



UNIVERSIDADE ESTADUAL DO NORTE DO PARANÁ
CAMPUS LUIZ MENEGHEL - CENTRO DE CIÊNCIAS TECNOLÓGICAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

MARCUS VINÍCIUS DORTA

**ESTIMANDO A DEPOSIÇÃO DE PRODUTOS
FITOSSANITÁRIOS FORA DA ÁREA ALVO: UMA
ABORDAGEM DE APRENDIZADO DE MÁQUINA**

BANDEIRANTES-PR

2017

MARCUS VINÍCIUS DORTA

**ESTIMANDO A DEPOSIÇÃO DE PRODUTOS
FITOSSANITÁRIOS FORA DA ÁREA ALVO: UMA
ABORDAGEM DE APRENDIZADO DE MÁQUINA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual do Norte do Paraná para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Bruno Squizzato Faiçal

BANDEIRANTES-PR

2017

DORTA, M. V.. **Estimando a deposição de produtos fitossanitários fora da área alvo: uma abordagem de aprendizado de máquina.** 51 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual do Norte do Paraná, Bandeirantes-PR, 2017.

RESUMO

O setor agrário possui diversas aplicações, entre elas se encontra a pulverização. A pulverização consiste em aplicar produtos fitossanitários nas lavouras, para controle de pragas e aumento de produção. Contudo inúmeras vezes a aplicação de produtos fitossanitários acontece de maneira errônea, tendo a deposição de produtos em locais indesejados, isso pode ocorrer por deriva. A deriva é a dispersão dos fluidos dos produtos causados pela corrente de ar, e também um dos maiores problemas encontrados na pulverização. Problema também que resulta em danos socioambientais, econômicos e aumento no custo de produção. Contudo existe uma área de Inteligência Artificial que visa o trabalho automáticos de máquinas, aprendizado de máquina, que consiste em fazer uma máquina aprender através de experiências passadas. Com dados ocorrentes de deriva durante uma pulverização, aplicados a técnicas de aprendizado de máquina, constrói-se a hipótese de que seja possível estimar uma curva de deposição de produtos, e apresentar essa estimativa ao usuário, podendo ele corrigir ou diminuir a deriva decorrente dos dados fornecidos durante a pulverização. O objetivo desse trabalho é estudar técnicas de aprendizado de máquina avaliando cada técnica em termos de acurácia. Com os estudos realizados sobre as técnicas abordadas, foi possível identificar que a técnica de redes neurais apresentou um melhor comportamento aos dados trabalhos, que obteve uma melhor função de regressão.

Palavras-chave: Aprendizado de Máquina, Estimativa de Deriva, Agricultura de Precisão.

DORTA, M. V.. **Title of the Work.** 51 p. Final Project (Bachelor of Science in Computer Science) – State University Northern of Parana , Bandeirantes–PR, 2017.

ABSTRACT

The agrarian sector has several applications, among them is spraying. The spraying consists in applying phytosanitary products in the crops, for pest control and increased production. However many times the application of plant protection products happens in an erroneous way, having the deposition of products in unwanted places, this can occur by drift. The drift is the dispersion of the fluids from the products caused by the air stream, and also one of the major problems encountered in spraying. This problem also results in socio-environmental, economic damages and an increase in the cost of production. However there is an area of Artificial Intelligence that aims at the automatic work of machines, machine learning, which is to make a machine learn through past experiences. With data on drift during spraying, applied to machine learning techniques, the hypothesis is constructed that it is possible to estimate a curve of product deposition, and present this estimate to the user, which can correct or reduce the drift provided during spraying. The objective of this work is to study machine learning techniques by evaluating each technique in terms of accuracy. With the studies carried out on the techniques, it was possible to identify that the neural network technique presented a better behavior to the data, which obtained a better regression function

Keywords: Machine Learning, Drift Spray, Precision Agriculture.

LISTA DE ILUSTRAÇÕES

Figura 1 – Demonstração da hierarquia de aprendizado	22
Figura 2 – Demonstração árvore de regressão	23
Figura 3 – Execução de uma árvore de regressão	23
Figura 4 – Exemplificação de uma folha de árvore de regressão	23
Figura 5 – Exemplo de funções de ativação	25
Figura 6 – Rede neural artificial	26
Figura 7 – Simulação de um painel utilizado em equipamento que implementa conceitos de AP	28
Figura 8 – Fase de seleção do modelo de algoritmo que irá compor o núcleo de processamento	32
Figura 9 – Exemplo de pré-processamento dos dados	33
Figura 10 – Elementos que compõe o sistema proposto em execução	33
Figura 11 – Esquema de coleta das amostras de deposição fora da área alvo.	37
Figura 12 – Exemplo da divisão em grupos da base dados	38
Figura 13 – Deposição incorreta de produtos em relação a distância	41
Figura 14 – Previsão da deposição incorreta de produtos em relação a distância gerados pela técnica de árvore de regressão	42
Figura 15 – Semelhança no comportamento dos dados no algoritmo de árvore de regressão.	42
Figura 16 – Previsão da deposição incorreta de produtos em relação a distância gerados pela técnica de redes neurais	43
Figura 17 – Semelhança dos valores preditos aos valores reais pela técnica de redes neurais	44

LISTA DE TABELAS

Tabela 1 – Previsões climáticas de chuva	20
Tabela 2 – Parâmetros utilizados na árvore de regressão	35
Tabela 3 – Parâmetros utilizados na rede neural	35
Tabela 4 – Tabela de erros	44
Tabela 5 – Resultados do p-valor obtidos pelo teste de Shapiro Wilk	45
Tabela 6 – Resultado do teste de Wilcoxon	45

LISTA DE ABREVIATURAS E SIGLAS

AP	Agricultura de precisão
AM	Aprendizado de máquina
IA	Inteligência artificial
PFT	Produtos fitossanitários
GPS	Sistema de posicionamento global
RNA	Rede neural artificial

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Contextualização e Motivação	15
1.2	Justificativa	17
1.3	Objetivos	18
1.3.1	Objetivo Geral	18
1.3.2	Objetivos específicos	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Aprendizado de Máquina	19
2.1.1	Árvores de Regressão	22
2.1.2	Redes Neurais Artificiais	24
2.2	Agricultura de Precisão	27
3	TRABALHOS RELACIONADOS	29
4	PROPOSTA	31
5	METODOLOGIA	35
5.1	Algoritmos Avaliados	35
5.2	Base de dados	36
5.3	Experimentos	37
6	RESULTADOS	41
7	DISCUSSÃO E TRABALHOS FUTUROS	47
	REFERÊNCIAS	49

1 INTRODUÇÃO

1.1 Contextualização e Motivação

Mudanças que revolucionarão a agricultura irão acontecer ao decorrer dos tempos, com o avanço rápido da tecnologia as práticas agrícolas se adequarão, dessa forma o tamanho das fazendas, a produção e o preço dos produtos agrícolas aumentarão [DALY, 1967]. No século passado, Daly [1967] previa que o setor agrícola teria um grande avanço, alterando o modo de produção agrícola, essa mudança no presente está ligada a tecnologia, denominada como Agricultura de Precisão (AP).

A AP possui o conceito de otimização agrícola, por meio da variabilidade do solo e clima implanta-se o processo de automação na lavoura através de informações relativas à área produtiva visando o controle de insumos de maneira correta. Além do cuidado existente sobre os insumos a serem aplicados, existe o cuidado sobre os recursos a serem utilizados para que a prática de AP seja efetuada da melhor maneira possível. Corá et al. [2004] afirma que a preocupação relacionada aos recursos naturais deve ser levada em conta, a produção agrícola precisa ser otimizada, visando à sustentabilidade e à segurança alimentar.

Segundo Mantovani, Queiroz e Dias [1998], essa preocupação passa pelas técnicas de agricultura de precisão que devem possuir formas de manejo sustentável, que devam permitir mudanças sem danos as reservas naturais, ao mesmo tempo em que os danos ao meio ambiente são minimizados.

A produção agrícola deve buscar o maior rendimento dos cultivos da maneira mais adequada possível, no qual a produção das culturas com menor consumo de insumos e mão de obra são as variáveis chaves para o melhor aproveitamento da lavoura. Nesse sentido Molin, Amaral e Colaço [2015] afirmam que a agricultura de precisão se mostra uma prática essencial, que ganha cada vez mais adeptos.

A agricultura de precisão envolve um conjunto de estratégias de gestão agrícola na qual se utiliza tecnologia de informação em tempo real – tecnologia de geoprocessamento, sistema de posicionamento global (GPS) entre muitos outros – baseada no princípio de atender as diferentes necessidades de cada micro-região do sistema produtivo e as suas variações naturais do solo na questão nutricional.

Quando se trata de variação natural, diferentes necessidades da lavoura, o objetivo é aperfeiçoar a aplicação de insumos potencializando a produção. Em Tschiedel e Ferreira [2002] conforme citado por Manzatto, Bhering e Simões [1999] a aplicação de insumos é um dos princípios básicos da agricultura de precisão, onde a gestão de insumos trata de

aplicar as coisas certas, no lugar certo, na quantidade correta e no momento exato.

Uma das linhas em aplicação de insumos na AP é a pulverização das lavouras, em que [Voll \[2000\]](#) aponta que nos sistemas de pulverização a precisão no momento da aplicação é primordial, tanto quanto no rendimento do insumo (custo/benefício), quanto para a preocupação do impacto no meio ambiente.

A prática da agricultura de precisão pode trazer diversos benefícios, englobando aspectos econômicos e ambientais. Do ponto de vista econômico, a aplicação localizada permite a priorização do investimento em insumos (fertilizantes, sementes, etc) nas áreas onde o potencial de produção seja mais efetivo, garantindo a possibilidade de maior retorno econômico com menores investimentos. O mesmo pode ser verificado no caso dos tratamentos fitossanitários, pois a aplicação localizada pode significar reduções de mais de 60% nas quantidades de defensivos utilizados. Considerando-se os aspectos ambientais, a racionalização e redução do uso de fertilizantes e defensivos deve ser avaliada como um dos principais benefícios da utilização da agricultura de precisão [[ANTUNIASSI; JÚNIOR, 2000](#)].

Sistemas de pulverização devem ser precisos, para que a deposição de produtos fitossanitários (PFT) ocorra de maneira correta, minimizando a perda de produtos e evitando que os mesmos não atinjam áreas não desejadas. Conforme [Cunha \[2008\]](#) na maioria das vezes as técnicas de pulverização não são levadas em conta, tendo pouca importância, mas é um fator fundamental para garantir que o produto alcance o alvo de forma eficiente, minimizando-se as perdas. Tendo esse conceito [Fernandes et al. \(2007 apud MATUO, 1990\)](#) completa com a seguinte afirmação: "tecnologia de aplicação de insumos agrícolas ou PFT's é de suma importância, ela é responsável pela correta colocação dos mesmos no alvo, cuidando da preservação do ambiente e da saúde do trabalhador".

Como em todas as vertentes relacionadas a AP, a pulverização de PFT possui um enorme problema chamado deriva, que [Hofman e Solseng \[2009\]](#) definem como o movimento de um produto fitossanitário através do ar, durante ou após a aplicação para um local diferente do alvo pretendido - gotículas em movimento através do vento, que pode ocorrer devido às características do produto sendo pulverizado, dos equipamentos e técnicas de aplicação e das condições climáticas. Com isso, a deriva ocasiona a deposição fora da área alvo. A deposição em local indesejada de PFT pode provocar danos econômicos como a perda do produto sendo aplicado, impactos no meio ambiente, danos a culturas adjacentes e danos à saúde humana [[GIL; SINFORT, 2005](#)]. Para ([CUNHA et al., 2003 apud SUMNER; SUMNER, 1999](#)) a deriva um dos maiores problemas existentes na agricultura.

Tendo em vista os problemas que a deposição fora a área alvo que ocorre por deriva podem ocasionar, e conhecendo as possibilidades que a AP proporciona, existem técnicas computacionais que unidas às práticas de AP possibilitam o aperfeiçoamento do sistema de pulverização. Técnicas de aprendizado de máquina (AM) aplicadas em sistemas de

pulverização que implementam AP podem possibilitar a gestão em tempo real de sistemas de pulverização podendo assim aperfeiçoar a aplicação de produtos fitossanitários.

A junção entre técnicas de AM implantadas as práticas de AP podem cooperar com a formação de novas práticas na agricultura. [McQueen et al. \[1995\]](#) apontam que uma vez que os dados forem manipulados corretamente resolvendo uma tarefa, isso indica que o aprendizado automático pode desempenhar um papel útil na resolução de problemas agrícolas em grande escala.

O setor agrícola tem avançado intensamente em recursos tecnológicos nas últimas décadas, quando maquinários agrícolas implementam conceitos de AP em seus sistemas. Implantar técnicas de AM em tais ferramentas podem proporcionar uma melhoria na gestão da lavoura. Aprendizado de máquina é um subcampo da inteligência artificial dedicado ao desenvolvimento de algoritmos, técnicas, que permitam ao computador aprender e aperfeiçoar seu desempenho. [Monard e Baranauskas \[2005\]](#) referem-se ao aprendizado de máquina como “um programa de computador que toma decisões baseadas em experiências acumuladas através da solução bem sucedida de problemas anteriores”.

Em AM o conceito fundamental é de um computador, máquina, adquirir experiência através de situações já apresentadas, isto é, prever novas situações através de experiências já vivenciadas. Sabendo disto, aplicar algoritmos de AM em sistemas de pulverização pode proporcionar um aperfeiçoamento do mesmo, onde uma técnica de AM implantado à um sistema de pulverização pode prever novas situações durante a aplicação de defensivos agrícolas.

Com as causas que a deposição fora da área alvo podem ocasionar, e tendo conhecimento das possibilidades que a computação oferece, o objetivo deste trabalho é estudar os algoritmos/técnicas de aprendizado de máquina já existentes, analisá-los, e apresentar qual técnica se adequou melhor aos dados de deposição fora a área alvo.

1.2 Justificativa

De acordo com o [PORTAL BRASIL \[2016\]](#), o setor agrícola cresce ao longo dos anos, tendo o agronegócio uma estimativa de crescimento em 2% no ano de 2017. [Teixeira \[2005\]](#), o crescimento do setor agrícola e a demanda por tecnologia também se multiplica.

[Ferreira \[2015\]](#), afirma que o cenário brasileiro em respeito à pulverização de produtos fitossanitários, tem causado risco ao meio ambiente e a saúde humana, causando contaminação aos recursos ambientais e ameaça a qualidade de vida das futuras gerações, que podem ocorrer devido às técnicas utilizadas na pulverização, com a inserção de novas tecnologias, cuja função é estimar a deriva do produto, esse risco pode ser minimizando.

[PORTAL BRASIL \[2017\]](#) menciona que cerca de 67% das propriedades agrícolas do

Brasil utilizam tecnologia, tanto na gestão quanto na produção. Com o aperfeiçoamento de um sistema de pulverização pela sistematização de agricultura de precisão, o sistema seria capaz de diminuir danos ambientais, econômicos, custo de produção e melhorar a qualidade de deposição dos produtos.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo deste trabalho é investigar algoritmos/técnicas de aprendizado de máquina capazes de estimar a curva de deposição incorreta resultante pela pulverização do produto sob determinadas condições, analisando a acurácia de cada técnica.

1.3.2 Objetivos específicos

Para atingir o objetivo geral deste trabalho, os seguintes objetivos específicos deverão ser atingidos:

- Pesquisar os algoritmos para estimar o volume depositado fora da área alvo;
- Realizar experimentos e análises nos algoritmos;
- Realizar análises sobre os resultados das previsões obtidas de cada algoritmo; e
- Contribuir com a geração de conhecimento para a agricultura de precisão.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Aprendizado de Máquina

Inteligência artificial (IA) é um ramo de pesquisa da ciência da computação, que através de códigos, símbolos computacionais, é possível construir mecanismos ou dispositivos que realizam tarefas ou problemas utilizando o conhecimento, isto é, uma máquina que pode fazer aquilo que era de exclusividade do gênero humano, agir racionalmente e executar tarefas que sejam necessárias a capacidade de pensar [TEIXEIRA, 2014].

Com IA é possível se criar programas que imitem a capacidade de raciocinar, de distinguir objetos e ações semelhantes a do ser humano, que só foi possível com a chegada dos “computadores modernos”, após a Segunda Guerra Mundial, e mais tardar na década de 1970 com a utilização dos conceitos de IA que foram empregados em soluções de problemas reais, se tornando Sistemas Especialistas ou Sistemas Baseados em Conhecimento.

Atualmente os problemas a serem resolvidos computacionalmente se tornaram mais complexos, devido ao grande volume de dados gerados por diferentes setores, requisitando ferramentas computacionais mais evoluídas, independentes, capazes de tomar decisões por si próprias, através de experiências passadas, resolver o problema de maneira autônoma [FACELI et al., 2015].

Por exemplo, em um site de compras online (e-commerce) onde o site sugere para o cliente produtos que “talvez” ele possa desejar, assimilando compras passadas e/ou pesquisas de outros clientes de produtos semelhantes, quem pesquisa o produto x também pesquisa o produto y. Esse processo de inferência através de conhecimento (experiências passadas) denomina-se Aprendizado de Máquina.

Aprendizado de máquina é um subcampo de IA cujo propósito é o desenvolvimento de ferramentas computacionais capazes de adquirir conhecimento de forma automática, que toma decisões de experiências de problemas anteriores aprendendo novas regras ou inferências[MICHALSKI; CARBONELL; MITCHELL, 2013].

Em AM, existe um princípio de inferência denominado indução, que permite obter conclusões genéricas a partir de um conjunto particular de exemplos. Na indução, um conceito é aprendido produzindo inferência indutiva sobre os exemplos apresentados, gerando assim algoritmos de AM que aprendem a induzir uma função ou hipótese capaz de resolver um problema a partir de dados que representam instancias dos problema a ser resolvido [MITCHELL, 1997]. Portanto as hipóteses geradas através da inferência podem ou não ser verdadeiras. Contudo a indução deve ser utilizada ponderadamente, pois se o

número de exemplos for insuficiente, ou caso os exemplos não forem bem selecionados, as hipóteses obtidas podem ser de pouco valor.

Existem diversas aplicações bem-sucedidas de técnicas de AM na solução de problemas reais, são eles:

- Reconhecimento de palavras faladas;
- Predição de taxas de cura de pacientes com diferentes doenças;
- Detecção do uso fraudulento de cartões de crédito;
- Condução de automóveis de forma autônoma em rodovias;
- Ferramentas que jogam gamão e xadrez de forma semelhante a campeões; e
- Diagnóstico de câncer por meio da análise de dados de expressão gênica.

A expansão das diversas aplicações existentes só pôde ser desenvolvida, devido ao desenvolvimento de algoritmos cada vez mais efetivos, e aos recursos computacionais existentes.

Para maior esclarecimento entre a relação de AM e indução de hipóteses, observe um conjunto de dados, tabela 1, denominado Clima que ilustra um exemplo, fictício, de estados climáticos para determinar se irá chover ou não.

Tabela 1 – Previsões climáticas de chuva

O clima estará ensolarado ?	O clima estará nublado ?	O clima estará úmido ?	Irá chover ?
SIM	NÃO	NÃO	NÃO
NÃO	SIM	NÃO	TALVEZ
NÃO	SIM	SIM	SIM

O conjunto de dados da Tabela 1 correspondem a uma relação de possível chuva dependente dos estados climáticos que possui os atributos (O clima está ensolarado?, O clima está nublado?, O clima está úmido?, Irá chover?).

Em tarefas de aprendizado, um dos atributos é considerado um atributo alvo (Irá chover?), cujo os valores podem ser estimados utilizando os valores dos demais atributos, denominados atributos de entrada (O clima está ensolarado?, O clima está nublado?, O clima está úmido?). Quando um algoritmo de AM está aprendendo ele utiliza um subconjunto de dados chamado treinamento, o qual através deste subconjunto ele constrói um modelo ou hipótese capaz de relacionar os valores dos atributos de entrada de um objeto do conjunto de treinamento ao valor de seu atributo alvo, ou seja, através de uma base construída, o algoritmo pode induzir um possível valor (uma hipótese) de seu

atributo alvo (Irá chover?) para um determinado dado utilizando apenas seus atributos de entrada (O clima está ensolarado?, O clima está nublado?, O clima está úmido?).

Voltando a quadro de exemplo Clima, o propósito principal é que um algoritmo de AM induza novas hipóteses capazes de determinar se irá chover ou não, assim quando elaborada uma nova hipótese válida, que ela sirva também para a criação de outra nova hipótese do conjunto de dados, e assim sucessivamente.

Para que um algoritmo de AM possa induzir uma hipótese a partir de um conjunto de dados, ele precisa utilizar uma forma de representação para descrever uma hipótese induzida, que pode ser redes neurais, árvores de decisão ou regras de decisão [FACELI et al., 2015].

A tarefa de aprendizado de um algoritmo de AM pode ser dividida em duas tarefas, as tarefas preditivas e tarefas descritivas. Em tarefas preditivas o objetivo é encontrar hipóteses através de um conjunto de treinamento, que preveja o valor do atributo alvo criando um novo exemplo a partir de seus atributos de entrada. Nesta tarefa, os algoritmos seguem um modelo de aprendizado supervisionado, que vem da supervisão de um agente externo, que seria um classificador, que determinaria a classe de novos exemplos não rotulados [MONARD; BARANAUSKAS, 2005].

Já em tarefas descritivas o objetivo é explorar ou descrever um conjunto de dados. O algoritmo analisa os exemplos e tenta determinar se alguns deles podem ser agrupados de alguma maneira. Neste tipo de tarefa os algoritmos não fazem uso do atributo de saída, portanto utilizam o modelo de aprendizado não supervisionado, uma tarefa de agrupamento por exemplo tem como objetivo encontrar grupos de objetos semelhantes ao conjunto de dados [MONARD; BARANAUSKAS, 2005]. A Figura 1 apresenta a hierarquia de aprendizado de acordo com os tipos de tarefas de aprendizado.

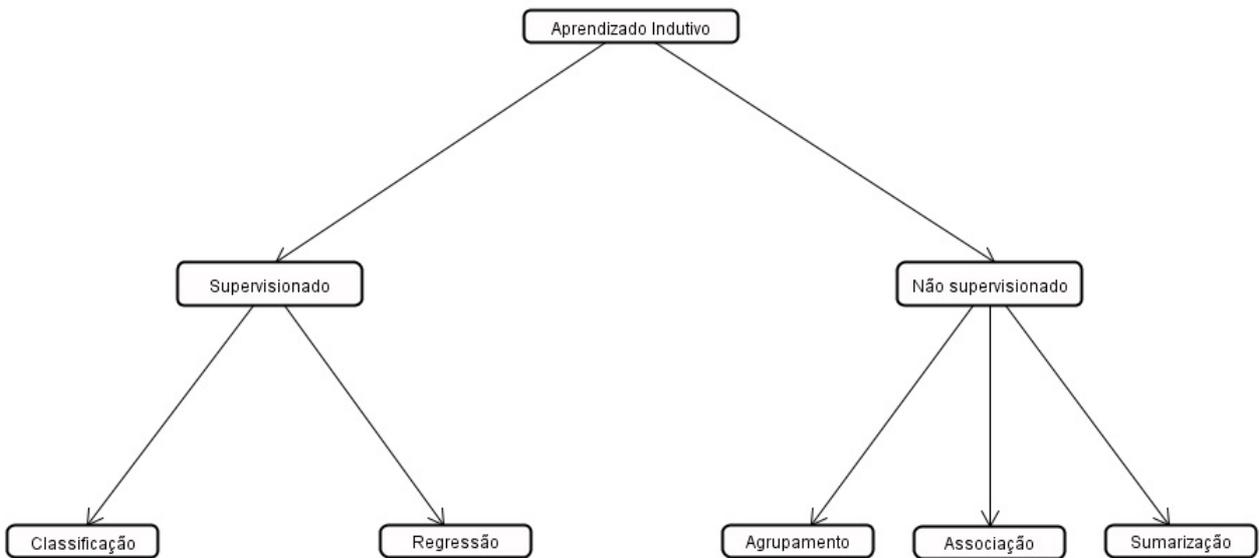


Figura 1 – Demonstração da hierarquia de aprendizado

Fonte: Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. FACELI, K; LORENA, A.C; GAMA, J; CARVALHO, A.P.L.F - Rio de Janeiro: LTC, 2015.

No topo encontra-se o aprendizado indutivo, em seguida o aprendizado supervisionado (preditivo) e não supervisionado (descritivo). As tarefas preditivas se diferem pelo rótulo de dados: discreto no caso de classificação, e contínuo no caso de regressão. As tarefas descritivas são genericamente divididas em: agrupamento, onde os dados são agrupados de acordo com a similaridade; associação, onde o objetivo é encontrar padrões frequentes de associações entre atributos de um conjunto de dados; e sumarização, que consiste em encontrar uma descrição simples e compacta para um conjunto de dados.

2.1.1 Árvores de Regressão

Quando nos referimos a aprendizado de máquina, fundamentalmente estamos nos referindo a técnicas cujo a idéia é procurar um possível candidato a um determinado problema em questão (indução), fazendo-se uso da aprendizagem. Uma das técnicas utilizadas são de métodos baseados em procura, o qual trabalhamos com modelos baseados em árvores (árvores de decisão e árvores de regressão), que são modelos que formulam uma máquina a procurar a solução de um problema na procura por possíveis candidatos em um espaço de possíveis soluções [Faceli et al. \[2015\]](#). Modelos baseados em árvores é uma representação de uma tabela, que através de informações contidas nesta tabela decisões são tomadas sob uma forma de árvore. [Mitchell \[1997\]](#) afirma que aprendizagem baseada em árvores de decisão é um dos métodos mais práticos para inferência indutiva. As árvores de regressão são idênticas a uma árvore de decisão, como mostra a [Figura 2](#), visto que uma árvore de regressão também é formada por um conjunto de nós de decisão.

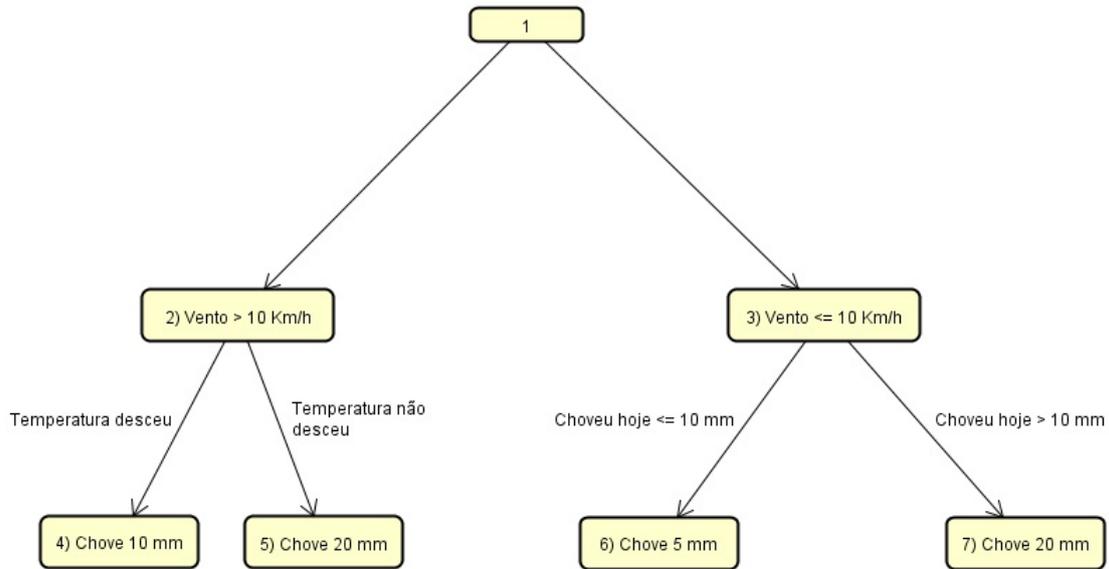


Figura 2 – Demonstração árvore de regressão

A diferença principal entre os modelos de árvores são que árvores de regressão as folhas das primeiras contém previsões numéricas, e árvores de decisão contêm decisões. De acordo com a Figura 3, é possível verificar como a árvore também pode ser interpretada, transformando-a em linhas de execução.

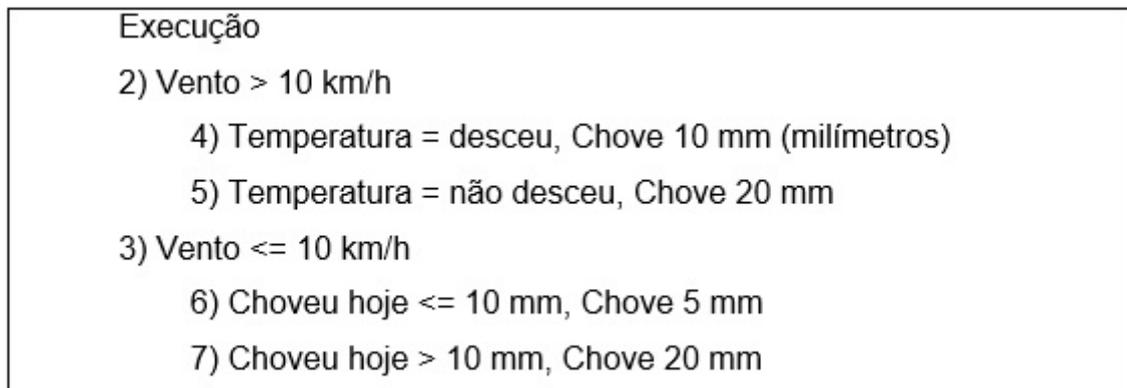


Figura 3 – Execução de uma árvore de regressão

A leitura da folha 7 da árvore pode ser interpretada conforme demonstra o Figura 4.

Se vento <= 10 Km/h E choveu hoje > 10 mm ENTÃO amanhã chove 20 mm

Figura 4 – Exemplificação de uma folha de árvore de regressão

Em árvores de regressão cada nó terminal ou folha contém uma constante, que no caso pode ser uma média ou uma equação para o valor previsto de um determinado conjunto de dados.

Para a seleção de um nó folha de uma árvore de regressão, algoritmos de AM utilizam um critério de parada, em que esse critério retorna uma constante, portanto, se o critério de parada for satisfeito, então o algoritmo retorna um nó folha rotulado com a constante que possui um valor previsto para um conjunto de dados a partir de seus atributos [FACELI et al., 2015].

Outro critério em algoritmos de AM, são os critérios de divisão, que nada mais são que atributos que “caracterizam” o conjunto de dados, que dão maior importância ao atributo alvo do conjunto de dados. Regras de divisão, para problemas de regressão, usualmente tratam minimizar o erro quadrático, como mostra a equação a baixo, que pode ser usado como uma medida de erro de previsão. Deste modo a constante de cada nó folha é a que minimiza o erro quadrático. Portanto, a constante associada à folha de uma árvore de regressão é a média dos valores do atributo alvo dos exemplos de treinamento que caem na folha [FACELI et al., 2015].

$$EQ = \sqrt{\sum_{i=1}^n (Y_i - \hat{Y})^2} \quad (2.1)$$

Na equação EQ é o erro quadrático, n é o número de amostras, Y_i é o valor fornecido pelo classificador para a i -ésima amostra e \hat{Y} é a média dos valores de todas as amostras.

Na construção da árvore existe uma etapa chamada poda, que para Faceli et al. [2015], a poda de árvores é utilizada para quando são classificados novos exemplos que não irão melhorar a capacidade preditiva da árvore. As estratégias de podas são divididas em duas: pré-poda e pós-poda. A pré-poda conta com regras de parada que previnem a construção de ramos desnecessários e tem a vantagem de que o tempo não é perdido na construção de uma estrutura que não é utilizada na árvore final. Já na pós-poda o critério a ser levado em conta é se devo ou não podar uma subárvore baseados no números de classificações incorretas.

2.1.2 Redes Neurais Artificiais

Ainda trabalhando no caso de modelos supervisionados existem métodos baseados em otimização, onde técnicas de AM recorrem a otimização de uma função em seu treinamento. Entre elas existem as Redes Neurais Artificiais (RNAs), que possuem inspiração em redes neurais biológicas presentes no cérebro que envolvem uma regra de correção de erros na qual recorrem a otimização de uma função quadrática do erro entre as respostas da RNA e os rótulos dos exemplos [FACELI et al., 2015]. A ideia principal é a de que

RNAs se assemelhem a forma de pensar do cérebro humano, que é constituído de neurônios que são interconectados por sinapse construindo uma rede de ligações neurais, na qual trabalham na aquisição de conhecimento [MEDEIROS, 2003]. Assim as RNAs são baseadas em modelos abstratos de como acredita-se que o cérebro (e os neurônios) do ser humano funciona.

O desenvolvimento de uma RNA se deu pela motivação de que este modelo imitasse a capacidade de alto processamento do cérebro humano em reconhecer e associar padrões [FRANCO et al., 2015]. A constituição de uma RNA é baseada por unidades de processamento simples do cérebro humano, neurônios artificiais, que aplicam funções matemáticas aos dados e são dispostos em camadas densamente interconectadas entre si [BINOTI; BINOTI; LEITE, 2014]. Tais unidades de processamento podem estar dispostas em uma ou mais camadas interligadas por diversas conexões.

A semelhança entre uma RNA e o cérebro humano se assemelha por dois aspectos: o conhecimento adquirido através do processo de aprendizagem, e as forças de conexão entre neurônios conhecidas como pesos que são utilizados para armazenar o conhecimento adquirido [HAYKIN, 2001]. Os pesos são ponderados e associados a cada entrada recebida pelos neurônios, que são dispostos por uma função matemática análogo ao processamento dos mesmos.

Finocchio [2014] explica que a saída de um neurônio é definido por meio da sua função de ativação, a qual processa o conjunto de entradas recebidas em estado de ativação, que podem ser valores binários (0 e 1), bipolares (-1 e 1) e valores reais, onde no estado de ativação são escolhidas as necessidades do problema em que a rede esteja trabalhando, a figura 5 mostra o formato das funções linear, limiar e sigmoidal.

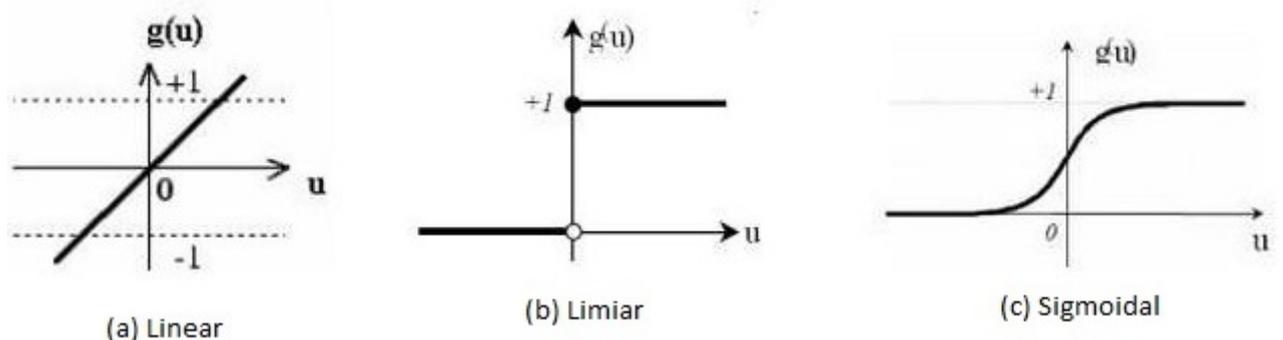


Figura 5 – Exemplo de funções de ativação

Fonte: Noções de redes neurais artificiais. Finocchio, M. A. F; Universidade Tecnológica Federal do Paraná - Campus Cornélio Procópio, 2014.

A função linear (figura 5(a)) implica em retornar como saída o valor de u . A função

limiar (figura 5(b)) apresenta que o neurônio assumirá valor 1 se o nível de atividade interna total for não negativo, isto é, quando a soma das entradas recebidas ultrapassa o limite (limiar) estabelecido, caso contrário assumirá o valor 0. A função sigmoideal (figura 5(c)) a saída do neurônio assumirá valores reais entre 0 e 1.

Faceli et al. [2015] esclarece que um RNA pode estar disposta em duas ou mais camadas (multicamadas), onde um neurônio pode receber em seus terminais de entradas valores de saída de neurônios da camada anterior e/ou enviar seu valor de saída a terminais de entradas de neurônios da camada seguinte. Uma rede neural multicamadas pode possuir camadas ocultas (camadas intermediárias) que ficam localizadas entre as camadas de entrada e saída, a diferenciação de uma rede com camadas ocultas é o não contato com o exterior.

Em redes multicamadas existem diversos meios de conexão entre os neurônios, são elas: completamente conectada, parcialmente conectada e localmente conectada [FACELI et al., 2015]. Em conexões completamente conectadas todos neurônios de uma camada estão conectados a todos os neurônios da camada anterior e da camada seguinte. Já as conexões parcialmente conectadas acontece quando os neurônios estão conectados a apenas alguns neurônios da camada anterior e da camada seguinte. E por fim a conexão localmente conectada acontece quando os neurônios que estão conectados se encontram em uma região bem definida. A figura 6 apresenta uma RNA multicamadas completamente conectada.

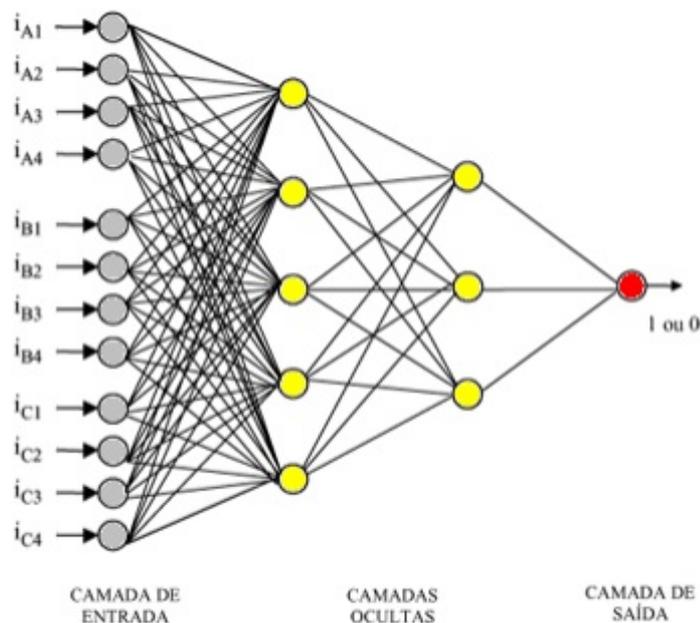


Figura 6 – Rede neural artificial

Fonte: Noções de redes neurais artificiais. Finocchio, M. A. F; Universidade Tecnológica Federal do Paraná - Campus Cornélio Procópio, 2014.

A camada de entrada recebe os atributos i_A , i_B e i_C com seus pesos ponderados, que são processados por uma função matemática f_a e dispostos a próxima camada até a camada de saída que gera os valores de saída.

O aprendizado de um algoritmo de RNA baseia-se fundamentalmente no ajuste de parâmetros: número de camadas, número de neurônios em cada camada, a presença ou não de retroalimentação, grau de conectividade (pesos). Contudo, o ajuste de parâmetros baseia-se principalmente na definição dos pesos associados às conexões que geram a maior acurácia preditiva do algoritmo. [Faceli et al. \[2015\]](#) divide o aprendizado de um algoritmo de RNA em quatro grupos: Correção de erros, que procuram ajustar os pesos da RNA afim de reduzir os erros cometidos pela rede; Hebbiano, que afirma a regra de Hebb, na qual se dois neurônios estão simultaneamente ativos, deve-se reforçar a conexão entre os mesmos; Competitivo, que define quais neurônios devem ter ajustes aos pesos associados; e Termodinâmico, que são algoritmos estocásticos baseados em princípios observados na metarlugia.

2.2 Agricultura de Precisão

Para [Lamparelli \[2016\]](#), agricultura de precisão(AP) é um conjunto de técnicas que permite o gerenciamento localizado dos cultivos, e prevê a redução dos gastos da produção agrícola. Já para [Bongiovanni e Bongiovanni e Lowenberg-DeBoer \[2004\]](#), a agricultura de precisão pode contribuir para a sustentabilidade a longo tempo, com a aplicação de fertilizantes e pesticidas apenas em locais onde eles são realmente necessários.

A AP trabalha sobre três vertentes na estratégia de gestão. A primeira é a de aumentar a produtividade, dentro dos limites da lucratividade. Outra estratégia é a redução de custos, com diminuição do uso de insumos por meio da racionalização guiada pela variabilidade espacial. E a terceira é a unidade de gestão diferenciada, a qual é entender que cada parte da lavoura necessita de uma estratégia, além dos padrões usuais [[MOLIN; AMARAL; COLAÇO, 2015](#)] .

[Inamasu e BERNARDI \[2014\]](#) definem AP como “um sistema de gestão que leva em conta a variabilidade espacial do campo com o objetivo de obter um resultado sustentável social, econômico e ambiental”.

Para se trabalhar com a agricultura de precisão é preciso que o produtor entenda as desigualdades de cada setor, tratando cada área de produção conforme suas exigências, isso faz que o ele tenha o conhecimento detalhado de toda sua propriedade, de cada metro quadrado [[TSCHIEDEL; FERREIRA, 2002](#)]. É uma estratégia que gerencia as particularidades das lavouras com variados níveis de complexidade.

O que difere o gerenciamento da AP é a integração de modernas tecnologia na coleta de dados, processamento e análise de múltiplas fontes de dados em alta resolução

espacial e temporal [COELHO, 2005].

Molin, Amaral e Colaço [2015] mencionam que a AP pode ter um forte ligamento com o GPS para se trabalhar com georreferenciamento de dados dependente das necessidades da lavoura. BERNARDI et al. [2014] afirmam que sistemas de informações geográficas auxiliam a camada de informações, ampliando a capacidade de interpretação dos dados, auxiliando a tomada de decisão.

A AP pode ser aplicada em diversos setores agrários, o diferencial é a junção da tecnologia ao trabalho agrário, que se torna mais eficaz e mais simples, podendo mostrar ao usuário os dados sendo coletados. A Figura 7 apresenta como exemplo a simulação de um painel, que permite que o usuário visualize as ações do equipamento em tempo real.

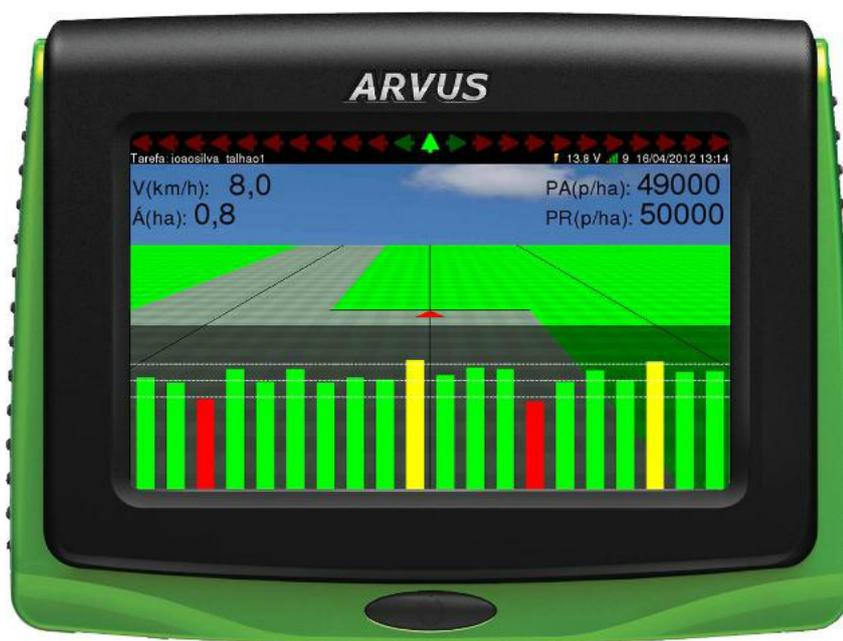


Figura 7 – Simulação de um painel utilizado em equipamento que implementa conceitos de AP

Ao utilizar de métodos da AP, trabalha-se com dados coletados dos diferentes tipos de lavouras, quanto mais dados disponíveis coletados, mais consistente é a informação gerada e o consequente diagnóstico referente a variabilidade existente na lavoura [MOLIN; AMARAL; COLAÇO, 2015].

3 TRABALHOS RELACIONADOS

Bueno et al. [2015] apresentam um trabalho que estabelece uma curva de deriva para cultura do feijão e da soja. Apresentam também uma curva com deriva com quatro pontas de pulverização fixas sendo elas: jato de plano simples com indução de ar para as pontas XR 11002 para gotas finas e AIXR 11002 para gotas grossas; e jato plano defletor simples e com indução de ar para as pontas TT 1102 para gotas médias, e TTI 1102 para gotas extremamente grossas. Dados os testes com as quatro pontas de pulverização, foi apresentado uma curva de deriva, e apresentado uma ponta de pulverização que menos ocorreu deriva, após isso foram aplicados os resultados aos modelos de estimativas de deriva “Modelo Alemão” e “Modelo Holandês” apresentando novos modelos de previsão de deriva. Os resultados apresentaram novos modelos de previsão de deriva, contudo são específicos para culturas e bicos de pulverização, além de ser muito complexo durante sua estruturação. Ao final é considerado que o "Modelo Holandês" se ajustou melhor aos dados, contudo não é indicado para previsões de deriva, pois possui um intervalo de confiança muito aberto, e tornando a técnica muito genérica.

Já Donkersley e Nuyttens [2011] apresentam um trabalho centrado as técnicas de pulverização, experimentos com técnicas reais, apresentando o volume de deriva de cada uma em relação à distância. No trabalho são utilizadas as técnicas de pulverização: a campo, laser, pomar folheado, sem folheto e túnel de vento. Cada método obteve variação sobre suas estimativas. Essas variações se devem a varios fatores, como: condições ambientais, sistema de cultivo, equipamento utilizado e rastreador utilizado. Cada técnica apresentou uma taxa de deriva depositada, entretanto, ambas as técnicas apresentaram a maior concentração de deriva entre os primeiro cinco metros.

Baetens et al. [2006] desenvolveram um modelo de dinâmica de fluidos computacionais (CFD), a fim de apresentarem um modelo de deriva de estado estacionário válido. Os experimentos foram realizados sobre um pulverizador com 54 bicos de pulverização à 50 cm do solo sobre uma pastagem a uma velocidade de 8 km/h. Os parâmetros utilizados para a realização do trabalho foram de fatores ambientais (velocidade do vento, direção do vento), e fatores mecânicos (altura da barra do pulverizador, velocidade do trator). O modelo criado apresenta uma sobre-predição, que é explicado devido aos resultados serem ruidosos devido a variação natural de seus parâmetros. A altura da barra, de acordo com o estudo, é o principal fator na variação da deriva, pois ela implica na altera: o comprimento de arrasto de uma partícula, e mudança na velocidade ambiental do vento. Contudo, é levado em consideração que a variação dos resultados do modelo é bem alta, que possui como principais causadores da deriva a altura da barra de pulverização e a velocidade do vento.

As técnicas existentes para estimativas de deriva geralmente são específicas para técnicas de aplicação e culturas, sendo muito robustas e complexas. Ambas as técnicas não envolvem conceitos de aprendizado de máquina para realizar as predições, portanto o trabalho tem por finalidade apresentar o desenvolvimento de técnicas de AM que possam ser utilizados em predições de deposição incorreta de produtos fitossanitários.

4 PROPOSTA

Métodos de aprendizado de máquina unidos com os princípios de agricultura de precisão, podem ajudar a solucionar problemas de deposição incorreta de produtos fitossanitários fora da área alvo. A junção de tais técnicas podem modernizar os sistemas de pulverização, a junção de tecnologia da informação à práticas agrícolas podem minimizar a perda de PFT que podem ser ocasionados pela deriva.

Técnicas de aprendizado de máquina se baseiam no princípio de que um sistema possa ser capaz de induzir novas hipóteses através de experiências já vivenciadas. Isto se torna uma prática viável se aplicada em sistemas de pulverização, onde dados coletados em tempo real possam ser aplicados à algoritmos de AM para prever informações futuras, em específico, prever o volume de PFT aplicados fora da área alvo. Contudo, existe a condição de que há diversas culturas e práticas agrícolas existentes, portanto tais técnicas devem apresentar aplicação para diversas culturas e diversas práticas de pulverização. O objetivo principal é estimar o volume do produto não aplicado adequadamente. Cada técnica utilizada deve apresentar uma estimativa de volume de produto com uma taxa de acerto plausível, podendo levar o usuário a tomar decisões para reduzir o volume de deriva.

O sistema proposto possui duas fases para sua implantação. A primeira refere-se a seleção da função de regressão gerado pelo algoritmo de AM. Enquanto que a segunda fase é o seu respectivo uso durante o manejo de proteção, provendo predições sobre o produto depositado em uma determinada região fora da área alvo. A Figura 8 apresenta os elementos e suas respectivas interações na primeira fase.

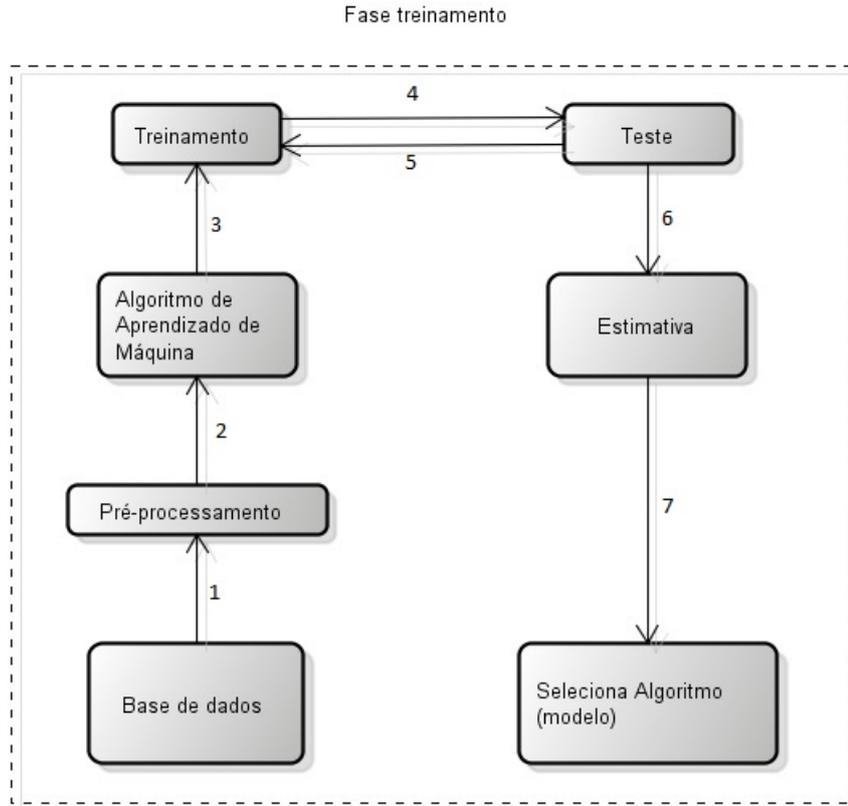


Figura 8 – Fase de seleção do modelo de algoritmo que irá compor o núcleo de processamento

A primeira etapa (1) é a da coleta dos dados, a qual é disponibilizada uma base de dados com todas as informações relevantes para o treinamento de um algoritmo de AM. Após a disponibilização da base de dados, a segunda etapa (2) é a de pré-processamento de dados, onde ocorreria a adequação dos dados, onde dados incompletos são removidos da base, informações redundantes são retiradas, dados simbólicos são convertidos para numéricos e é mantido uma padronização no modelo da base de dados. Um exemplo seria a alteração de atributos nominais para numéricos, como mostra a Figura 9. A próxima etapa (3) diz respeito ao algoritmo de aprendizado de máquina, onde é selecionado um algoritmo e feito suas configurações, seus parâmetros são ajustados. A etapa 4 e 5 são realizados o treinamento e testes do algoritmo. O treinamento é constituído por uma parte da base de dados (a maior parte), e o teste o restante da base. A base é treinada pelo algoritmo de AM, e então testada nos dados restantes. Esse processo é repetido n vezes, a cada ciclo o teste é realizado com novos dados da base. Após treinado o algoritmo, é analisado sua estimativa (6), é verificado sua acurácia, o quão precisas foram as estimativas. O ultimo passo (7) seleciona o algoritmo que possui maior acurácia, o que possui maior taxa de acerto nas predições.

Tabela original			Tabela pré-processada		
Tipo da cultura	Bico de pulverização	Altura da barra (m)	Tipo da cultura	Bico de pulverização	Altura da barra (m)
Soja	ATR	0.5	1	1	0.5
Milho	ATR	0.3	2	1	0.3
Soja	XR	0.5	1	2	0.5
Milho	XR	0.3	2	2	0.3

Figura 9 – Exemplo de pré-processamento dos dados

Após o modelo selecionado, foi proposto um sistema, onde o modelo selecionado é implantado em um sistema em execução para realizar previsões em momento real. A Figura 10 mostra o sistema proposto em execução do modelo selecionado.

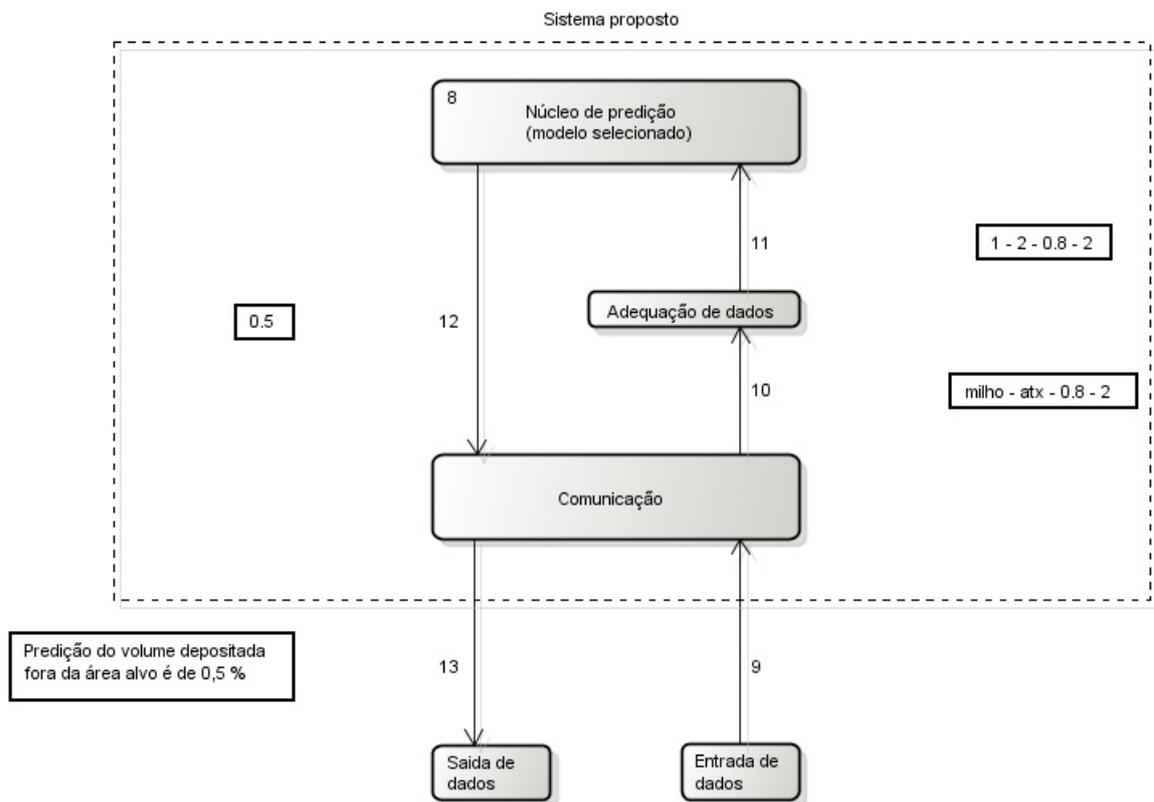


Figura 10 – Elementos que compõem o sistema proposto em execução

Após o modelo selecionado, o mesmo é implantado dentro de um núcleo de predição no esquema do sistema proposto, passo 8. O próximo passo (9) seria o de entrada de dados, onde em um sistema de pulverização, um agente externo coletaria dados de possíveis causas de deposição fora da área alvo em tempo real, que ocorre por deriva. Por exemplo: altura da barra, pressão dos bicos de pulverização, direção do vento, velocidade do vento, entre outros. Após a coleta dos dados, eles seriam passados a um sistema de comunicação (10), o qual é a interface de comunicação entre o usuário e a aplicação, que padronize os

dados para que sejam legíveis tanto para o sistema de pulverização quanto para o algoritmo de predição. Neste caso, a Figura 10 exemplifica o passo 10. Um agente externo entregaria as seguintes informações: milho - atx - 0.8 (m) - 2 (m). Onde o dado "milho" se refere a cultura, "atx" ao bico de pulverização, "0.8 (m)" a altura da barra de pulverização em metros e "2 (m)" que se refere a distância que ocorre a deposição. O próximo passo (11) é realizado a adequação dos dados, semelhante ao passo 2 da Figura 8. Onde os dados sejam adequados ao algoritmo de predição, no caso da Figura 10, os dados passados pelo passo 10 (milho - atx - 0.8 (m) - 2 (m)), seriam adequados para: 1 - 1 - 0.8 - 2, que qual seriam os dados da cultura, bico de pulverização, altura da barra e a distância que está ocorrendo a deposição, respectivamente. Esse dados seriam passados a etapa 11, núcleo de predição, que é composto pelo modelo selecionado no passo 7 da Figura 8. O núcleo de predição é dito como oitavo passo, pois após selecionado o modelo, o mesmo é incluso ao núcleo de predição. Essas predições são realizadas sobre a deposição fora da área alvo a uma determinada distância. Após realizada a predição, é então retornado um valor, passo 12, o qual se refere ao volume de deposição incorreta a uma determinada distância, a 2 metros de distância segundo a Figura 10. Então esse dado é passado pela camada de comunicação (13), onde é entregue a outro agente, que mostraria ao usuário informando a deposição fora da área alvo, como mostra a Figura 10 é mostrado ao usuário: "Predição do volume depositado fora da área alvo é de 0.5%".

Com a ideia de um sistema em execução, implementando conceitos de AM, seria possível prever situações que causassem a aplicação errada de PFT, situações que ocorram por deriva. Desse modo, os dados a serem coletados por agentes externos, seriam dados relativos a deriva de PFT.

Desta forma, técnicas de aprendizado de máquina podem estimar a deposição de produtos fora da área alvo que ocorrem por deriva, para que posteriormente, perante ações tomadas pelo operador do sistema de pulverização, possam ser tomadas decisões que minimizem a deriva. Isto estaria ocasionando diversos benefícios, como: a perda desnecessária de produtos a serem aplicados, impactos ambientais, e deposição em áreas indesejadas.

Portanto a proposta do trabalho é estudar algoritmos de aprendizado de máquina, realizar experimentos e análises sobre cada técnica abordada, analisar os dados de cada algoritmo preditor. E ao final do trabalho, apresentar qual técnica apresenta maior acurácia do dados trabalhados.

5 METODOLOGIA

5.1 Algoritmos Avaliados

O trabalho focou na construção do núcleo de predição, que é constituído por um algoritmo preditor. Tal avaliação foi realizada no ambiente de desenvolvimento RStudio para linguagem R, a qual é um software livre. A versão da linguagem R utilizada é a 3.3.3.

Com o objetivo de encontrar um algoritmo que apresentasse maior acurácia nas predições, foram avaliados dois algoritmos distintos, os quais possuem o viés de busca e forma de aprendizado diferentes uns dos outros. As técnicas utilizadas são: árvores de regressão e redes neurais artificiais. Enquanto árvores de regressão trabalham com métodos baseados em procura, redes neurais artificiais trabalham com métodos baseados em otimização. Para o ajuste dos parâmetros, foram realizados pré-testes para configurar os algoritmos. A técnica de árvores de regressão utilizou o pacote `rpart`, que teve seus parâmetros ajustados como mostra a Tabela 2.

Tabela 2 – Parâmetros utilizados na árvore de regressão

Parâmetros	Valor
<code>minsplrit</code>	2
<code>cp</code>	0
<code>maxcompete</code>	10

O parâmetro `minsplrit` equivale ao número mínimo de observações que devem existir em um nó antes de tentar uma divisão. Já o parâmetro `cp` diz respeito a pré-poda da árvore, onde são eliminadas divisões que não melhoram a precisão preditiva da árvore, assim economizando tempo de construção. E por fim o `maxcompete`, que é o número de divisões concorrentes retidos a cada nó, o qual é importante saber qual divisão foi escolhida, mas quais variáveis vieram em seguida.

A segunda técnica utilizada foi a de RNA utilizando o pacote `neuralnet`, o qual teve as seguintes configurações, Tabela 3.

Tabela 3 – Parâmetros utilizados na rede neural

Parâmetros	Valor
Função de ativação	Sigmoidal
<code>hidden</code>	7 e 5
<code>linear.output</code>	TRUE

O parâmetro função de ativação diz respeito a saída de um neurônio, como foi mencionado na seção 4.2. O parâmetro hidden diz respeito a especificação do número de neurônios escondidos em cada camada, no caso do parâmetro utilizado (7 e 5), é a concatenação de sete neurônios na primeira camada oculta e cinco na próxima camada oculta. Por fim o parâmetro linear.output que define se a função deve ser ou não aplicada aos neurônios de saída, neste caso, a função de ativação não é aplicada.

O ajuste dos parâmetros foram realizados de forma empírica, onde eram realizados pré-testes e conforme a acurácia alcançada, as configurações eram alteradas para cada algoritmo. As configurações apresentadas são as melhores encontradas. Assim que as acurácias das predições alcançavam um limiar satisfatório (ou seja, o erro médio das predições era baixo) a respectiva configuração foi considerada para os experimentos.

5.2 Base de dados

A base de dados utilizada neste trabalho foi disposta por [Bueno et al. \[2015, p. 29\]](#), na qual os dados foram recolhidos através da pulverização da cultura do feijão. A base apresenta a deposição fora da área alvo a cada 2,5 metros de distância até os 50 metros. O trabalho foi realizado sob a norma ISO 22866 (ISO, 2005) que determina as condições climáticas de pulverização do produtos fitossanitários, que são: temperatura deve estar entre 5 e 35 °C, a velocidade mínima do vento deve ser de $1,0 \text{ ms}^{-1}$ ($3,6 \text{ kmh}^{-1}$) e a direção do vento dentro dos limites $90^\circ \pm 30^\circ$ em relação a linha de pulverização. Os teste realizados consideram quatro modelos de pontas de pulverização: XR, TT, AIXR e TTI; estando a 0,5 metros entre si, e a 0,5 metros de altura em relação a cultura. As aplicações foram realizadas por um pulverizador hidráulico, com pressão de 3 bar (300 kPa) para cada bico, e a velocidade de pulverização foi de 6,5 km/h.

Os dados que que constituem a base de dados foram recolhidos por [Bueno et al. \[2015\]](#) através de amostras que coletaram a deposição incorreta que ocorreu por deriva. As amostras se encontravam em blocos até umas distância total de 50 metros da área pulverizada, contabilizando a deposição a cada 2,5 metros. Foi realizado um esquema de 4 x 20, onde foram distribuídos coletores a cada 2,5 metros de distância até a distância máxima de 50 metros. A cada distância (2,5 metros) foram colocados quatro coletores alinhados lado a lado com 1,5 metros de distância cada. A Figura 11 demonstra o esquema utilizado na coleta das amostras.

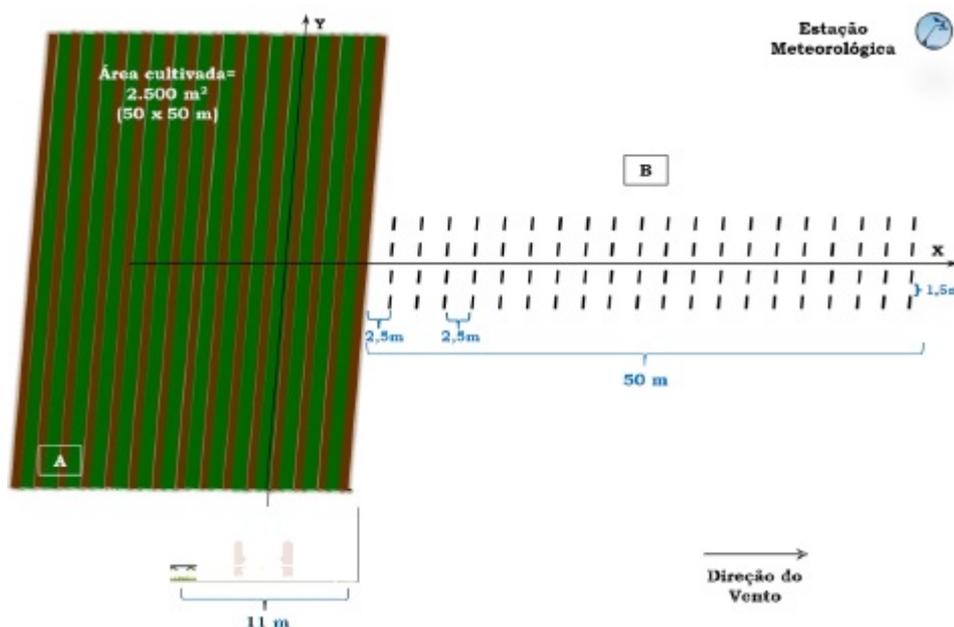


Figura 11 – Esquema de coleta das amostras de deposição fora da área alvo.

Fonte: Deriva e análise de riscos das aplicações de fungicidas, herbicidas e inseticidas nas culturas do feijão e da soja. BUENO, M.R; Uberlândia - Minas Gerais, 2015.

Para a coleta das amostras foram realizadas 10 repetições, 10 pulverizações para cada ponta utilizada. Obteve-se então 40 amostras de deposição incorreta, das mesmas fora feito a média para cada distância demarcada. A média foi realizada sobre a porcentagem que foi coletado nas amostra por hectare (ha) pulverizado. A base de dados que está sendo utilizada não possui replicação de dados, mas é composta pela média das replicações.

Esta base de dados é composta pelos seguintes atributos: distância (m) em relação a área pulverizada, pressão da barra de pulverização (bar), altura da barra de pulverização (m), pontas de pulverização (XR, TT, AIXR e TTI), e a deriva depositada. A base de dados pode ser encontrada no [github](#)¹.

Pode se observar que a base apresenta uma diminuição da porcentagem de deriva em relação ao aumento da distância entre o bico de pulverização e o local de deposito. Pode-se observar também que a partir dos 12,5 metros, a deriva de todas as pontas de pulverização foram menores que 1%, comparado a quantidade de liquido pulverizado.

5.3 Experimentos

A base de dados utilizada sofreu um pré-processamento, onde dados nominais como as pontas de pulverização (XR, TT, AIXR e TTI), foram convertidos para atributos nu-

¹ github.com/MarcusVDorta/Aprendizado-de-M-quina

méricos (1, 2, 3 e 4) respectivamente. Dados incompletos e redundantes foram removidos da base.

Para ambas as técnicas foi utilizado o método de avaliação dos modelos preditivos, validação cruzada (*cross-validation*). No método de validação cruzada a base de dados é dividida em r subconjuntos aproximadamente iguais, então a base é dividida em $r - 1$ conjuntos para fase de treinamento de um preditor, que posteriormente é testado no conjunto restante. Esse processo é repetido n vezes, utilizando em cada ciclo uma partição (grupo) diferente para teste, o desempenho do preditor é dado pela média dos desempenhos observados por cada subconjunto de teste. Para realização do trabalho a base contendo 80 (oitenta) itens de dados amostrais, com os dados sorteados, com volume até 50 metros foi subdividida em 10 (dez) grupos, cada grupo contendo 8 dados, a figura 12 exemplifica esse processo. Para a validação do preditor está sendo utilizado $k=9$, onde são postos 9 grupos (72 itens de dados) para treinamento do preditor e 1 grupo (8 itens de dados) para teste, e a cada ciclo é retirado um grupo diferente para teste, e o restante utilizado para treinamento do preditor até que todos os grupos, um por vez, tenham sido utilizados no teste do preditor.

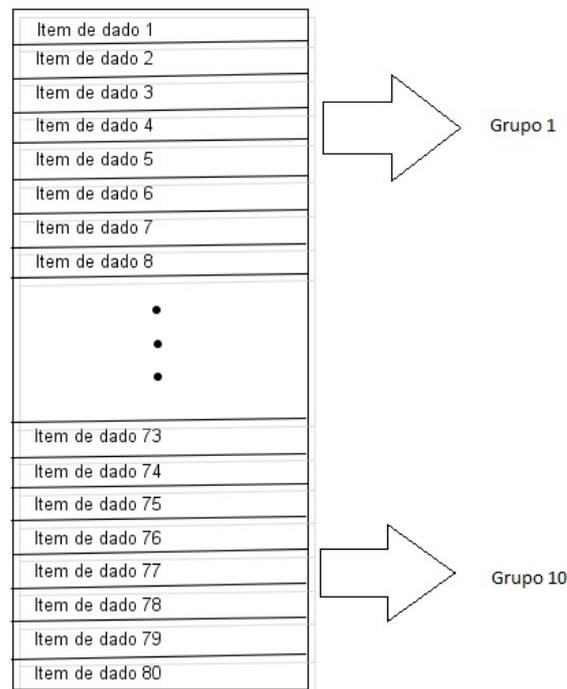


Figura 12 – Exemplo da divisão em grupos da base dados

Nas técnicas implementadas, ao passar os dados para os preditores, eram especificados quais atributos eram utilizados para a construção dos preditores e qual atributo é utilizado como atributo alvo. Neste caso era tomado como atributo alvo "deriva depositada" no teste, e no treino eram passados todos os atributos para construção do preditor.

Ao final dos ciclos da técnica de validação, é gerada uma lista com as respectivas previsões dos itens de dados testados. Logo após são criados mais dois atributos: dados previstos e erro. O erro é a diferença entre a deposição real e a prevista. A acurácia do preditor se dá pela média do erro.

6 RESULTADOS

O objetivo de cada algoritmo era prever a deposição incorreta do produto aplicado, e analisar o comportamento de cada técnica. As técnicas de árvores de regressão e redes neurais tiveram a mesma técnica de avaliação mantida. Cada técnica teve suas previsões geradas, e logo após era feita a diferença entre os dados reais e os exemplos previstos. A base de dados apresenta a diminuição na deposição de produto de acordo com o afastamento do mesmo que ocorre por deriva, uma curva de deposição. A Figura 13 apresenta um gráfico da base de dados utilizada.

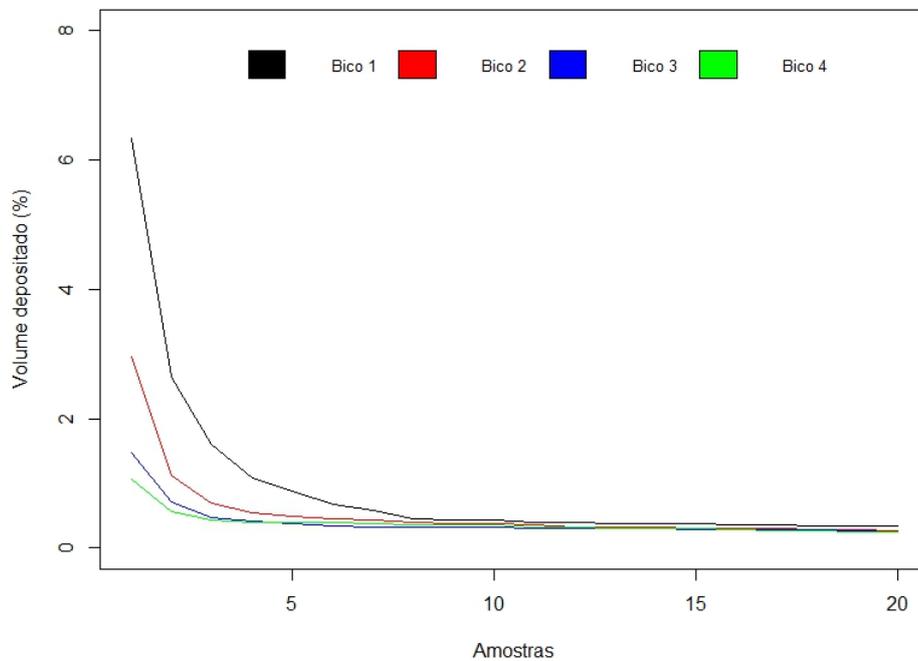


Figura 13 – Deposição incorreta de produtos em relação a distância

Com o gráfico dos dados reais gerado acima, pode-se observar que com o aumentar da distância, a deposição do produto diminui, a partir da 10^a amostra, 25 metros de distância, a deposição do produto é menor que 0% da área pulveriza por hectare. A primeira técnica utilizada foi a de árvore de regressão, que gerou uma lista com os dados previstos, que apresenta a deposição incorreta do produto de acordo com a distância. Logo após foi feita a diferença entre os dados reais e os previstos pela técnica. O algoritmo de árvore de regressão teve o erro médio de 0,1298% para as previsões de produto não aplicado adequadamente. A Figura 14 ilustra a curva de deposição que a técnica previu para os bicos de pulverização. A Figura 15 apresenta uma comparação da deriva depositada (dado real) com as previsões realizadas, apresenta também o erro dos dados gerado pela técnica.

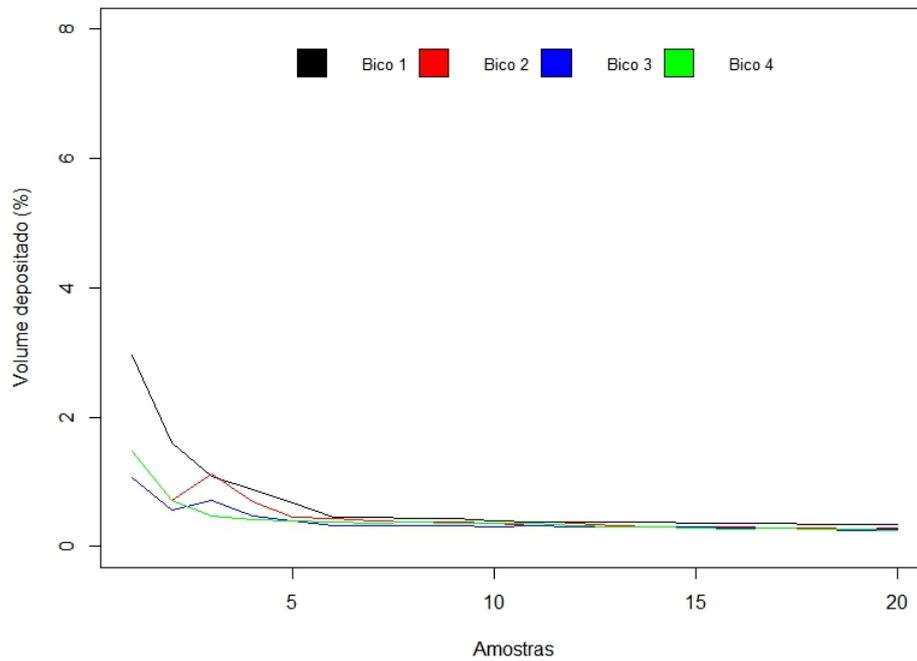


Figura 14 – Previsão da deposição incorreta de produtos em relação a distância gerados pela técnica de árvore de regressão

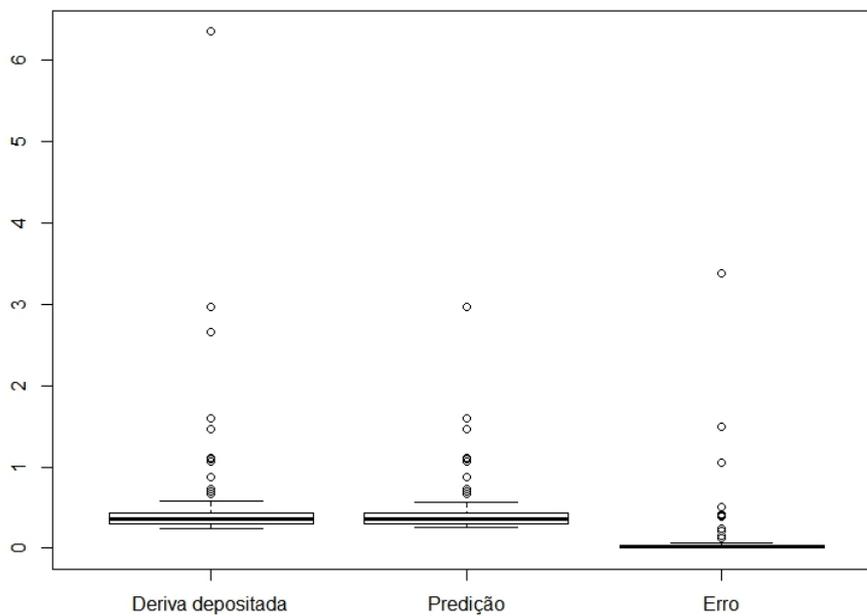


Figura 15 – Semelhança no comportamento dos dados no algoritmo de árvore de regressão.

Como pode-se observar na Figura 14, existe uma semelhança entre a deriva depositada com as previsões geradas pela técnica de árvore de regressão após a oitava (8ª)

amostra. A segunda técnica abordada foi a técnica de redes neurais artificiais que após a realização dos passos citados seção 6, foi feita a diferença entre os dados reais e os previstos. A técnica de redes neurais teve como erro médio 0,0579% entre a diferença entre os dados reais dos previstos. A Figura 16 apresenta uma curva de deposição que comparada a Figura 13 apresenta um comportamento mais semelhante do que a técnica de árvore de regressão.

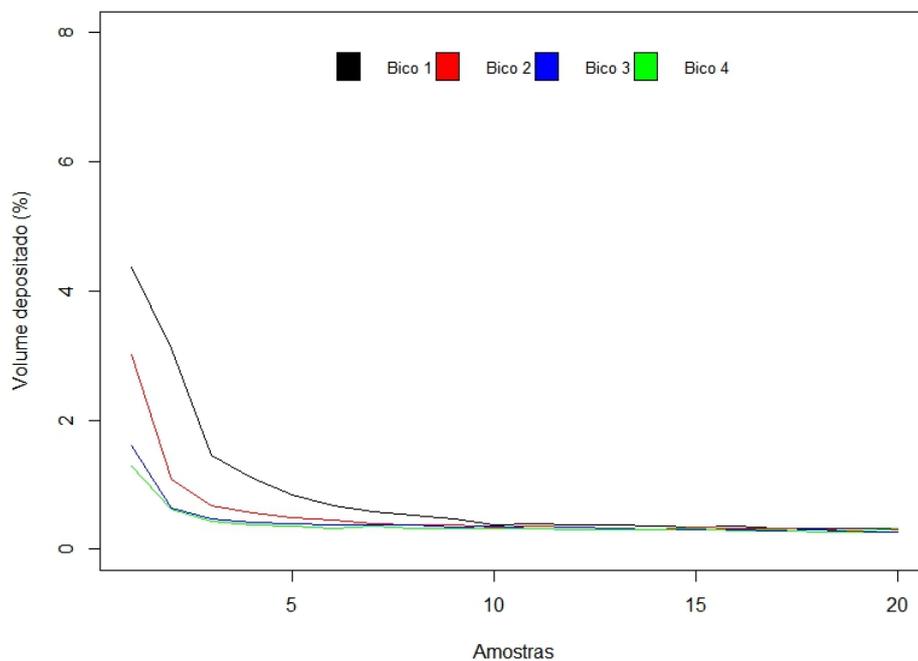


Figura 16 – Previsão da deposição incorreta de produtos em relação a distância gerados pela técnica de redes neurais

A técnica de redes neurais teve maior acurácia sobre a técnica de árvore de regressão, pode-se observar que os dados previstos até a quinta (5ª) amostra são mais semelhantes quando comparado a técnica de árvore de regressão. A técnica de árvore de regressão apresenta um comportamento inesperado próximo a 3ª amostra, o que não acontece na técnica de redes neurais. A Figura 17 ilustra o comportamento dos dados na técnica de redes neurais.

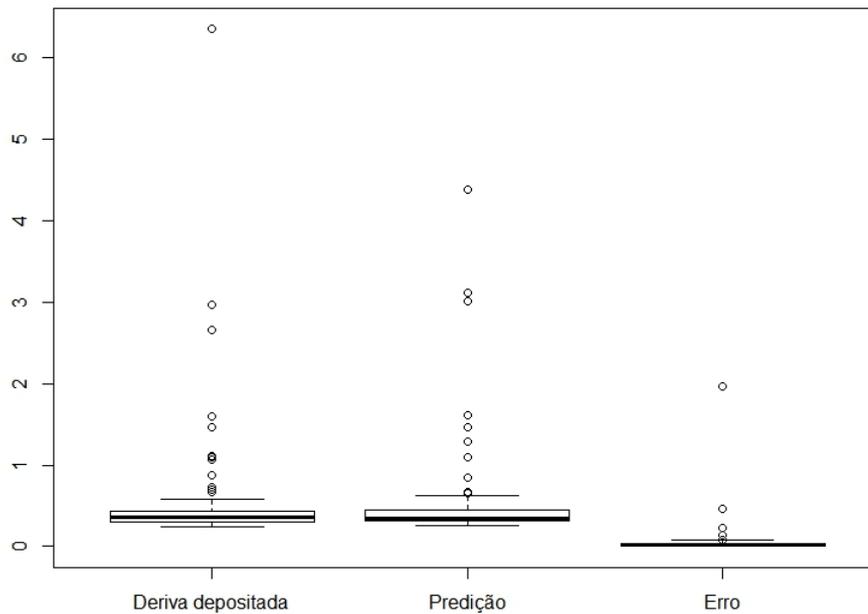


Figura 17 – Semelhança dos valores preditos aos valores reais pela técnica de redes neurais

Analisando a figura acima (Figura 17), os dados preditos pela técnica de redes neurais apontam um comportamento mais semelhante aos dados reais do que pela técnica de árvore de regressão. Pode-se observar também o erro apresentado pela Figura 17 também é menor comparado a outra técnica. Ambas as técnicas apresentam um comportamento semelhante a partir da décima amostra, onde os dados previstos se aproximam de 0. Pode-se notar na Tabela 4 que a técnica rede neural apresentou menor erro mínimo, máximo e médio em relação a técnica árvore de regressão. Tal comportamento demonstra que a rede neural proporcionou uma função de regressão mais adequada ao conjunto de dados disponível para o treinamento.

Tabela 4 – Tabela de erros

Técnica	Erro mínimo (%)	Erro máximo (%)	Erro médio (%)
Árvore de regressão	0.0002	3.3802	0.1298
Rede neural	8^{-5}	1.9711	0.0579

No trabalho de [Bueno et al. \[2015\]](#), é apontado o modelo IMAG de deriva de regressões exponenciais de quatro parâmetros (Modelo Holandês), como o que melhor se ajustou aos dados tanto para a cultura do feijão quanto da soja, com ajuste acima de 99%. Essa metodologia é realizada sobre cada bico de pulverização, sendo que o ajuste dos mesmos utilizam coeficientes diferentes para o ajuste da técnica, tornando uma técnica muito complexa por existir diversas configurações e ajustes, e robusta por ser específica

para cultura e método de aplicação. Já quando trabalhamos com aprendizado de máquina, uma técnica é capaz de realizar previsões a partir de uma única configuração, não sendo específica para cultura ou práticas agrícolas. Contudo, [Bueno et al. \[2015\]](#) também afirma que o Modelo Holandês não é indicado para estimar a deposição em deriva na cultura do feijão e da soja, pois os coeficientes utilizados na equação apresentam variação nos intervalos de confiança, tornando esse modelo excessivamente genérico.

Para realizar a avaliação dos dados estatisticamente primeiramente os dados previstos pelas técnicas foram submetidos ao teste de Shapiro Wilk de distribuição normal. O teste de Shapiro resulta em um p-valor. Caso p-valor seja maior que 0.05, a normalidade dos dados é aprovada, caso o p-valor dos dados seja menor que 0.05, a normalidade dos dados é reprovada. Caso p-valor for menor que 0.05, os dados são dispostos ao teste de Wilcoxon, que compara técnicas, os dados previstos das duas técnicas, e afirma se existe diferença estatística entre elas. Caso o teste de Wilcoxon obtiver p-valor maior que 0.05, não existe diferença estatística entre as técnicas, caso p-valor menor que 0.05, existe diferença estatística entre as técnicas. A Tabela 5 apresenta o p-valor obtido pelo teste de Shapiro Wilk, e a Tabela 6 valor obtido pelo teste de Wilcoxon comparando as duas técnicas de aprendizado de máquina (árvore de regressão e redes neurais).

Tabela 5 – Resultados do p-valor obtidos pelo teste de Shapiro Wilk

Técnica	p-valor
Árvore de regressão	0.000
Redes neurais	0.000

Tabela 6 – Resultado do teste de Wilcoxon

Teste	p-valor
Wilcoxon	0.8444

Ambas as técnicas de predição obtiveram normalidade reprovada pelo teste de Shapiro, onde Árvore de regressão teve como p-valor 0.0000, e redes neurais obteve o p-valor igual a 0.0000. Sendo assim, as técnicas foram submetidas ao teste de Wilcoxon, que obteve um p-valor igual a 0.8444, indicando que as técnicas não possuem diferença estatística significativa.

7 DISCUSSÃO E TRABALHOS FUTUROS

Este trabalho teve como objetivo estudar técnicas de aprendizados e máquina, as quais foram implementadas e analisadas. A partir da implementação das técnicas de aprendizado de máquina, árvores de regressão e redes neurais, foi possível prever curvas de deposição de produtos que acontecem através da deriva. As análises realizadas sobre os algoritmos foram efetuadas para relatar se ambos seriam eficazes ao realizar a previsão do produto depositado fora da área alvo. Os algoritmos podem ser implantados a um sistema de manejo de proteção agrícola, que através de um agente externo, os dados são coletados em tempo real e ajustados para o uso do núcleo preditor, composto por alguma das técnicas. Por fim, os dados transpassados à técnica de previsão apresentaria ao usuário, operador do sistema de pulverização, uma estimativa, uma curva de deposição incorreta do produto aplicado, assim sendo capaz de tomar medidas que reduziriam a perda do produto.

Com os testes realizados sobre as técnicas de árvores de regressão e redes neurais, teve como resultado que a técnica de redes neurais possui maior acurácia nos ajustes dos dados, tendo o erro médio das previsões como menor erro quando comparado a outra técnica. Pode-se observar também que a técnica de redes neurais teve o comportamento dos dados previstos mais semelhantes aos dados reais, um melhor comportamento segundo as previsões, Figura 17. Portanto conclui-se que a técnica de redes neurais se mostrou mais adequado, tendo a função de regressão que melhor se ajustou aos dados reais.

A análise estatística realizada sobre os dados gerados pelas técnicas de regressão já mencionadas, determinou que as técnicas não possuem diferença estatisticamente significativa estatística. Contudo, é possível que devido aos dados terem sido analisados como um todo, regiões onde a variação de deposição era maior não houve a análise detalhada. Como é possível observar na Figura 13, a partir da 5ª amostra (12.5 metros) os dados mantêm um mesmo comportamento, próximo a 0%, o que estaria ocultando o comportamento das amostras até a 5ª amostra, a qual possui maior variação, o que necessitaria de uma análise mais específica. Uma possível solução para essa situação seria a avaliação do comportamento das técnicas por regiões, locais onde as amostras apresentam maior variação.

Devido aos resultados obtidos é possível almejar novos trabalhos, tais como: (i) realizar tal estudo com base de dados maiores e com mais atributos; (ii) avaliar técnicas de aprendizado de máquina com vieses diferentes, como por exemplo: random forest, knn para regressão e naive bayes; (iii) realizar testes estatísticos orientados por região de deposição.

REFERÊNCIAS

- ANTUNIASSI, U. R.; JÚNIOR, C. Aplicação localizada de produtos fitossanitários. *Simpósio de Agricultura de Precisão, Viçosa*, p. 181–202, 2000.
- BAETENS, K. et al. The relative importance of environmental and field sprayer parameters for reducing drift: a cfd sensitivity study. *Aspects of Applied Biology*, AAB; 1999, v. 77, n. 2, p. 303, 2006.
- BERNARDI, A. d. C. et al. Agricultura de precisão: resultados de um novo olhar. *Embrapa Instrumentação-Livros técnicos (INFOTECA-E)*, Brasília, DF: Embrapa, 2014. 596 p., 2014.
- BINOTI, D. H. B.; BINOTI, M. L. M. da S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. *Revista Ciência da Madeira (Brazilian Journal of Wood Science)*, v. 5, n. 1, p. 10–12953, 2014.
- BONGIOVANNI, R.; LOWENBERG-DEBOER, J. Precision agriculture and sustainability. *Precision agriculture*, Springer, v. 5, n. 4, p. 359–387, 2004.
- BUENO, M. R. et al. Deriva e análise de risco das aplicações de fungicidas, herbicidas e inseticidas nas culturas do feijão e da soja. Universidade Federal de Uberlândia, 2015.
- COELHO, A. M. Agricultura de precisão: manejo da variabilidade espacial e temporal dos solos e culturas. *Embrapa Milho e Sorgo-Documentos (INFOTECA-E)*, Sete Lagoas: Embrapa Milho e Sorgo, 2005., 2005.
- CORÁ, J. et al. Variabilidade espacial de atributos do solo para adoção do sistema de agricultura de precisão na cultura de cana-de-açúcar. *Revista Brasileira de Ciência do Solo*, Sociedade Brasileira de Ciência do Solo, v. 28, n. 6, 2004.
- CUNHA, J. et al. Avaliação de estratégias para redução da deriva de agrotóxicos em pulverizações hidráulicas. *Planta Daninha*, SciELO Brasil, v. 21, n. 2, p. 325–332, 2003.
- CUNHA, J. P. A. R. da. Simulação da deriva de agrotóxicos em diferentes condições de pulverização. *Ciênc. Agrotec*, Lavras, v. 32, n. 5, p. 1616–1621, 2008.
- DALY, R. F. Agriculture: Projected demand, output and resource structure. *Implications of Changes (Structural and Market) on Farm Management and Marketing Research*, p. 82–119, 1967.
- DONKERSLEY, P.; NUYTENS, D. A meta analysis of spray drift sampling. *Crop Protection*, Elsevier, v. 30, n. 7, p. 931–936, 2011.
- FACELI, K. et al. *Inteligência Artificial: Uma abordagem de aprendizado de máquina*. [S.l.]: Rio de Janeiro : LTC, 2015. 378 p.
- FERNANDES, A. P. et al. Caracterização do perfil de deposição e do diâmetro de gotas e otimização do espaçamento entre bicos na barra de pulverização. *Engenharia Agrícola*, SciELO Brasil, p. 728–733, 2007.

- FERREIRA, M. L. P. C. A pulverização aérea de agrotóxicos no Brasil: cenário atual e desafios. *Revista de Direito Sanitário*, v. 15, n. 3, p. 18–45, 2015.
- FINOCCHIO, M. A. F. Noções de redes neurais artificiais. *Universidade Tecnológica Federal do Paraná - Campus Cornélio Procopio*, 2014.
- FRANCO, D. P. et al. Uma ferramenta computacional forense para verificação de autenticidade de assinaturas manuscritas através de processamento digital de imagens e redes neurais artificiais. *Anais do Computer on the Beach*, p. 121–130, 2015.
- GIL, Y.; SINFORT, C. Emission of pesticides to the air during sprayer application: A bibliographic review. *Atmospheric Environment*, Elsevier, v. 39, n. 28, p. 5183–5193, 2005.
- HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2001.
- HOFMAN, V.; SOLSENG, E. Reducing spray drift. 2009.
- INAMASU, R. Y.; BERNARDI, A. d. C. Agricultura de precisão. *Embrapa Instrumentação-Capítulo em livro técnico-científico (ALICE)*, In: BERNARDI, AC de C.; NAIME, J. de M.; RESENDE, AV de; BASSOI, LH; INAMASU, RY (Ed.). Agricultura de precisão: resultados de um novo olhar. Brasília, DF: Embrapa, 2014. p. 21-33., 2014.
- LAMPARELLI, R. *Agricultura de precisão*. 2016. Acessado em: 26/06/2017. Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_72_711200516719.html>.
- MANTOVANI, E.; QUEIROZ, D.; DIAS, G. Máquinas e operações utilizadas na agricultura de precisão. In: IN: CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA, 27., 1998, POÇOS DE CALDAS, MG. MECANIZAÇÃO E AGRICULTURA DE PRECISÃO. LAVRAS: UFLA/SBEA, 1998. P. 109-157. *Embrapa Milho e Sorgo-Artigo em anais de congresso (ALICE)*. [S.l.], 1998.
- MANZATTO, C.; BHERING, S.; SIMÕES, M. Agricultura de precisão: propostas e ações da Embrapa Solos. *EMBRAPA Solos*, 1999.
- MATUO, T. *Técnicas de aplicação de defensivos agrícolas*. [S.l.]: Funep, 1990.
- MCQUEEN, R. J. et al. Applying machine learning to agricultural data. *Computers and electronics in agriculture*, Elsevier, v. 12, n. 4, p. 275–293, 1995.
- MEDEIROS, L. d. *Redes neurais em delphi*. [S.l.: s.n.], 2003.
- MICHALSKI, R. S.; CARBONELL, J. G.; MITCHELL, T. M. *Machine learning: An artificial intelligence approach*. [S.l.]: Springer Science & Business Media, 2013.
- MITCHELL, T. M. *Machine Learning*. [S.l.]: McGraw-Hill Science/Engineering/Math, 1997. 432 p.
- MOLIN, J. P.; AMARAL, L. R. do; COLAÇO, A. *Agricultura de precisão*. [S.l.]: Oficina de Textos, 2015.

- MONARD, M. C.; BARANAUSKAS, J. A. *Conceitos sobre aprendizado de máquina*. [S.l.]: In: Rezende, S.O. *Sistemas Inteligentes-Fundamentos e Aplicações*. Barueri, SP, 2005. 86–114 p.
- PORTAL BRASIL. *Agronegócio deve ter crescimento de 2% em 2017*. 2016. Acessado em: 26/06/2017. Disponível em: <<http://www.brasil.gov.br/economia-e-emprego/2016/12/agronegocio-deve-ter-crescimento-de-2-em-2017>>.
- PORTAL BRASIL. *Mais de 67% das lavouras usam tecnologia na produção agrícola*. 2017. Acessado em: 26/06/2017. Disponível em: <<http://www.brasil.gov.br/infraestrutura/2017/04/mais-de-67-das-lavouras-usam-tecnologia-na-producao-agricola>>.
- SUMNER, P.; SUMNER, S. Comparison of new drift reduction nozzles. *St. Joseph: ASAE*, 1999.
- TEIXEIRA, J. *Inteligência Artificial*. [S.l.: s.n.], 2014.
- TEIXEIRA, J. C. Modernização da agricultura no brasil: impactos econômicos, sociais e ambientais. *Revista Eletrônica AGB-TL*, v. 1, n. 2, p. 21–42, 2005.
- TSCHIEDEL, M.; FERREIRA, F. M. Introdução à agricultura de precisão: Conceitos e vantagens. *Ciência Rural*, Universidade Federal de Santa Maria, v. 32, n. 1, p. 159–163, 2002.
- VOLL, E. Agricultura de precisão: Manejo de plantas daninhas. *ALUÍSIO BORÉM e outros. Agricultura de Precisão. Viçosa: Giúdice e Borém*, 2000.