



**UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA AGRÍCOLA**

FERNANDO D'ARMADA CUNHA

**ESTUDO DE CASO EM UMA UNIDADE DE
BENEFICIAMENTO DE SEMENTE UTILIZANDO
MINERAÇÃO DE DADOS**

CAMPINAS
DEZEMBRO DE 2019



**UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA AGRÍCOLA**

FERNANDO D'ARMADA CUNHA

**ESTUDO DE CASO EM UMA UNIDADE DE
BENEFICIAMENTO DE SEMENTE UTILIZANDO
MINERAÇÃO DE DADOS**

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

Orientador: Rafael Augustus de Oliveira

CAMPINAS
DEZEMBRO DE 2019

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

C914e Cunha, Fernando d'Amada, 1995-
Estudo de caso em uma unidade de beneficiamento de semente utilizando mineração de dados / Fernando d'Amada Cunha. – Campinas, SP : [s.n.], 2019.

Orientador: Rafael Augustus de Oliveira.
Trabalho de Conclusão de Curso (graduação) – Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola.

1. Secagem. 2. Árvore. 3. Mineração de dados. I. Oliveira, Rafael Augustus de, 1979-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Agrícola. III. Título.

Informações adicionais complementares

Titulo em outro idioma: Case study in a seed beneficiation unit using data mining

Palavras-chave em inglês:

Drying

Tree

Data mining

Titulação: Engenheiro Agrícola

Banca examinadora:

Rafael Augustus de Oliveira [Orientador]

João Domingos Biagi

Luis Henrique Antunes Rodrigues

Data de entrega do trabalho definitivo: 20-12-2019



ESTUDO DE CASO EM UMA UNIDADE DE BENEFICIAMENTO DE SEMENTE UTILIZANDO MINERAÇÃO DE DADOS

Fernando D'Armada Cunha

BANCA EXAMINADORA

Prof. Dr. Rafael Augustus de Oliveira
Orientador

Prof. Dr. Luiz Henrique Antunes Rodrigues

Prof. Dr. João Domingos Biagi

DEDICATÓRIA

Dedico este trabalho a meus pais, irmãs, familiares e amigos que me acompanharam neste trajeto. E para todos aqueles que de alguma forma fizeram parte dessa jornada.

AGRADECIMENTOS

Agradeço primeiramente a Deus pelo dom da vida.

Aos meus pais, Virgílio e Gleyd, por sempre me apoiarem e darem a vida para me ajudar.

Às minhas irmãs, uma eterna gratidão.

Ao meu avô Alfredo, pela amizade e companheirismo que de onde quer que ele esteja, está me acompanhando.

Ao professor Dr. Rafael, por toda a atenção e capricho no desenvolvimento desse projeto.

A Pâm por ser essa pessoa muito talentosa e atenciosa.

Aos meus colegas e amigos de graduação que de alguma forma me ajudaram.

Ao Nicolas e o Alrimar pelo convívio e ajuda nesses dois últimos anos.

A Reina, pela confiança, disponibilidade e dedicação.

**“FAÇA O TEU MELHOR, NA CONDIÇÃO QUE VOCÊ TEM,
ENQUANTO VOCÊ NÃO TEM CONDIÇÕES MELHORES,
PARA FAZER MELHOR AINDA.”**

MARIO SERGIO CORTELLA

RESUMO

A Reina Sementes LTDA é uma unidade de beneficiamento de semente de soja. Dentro da unidade, a secagem é um processo importante para assegurar a umidade do produto em baixos níveis mantendo as atividades metabólicas da semente controladas, a fim de não haver perdas e danos do produto nos demais processos da unidade beneficiadora. Para isso, saber como está sendo feita a tomada de decisões entre os tipos de secagem (contínua ou intermitente) pela empresa é uma forma de garantir um melhor aproveitamento desses processos. Sendo assim, foi realizada a modelagem em R por meio de uma árvore de decisão com o objetivo de verificar, através da análise dos dados, se a Reina está realizando a secagem conforme o indicado na literatura. A partir da árvore de decisão os parâmetros para tomada de decisão foram a umidade inicial menor ou maior que 13% e a temperatura da massa da semente menor ou maior que 34°C. Quando a umidade era superior a 13% realizava-se a secagem do tipo intermitente (66% do total de casos), e quando a umidade era inferior a 13% com temperatura da massa menor que 34°C utilizava-se a secagem contínua. Obteve-se para os resultados uma acurácia de 80,57%. A modelagem realizada foi importante para verificar e analisar os parâmetros da tomada de decisão da empresa, e comparar com a literatura que geralmente relaciona a realização da secagem intermitente para umidades de semente mais elevadas e a secagem contínua para umidades mais baixas. Sendo assim, a REINA está realizando sua tomada de decisões para a forma de secagem de maneira pertinente, dentro dos parâmetros analisados neste projeto.

Palavras chaves: secagem, árvore, mineração de dados.

ABSTRACT

Reina Sementes LTDA is a soybean seed processing unit. Within the unit, the drying process is important to protect the product at low levels while keeping the metabolic activities of the seed under control, in order to avoid loss and damage to the product in the other processes of the seed production unit. For this, knowing how a company makes the decision concerning the types of drying (continuous or intermittent) is a way to ensure the best use of these processes. Therefore, an R tree modeling was performed using a decision tree to use data analysis to see if Reina is performing the drying as indicated in the literature. From the decision tree the parameters for decision making were the initial humidity below or above 13% and the seed mass temperature below or above 34°C. When the humidity was higher than 13%, intermittent drying was performed (66% of the total cases), and when the humidity was below 13% with mass temperature below 34°C, continuous drying was used. An accuracy of 80.57% was obtained for the results. The modeling performed was important to verify and analyze the parameters of the company's decision making, and to compare with the literature that generally relates to intermittent drying for higher seed humidity and continuous drying for lower humidity. Therefore, REINA is making its decisions regarding the form of drying in a relevant way, within the parameters analyzed in this project.

Keywords: drying, machine learning, tree.

SUMÁRIO

1. Introdução	10
2. Objetivos	11
2.1. Objetivo Geral	11
2.2. Objetivos Específicos	11
3. Revisão Bibliográfica	12
3.1. Sementes	12
3.2. <i>Soja</i>	12
3.3. Parâmetros qualitativos pós-colheita de sementes	13
3.4. <i>Teor de água</i>	13
3.5. <i>Pureza</i>	13
3.6. <i>Danos mecânicos</i>	14
3.7. <i>Sanidade</i>	14
3.8. Fatores que influenciam a qualidade das sementes	15
3.9. <i>Temperatura</i>	15
3.10. <i>Umidade</i>	15
3.11. Secagem de sementes	16
3.12. <i>Secagem intermitente</i>	17
3.13. <i>Secagem contínua</i>	18
3.14. Mineração de Dados	20
3.15. <i>A potencialidade da mineração</i>	20
3.16. <i>As principais linguagens de programação utilizadas em mineração de dados</i>	20
4. Material e Métodos	21
4.1. A Reina Sementes Ltda	21
4.2. Variáveis Analisadas	22
4.3. Modelagem	24
5. Resultados e Discussão	27
6. Conclusão	31
Referências Bibliográficas	32
Apêndice	35

1 INTRODUÇÃO

Em uma Unidade de Beneficiamento existem diferentes processos até que a semente seja embalada e expedida. Recepção, pré limpeza, secagem, beneficiamento, tratamento e armazenagem são os principais processos dentro de uma Unidade de Beneficiamento de Sementes (UBS).

Após a colheita, as sementes ficam armazenadas no campo, sujeitas a condições que geram potenciais perdas qualitativas e quantitativas, provenientes dos processos iniciais de pós-colheita. Sendo assim, os produtores têm optado por retirar as sementes mais rapidamente do campo (Reina Sementes LTDA), ou seja, com um teor de água mais elevado que, segundo NUNES (2016), associado à manipulação por um período de tempo elevado, colabora para o processo de deterioração das sementes, visto que favorece as atividades metabólicas, consumo de reservas, dentre outros fatores que desfavorecem o vigor da mesma.

Dois motivos facilitam o entendimento desta decisão pelos produtores, primeiro a possibilidade de colher mais horas por dia e mais dias por safra e, segundo, menor perda de sementes por deiscência/degrane natural (NUNES, 2016).

Segundo dados da Reina Sementes LTDA, essas decisões ocasionam uma recepção de sementes com maior teor de água. Prova disso foi o aumento significativo no percentual de secagem em que no ano de 2017, 42% de toda a semente recebida possuía mais de 13% de umidade (necessitando, portanto, de secagem) e, no ano de 2019 mais de 80% da semente recebida chegou para empresa com teor de água acima de 13%, ou seja, em dois anos foi dobrada a quantidade necessitada de secagem.

Atualmente, a Reina Sementes LTDA recebe cinco cultivares de soja para o tratamento, sendo elas 7709, 6906, 7667, 7901 e 8338, resultando em cerca de 19.588 mil toneladas de sementes.

Sendo assim, é interessante analisar a tomada de decisão da empresa para verificar se a secagem está sendo realizada de forma adequada.

2 OBJETIVOS

2.1 Objetivo Geral

O presente trabalho tem por objetivo a elaboração de uma árvore de decisão do processo de secagem de soja realizado pela Empresa Reina Sementes LTDA, localizada na Rodovia Federal Br 365, km 417 em Patos de Minas - MG e verificar se a decisão está sendo tomada de forma pertinente.

2.2 Objetivos Específicos

- Adquirir e tratar os dados de secagem oferecidos pela empresa das safras de 2017, por meio do programa estatístico RStudio;
- Criar uma modelagem computacional a partir do tratamento dos dados;
- Verificar se os parâmetros da tomada de decisão utilizados pela empresa estão pertinentes com a literatura.

3 REVISÃO BIBLIOGRÁFICA

3.1 Sementes

Uma semente precisa estar viva e capaz de germinar sob condições adequadas, enquanto o grão é utilizado como alimento (MATOS, 2013). Entretanto, segundo KRZYZANOWSKI, FRANÇA-NETO e HENNING (2018), a semente não é necessariamente um grão que germina e, portanto, possui diversos atributos de qualidade em nível genético, físico, fisiológico, entre outros, que um grão não possui.

A semente de alta qualidade é aquela que apresenta características como: altas taxas de germinação, de vigor e de sanidade. Além disso, possui pureza física e de variedade (não contendo sementes de erva daninha). No campo essas sementes resultam em plântulas fortes e bem desenvolvidas, com maior velocidade de emergência e desenvolvimento. Ademais, áreas plantadas originadas de sementes de alta qualidade, em situações de estresse (secas, baixas temperaturas, etc.), sofrem menos danos que aquelas que utilizam sementes de menor qualidade (FRANÇA NETO, KRZYZANOWSKI & HENNING, 2010). Sendo assim, uma semente de alta qualidade é o primeiro passo para uma produção bem sucedida.

3.1.1 Soja

A soja (*Glycine max L.*), pertencente à família *Fabaceae* é originária da Ásia. Seus primeiros registros no Brasil foram no século 18, em 1882 na Bahia (MANDARINO, 2017) daí em diante houve expansão desse produto, fazendo com que o Brasil se tornasse o segundo maior produtor, ficando atrás apenas dos EUA (EMBRAPA, 2019).

Segundo o levantamento realizado pela Companhia Nacional de Abastecimento (CONAB), o Brasil terá produção de grãos recorde na safra de 2019/2020 (Agência Brasil, 2019). Além disso, o estudo destaca que a soja apresenta crescimento na área plantada (2,3% maior que a última temporada) e aumento de 5,1% em relação a produção de 2018/2019, chegando a 120,9 milhões de toneladas (CONAB, 2019).

Essa produção crescente, ano a ano, aumenta a procura por tecnologias que viabilizem uma maior produção em menor área. O uso de sementes de alta qualidade que apresentem alto poder germinativo e potencial genético para a produção de grãos de soja cada vez melhores e resistentes a pragas e doenças vem crescendo, e instigam o mercado a investir na produção e beneficiamento dessas sementes (MAPA, 2019).

A produção de sementes de soja no Brasil foi de 2,3 milhões de toneladas na safra de 2014/2015 (63% do total de semente produzida), sendo 65% oriundas de produtores regularizados e 35% do mercado informal (EMBRAPA, 2015).

Vale destacar que pesquisas realizadas pela Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA, revelam que a utilização de sementes de alta qualidade pode aumentar a produtividade em cerca de 15% (MAPA, 2019).

3.2 Parâmetros qualitativos pós-colheita de sementes

Um fator crucial para o bom desenvolvimento germinativo é a qualidade da semente. Quando plantada, uma semente de alta qualidade terá maiores chances e condições de superar as adversidades ambientais que podem interferir na produtividade final (PESKE, 2015).

Segundo PESKE (2015), aspectos como: germinação, vigor, pureza física, genética e sanidade, estão diretamente ligadas a boa qualidade da semente.

3.2.1 Teor de água

Teor de água é a relação entre a massa de água e a massa seca de um produto podendo ser expressa tanto em base seca, quanto em base úmida (TRINDADE, 2013). Esta primeira, demonstrada na equação 1, é expressa em relação à massa seca; já a segunda, representada pela equação 2, é em relação à massa total do produto:

$$U(\%) = 100 * \frac{Mu}{Ms} \quad [1]$$

em que, Mu é igual a massa úmida e Ms é igual a massa seca.

$$U(\%) = 100 * \frac{Mu}{Mt} \quad [2]$$

em que, Mu é igual a massa úmida e Mt é igual a massa total do produto.

3.2.2 Pureza

O teste de pureza deve ser realizado na recepção do produto. Uma amostra da carga é retirada, pesada e certificada. Após esta primeira verificação, as sementes devem ser examinadas e separadas entre sementes puras, materiais inertes e outras sementes. Para serem consideradas puras, as sementes devem pertencer à espécie da análise demandada pelo solicitante ou por ser a predominante na amostra (MAPA, 2009).

Para a verificação de outras sementes devem ser incluídas a proporção de divergência com a espécie analisada (semente pura). Ainda segundo MAPA (2009), são considerados materiais inertes aqueles que não são definidos como semente pura e outras sementes.

Além disso, é realizado o teste VOC (verificação de outros cultivares), sendo feito por um especialista familiarizado com o cultivar na intenção de verificar a não compatibilização do mesmo (MAPA, 2009).

3.2.3 *Danos mecânicos*

A soja é muito suscetível a danos mecânicos (ruptura de tegumento) devido ao fato das partes vitais do eixo embrionário estarem situadas sobre um tegumento delgado (FLOR, 2003). Por isso, é importante a verificação da porcentagem de danos mecânicos sofridos pela semente ao longo da cadeia de produção.

Um dos principais testes realizados para esta verificação é o teste do hipoclorito de sódio, no qual, rapidamente, é identificado esse porcentual de injúria. Este teste consiste na contagem de pelo menos 100 sementes e em seguida na submersão das mesmas em uma solução com 25 ml de hipoclorito de sódio e 975 ml de água. Após 10 minutos, são retiradas as sementes e espalhadas sobre um papel toalha fazendo a análise do quanto a semente absorve da massa dessa solução, ressaltando que porcentagens acima de 10 indicam que a semente está muito danificada (KRZYZANOWSKI, 2004).

3.2.4 *Sanidade*

A sanidade de uma semente diz respeito à presença ou ausência de agentes patogênicos como fungos, vírus, bactérias e nematóides. Os testes de sanidade são feitos com o intuito de identificar quais são esses agentes e comparada com o Manual de Análise Sanitária da Semente (MAPA, 2009).

3.2.5 *Germinação*

Germinação por definição é a aptidão para produzir uma planta normal em condições usuais de campo. Segundo o Ministério da Agricultura, Pecuária e Abastecimento (MAPA), os parâmetros utilizados para verificar se a plântula continuará seu desenvolvimento até dar origem a uma planta normal são: gemas terminais, sistema radicular, cotilédones e a parte aérea (MAPA, 2009).

O teste do tetrazólio é um bom exemplo para certificar a viabilidade das sementes quando o resultado de testes germinativos apresentarem uma alta porcentagem de sementes não germinadas (MAPA, 2009). Segundo FRANÇA NETO, KRZYZANOWSKI & COSTA (1998), este teste possui como indicativo de viabilidade a coloração resultante das reações das enzimas desidrogenases.

3.3 Fatores que influenciam a qualidade das sementes

3.3.1 Temperatura

A temperatura é um fator importante e deve sempre ser controlada, visto que é preponderante para diversos problemas. Na Tabela 1, são mostrados os principais aspectos de temperaturas altas e baixas e suas consequências para as sementes em diferentes etapas.

<i>Aspectos de temperaturas não adequadas sobre as sementes</i>		
Etapa do Processo	Temperatura	
	Baixa	Alta
<i>Desenvolvimento da cultura a campo</i>	Formação de grãos chochos e estéreis	Abortamento de vagens, alterações no ciclo da cultura
<i>Secagem</i>	Ar desumidificado e frio possui capacidade de retirar água das sementes	Temperatura da massa de grãos não pode ultrapassar 43°C
<i>Armazenamento</i>	Desejável	Desenvolvimento de patógenos; Aumento da respiração e da deterioração
<i>Logística de transporte</i>	Desejável	Se ultrapassar 24 horas pode haver comprometimento da qualidade

Tabela 1: Aspectos de temperatura não adequadas sobre as sementes (Fonte: MENEGHELLO, 2014).

3.3.2 Umidade

Segundo MENEGUELLO (2014), o conhecimento sobre a propriedade higroscópica, capacidade que o produto tem de ganhar ou perder água em um processo dinâmico de acordo com a umidade relativa e temperatura, tem total relevância quando se fala de controle de umidade.

Na secagem, deseja-se que as sementes apresentem baixa umidade pensando no controle das suas atividades metabólicas, o que irá garantir a qualidade da semente. Alguns reflexos da falta ou do excesso de umidade das sementes em suas diferentes etapas do processo podem ser observadas na Tabela 2:

Reflexos da falta e do excesso de umidade sobre as sementes		
Etapa do Processo	UMIDADE	
	Falta	Excesso
Desenvolvimento da cultura a campo	Emergência desuniforme; Mau desenvolvimento da cultura	Mau desenvolvimento da cultura, aumento da incidência de doenças
Colheita	Aumento de danos mecânicos	Aumento da velocidade de deterioração, Germinação na espiga/vagem
Secagem	Desejável	Secagem OBRIGATÓRIA
Armazenamento	Desejável	Desenvolvimento de patógenos, aumento da respiração e da deterioração
Semeadura	Não desencadeia a germinação	Dificulta a emergência
Tratamento	Dificuldade de distribuição uniforme do produto for usada pouca calda	Aumento na umidade se for usado muita calda

Tabela 2: Reflexo da falta e do excesso de umidade sobre as sementes (Fonte: MENEGHELLO, 2014).

3.4 Secagem de sementes

Segundo GARCIA *et al.* (2002), a secagem, quando aplicada às sementes, contribui para a preservação da qualidade durante o armazenamento, possibilitando também que a colheita seja feita antecipadamente, quando necessário, impedindo eventuais perdas.

Além disso, a secagem pode ser definida como um processo fundamental na produção de sementes de alta qualidade, pois reduz o teor de água (e conseqüentemente a atividade metabólica), preservando as sementes das alterações químicas e físicas, contribuindo para manutenção da qualidade durante o armazenamento (BAUDET & VILLELA, 1999).

Segundo RADKE, BABATA & VILLELA (2017), durante a fase de armazenamento tanto a umidade do ar (do ambiente) como a umidade da semente devem ser monitoradas e controladas, para que a qualidade da semente seja garantida até o final do processo dentro da unidade de beneficiamento de semente.

Durante a escolha do método de secagem fatores como quantidade de sementes a serem secadas e qualidade desejada são determinantes (GARCIA *et al.*, 2004). A secagem natural (Figura 1), realizada na própria lavoura, não permite a obtenção de um produto com nível de qualidade elevado, pois está sujeita às intempéries climáticas, além da desuniformidade causada pelos diferentes graus de maturação entre as sementes de uma planta e desta em relação a outras plantas (PERES & PESKE, 2016).

Para a secagem de grandes quantidades de semente é indicada a utilização da secagem artificial (Figura 2), que por sua vez tem seus custos em função de fatores como volume de sementes, temperatura do ar, umidade, velocidade de secagem, entre outros (GARCIA *et al.* 2004). Sendo a secagem com ar forçado fundamental para a obtenção de sementes de alta qualidade (PERES & PESKE, 2016).

Figura 1: Exemplo de secagem natural aplicada a sementes de soja.



Fonte: NUNES (2016).

Figura 2: Exemplo de secador.



Fonte: AVELAR, PESKE & VILLELA (2012).

Segundo NUNES (2016), a secagem artificial é realizada a partir da exposição das sementes a um fluxo de ar, aquecido ou não, e ela se diferencia e caracteriza conforme o tipo de secagem, podendo ser estacionária, contínua ou intermitente.

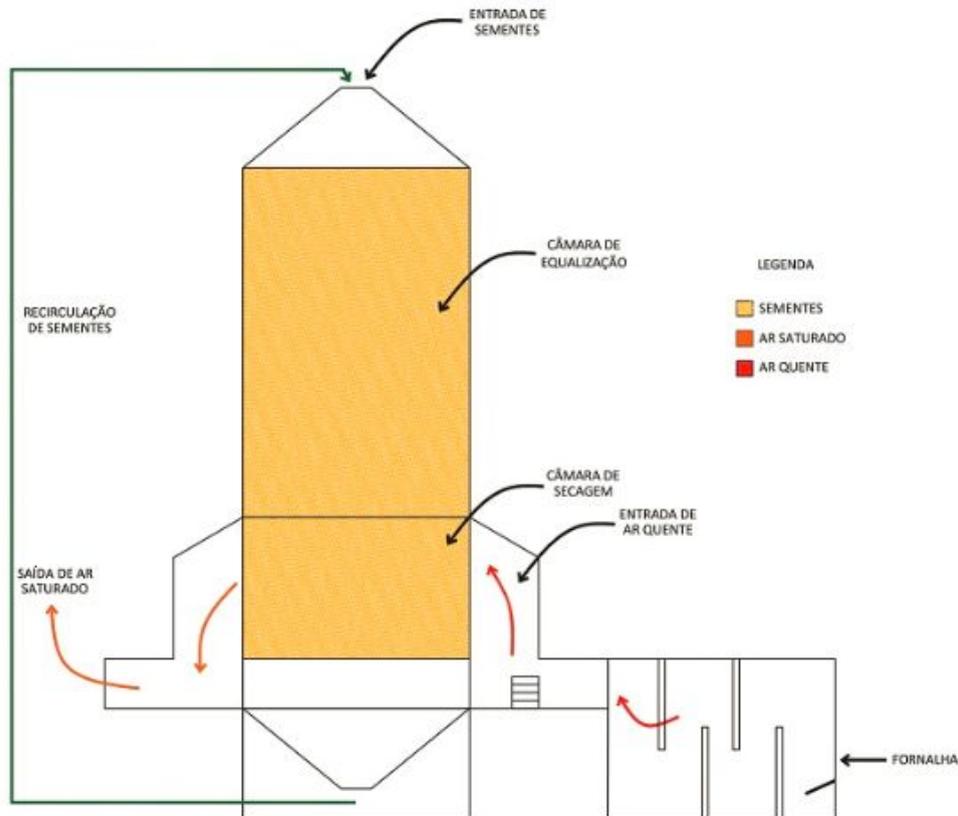
3.4.1 *Secagem intermitente*

Os secadores intermitentes (Figura 3) são geralmente compostos por uma câmara de secagem (onde ocorrem as trocas de calor e água, entre as sementes e o ar aquecido) e à câmara de equalização (onde ocorre a movimentação do produto, porém sem a passagem de ar) (PERES & PESKE, 2016).

Segundo GARCIA *et al.* (2004), a secagem intermitente é caracterizada justamente pelo contato das sementes com o ar aquecido durante curtos períodos de tempo, intercalados com períodos sem exposição ao fluxo de ar aquecido na câmara de equalização. Isso permite que haja a homogeneização da umidade e temperatura nas sementes quando as mesmas estão

passando pelas partes do sistema onde não recebem ar aquecido. Além disso, essa intermitência possibilita que ocorra o transporte de água do interior para a superfície da semente durante o período de equalização, diminuindo o gradiente de concentração dentro da semente (NUNES, 2016). Em seguida, essa água é removida quando o produto passa novamente pela câmara de secagem, e o processo se repete até que seja alcançada a umidade desejada (PERES & PESKE, 2016).

Figura 3: Exemplo de secador intermitente.



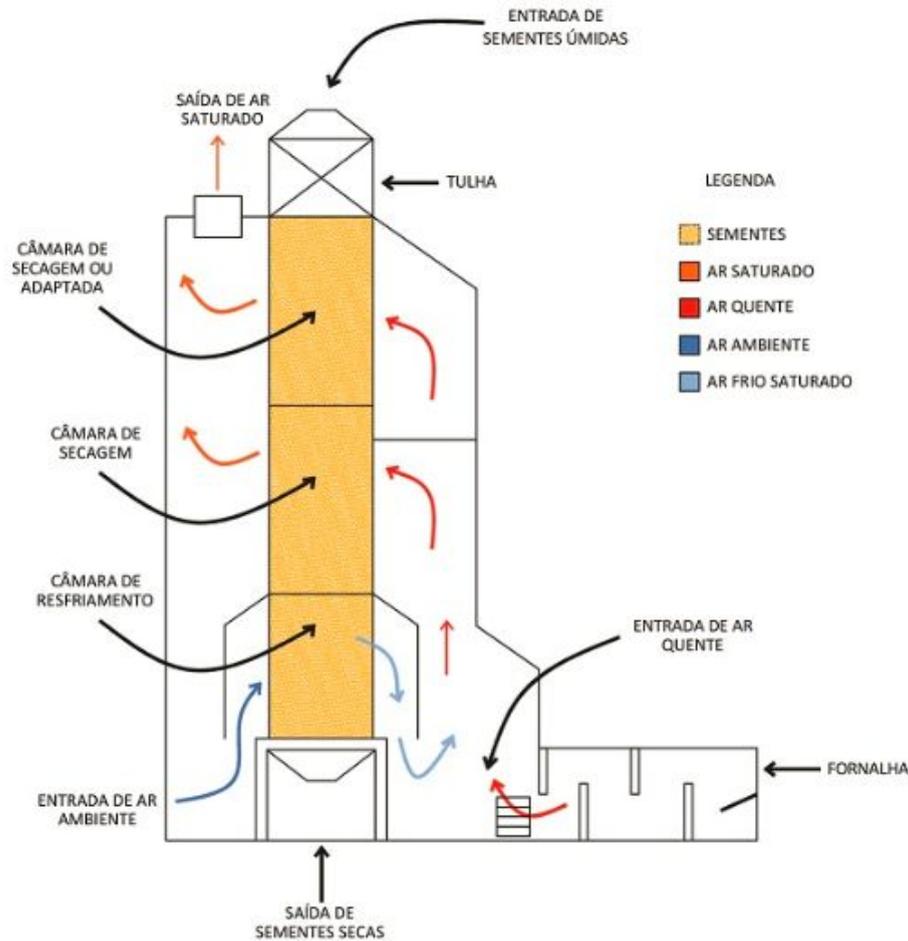
Fonte: PERES & PESKE (2016).

3.4.2 Secagem contínua

A secagem contínua acontece quando o produto possui uma entrada contínua no secador. Esses secadores contínuos geralmente são compostos por duas câmaras, uma de secagem e outra de resfriamento (Figura 4). Durante o processo, as sementes passam pela câmara de secagem, entrando úmidas no topo e saindo secas na base do secador. Para que seja possível a secagem durante essa passagem única pelo secador, é necessária a elevação da temperatura e/ou a diminuição do fluxo de sementes que passam pela câmara (aumento do tempo de exposição ao ar aquecido) o que pode causar danos térmicos às sementes, sendo,

portanto, necessária atenção na execução de projeto desses secadores (PERES & PESKE, 2016; NUNES, 2016).

Figura 4: Exemplo de secador de fluxo contínuo.



Fonte: PERES & PESKE (2016).

3.4.3 O Fluxo de ar

Devido ao fato da secagem envolver processos de transferência de energia e massa entre o ar e o produto a ser secado, de forma que ambos entrem em equilíbrio (térmico e higroscópico) é necessário que o produto esteja envolto com uma massa de ar capaz de absorver e transportar a água removida do mesmo (PERES & PESKE, 2016).

O fluxo de ar é um parâmetro importante durante o processo de secagem e influencia diretamente o tempo da mesma. Segundo BAUDET & VILLELA (1999), ele deve ser suficiente de forma que não se sature antes de sair do secador, além de ser capaz de absorver a água evaporada das sementes. Desta forma, esse fluxo de ar quando adequado permite o transporte de calor e água, sendo essa transferência de energia do ar dependente das

condições de temperatura, umidade, vazão de ar e da área superficial das sementes (PERES & PESKE, 2016).

3.5 Mineração de Dados

3.5.1 A potencialidade da mineração

Com o avanço da tecnologia em diferentes segmentos, o número de dados gerados, coletados e armazenados aumentou de forma significativa. John Naisbitt disse que “estamos nos afogando em informação, mas com sede de conhecimento”, com isso entende-se a necessidade do mercado em conseguir analisar e tirar conclusões das diversas informações contidas nesses bancos de dados (SANTOS, s.d.).

De acordo com CHAVES [s.d.], em se tratando de agricultura, decisões de tempo de plantio, colheita, adubação do solo, agricultura de precisão, alimentação de gado, entre outros, são alguns dos muitos exemplos que podem ser citados a respeito da mineração de dados aplicada ao setor agrícola.

3.5.2 As principais linguagens de programação utilizadas em mineração de dados

Dentre as diversas linguagens de programação existentes, as que mais se destacam para a ciência de dados são: *R* (*Rstudio*®) e *Python* (GRANATYR, 2017).

De acordo com TRINDADE (2017) a linguagem R foi criada por estatísticos e possui pacotes excelentes para diversas áreas de aplicação principalmente em se tratando de resolução com complexidade matemática. Possui ainda uma maior complexidade na escrita dos códigos um pouco mais acentuada do que o Python, que por sua vez é uma excelente interface orientada a objetos considerada, também, umas das linguagens mais rápidas.

4 MATERIAL E MÉTODOS

4.1 A Reina Sementes Ltda

Para atender a necessidade de secagem, a Reina (Figura 5) possui três secadores colunados do tipo SCR-30 Rotasilos (Figura 6), sendo este uma torre de secagem com difusores, escada amortecedora de sementes, funil de descarga para fitas transportadoras, controle de temperatura na massa e, na entrada do ar, controle de nível rotativo e variador de rotação por inversor de frequência na descarga (ROTASILOS, 2018).

Figura 5: Foto da Reina Sementes.



Fonte: Autoria própria.

Figura 6: (a) Secador 1, (b) Secador 2 e 3 (ao fundo).



A



B

Fonte: Autoria própria.

A Empresa, em períodos de safra, opera em três turnos de trabalho, conseguindo efetuar no máximo 180 toneladas por secador por dia. Cada secador possui capacidade de 30 toneladas por hora e trabalha, pela queima de combustível, em uma faixa de temperatura de secagem que varia entre 60 a 65°C com isso, consegue retirar entre 0,8 a 1,2% de umidade da semente por hora por secador, trabalhando entre secagem contínua e intermitente.

4.2 Variáveis Analisadas

A Reina Sementes disponibilizou para o presente estudo uma planilha com todos os dados de secagem realizados nesta safra. Na mesma, continham os seguintes campos:

- Cultivar da soja;
- Número do secador utilizado para secagem;
- Data enchimento do secador;
- Hora fim do enchimento;
- Data do esvaziamento;
- Hora fim do esvaziamento;
- Verificações (explicada posteriormente);
- Temperatura de entrada no secador;
- Temperatura do secador;
- Tipo de secagem;
- Umidade inicial da semente;
- Umidade ambiente;
- Temperatura ambiente;
- Temperatura da massa da semente;
- Dano mecânico final (porcentagem do quanto a semente danificou);
- Umidade da semente final;
- Temperatura da massa da semente final.

Além dos dados disponibilizados, foi acrescentado o tempo total de secagem, calculado a partir da equação 3:

$$\text{Tempo Total de Secagem} = \text{hora do fim do esvaziamento} - \text{hora do fim do enchimento} \quad [3]$$

Com o intuito de analisar a relação entre as variáveis e os dados computados pela Reina, foram geradas tabelas pela função *prop.table*. A Tabela 3 mostra a relação em

porcentagem de secagem e cultivar recebido pela empresa, sendo o volume recebido do cultivar 7709 o de valor mais expressivo (52%), devido a localidade da Reina, ou seja, devido ao número de produtores que plantam esse tipo de soja na região.

Cultivar	Soja Seca (%)
6906	10,2
7667	3,8
7709	29,5
7901	52,0
8338	4,5

Tabela 3: Porcentagem de soja seca por cultivar (Fonte: Autoria própria).

Dentre os secadores, percebe-se pela Tabela 4 que o secador 1 foi o mais utilizado, o que é explicado devido a sua sequência com a linha de beneficiamento da empresa, ou seja, a escolha do mesmo já é predisposta a acontecer.

Número do Secador	Utilização de secagem (%)
1	41,0
2	31,0
3	28,0

Tabela 4: Porcentagem de soja seca por secador (Fonte: Autoria própria).

Entre o cultivar e o número do secador, percebe-se que em 100% do volume recebido não houve secagem dos cultivares 7667 e 8338 no secador 1 e o maior volume seco foi no secador 2, correspondente ao tipo 7901, como mostrado na Tabela 5:

Cultivar	Secador 1 (%)	Secador 2 (%)	Secador 3 (%)
6906	5,32	4,89	0
7667	0	1,87	1,95
7709	20,86	1,15	7,5
7901	14,74	22,52	14,74
8338	0	0	4,46

Tabela 5: Relação entre o tipo de cultivar e a porcentagem do secador utilizado (Fonte: Autoria própria).

Para quantificar a porcentagem entre os tipos de secagem (contínuo ou intermitente) e o tipo do cultivar, na Tabela 6, verifica-se que 60% do volume seco é resultante da secagem intermitente e, o restante, do processo contínuo.

Tipo de Secagem	Cultivar 6906	Cultivar 7667	Cultivar 7709	Cultivar 7901	Cultivar 8338
Contínuo	4,10	0,94	12,04	21,65	1,18
Intermitente	6,05	2,86	18,66	29,32	3,20

Tabela 6: Relação entre o tipo de cultivar e a porcentagem do secador utilizado (Fonte: Autoria própria).

Ainda para analisar e identificar as relações entre as variáveis, utilizou-se a função *ggplot*, para gerar determinados gráficos, como *Tipo de Secagem x Umidade Inicial e Umidade Final x Densidade* que serão discutidos no decorrer deste trabalho.

Vale destacar que, durante a plotagem dos gráficos, percebeu-se a existência de *outliers* (valores inconsistentes), possivelmente gerados por erros humanos durante a coleta dos dados, visto que eles circulam entre duas pessoas, a primeira que anota os resultados obtidos em uma folha de papel e a segunda que os transfere para o computador, conforme a logística informada pela empresa.

4.3 Modelagem

Os dados que inicialmente eram em formato *.xlsx* (*Excel*®) foi alterado para *.csv* para então ser computado no *Rstudio*®. Em Anexo 1 está demonstrado todo o *script* realizado.

Primeiramente, foi realizado um tratamento para que não houvesse excedente de informações ou problemas durante a modelagem. Colunas como data, horários de início de enchimento, fim do esvaziamento e número de verificações (coletas de amostra para análise de umidade da semente e danos mecânicos por meio do teste de hipoclorito) não traria correlações necessárias para a análise, por isso foram deletadas.

Depois de analisar a correlação dos dados da secagem da empresa, como mostrado anteriormente, foi feita a modelagem do sistema por meio de uma árvore de decisão para identificar possíveis parâmetros para comparar com a decisão real (realizada pela Reina) da secagem.

A árvore de decisão (*decision tree*) é um modelo de aprendizagem de máquina supervisionado, no qual forma-se um gráfico no formato de uma árvore. Ela mostra as

condições e probabilidades em que foram realizadas as tomadas de decisão, como se fossem caminhos percorridos, servindo tanto para problemas de regressão quanto para classificação (PRATES, 2018). Ainda segundo Prates (2018), “construir uma árvore de decisão se trata de encontrar regras sobre as variáveis do modelo que retornam o maior ganho de informação, isto é, que tornam os ramos da árvore mais homogêneos, com menor entropia”, sendo a entropia a probabilidade do evento ocorrer.

Primeiramente, foi definida a variável “*tipo_de_secagem*” como a variável *target*, ou seja, a que será analisada. Observando no conjunto de dados, o tipo de secagem varia entre contínuo e intermitente. Então, alterou-se o nome desses parâmetros respectivamente para 0 e 1, como mostrado na Figura 7, para se obter um *target* binário. Com isso, a partir de uma seleção aleatória, o restante dos dados foi utilizado para medir a probabilidade da ocorrência do evento (intermitente = 1, contínuo = 0).

Figura 7: Script de mudança para resultado binário.

```
dados$tipo_de_secagem[dados$tipo_de_secagem == "Intermitente"] = 1
dados$tipo_de_secagem[dados$tipo_de_secagem == "Contínuo"] = 0
```

Fonte: Autoria própria.

Na sequência, foi realizada a separação entre conjunto de treino e conjunto de teste (Figura 8). O objetivo do conjunto de treino, como o próprio nome diz, é treinar o modelo realizado com o banco de dados para que, então, o modelo seja testado pelo conjunto de teste. A proporção entre esses conjuntos foi definida em 80% para o conjunto de treino, chamado na modelagem de “*base_treinamento*” entre valores aleatórios (*sample.split*) da coluna escolhida e 20% dos dados “*tipo_de_secagem*” para o conjunto de teste.

Figura 8: Script da separação dos dados entre treino e teste.

```
set.seed(160150)
indice_treino = sample.split(dados$tipo_de_secagem, SplitRatio = 0.8)
base_treinamento = subset(dados, indice_treino == TRUE)
base_teste = subset(dados, indice_treino == FALSE)
is.data.frame(base_treinamento)
head(base_treinamento)
```

Fonte: Autoria própria.

Após essa separação dos conjuntos, fez-se o modelo de árvore de decisão utilizando a biblioteca “*rpart*”, mostrada na Figura 9. Este modelo foi então treinado utilizando os dados da coluna “*tipo_de_secagem*” juntamente com todos os outros do conjunto de treinamento.

Em seguida, definiu-se três parâmetros dentro do modelo: *minbucket* (mínimo de correlações), *maxdepth* (máximo de profundidade da árvore) e *method* (método de classificação).

Figura 9: Script da árvore de decisão.

```
set.seed(160150)
classificador_tree= rpart(tipo_de_secagem ~.,
                          base_treinamento,
                          control = rpart.control(
                            minbucket = 15,
                            maxdepth = 5,
                            method = 'class'))
classificador_tree

previsoes_tree = predict(classificador_tree, base_teste,
                        type = 'class')

acuracia_tree = mean(previsoes_tree == base_teste$tipo_de_secagem)
acuracia_tree
table(previsoes_tree, base_teste$tipo_de_secagem)
rpart.plot(classificador_tree)
```

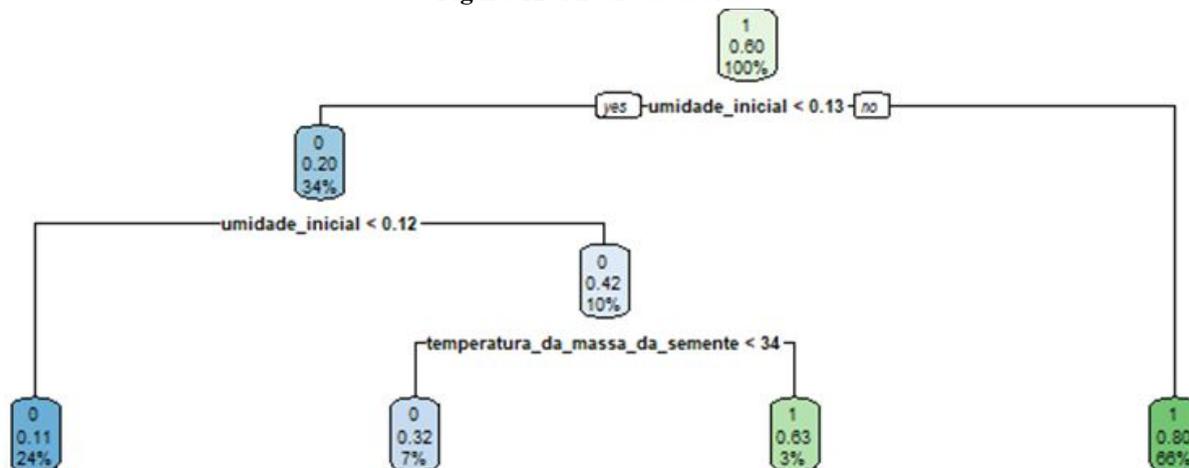
Fonte: Autoria própria.

Logo após, foi plotada a matriz de confusão da árvore de decisão, pois é uma forma efetiva de validar modelos de classificação, mostrando a relação entre o número de acertos com as classificações preditas corretamente por cada classe, verdadeiro positivo e verdadeiro negativo (MONARD & BARANAUSKAS, 1999). Por fim, verificou-se a acurácia (relação entre verdadeiros e o total analisado) do modelo.

5 RESULTADOS E DISCUSSÃO

O primeiro resultado da modelagem foi a árvore de decisão mostrada na Figura 12.

Figura 12: Árvore de decisão.



Fonte: Autoria própria.

Observa-se que o modelo definiu como *setpoint* (referência ou nó inicial da árvore) a secagem intermitente, possivelmente devido ao número de repetições, visto que, de todo volume secado pela empresa, 60% se refere a secagem intermitente.

Além disso, a primeira variável analisada corresponde a que tem uma maior relação de entropia (maior ganho de informação). Logo, o que divide a primeira tomada de decisão seria a umidade inicial da semente (umidade inicial < 0,13). Caso a umidade fosse maior que 13% a primeira e única decisão seria a secagem intermitente (o que correspondeu em 66% dos casos), do contrário, seria a secagem contínua, com mais passos para a verificação.

No nó da secagem contínua, o modelo ainda utiliza a umidade inicial da semente como critério de decisão, considerando que, quando menor que 12%, a secagem seria primeira e unicamente do tipo contínua (o que correspondeu a 24% dos casos). Do contrário, é necessária uma terceira verificação, usando como critério de decisão a temperatura da massa da semente (novo *setpoint*) que, quando menor que 34°C (7% dos casos), corresponde a secagem contínua, e, quando maior (3% dos casos), corresponde à secagem intermitente.

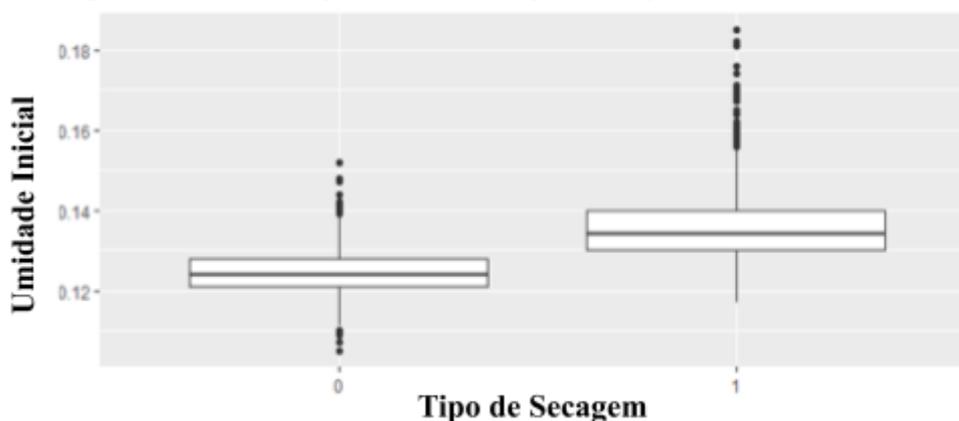
Na matriz de confusão mostrada na Tabela 5, as linhas estão representando as predições e as colunas o realizado, o modelo acertou em 66 casos (verdadeiro positivo) que a secagem era do tipo contínua e 158 casos (verdadeiro negativo) que era intermitente. Para a secagem contínua, a validação foi de 59,45% dos casos, em contrapartida, para a intermitente, correspondeu a 94,61% do casos. Assim, a acurácia obtida foi de 80,57%.

Matriz de Confusão		Classe Real	
		Contínuo	Intermitente
Predições	Contínuo	66	9
	Intermitente	45	158

Tabela 5: Matriz de confusão (Fonte: Autoria própria).

A fim de justificar a importância dada a umidade inicial da semente pela empresa, foi plotado um gráfico (Figura 13) para validação da primeira linha de decisão da árvore analisada, comprovando que a empresa realmente secou de maneira intermitente a maioria das sementes com umidade maior que 13%.

Figura 13: Gráfico boxplot relacionando tipo de secagem com umidade inicial.



Fonte: Autoria própria.

Além disso, observou-se que não houve relação entre a umidade relativa do ar e a temperatura externa com a secagem, o que pode indicar equívocos no resultado da árvore, pois a secagem quando realizada com a umidade relativa alta e/ou temperatura externa baixa, não é tão eficiente (sendo o gasto energético maior, exemplo: baixa eficiência de secagem durante a noite), havendo assim relação direta entre as condições de temperatura e umidade externa e o resultado da secagem, segundo relatos da empresa.

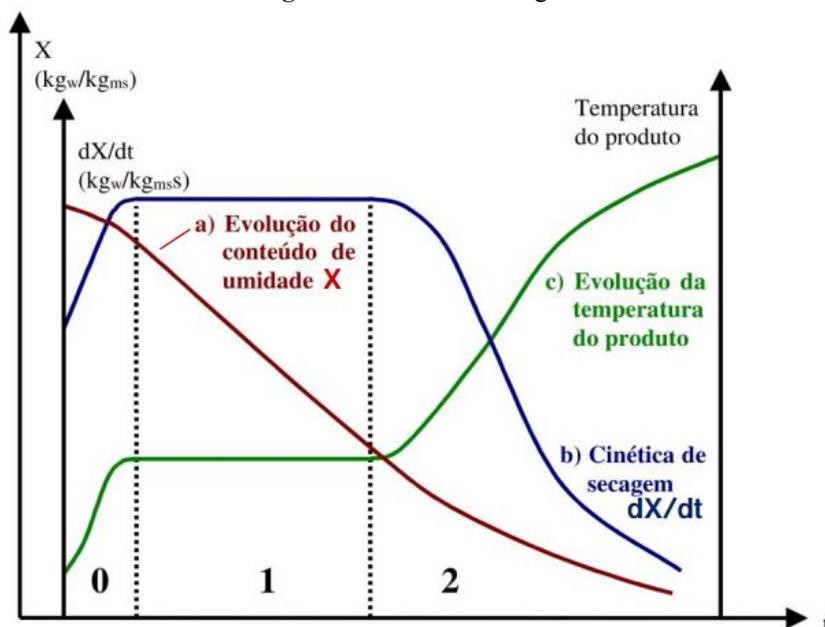
Como dito anteriormente, a empresa utilizou a secagem intermitente nos casos em que a umidade inicial foi maior que 13%, correspondendo a 66% do total de casos. Segundo PESKE, ROSENTHAL & ROTA (2012), a secagem intermitente é capaz de retirar no máximo 1,5 p.p (pontos percentuais) de umidade por hora na semente de soja, sem causar danificações às mesmas. Portanto, a utilização desse método de secagem é pertinente, pois proporciona melhor aproveitamento do processo, ou seja, uma retirada de teor de água maior

(no caso da empresa 1,2 p.p), em razão do processo de deslocamento da água do interior para a superfície da semente no tempo de equalização.

Para as sementes com umidade inferior a 13% e temperatura da massa inferior a 34°C (correspondente a 31% do total de casos), foi utilizada a secagem contínua. Segundo PESKE, ROSENTHAL & ROTA (2012) a secagem contínua é aplicável para umidades baixas, pois oferece uma maior exposição ao ar de secagem, retirando o teor de água com maior facilidade devido a exposição ao ar de secagem que força a migração mais rápida do fluxo de água para a sua superfície, sem que haja um aumento significativo da temperatura do ar artificial, reduzindo riscos térmicos.

Segundo PARK et al. (2007), os dois parâmetros analisados para a secagem neste projeto, umidade inicial e temperatura da massa da semente, estão relacionados entre si como mostra a Figura 14. Nela, percebe-se que a temperatura da massa do produto (linha verde) tende a aumentar, em contrapartida, a umidade do produto (linha vermelha) tende a diminuir ao longo do tempo da secagem. A curva azul representa a energia gasta durante o processo.

Figura 14: Curva de secagem.

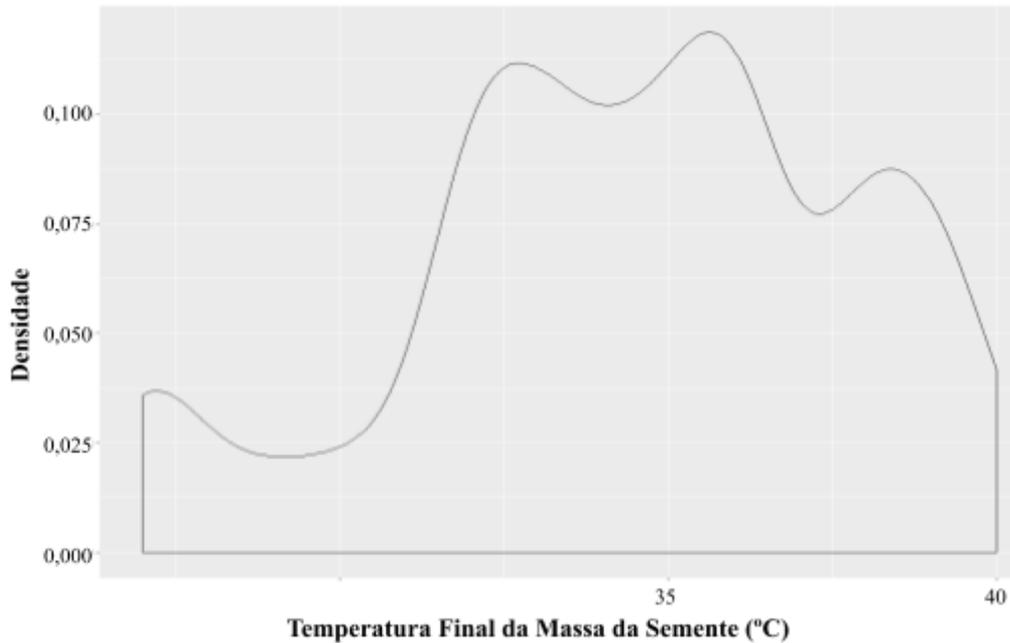


Fonte: PARK et al. (2007).

Para analisar o comportamento da temperatura da massa da semente, não foram encontradas referências na literatura com base nos 34°C utilizados para a escolha do tipo de secagem. Porém, acredita-se que esse *setpoint* foi utilizado como forma de prevenir os riscos a altas temperaturas no produto em detrimento do fluxo contínuo de ar de secagem, ou seja,

quando o mesmo for atingido utiliza-se secagem intermitente devido ao tempo em que o produto não fica submetido ao ar de secagem reduzindo este risco. De acordo com MENEGUELLO (2014) a temperatura da massa da semente não pode ultrapassar 43°C, pois compromete a qualidade das sementes. Com isso, para verificar os dados da empresa, plotou-se o gráfico (Figura 15) e verificou-se que nenhuma amostra teve sua temperatura acima de 40°C.

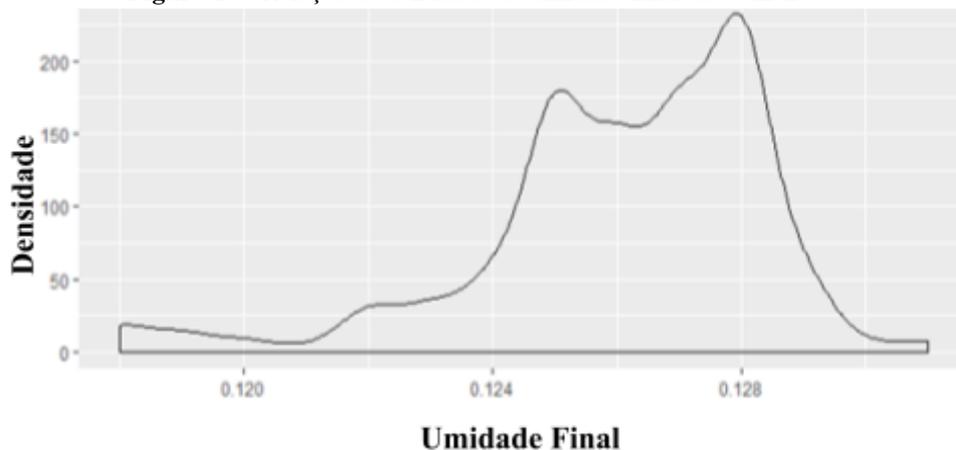
Figura 15: Temperatura da massa final das sementes.



Fonte: Autoria própria.

Por fim, analisando as umidades finais, a Reina obteve as umidades variando de 11 a 13%, resultado mostrado na Figura 16 que, de acordo com FONTES e MANTOVANI [s.d.], é a variação ideal para prevenção de fungos para o armazenamento da mesma.

Figura 16: Relação da densidade de umidade final das sementes.



Fonte: Autoria própria.

6 CONCLUSÃO

A modelagem realizada foi importante para verificar e analisar os parâmetros da tomada de decisão da empresa, e comparar com a literatura que geralmente relaciona a realização da secagem intermitente para umidades de semente mais elevadas e a secagem contínua para umidades mais baixas. Sendo assim, a REINA está realizando sua tomada de decisões para a forma de secagem de maneira pertinente, dentro dos parâmetros analisados neste projeto.

Seria interessante captar o tempo de equalização do secador durante a secagem intermitente para comparar os tempos gastos nos dois métodos de secagem. Além disso, a realização da automação dos secadores, por conta da melhoria e confiabilidade dos dados e do processo, que hoje são feitos manualmente, diminuiria assim os erros. Por conta disso, para a safra de 2020, a Reina vai adquirir um controle de temperatura do secador, controlador de chama, chave de ignição, ionizador e total integração com CLP (controlador lógico programável).

Sendo assim, os próximos passos a serem dados a este estudo de caso, seria uma cartilha de recomendação para um modelo de redução do tempo de secagem visando a otimização dos custos neste processo.

REFERÊNCIAS BIBLIOGRÁFICAS

AGÊNCIA BRASIL. Brasil deverá ter produção recorde de grãos na safra 2019/2020. Brasília, 2019. Disponível em: <<http://agenciabrasil.ebc.com.br/economia/noticia/2019-10/primeiro-levantamento-de-safra-20192020-indica-producao-recorde>>. Acesso em: 20 Nov. 2019.

AVELAR, S. A. G.; PESKE, S. T.; VILLELA, F. A. Avanços na secagem de sementes. Seednews, 16^aed, jul. 2012. Disponível em: <<https://seednews.com.br/edicoes/artigo/908-avancos-na-secagem-de-sementes-edicao-julho-2012>>. Acesso em: 23 Nov. 2019.

BAUDET, L.; VILLELA, F. A. Princípios de secagem. Seednews, 3^aed, 1999. Disponível: <<https://seednews.com.br/edicoes/artigo/2337-principios-de-secagem-edicao-marco-1999>>. Acesso em: 23 Nov. 2019.

CHAVES, S. Como a mineração de dados pode auxiliar o agronegócio. [s.d.]. Disponível em: <<https://excelenciaempauta.com.br/mineracao-de-dados-auxilia-o-agronegocio/>>. Acesso em: 23 Nov. 2019.

CONAB. Acompanhamento da Safra Brasileira de Grãos. , v. 7 Safra 2019/20 - Segundo levantamento, Brasília, p. 1-25. 2019. ISSN 2318-6852. Disponível em: <<https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>>. Acesso em: 22 Nov. 2019.

EMBRAPA. Semente de soja de qualidade é primeiro passo para sucesso da safra. 2015. Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/3394051/semente-de-soja-de-qualidade-e-primeiro-passo-para-sucesso-da-safra>>. Acesso em: 22 Nov. 2019.

EMBRAPA. Soja em números (safra 2018/2019). 2019. Disponível em: <<https://www.embrapa.br/soja/cultivos/soja1/dados-economicos>>. Acessado em 08 de Dezembro de 2019.

FLOR, E. P. O. Avaliação de danos mecânicos em sementes de soja por meio da análise de imagens. Tese. ESALQ, 2003. Disponível em: <<https://teses.usp.br/teses/disponiveis/11/11136/tde-26112003-161958/publico/ebert.pdf>>. Acesso em: 23 nov. 2019.

FONTES, R. A.; MANTOVANI, B. H. M. Armazenamento das sementes. EMBRAPA, s.d. Disponível em:

<<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/57363/1/Circ-19-Armazenamento-sem-entes.pdf>>. Acesso em: 23 nov. 2019.

FRANÇA NETO, J.B.; KRZYZANOWSKI, F.C.; COSTA, N.P. da. O teste de tetrazélio em sementes de soja. Londrina : EMBRAPA-CNPSO, 1998. 72p. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/162988/1/O-TESTE-DE-TETRAZOL-IO-EM-SEMENTES-DE-SOJA.pdf>>. Acesso em: 23 nov. 2019.

FRANÇA-NETO, J. B.; KRZYZANOWSKI, F. C.; HENNING, A. A. A importância do uso de semente de soja de alta qualidade. EMBRAPA, 2010. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/bitstream/doc/661047/1/ID30537.pdf>>. Acesso em: 22 nov. 2019.

GARCIA, D. C. et al. A secagem de sementes. Ciência Rural, Santa Maria, v.34, n.2, p.603-608, mar-abr, 2004. Disponível em: <<http://www.scielo.br/pdf/cr/v34n2/a45v34n2.pdf>>. Acesso em: 18 nov. 2019.

GRANATYR, J. 3 Linguagens para Inteligência Artificial. 2017. Disponível em: <<https://iaexpert.com.br/index.php/2017/04/05/3-linguagens-para-inteligencia-artificial/>>. Acesso em: 23 Nov. 2019.

KRZYZANOWSKI, F. C.; FRANÇA-NETO, J. B.; HENNING, A. A. A alta qualidade da semente de soja: fator importante para a produção da cultura. Circular Técnica 136. Londrina - PR, 2018. ISSN 2176-2864.

KRZYZANOWSKI, F. C. Teste do hipoclorito de sódio para semente de soja. Circular Técnica 37. Londrina -PR, 2004. ISSN 1516-7860

MANDARINO, J. M. G. Origem e história da soja no Brasil. Embrapa Soja, 2017. Disponível em: <<https://blogs.canalrural.uol.com.br/embrapasoja/2017/04/05/origem-e-historia-da-soja-no-brasil/>>. Acessado em 08 de Dezembro de 2019.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento. Sementes de alta qualidade aumentam produção de soja entre 10% e 15%. 2019. Disponível em: <<http://www.agricultura.gov.br/noticias/sementes-de-alta-qualidade-aumentam-producao-de-soja-entre-10-e-15>>. Acesso em: 22 Nov. 2019.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento. Regras para análise de sementes. 1ªed. 2009. Disponível em: <http://www.agricultura.gov.br/assuntos/insumos-agropecuarios/arquivos-publicacoes-insumos/2946_regras_analise_sementes.pdf>. Acesso em: 23 Nov. 2019.

MATOS, P. C. T. Sementes comestíveis. ESALQ, 2013. Disponível em: <http://www.esalq.usp.br/acom/clipping/arquivos/21-02-2013_Sementes_comestiveis_Agrolink_AL.pdf>. Acesso em: 20 Nov. 2019.

MENEGHELLO, G. E. Qualidade de sementes: umidade e temperatura. Seednews, 18ªed. 2014. Disponível em: <<https://seednews.com.br/edicoes/artigo/258-qualidade-de-sementes:-umidade-e-temperatura-edicao-novembro-2014>>. Acesso em: 23 Nov. 2019.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. Aprendizado supervisionado: Conceitos e definições. Cap.4. 1999. Disponível em: <<http://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>>. Acesso em: 23 Nov. 2019.

NUNES, J. L. S. Tecnologia de sementes - Secagem, Beneficiamento e Armazenagem. 2016. Disponível em: <https://www.agrolink.com.br/sementes/tecnologia-sementes/secagem--beneficiamento-e-armazenagem_361343.html>. Acesso em: 07 out. 2019.

PARK, K. J.; ANTONIO, G. C.; OLIVEIRA, R. A.; PARK, K. J. B. Conceitos de processo e equipamentos de secagem. Campinas, 2007. Acesso em: 18 dez. 2019.

PERES, W. B.; PESKE, S. T. Secagem de sementes. Seednews, 2ªed, 2016. Disponível em: <<https://seednews.com.br/edicoes/artigo/816-secagem-de-sementes:-edicao-novembro-2016>>. Acesso em: 24 out. 2019.

PESKE, S. Qualidade de semente de soja. Agronegócio em foco, Jun. 2015. Disponível em: <<http://www.pioneersementes.com.br/blog/44/qualidade-de-semente-de-soja>>. Acesso em: 24 out. 2019.

PESKE, S.T.; ROSENTHAL, M.D.; ROTA, G.R.M. Sementes: Fundamentos científicos e tecnológicos. 3ª edição. Pelotas: Editora rua Pelotas, 2012. 573p.

PRATES, W. R. O que é árvore de decisão (decision tree)? Exemplos em R. 2018. Disponível em: <<https://www.wrprates.com/o-que-e-arvore-de-decisao-decision-tree-linguagem-r/>>. Acesso em: 18 nov. 2019.

RADKE, A. K.; BABATA, M. M.; VILLELA, F. A. Secagem intermitente e à qualidade de sementes de soja associada à operação da estrutura. Tecnologia & Ciência Agropecuária, João pessoa, v.11, n.4, p.0-1, out. 2017. Disponível em:

<<http://revistatca.pb.gov.br/edicoes/volume-11-2017/v-11-n-4-novembro-2017/12-secagem-intermitente-e-a-qualidade-de-sementes.pdf>>. Acesso em: 24 out. 2019.

ROTASILOS do Brasil. 2018. Disponível em: <<http://www.rotasilos.com.br/equipamentos.php?equipamentos=18>>. Acesso em: 07 out. 2019.

SANTOS, R. Introdução à Mineração de Dados com Aplicações em Ciências Ambientais e Espaciais. LAC, [s.d]. Disponível em: <<http://www.lac.inpe.br/~rafael.santos/Docs/ELAC/2012/ELAC-DM.pdf>>. Acesso em: 22 Nov. 2019.

TRINDADE, C. Comparativo: R vs. Python para data science. 2017. Disponível em: <<https://imasters.com.br/back-end/comparativo-r-vs-python-para-data-science>>. Acesso em: 23 nov. 2019.

TRINDADE, M. S. Secagem de Soja em camada espessa: modelagem matemática e simulação numérica. Dissertação. UNIJUÍ, 2013. Disponível em: <<https://docs.google.com/document/d/1rF5iFYyuBkxsV5EqaZtBamT0d6E1aXG-8hek4Z75-Bc/edit>>. Acesso em: 23 nov. 2019.

APÊNDICE

```
#Fernando dArmada Cunha 160150
```

```
require(data.table)
require(dplyr)
require(lubridate)
require(janitor)
require(ggplot2)
require(naniar)
require(stringr)
require(caret)
require(caTools)
require(randomForest)
require(rpart)
require(rpart.plot)
```

```
#Setando o diretorio
setwd("C:/Users/Fernando/Desktop")
```

```
#Lendo os dados
dados = fread('DADOS_REINA.csv')
```

```
#Manipulando os nomes das variaveis
dados = dados %>% clean_names()
names(dados)
```

```
#Apagando colunas
dados$verificacao = NULL
dados$data_do_fim_de_enchimento = NULL
dados$data_do_fim_de_esvaziamento = NULL
dados$horario_fim_de_enchimento = NULL
```

```

dados$horario_fim_de_esvaziamento = NULL

#Mudando os valores da coluna tipo_de_secagem
dados$tipo_de_secagem[dados$tipo_de_secagem == "Intermitente"] = 1
dados$tipo_de_secagem[dados$tipo_de_secagem == "Continuo"] = 0

#verificacao de dados faltantes
vis_miss(dados)

#Verificando o tipo dos dados e alterando o que se julga necessario
lapply(dados, class)
dados$cultivar = as.character(dados$cultivar)
dados$n°_do_secador = as.character(dados$n°_do_secador)
dados$tipo_de_secagem = as.numeric(dados$tipo_de_secagem)

#Analisando por tabela os dados e correlações entre as variaveis
prop.table(table(dados$cultivar))
prop.table(table(dados$n°_do_secador))
prop.table(table(dados$dano_mecanico))
prop.table(table(dados$tipo_de_secagem))
prop.table(table(x = dados$n°_do_secador, y = dados$cultivar))
prop.table(table(x = dados$tipo_de_secagem, y = dados$cultivar))
prop.table(table(x = dados$tipo_de_secagem, y = dados$n°_do_secador))
prop.table(table(x = dados$tempo_de_secagem, y = dados$tipo_de_secagem))
prop.table(table(x = dados$umidade_inicial, y = dados$tipo_de_secagem))
matriz = cor(select_if(dados, is.numeric))

#Correlacao dos dados por grafico de density e point
ggplot(dados, aes(x = tempo_de_secagem)) + geom_density()
ggplot(dados, aes(x = n°_do_secador)) + geom_density()
ggplot(dados, aes(x = umidade_da_semente_final)) + geom_density()

```

```

ggplot(dados, aes(x = tempo_de_secagem)) + geom_density() +
facet_grid(~tipo_de_secagem)
temp = dados[, c("tempo_de_secagem", "umidade_da_semente_final", "umidade_inicial")]
temp$dif_umidade = temp$umidade_da_semente_final - temp$umidade_inicial
ggplot(temp, aes(x=tempo_de_secagem, y = dif_umidade)) + geom_point()
temp <- temp[temp$dif_umidade > -0.15]

ggplot(dados, aes(x=tempo_de_secagem, y = umidade_inicial)) + geom_point() +
facet_grid(~tipo_de_secagem)
dados <- dados[dados$umidade_inicial < 0.30]
dados <- dados[dados$tempo_de_secagem < 40]
ggplot(dados, aes(x= tempo_de_secagem, y = temperatura_da_massa_da_semente)) +
geom_point()
dados <- dados[dados$temperatura_da_massa_da_semente < 45]
ggplot(dados, aes(x=umidade_ambiente, y = umidade_inicial)) + geom_point()
dados <- dados[dados$umidade_ambiente < 4]
dados <- dados[dados$umidade_ambiente > 0.3]
ggplot(dados, aes(x=umidade_ambiente, y = temperatura_ambiente)) + geom_point()
dados <- dados[dados$temperatura_ambiente < 50]
ggplot(dados, aes(x=umidade_ambiente, y = temperatura_do_ar_do_secador )) +
geom_point()
dados <- dados[dados$temperatura_do_ar_do_secador > 28]
ggplot(dados, aes(x= tempo_de_secagem, y = umidade_inicial )) + geom_point()
ggplot(dados, aes(x =tipo_de_secagem, y = umidade_inicial)) + geom_boxplot()
ggplot(dados, aes(x =tipo_de_secagem, y = umidade_inicial)) + geom_bin2d()
ggplot(dados, aes(x =tipo_de_secagem, y = umidade_inicial)) + geom_raster()
ggplot(dados, aes(x =tipo_de_secagem, y = umidade_inicial)) + geom_raster()

#Separando o conjunto de dados em treino e teste
set.seed(160150)
indice_treino = sample.split(dados$tipo_de_secagem, SplitRatio = 0.8)
base_treinamento = subset(dados, indice_treino == TRUE)

```

```
base_teste = subset(dados, indice_treino == FALSE)
is.data.frame(base_treinamento)
head(base_treinamento)
#Modelagem: arvore de decisao para indicar a maneira secada
set.seed(160150)
classificador_tree= rpart(tipo_de_secagem ~.,
                          base_treinamento,
                          control = rpart.control(
                            minbucket = 15,
                            maxdepth = 5,
                            cp = 0.005),
                          method = 'class')
classificador_tree
previsoes_tree = predict(classificador_tree, base_teste,
                        type = 'class')
acuracia_tree = mean(previsoes_tree == base_teste$tipo_de_secagem)
acuracia_tree
table(previsoes_tree, base_teste$tipo_de_secagem)
rpart.plot(classificador_tree)
```