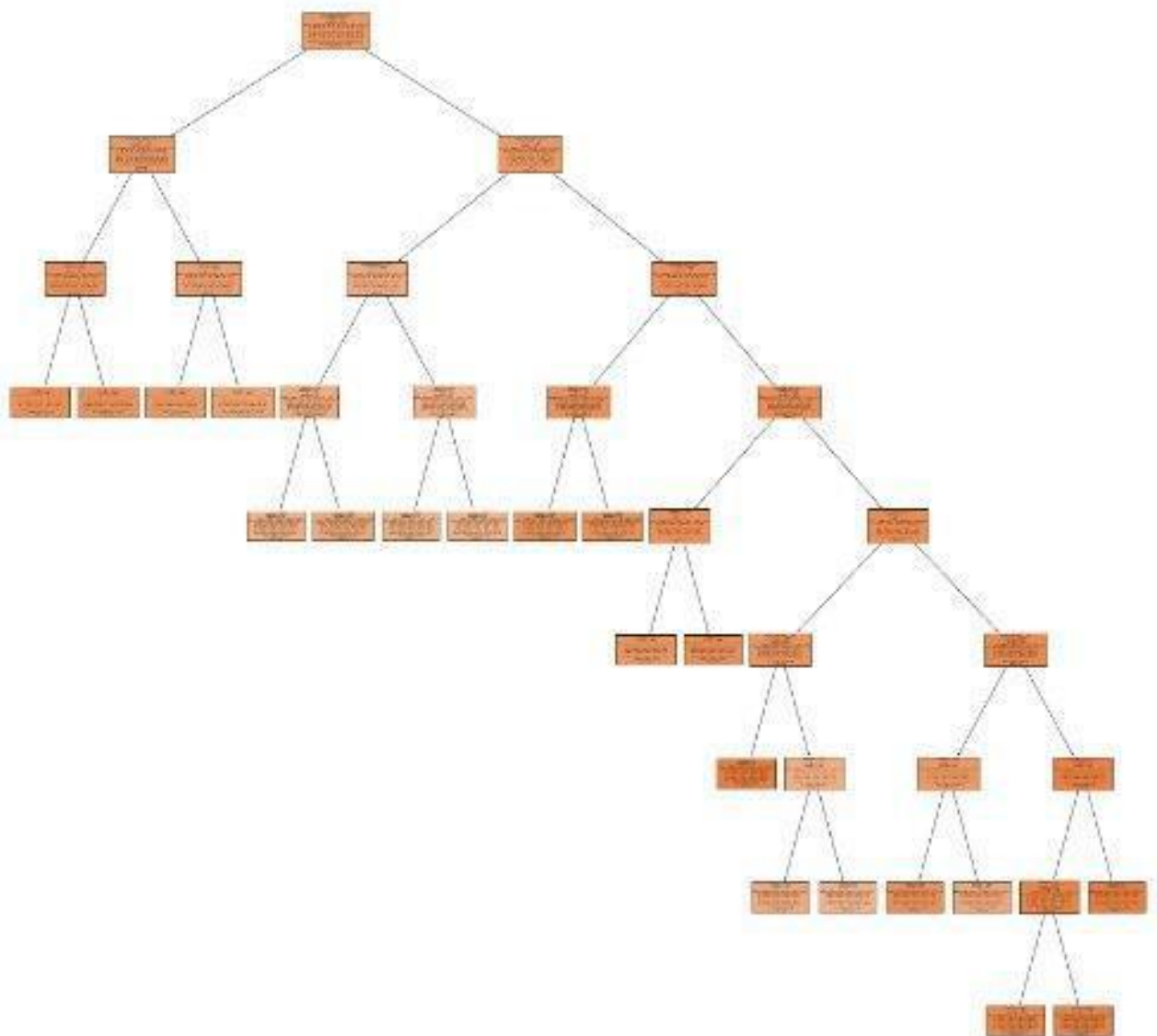
Out[13]: DecisionTreeClassifier(criterion='entropy')
```

```
In [14]: # usado para treinar o modelo
          arvore = arvore.fit(x_train,y_train)
          arvore

Out[14]: DecisionTreeClassifier(criterion='entropy')
```

```
In [15]: fig = plt.figure(figsize=(30,30))
          fig = tree.plot_tree(arvore, feature_names=x_train.columns,class_names=df.columns[5], filled=True)
          fig.savefig("decisiontree.png")
```

**Figura 12: Construção de uma análise de árvore de decisão.**



**Figura 13: Árvore de decisão.**

## 4. Modeling e Evaluation

A avaliação do modelo é parte fundamental do processo de desenvolvimento de um algoritmo de análise de dados em data Science, nos ajuda a encontrar o melhor modelo que representa os dados e como esse modelo funcionará no futuro.

Existem dois métodos de avaliação de modelos em ciência de dados, Percentege Split e Cross-Validation.

### Hold-Out

Neste método, o grande conjunto de dados é dividido aleatoriamente em três subconjuntos: O conjunto de treinamento é um subconjunto do conjunto de dados usado para construir modelos preditivos.

O conjunto de validação é um subconjunto do conjunto de dados usado para avaliar o desempenho do modelo construído na fase de treinamento. Ele fornece uma plataforma de teste para ajustar os parâmetros do modelo e selecionar o modelo de melhor desempenho. Nem todos os algoritmos de modelagem precisam de um conjunto de validação.

O conjunto de testes ou exemplos não vistos é um subconjunto do conjunto de dados para avaliar o provável desempenho futuro de um modelo. Se um modelo se encaixa muito melhor no conjunto de treinamento do que no conjunto de teste, provavelmente o overfitting é a causa.

### Validação cruzada ( Cross-Validation )

Quando apenas uma quantidade limitada de dados está disponível, para obter uma estimativa imparcial do desempenho do modelo, usamos a validação cruzada k-fold. Na validação cruzada k-fold, dividimos os dados em k subconjuntos de tamanhos iguais. Construímos modelos k vezes, cada vez deixando de fora um dos subconjuntos do treinamento e o usamos como conjunto de teste

É muito importante na classificação que, ao observar os resultados da avaliação, você só obtenha resultados confiáveis da avaliação se os dados de teste forem diferentes dos dados de treinamento.

### Percentege Split

Para apenas um conjunto de dados, deve dividi-lo em duas partes: usar parte dele para treinamento e parte dele para teste – é comum usar dois terços dele para treinamento e um terço para teste. É muito importante que os dados de treinamento sejam diferentes dos dados de teste.

Os conjuntos de treinamento e teste são produzidos por amostragem independente de uma população infinita. Se forem os mesmos dados, os resultados de sua avaliação são enganosos.

Com base nos dados avaliados, vários testes de validação foram feitos para obtermos respostas que indicassem as situações de melhor produtividade e mínimas perdas.

Foram seccionadas colunas de dados que consideramos como “colunas de respostas” que viessem a nos indicar as situações de maior benefício ou o exatamente inverso, que nos mostrasse situações de mínimas perdas, no que chamamos de fator de análise.

Toda classificação foi identificada por sigla do talhão e sigla da fazenda para que pudéssemos identificar o local das atividades e características.

Avaliamos:

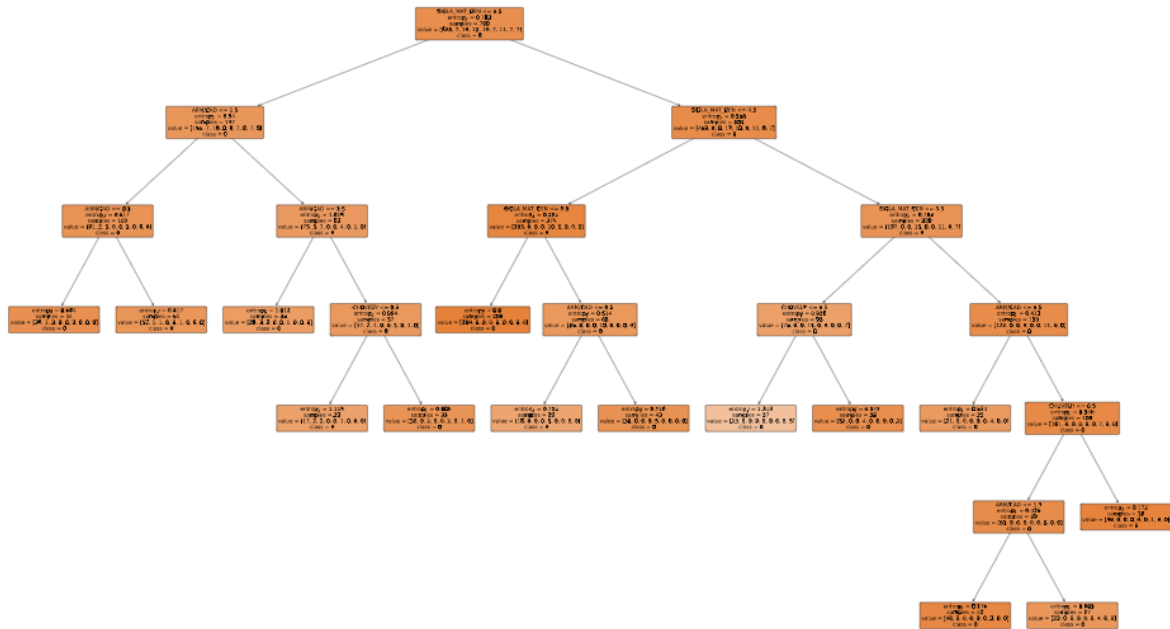
- Material genético utilizado
- Capacidade hídrica do solo
- Espaçamentos entre a “matriz” de árvores plantadas
- Temporalidade (época/mês de plantio)
- E situação climática entre atividades de manejo
  - Ter chovido ou não antes e depois do manejo de atividades.

Buscamos com isso respostas que nos conduzisse a entender em que condições consegue-se maior produção:

- Número de fustes por hectare
- Altura, diâmetro, volume médio percentuais
- Mortes, falhas, quebradas e tombadas percentuais.



Nossas análises não foram satisfatórias e não conseguimos chegar a modelos de decisão que nos garantissem uma análise conclusiva.



Permanecemos nas análises do algoritmo a fim de conseguirmos estabelecer uma configuração assertiva para conclusão da aprendizagem e testes.

## 5. CONCLUSÃO E CONSIDERAÇÕES FINAIS

Conclui-se até momento das pesquisas que a análise de árvore de decisão seguindo os conceitos e valorações atribuídos até momento não trazem a clareza desejada para a efetivação da análise.

Ainda não foram analisados algoritmos de cluster, onde esperamos conseguir “acumular” comportamentos similares nas atividades versus situações climáticas e material genético utilizado no plantio.

As análises elaboradas não foram suficientes para conclusões específicas.

## 6. GLOSSÁRIO

**Talhão** – subdivisão de área de plantio

**Hectare** – área equivalente a 10.000 m<sup>2</sup>

**Inteligência Artificial (IA)** – parte da ciência da computação que se propõe a elaborar sistemas que simulem a capacidade humana de raciocinar, perceber, tomar decisões e resolver problemas.

**Fuste** – árvore em condição de colheita

**Aceiros** – áreas de desbastes com a finalidade de inibir o avanço de incêndios

## 7. REFERÊNCIAS

<sup>1</sup>KERSYS – SOLUÇÕES PARA GESTÃO FLORESTAL E AGRICULTURA. Disponível em: <https://kersys.com.br/>. Acesso em: 3 out. 2020.

A IMPORTÂNCIA da tecnologia da informação para o agronegócio brasileiro. MOLIN, Rafael Dal, 22 jan. 2020. Disponível em: <https://www.agroplanning.com.br/2020/01/22/a-importancia-da-tecnologia-da-informacao-para-o-agronegocio-brasileiro/>. Acesso em: 4 out. 2020.

LEARN PANDAS TUTORIAL. Disponível em: <https://www.kaggle.com/learn/pandas>. Acesso em: 01 out. 2020

KEGGLE - CREATING, READING AND WRITING. Disponível em: <https://www.kaggle.com/residentmario/creating-reading-and-writing>. Acesso em: 01 out. 2020