

**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**

[**1.**](#_1fob9te) **INTRODUÇÃO** 3

[**2.**](#_2et92p0) **BUSINESS UNDERSTANDING** 4

[**2.1. 2.**](#_tyjcwt) **KERSYS – SOLUÇÕES PARA GESTÃO FLORESTAL E AGRICULTURA** 4

[**2.2. SOLUÇÕES** 5](#_1t3h5sf)

[**2.3. KIA - KERSYS INTELIGÊNCIA ARTIFICIAL** 6](#_2s8eyo1)

[**2.4. DESAFIO** 6](#_17dp8vu)

[**2.5. REQUISITOS** 6](#_3rdcrjn)

[**3.**](#_lnxbz9) **DATA UNDERSTANDING** 7

[**3.1. DISPOSIÇÃO DO TALHÃO** 7](#_35nkun2)

[**3.2. ANÁLISE DOS DADOS** 8](#_44sinio)

[**4.**](#_4i7ojhp) **DATA PREPARATION** 12

[**5.**](#_3whwml4) **MODELING E EVALUATION** 15

[**5.1.**](#_2bn6wsx) **MACHINE LEARNING** 15

[**5.2.**](#_qsh70q) **OS ALGORITMOS DE MACHINE LEARNING** 15

[**5.2.1.**](#_3as4poj) **APRENDIZAGEM SUPERVISIONADA** 15

[**5.2.1.**](#_1pxezwc) **APRENDIZAGEM NÃO SUPERVISIONADA** 16

[**5.3.**](#_49x2ik5) **TÉCNICA DE ESTATÍSTICA DE AMOSTRAGEM UTILIZANDO CLUSTERS** 16

[**5.4.**](#_2p2csry) **ALGORITMO DE AGRUPAMENTO (CLUSTERING)** 18

[**5.5.**](#_147n2zr) **GRÁFICO DOS CLUSTERS** 19

[**5.6.**](#_23ckvvd) **ÁRVORE DE DECISÃO** 21

[**6.**](#_ihv636) **CONCLUSÃO E CONSIDERAÇÕES FINAIS** 24

[**7.**](#_1hmsyys) **GLOSSÁRIO** 25

[**8.**](#_2grqrue) **REFERÊNCIAS** 26

# **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 doagronegócio 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.

# **BUSINESS UNDERSTANDING**

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

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 certeira. 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 desta¹.

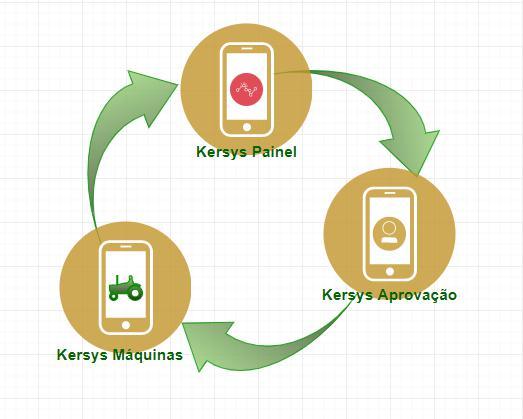
A Kersys ainda possui uma trajetória cheia de conquistas e evolução, incluindo a transferência da sede para São José dos Campos e uma parceira com o Parque da Tecnológico da cidade. Na figura a seguir, podemos visualizar um pouco mais de sua história:



## **2.2. 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, paracadastro, planejamento da operação e análise e previsões climáticas, estoque, insumos entre outros fatores determinantes.
* **Kersys Agronegócios** –Solução WEB para gestão e fazendas, coloca de formafácil, 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 projetoflorestal, 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.



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 queos 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 mesmaferramenta, 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** *(****K****ersys* ***I****nteligência* ***A****rtificial)*, que trabalha para colocar nas mãos do gestor do negócio um quadro mais assertivonas projeções e tomadas de decisões.

## **2.3. 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.

## **2.4. 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.

## **2.5. 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.

# **DATA UNDERSTANDING**

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.

## **3.1. 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.



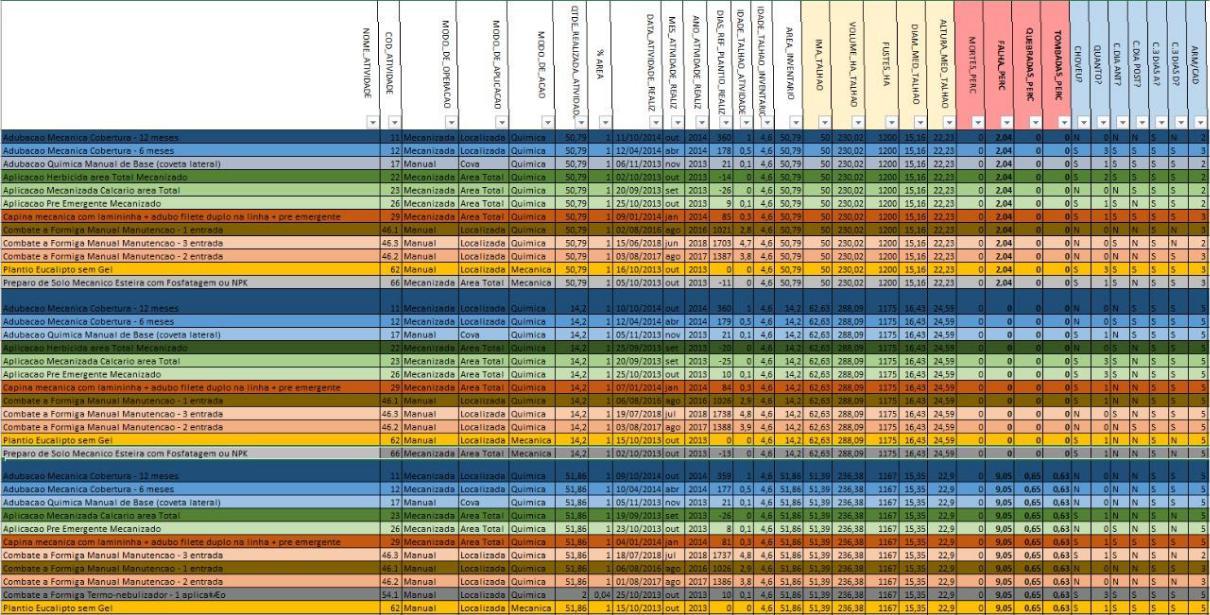
## **3.2. ANÁLISE DOS DADOS**

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

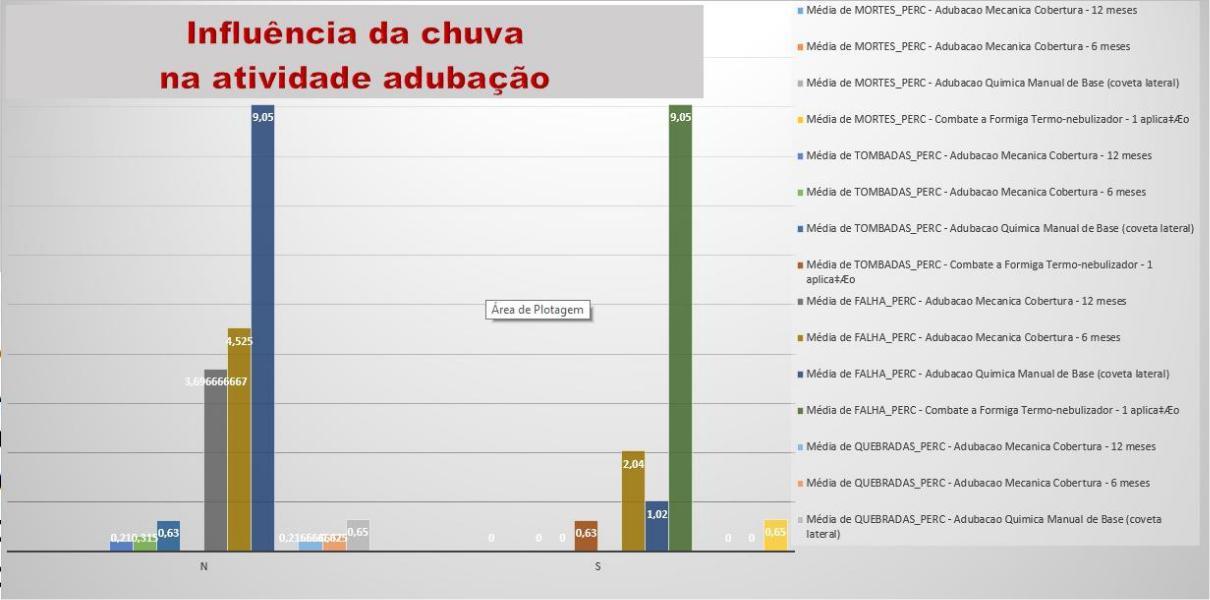
|  |  |
| --- | --- |
| **COLUMNS** | **TYPE** |
| Z | int64 |
| CODFAZENDA | int64 |
| SIGLA\_FAZENDA | object |
| SIGLA\_PROJETO | object |
| CODTALHAO | int64 |
| SIGLA\_TALHAO | object |
| AREA\_TALHAO | object |
| CICLO | int64 |
| ROTACAO | int64 |
| ESPACAMENTO | object |
| SIGLA\_MAT\_GEN | object |
| DATA\_PLANTIO | object |
| MES\_PLANTIO | object |

|  |  |
| --- | --- |
|  |  |
| 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) |  |

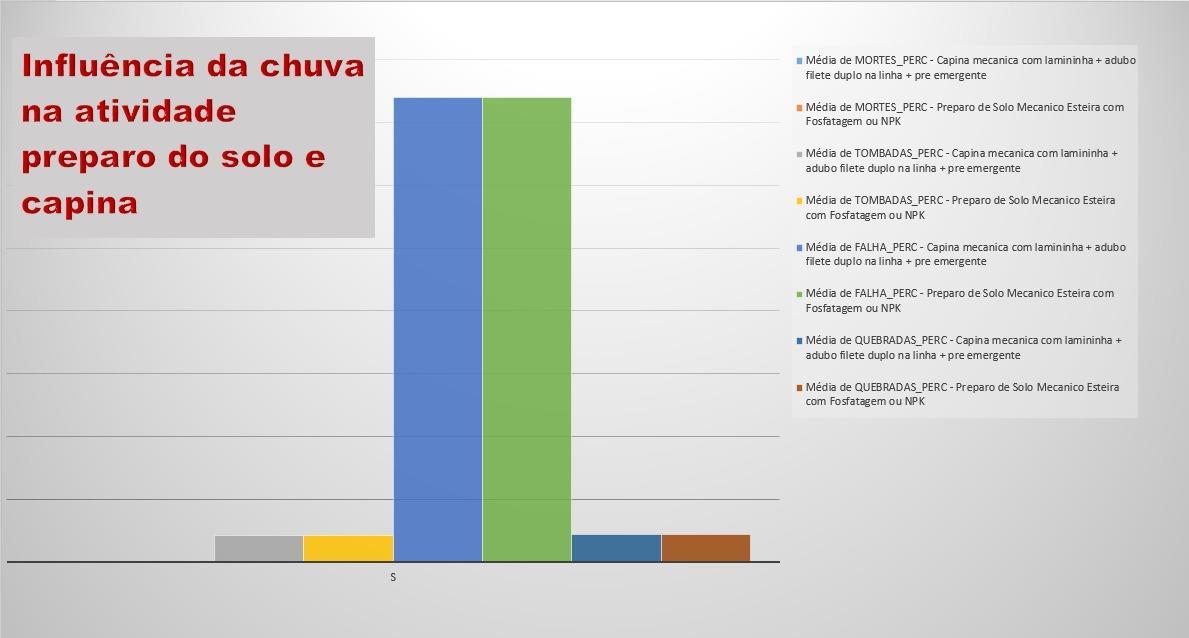
**Análise de características por talhão:**



**Influência das chuvas na atividade:**



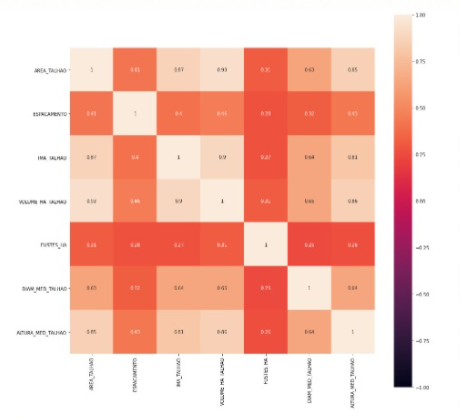
**Influência da chuva na atividade de preparo do solo:**

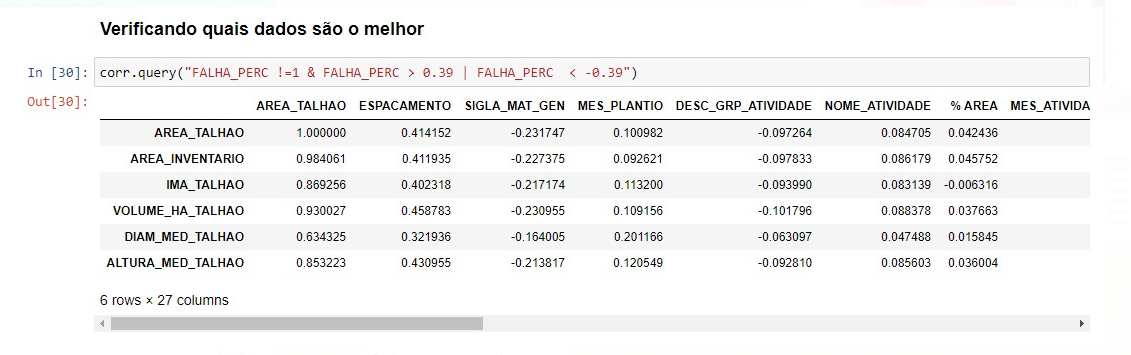


**Influência da chuva na atividade de Combate às formigas:**



**Análise de correlação dos dados:**



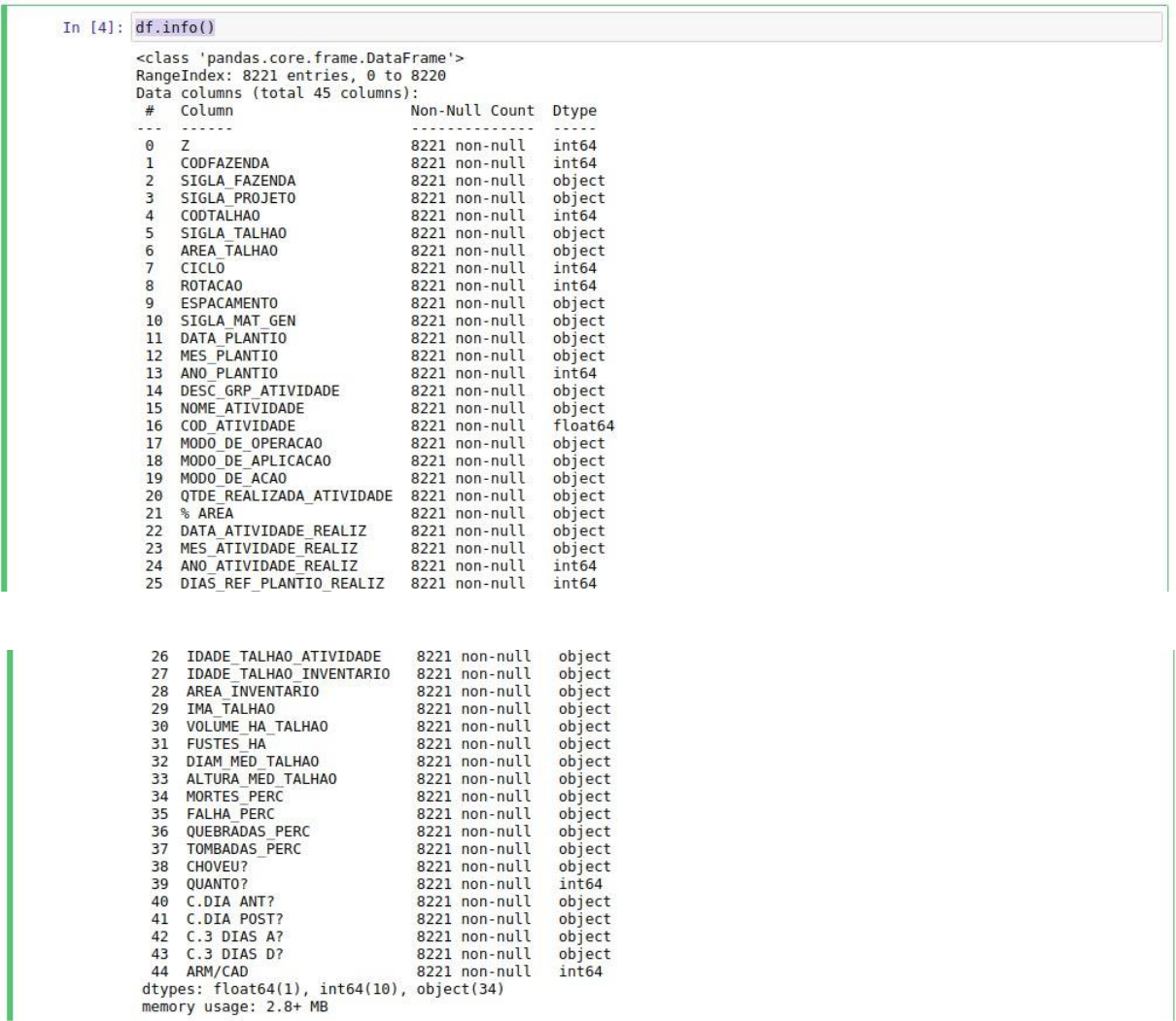


# **DATA PREPARATION**

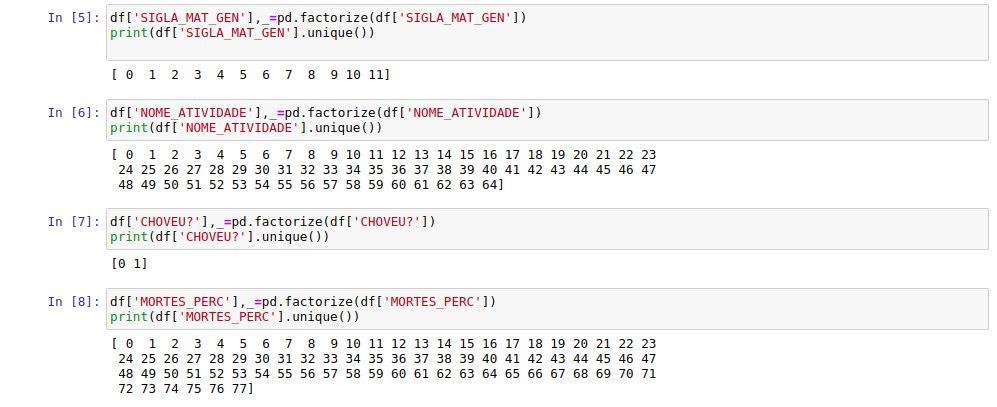
Nesta etapa fizemos a preparação dos dados para análises e interpretações. Colunas foram renomeadas, alguns dados foram convertidos para numéricos, alguns dados são ordinais e outros nominais, classificados os números que tem pouca relevância e outros alta relevância, e os dados booleanos foram alterados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

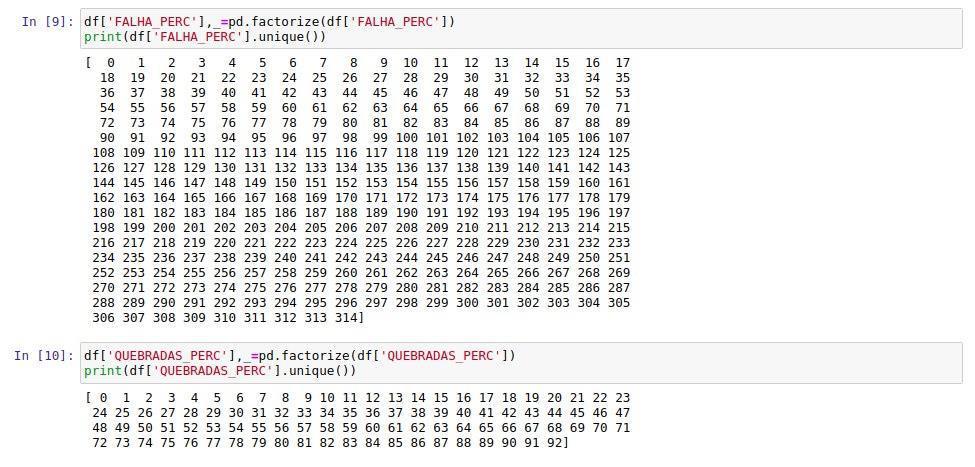
Abaixo a classificação dos dados fazendo uso da ferramenta Jupyter Notebook:

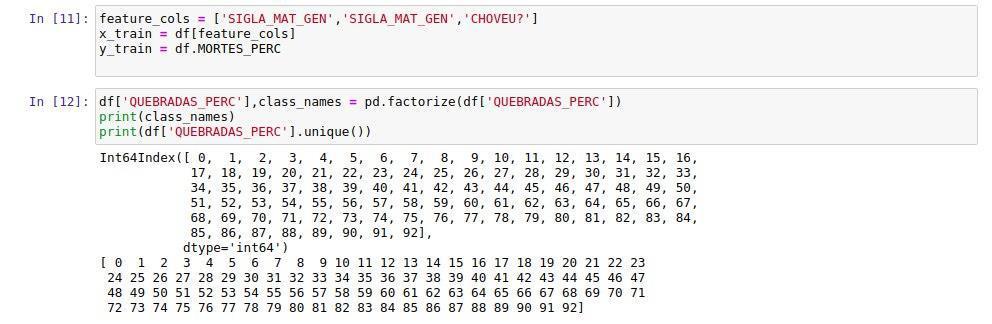
**Estrutura do Banco de dados a ser analisado:**



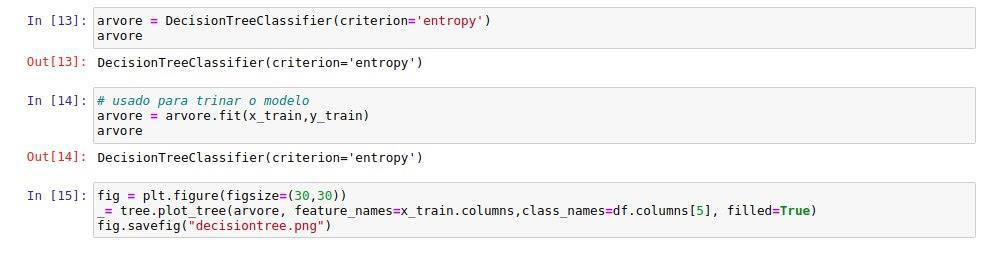
**Adequação dos dados para análise:**

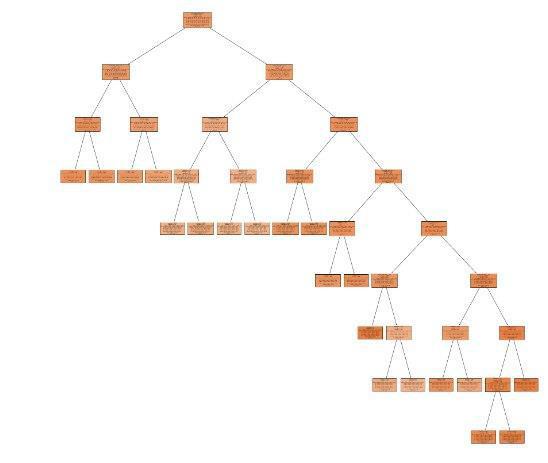






**Construção de uma análise de árvore de decisão:**





# **MODELING E EVALUATION**

## **MACHINE LEARNING**

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.

Quando se é inviável mapear todas as variáveis e possibilidades para a resolução das questões envolvidas na análise de dados, pode-se utilizar os algoritmos de Machine Learning. Pensando nesse contexto, os algoritmos analisam dados de entrada, processam e preveem saídas possíveis dentro de um intervalo definido. Ainda neste processo, são testadas diferentes abordagens, otimizando também a capacidade de se alcançar um resultado satisfatório.

Os sistemas artificiais no Machine Learning possuem duas fases: a de treino e a de execução. Na primeira, recebem um conjunto de entrada que tenham relação com a situação referente ao problema. E na segunda fase, é iniciado a dedução dos possíveis caminhos de solução a partir do que se aprendeu na etapa anterior. Essas duas fases podem variar, de acordo com a categoria do algoritmo sendo utilizado.

## **OS ALGORITMOS DE MACHINE LEARNING**

Os algoritmos de Machine Learning podem ser divididos em três categorias: Aprendizagem Supervisionada, Aprendizagem Não Supervisionada e Aprendizado por Reforço.

### **APRENDIZAGEM SUPERVISIONADA**

A Aprendizagem Supervisionada é utilizada nos casos em que uma propriedade (rótulo) está disponível para um determinado conjunto de dados (conjunto de treinamento). Todas as entradas e saídas nesse algoritmo são conhecidas, mas precisam ser previstas para serem utilizadas em outras instâncias.

Um exemplo para esse caso seria ensinar uma máquina a reconhecer e categorizar e-mails, fazendo a separação entre os que são relevantes e os que são tidos como SPAM. Na fase de treinamento, são enviados para o algoritmo amostras de ambos os casos, tendo em vista que os dois blocos de informação são conhecidos. Já na fase de execução, o sistema terá a capacidade de determinar se um e-mail é novo e não uma mensagem indesejada. Ou seja: mesmo que o software já saiba quais entradas são associadas com quais saídas, é necessário que ele aprenda um meio de entender essa associação. Nesse processo, ele tenta identificar padrões e estabelece previsões que ajudam a otimizar a abordagem.

A ferramenta de apoio que utilizamos que se enquadra em Aprendizagem Supervisionada é a **Árvore de Decisão,** que usa um gráfico ou modelo de decisões e suas possíveis consequências, incluindo resultados de eventos não planejados, custos de recursos e utilidade. Os pontos de decisão da árvore são chamados de “nós” e em cada um deles o resultado será seguir por determinado caminho ou por outro. Os caminhos existentes são intitulados “ramos”. Os nós são responsáveis pelas conferências que irão indicar um ramo ou outro para sequência do fluxo.

### **APRENDIZAGEM NÃO SUPERVISIONADA**

Já na aprendizagem não supervisionada, o desafio é descobrir relações implícitas em um conjunto de dados não rotulados. O algoritmo se torna então responsável por identificar padrões para rotular os dados. Um exemplo que se encaixa nesse caso, é quando uma empresa fornece como base para um sistema de ML um conjunto de dados sobre clientes, esperando que o sistema seja capaz de identificar possíveis atributos comuns e padrões de comportamento para criar ofertas específicas e segmentadas. No início, não se conhecerá de forma satisfatória as informações, a ponto de se saber as possíveis saídas.

A ferramenta que utilizamos, categorizada como Aprendizagem Não Supervisionada é o **Algoritmo de Agrupamento (Clustering).** Como o próprio nome diz, essa técnica agrupa um conjunto de objetos de tal forma que os do mesmo grupo (cluster) são mais semelhantes uns aos outros do que aqueles em outros grupos.

## **TÉCNICA DE ESTATÍSTICA DE AMOSTRAGEM UTILIZANDO CLUSTERS**

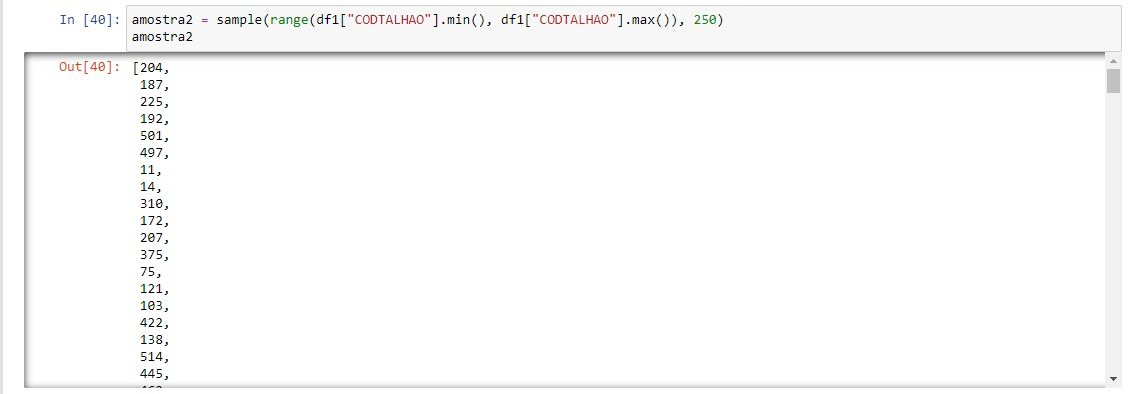
A técnica de Estatística de Amostragem (ou amostra de conglomerados) explora a existência de grupos (clusters) entre os dados. Esses grupos representam adequadamente os dados em sua totalidade, em relação ao item que se deseja analisar. Em resumo, esses clusters contém variabilidade dos dados completos.

A amostra a ser analisada deve ser representativa, ou seja, deve ser possível ter uma boa visibilidade da população inteira (Base de Dados) a partir daquele grupo. Levando isso em consideração, foram analisadas três amostras de nossa Base de Dados, a fim de averiguar um padrão de comportamento da população.

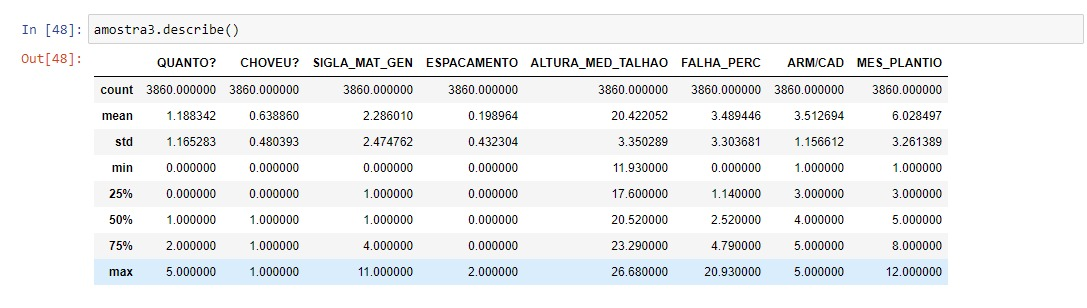
Antes da separação das três amostras, os dados passaram por uma preparação, para que a análise pudesse ser feita e os algoritmos fossem capazes de compreender as informações de acordo com o esperado.

***Figura 14: Amostra um de dados, do primeiro sorteio, pelos códigos de talhão.***

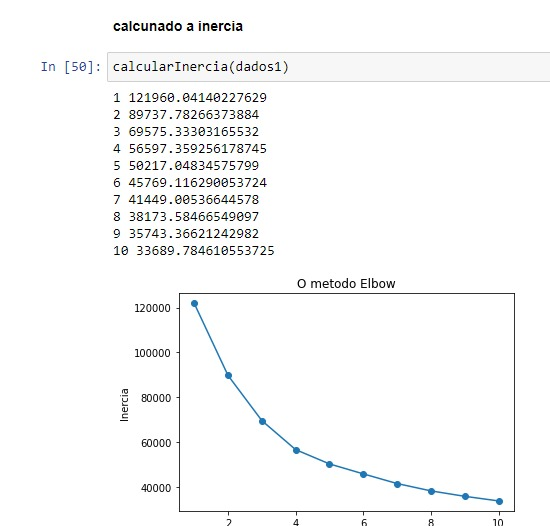


***Figura 15: Amostra dois de dados, do segundo sorteio, pelos códigos de talhão***



***Figura 16: Média, desvio padrão, menor e maior valor na Amostra Três, e valores que estão em cada intervalo, 25%, 50%, 75%.***

Cria-se um teste não supervisionado para cada agrupamento. Em cada uma dessas amostras é separado os dados de teste e treinamento, na proporção de 30% para testes e 70% para treinamentos, com a finalidade de que as análises não sejam camufladas por serem feitas em cima das mesmas informações.



***Figura 17: Calculando a inercia.***

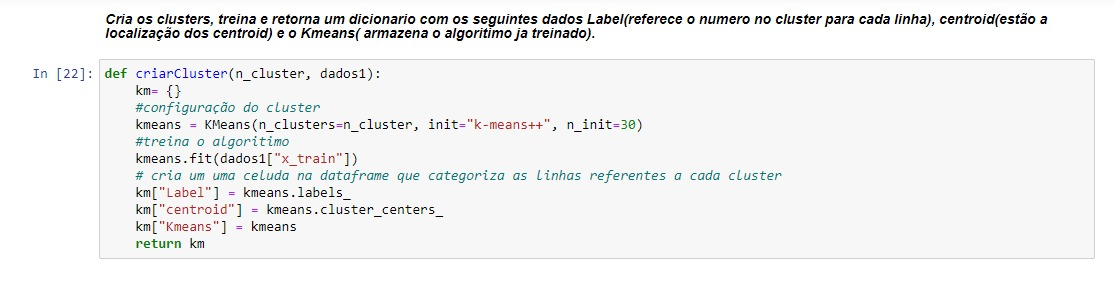
## **ALGORITMO DE AGRUPAMENTO (CLUSTERING)**

Cria-se o Algoritmo de Agrupamento (clustering), utilizando-se dos dados para se treinar o mesmo. O retorno obtido é um dicionário com os rótulos, labels (índice de cada cluster) referentes as linhas dos dados de treino. Outro rótulo são os centroides, que são a referência da localização dos dados no gráfico de cada cluster, e por fim, há ainda o rótulo kmeans, que guarda o próprio algoritmo.



***Figura 18: Criação e treinamento do Algoritmo de Clustering.***

No Algoritmo de Clustering, foram usados três parâmetros para sua construção.



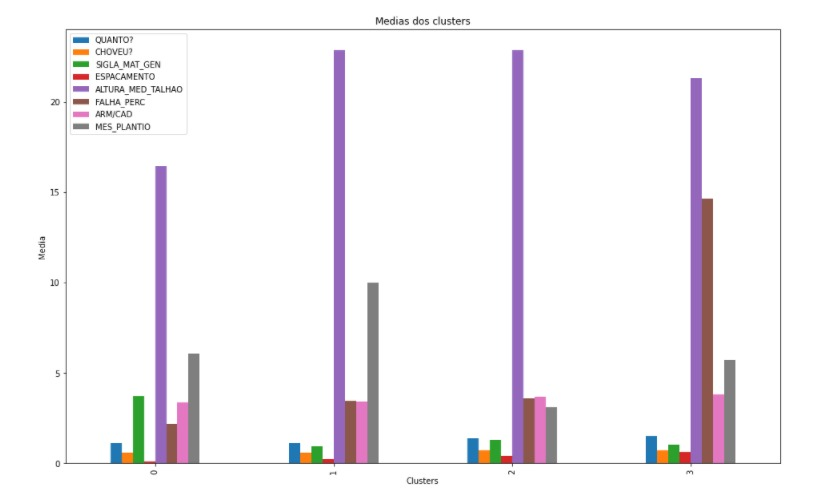
***Figura 19: Parâmetros: n\_init, init, n\_clusters.***

Foram utilizados três parâmetros: n\_init, init e n\_clusters:

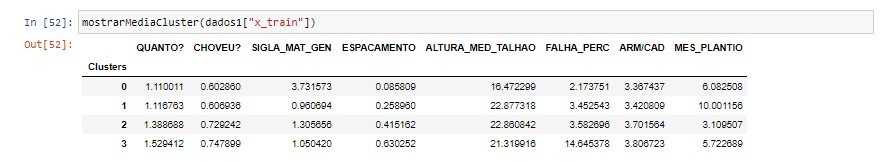
* **n\_clusters:** determina quantos clusters serão usados.
* **init:** gera randomicamente ou mais bem distribuídos os centroides. Seus parâmetros são:
  + **kmeans++:** distribui os centroides da melhor maneira possível. Se os centroides randomicamente escolhidos ficarem muito próximos, é determinada uma melhor distribuição.
  + **Randon:** localiza os centroides aleatoriamente.
* **n\_init:** é o número de vezes que o kmeans é executado para que os centroides fiquem distribuídos de forma mais coerente com os dados do cluster. Foi usado no nosso algoritmo um n\_init=30.

## **GRÁFICO DOS CLUSTERS**

A seguir, temos alguns exemplos de gráficos dos Clusters:



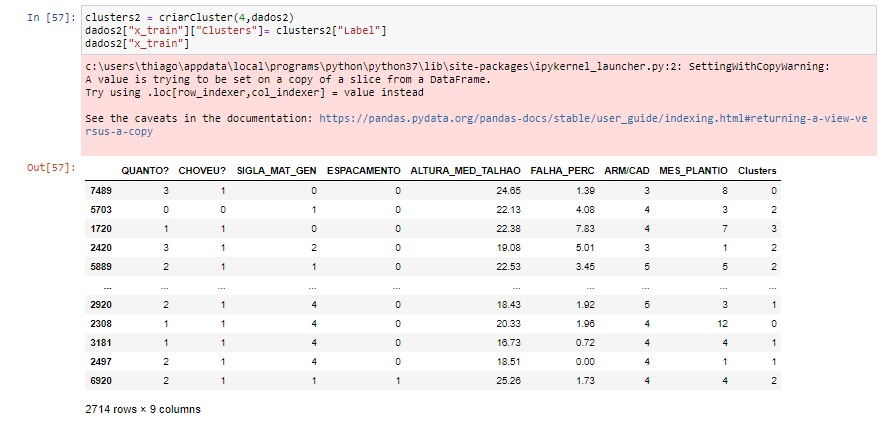
***Figura 20: Médias dos Clusters***



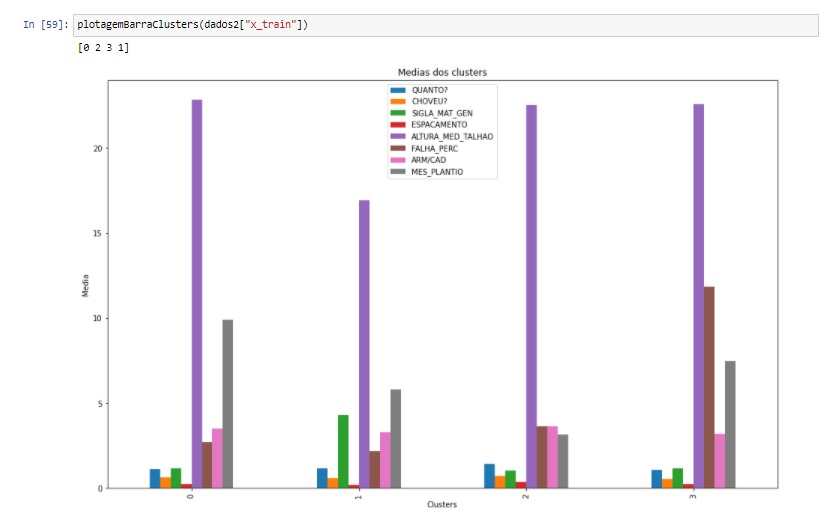
***Figura 21: Amostra 1: Média dos dados dos Clusters.***



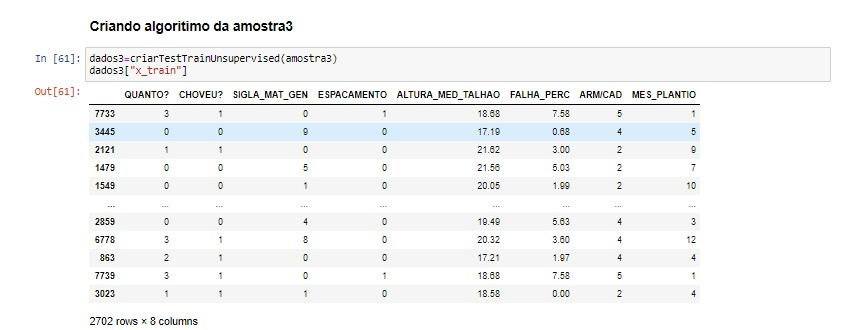
***Figura 22: Criando algoritmo da Amostra 2.***



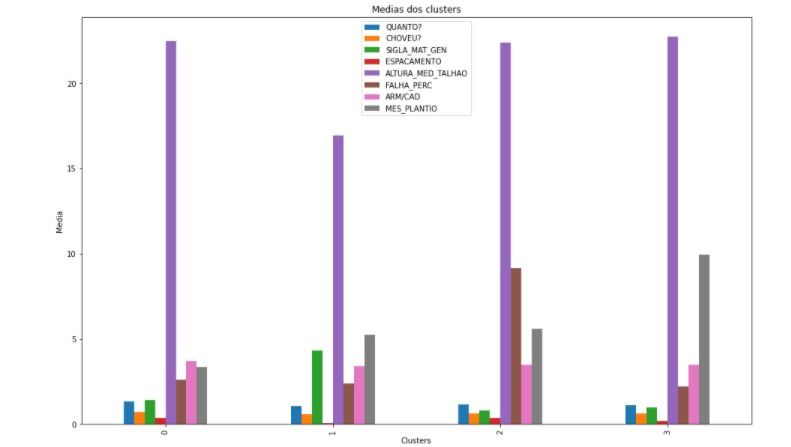
***Figura 23: Amostra 2.***



***Figura 24: plotagem da Amostra 2.***



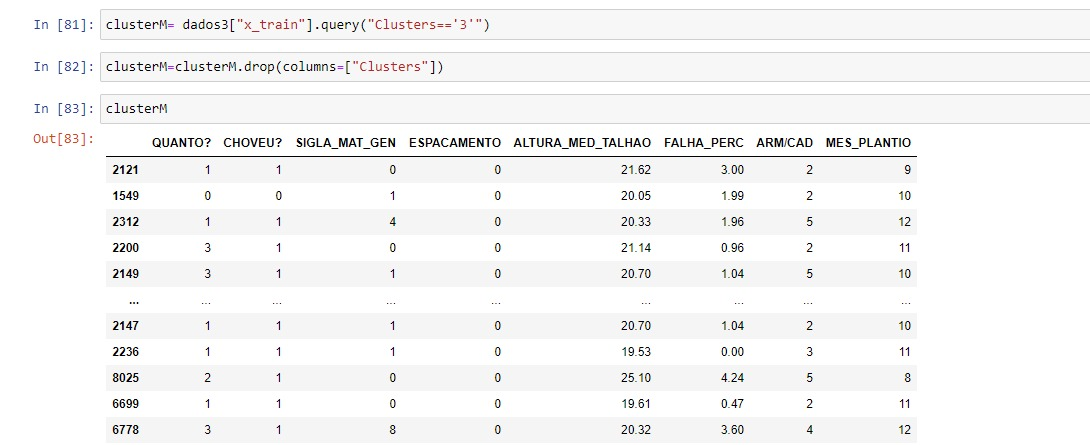
***Figura 25: Criando algoritmo da Amostra 3.***



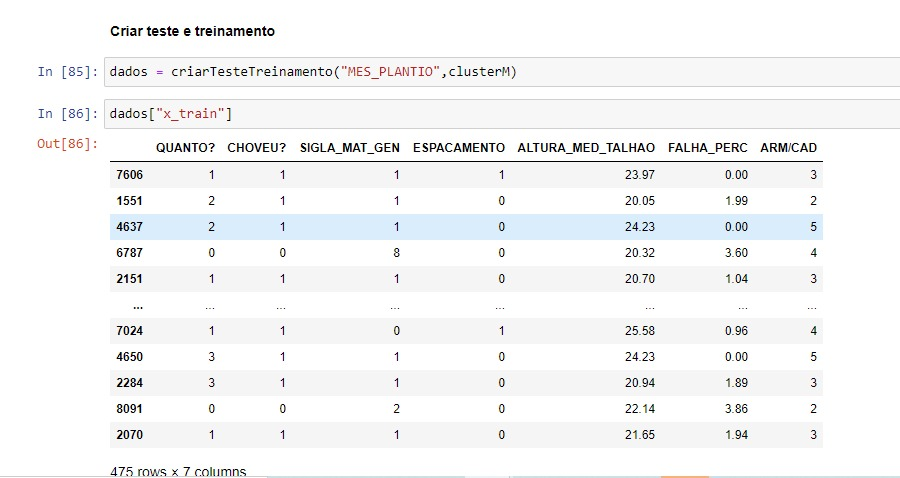
***Figura 26: Gráfico da Amostra 3.***

## **ÁRVORE DE DECISÃO**

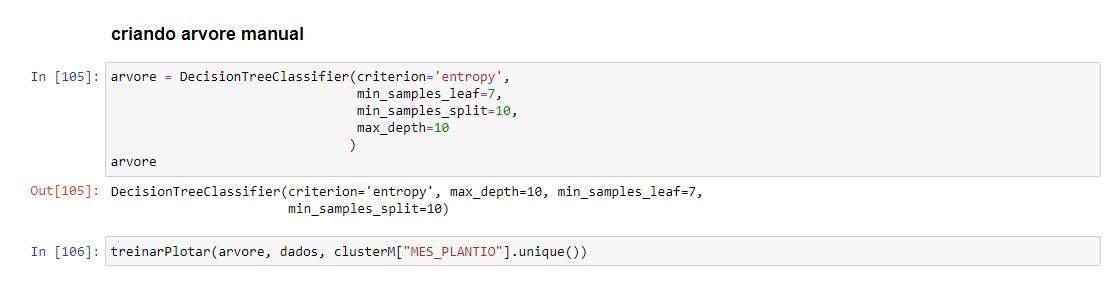
O critério de decisão para se construir a árvore teve como base o cluster que possui as médias dos índices dos meses entre 9 a 12. Observação: foi excluída a coluna de label do cluster para não se poluir a análise.



***Figura 27:Cluster contendo a média dos índices dos meses entre 9 e 12.***



***Figura 28: Criando teste e treinamento supervisionado com o atributo target como mês de plantio.***



***Figura 29: Criando e plotando a árvore manual.***

**Parâmetros da árvore:**

*criterion='entropy',*

*min\_samples\_leaf=7,*

*min\_samples\_split=10,*

*max\_depth=10*

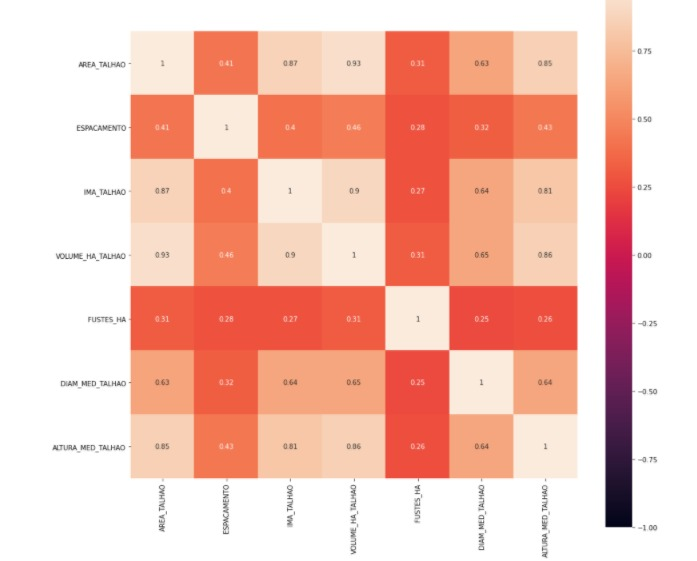
Sendo:

* **entropy:** critério da árvore de decisão que separa os dados usando o princípio da pureza dos dados. Quanto mais próximo do zero, maior é a pureza.
* **min\_samples\_leaf:** define o tamanho mínimo da folha a ser criada.
* **min\_samples\_split:** define o mínimo de elementos que a folha precisa ter para que um novo nó (galho) seja criado.
* **max\_depth:** determina a profundidade em que os nós podem estar dentro da árvore, máxima profundidade da árvore.

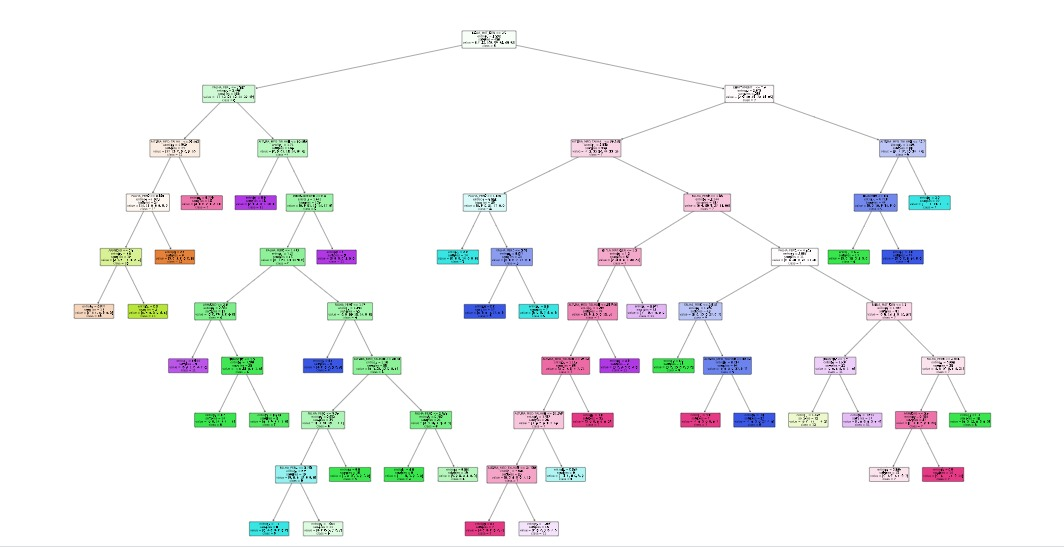
**Segunda análise:** verificando correlação entre espaçamentos e produção.



***Figura 30: Melhores correlações entre o espaçamento do talhão.***



***Figura 31: Matriz de correlação da análise do espaçamento.***



***Figura 32: Árvore de Decisão.***

# **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 deseja para a efetivação da análise.

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

As análises elaboradas não foram suficientes para conclusões específicas.

# **GLOSSÁRIO**

**Talhão** –subdivisão de área de plantio

**Hectare** –área equivalente a 10.000 m²

**Inteligência Artificial (IA)** –parte da ciência da computação que se propõe a elaborarsistemas 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

# **REFERÊNCIAS**

¹KERSYS – SOLUÇÕES PARA GESTÃO FLORESTAL E AGRICULTURA. Disponível em: [https://kersys.com.br/.](https://kersys.com.br/.%20) 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/.](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.](https://www.kaggle.com/residentmario/creating-reading-and-writing) Acesso em: 01 out. 2020

10 ALGORITMOS DE MACHINE LEARNING QUE VOCÊ PRECISA CONHECER. 18 dez. 2018. Disponível em: <https://semantix.com.br/10-algoritmos-de-machine-learning-que-voce-precisa-conhecer/>. Acesso em: 09 dez. 2020.

PRINCIPAIS ALGORITMOS DE MACHINE LEARNING. GOMES, Pedro César Tebaldi, 18 fev, 2019. Disponível em: <https://www.datageeks.com.br/algoritmos-de-machine-learning/>. Acesso em: 09 dez. 2020.

COMO FUNCIONA O ALGORITMO ÁRVORE DE DECISÃO. Disponível em: <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/>. Acesso em 09 dez. 2020.

AMOSTRA PROBABILÍSTICA: AMOSTRA POR CONGLOMERADOS. OCHOA, Carlos, 30 set. 2015. Disponível em: <https://www.netquest.com/blog/br/blog/br/amostra-conglomerados#:~:text=A%20amostra%20por%20conglomerados%20%C3%A9,cont%C3%AAm%20variabilidade%20da%20popula%C3%A7%C3%A3o%20inteira>. Acesso em: 09 dez. 2020.

SKLEARN.CLUSTER.KMEANS. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>. Acesso em: 09 dez. 2020.