UNIVERSIDADE FEDERAL DO PAMPA

LEONARDO GAUER SCHULTE

SUPORTE À DECISÃO EM PASTAGENS: ANÁLISE ESPAÇO-TEMPORAL E APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA DISPONIBILIDADE DE FORRAGEM NO CONTEXTO DE SMART FARMING

Bagé
2019

LEONARDO GAUER SCHULTE

SUPORTE À DECISÃO EM PASTAGENS: ANÁLISE ESPAÇO-TEMPORAL E APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA DISPONIBILIDADE DE FORRAGEM NO CONTEXTO DE SMART FARMING

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Orientador: Naylor Bastiani Perez
Coorientador: Leonardo Bidese de Pinho

Bagé
2019

Ficha catalográfica elaborada automaticamente com os dados fornecidos
pelo(a) autor(a) através do Módulo de Biblioteca do
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

S386s Schulte, Leonardo Gauer
Suporte à Decisão em Pastagens: Análise Espaço-temporal e Aprendizado de Máquina para Predição da Disponibilidade de Forragem no Contexto de Smart Farming / Leonardo Gauer Schulte.
103 p.

Dissertação(Mestrado)-- Universidade Federal do Pampa, MESTRADO EM COMPUTAÇÃO APLICADA, 2019.
"Orientação: Naylor Bastiani Perez".

1. Modelos computacionais. 2. Pecuária de precisão. 3. Ajuste de lotação. 4. Oferta de forragem. 5. Manejo de pastagem. I. Título.

LEONARDO GAUER SCHULTE

SUPORTE À DECISÃO EM PASTAGENS: ANÁLISE ESPAÇO-TEMPORAL E APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA DISPONIBILIDADE DE FORRAGEM NO CONTEXTO DE SMART FARMING

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Área de concentração: Tecnologias para a produção agropecuária.

Dissertação defendida e aprovada em: 01 de Março de 2019.

Banca examinadora:

Prof. Dr. Naylor Bastiani Perez
Orientador

Profª. Drª. Ana Paula Lüdtke Ferreira
UNIPAMPA

Prof. Dr. Gustavo Trentin
Embrapa Pecuária Sul

Profª. Drª. Teresa Cristina Moraes Genro
Embrapa Pecuária Sul

AGRADECIMENTO

Ao pesquisador Naylor Bastiani Perez e ao professor Leonardo Bidese de Pinho pelas orientações e discussões sobre o manejo das pastagens naturais sujeitas à infestação, a definição das questões de pesquisa, desenvolvimento da metodologia e análise dos resultados, sendo estas fundamentais na construção do trabalho.

Ao pesquisador Gustavo Trentin pela disponibilização dos dados da estação meteorológica e discussões acerca de aspectos de agrometeorologia aplicados ao escopo do trabalho, bem como à equipe do setor de campos experimentais da Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul pela execução dos experimentos.

À professora Ana Paula Lüdtke Ferreira pelas contribuições na qualificação do trabalho e na dissertação, além do desenvolvimento e auxílio nas correções no modelo de escrita do texto, à professora Teresa Cristina Moraes Genro pelas contribuições realizadas na dissertação e, ao professor Sandro da Silva Camargo pela contribuição na escolha e instalação da plataforma de exploração dos dados.

À Universidade Federal do Pampa e à Empresa Brasileira de Pesquisa Agropecuária por disponibilizarem a infraestrutura e equipamentos necessários ao desenvolvimentos dos experimentos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pelo apoio financeiro.

RESUMO

Na pecuária extensiva de corte é importante saber a quantidade de alimento disponível na pastagem para que se possa estimar a quantidade ideal de animais que devem ocupar aquela área. Na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul, a massa de forragem disponível em determinadas áreas experimentais é estimada por um método direto, onde são usadas gaiolas para isolar determinada área da pastagem e, a cada mês, a vegetação dessa área é cortada, seca e depois pesada. Dessa forma, obtém-se a massa seca da amostra coletada que é utilizada para a estimar a taxa de acúmulo de pasto para o mês posterior. Este trabalho propõe um modelo computacional baseado em Redes Neurais Artificiais *Long Short-Term Memory* com o intuito de melhorar a estimativa de disponibilidade de pasto feita atualmente na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul, com base em dados amostrais da disponibilidade de pasto, coletados por meio do método direto, e meteorológicos. A metodologia é baseada em pesquisa exploratória e experimental interdisciplinar, e seu desenvolvimento, realizado na forma de estudo de caso, apontou alguns problemas no processo de amostragem de disponibilidade de pasto que vinha sendo realizado. Buscando contornar os problemas identificados, foi definida uma arquitetura onde os dados amostrais de disponibilidade de pasto e meteorológicos, são persistidos em um banco de dados geográfico, após um processo de *Extract, Transform, and Load*. Depois de carregados no banco de dados geográfico, os dados passam por um pré-processamento que os prepara para o processo de treinamento da Redes Neural Artificial *Long Short-Term Memory* desenvolvida. Os resultados mostram que a metodologia proposta é capaz de estimar com significativa acurácia a disponibilidade de pasto instantânea e do mês posterior, uma vez que o erro médio verificado é de mesma magnitude do erro associado ao método de amostragem utilizado.

Palavras-chave: Modelos computacionais. Pecuária de precisão. Ajuste de lotação. Oferta de forragem. Manejo de pastagem.

ABSTRACT

In extensive herding, it is important to know the amount of food available in the pasture so that the ideal amount of animals that occupy that area can be estimated. In Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul, the forage mass available in certain experimental areas is estimated by a direct method, where cages are used to isolate a particular area of the pasture and, every month, the vegetation of this area is cut, dried and then weighed. Thus, the dry mass of the collected sample is obtained, which is used to estimate the pasture accumulation rate for the subsequent month. This work proposes a computational model based on Long Short-Term Memory Artificial Neural Networks with the aim of improving the estimation of pasture availability currently done at Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul, based on data from pasture availability, collected through the direct method, and meteorological. The methodology is based on exploratory and experimental interdisciplinary research, and its development, carried out in the form of a case study, pointed out some problems in the pasture availability sampling process that had been carried out. In order to circumvent the problems identified, an architecture was defined where the pasture availability and meteorological data are persisted in a geographic database, after an Extract, Transform, and Load process. After being loaded into the geographic database, the data goes through a pre-processing that prepares them for the training process of the Long Short-Term Memory Artificial Neural Network developed. The results show that the proposed methodology is capable of estimating with significant accuracy the availability of instantaneous pasture and the subsequent month, since the average error verified is of the same magnitude of the error associated to the sampling method used.

Keywords: Computational models. Precision farming. Stocking rate. Forage offer. Grazing management.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ANFIS	*Adaptative Network based Fuzzy Inference Systems*
ER	Entidade-Relacionamento
ETL	*Extract, Transform, and Load*
FMIS	*Farm Management Information Systems*
GMD	Ganho Médio Diário
IDC	*International Data Corporation*
IoT	*Internet of Things*
LSTM	*Long Short-Term Memory*
NOAA	*National Oceanic and Atmospheric Administration*
PBI	*Pasture Base Ireland*
PGSUS	*Pasture Growth Simulation Using Smalltalk*
RNA	Rede Neural Artificial
SAD	Sistema de Apoio à Decisão
SGBD	Sistema Gerenciador de Banco de Dados
SIG	Sistemas de Informação Geográficas
TIC	Tecnologias da Informação e da Comunicação

SUMÁRIO

1 INTRODUÇÃO

Em sistemas produtivos com foco em pecuária extensiva de corte, é importante saber a quantidade de alimento disponível na pastagem para que se possa estimar o número ideal de animais que devem ocupar aquela área (capacidade de suporte). No Brasil os principais problemas relacionados ao manejo de bovinos em pastagens ocorrem devido ao desperdício de forragem, causado pela falta de planejamento (SALMAN; SOARES; CANESIN, 2006). A alimentação representa de 50% a 80% dos custos dentro de um sistema de produção animal, sendo que para os ruminantes, cuja dieta é baseada em alimentos volumosos, o uso das gramíneas nas pastagens representa a forma mais econômica para alimentação animal. Portanto, a correta exploração da forrageira pode assegurar a alimentação adequada do rebanho e permitir a redução dos custos de produção (JAYARAMAN *et al.*, 2016).

O uso eficiente das forrageiras sob pastejo na alimentação animal tem uma relação direta com a produtividade do sistema. Para que possam expressar seu potencial genético, bovinos em pastejo devem ter acesso às pastagens com disponibilidade e valor nutritivo adequados, que lhe permitam consumir quantidades suficientes com boa qualidade. Isto ocorre porque os bovinos têm capacidade de selecionar sua própria dieta durante o pastejo, desde que haja condições na pastagem para isso, sendo essa seleção tanto para espécies de plantas quanto para partes das plantas (SALMAN; SOARES; CANESIN, 2006). Segundo Nabinger (2006), para otimizar o ganho médio diário (GMD) de massa dos animais, é necessário que esteja disponível, diariamente, uma quantidade de forragem igual a 13,5% do peso do animal, enquanto que para otimizar o ganho de massa por hectare (G/ha), essa relação é de 11,5%. Grande parte dos erros de manejo da pastagem está relacionada com o desconhecimento das várias inter-relações solo-planta-animal que impedem a obtenção de níveis de produção animal satisfatórios. A definição de critérios de avaliação para o estabelecimento de um programa de utilização e manejo da pastagem, no entanto, depende do conhecimento de parâmetros quantitativos e qualitativos da vegetação. A determinação precisa da quantidade de forragem disponível é importante para calcular a velocidade de crescimento da planta e a capacidade de suporte da pastagem para evitar desperdícios (SALMAN; SOARES; CANESIN, 2006).

A Empresa Brasileira de Pesquisa Agropecuária possui diversas unidades descentralizadas de pesquisas. Entre elas está a unidade de pesquisa ecorregional denominada Pecuária Sul, a qual desenvolve pesquisas em bovinocultura de corte e leite,

ovinocultura e forrageiras nos Campos Sul-brasileiros, compreendidos pelos Estados do Rio Grande do Sul, Santa Catarina e Paraná. Em particular, na Pecuária Sul são realizados estudos de longo prazo onde a forragem disponível é estimada por um método direto no qual são usadas gaiolas para isolar determinada área da pastagem e, a cada mês, a vegetação dessa área é cortada, seca e depois pesada. Dessa forma, obtém-se a massa de forragem da amostra coletada que é utilizada para a estimativa do crescimento da pastagem entre os períodos amostrados. A partir desta estimativa, é ajustada a taxa de lotação buscando manter uma quantidade de forragem igual a 13,5% do peso do animal. Nesse procedimento, assume-se que o crescimento do pasto no mês atual é igual ao crescimento do mês anterior. Contudo, essa estimativa é imprecisa, uma vez que não considera, de maneira objetiva, o efeito das variações meteorológicas que interferem na fisiologia das plantas e, por conseguinte, influenciam no crescimento da vegetação.

Assim, faz-se relevante aprimorar a estimativa de disponibilidade de massa de forragem em pastagens feita, atualmente, na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul com base em dados históricos da disponibilidade de massa de forragem, estimados por meio do método direto, associando-os aos dados meteorológicos (temperatura, umidade relativa do ar, precipitação, radiação solar, direção e velocidade do vento, soma térmica, balanço hídrico, etc) obtidos a partir da estação meteorológica de superfície automática, do Instituto Nacional de Meteorologia, instalada na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul, próxima à área experimental (INMET, 2019). Nesse contexto, existe significativo potencial para aplicação de técnicas de aprendizado de máquina, que visam predizer determinado comportamento com base no aprendizado sobre um conjunto de dados (MIRKIN, 2011; KAVAKIOTIS *et al.*, 2017).

Com base no conjunto de dados amostrado periodicamente pelos estudos realizados pela Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul, o que fornece uma representação espaço-temporal das áreas de pastagem estudadas ao longo de aproximadamente quatro anos (2014 até 2018), bem como dos dados meteorológicos diários correspontendes a este período, associado às técnicas de aprendizado de máquina, tais como Redes Neurais Artificiais (RNA), hipoteticamente seria possível embasar o desenvolvimento de um Sistema de Apoio à Decisão (SAD) qualificado para otimizar a ocupação da pastagem e proporcionar maior controle sobre a produtividade dos animais. Assim, o desenvolvimento de um modelo baseado em RNA para estimar a massa de forragem, baseando-se em dados especializados de áreas amostrais, pode tornar mais eficiente o processo de tomada de decisão do produtor rural que busque alternativas para

maximizar a lucratividade em um sistema de pecuária de corte com manejo extensivo.

1.1 Questões de Pesquisa

Dado o contexto do trabalho, surgem duas principais questões de pesquisa que motivam a sua realização, sendo elas:

- É possível construir um método baseado em técnicas de aprendizagem de máquina, a partir de séries temporais de dados amostrais da pastagem e meteorológicos, que seja capaz de estimar com maior acurácia que o método tradicional aplicado na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul, a massa de forragem ao longo do tempo?
- É possível adaptar o modelo para que, a partir de um conjunto de dados espaço-temporais acrescidos da previsão meteorológica para aproximadamente 30 dias, seja capaz de predizer a massa de forragem no período subsequente, com maior eficácia que o método tradicional, baseado na taxa de acúmulo do mês anterior?

1.2 Objetivos

O objetivo geral deste trabalho é construir um modelo computacional baseado em RNA, capaz de realizar uma estimativa eficaz da massa de forragem a partir de dados históricos do crescimento da pastagem coletados convencionalmente por meio do método direto de amostragem, associando-os aos dados meteorológicos: umidade relativa do ar, precipitação, radiação solar, velocidade do vento, soma térmica e balanço hídrico, com o intuito de otimizar a ocupação animal na pastagem, por meio da predição da massa de forragem para um período subsequente a fim de proporcionar maior eficiência do sistema produtivo.

São objetivos específicos deste trabalho:

- descrever os processos realizados atualmente na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul para a coleta de dados amostrais de disponibilidade de massa de forragem;
- identificar melhorias no processo de coleta de dados de pastagens realizado atualmente na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul;

- desenvolver um banco de dados georreferenciado como modelo de persistência para migrar os dados atualmente registrados em planilhas;
- desenvolver procedimentos para migrar os dados contidos nas planilhas para o banco de dados geográfico desenvolvido;
- identificar e aplicar técnicas capazes de melhorar a qualidade dos dados obtidos na experimentação com pastagens;
- identificar e aplicar técnicas de descoberta de conhecimento e aprendizado de máquina sobre os dados obtidos na experimentação com pastagens;
- validar o modelo desenvolvido.

1.3 Estrutura do Trabalho

O restante do trabalho está dividido em quatro capítulos. No Capítulo 2 é apresentado o referencial teórico relacionado às tecnologias de *Smart Farming*, necessárias para embasar a identificação, o projeto e a implementação do modelo desenvolvido. Além disso, são apresentadas as ferramentas utilizadas, bem como os desafios e benefícios desse tipo de projeto e os trabalhos correlatos. No Capítulo 3, é detalhado o desenvolvimento do trabalho, com base no estudo de caso realizado, seguido da metodologia proposta. No Capítulo 4 são expostos resultados, caracterizados como contribuições do trabalho. No Capítulo 5, são feitas as considerações finais, onde são analisados os resultados obtidos e apresentadas as expectativas para trabalhos futuros. Por fim, são apresentadas as referências do trabalho e os apêndices, que apresenta os programas desenvolvido em Python, para o processo de ETL, pré-processamento, modelagem da rede neural e a geração de gráficos para a visualização dos resultados.

2 REFERENCIAL TEÓRICO

Neste capítulo são apresentados, com base em revisão de literatura, a contextualização do trabalho seguida de alguns conceitos relacionados ao tema de pesquisa, tais como, *Smart Farming*, *Internet of Things*, *Big Data*, coleta de dados, qualidade de dados, sistemas de apoio a decisão, etc. Esses conceitos servirão como base para a escolha do modelo computacional utilizado no projeto da solução para a discussão dos resultados.

A análise bibliográfica seguiu uma abordagem sistemática adaptada da proposta por Jesus, Resende e Zambalde (2013), e envolveu três etapas: coleta de trabalhos relacionados, filtragem dos trabalhos relevantes e revisão e análise detalhada dos trabalhos relacionados ao estado da arte. Na primeira etapa, uma busca por palavras-chave tais como, estimativa de forragem, *Farm Management Information System* (FMIS), ajuste de taxa de lotação, pecuária de precisão, predição de series temporais, Redes Neurais Artificiais (RNA) *Long Short-Term Memory* (LSTM), etc. e seus termos correspondentes em inglês, em livros e artigos científicos foi realizada a partir das bases de dados *ACM Digital Library*, *Multidisciplinary Digital Publishing Institute* (MDPI), *IEEE Xplore* e *ScienceDirect*, bem como dos serviços de indexação *Web of Science* e *Google Scholar*. A busca de artigos foi restrita aos idiomas Português e Inglês. Na segunda etapa, foi feita uma leitura superficial com ênfase no resumo, introdução e conclusão com o intuito de verificar o objetivo do trabalho e os resultados alcançados pelo mesmo. Na etapa final, os artigos filtrados com base no critério de maior relevância para o tema do trabalho foram analisados em sua totalidade, considerando o problema que abordaram, fontes de dados empregadas, solução proposta, impacto atingido (se mensurável), ferramentas, sistemas, algoritmos e metodologias utilizadas (KAMILARIS; KARTAKOULLIS; PRENAFETA-BOLDÚ, 2017).

2.1 Contextualização

Estima-se que, até 2050, a população mundial tenha um aumento superior a um terço em relação à quantidade registrada em 2010, sendo que essa explosão populacional será vista principalmente nos países em desenvolvimento. Esse cenário previsto implica também um aumento proporcional na demanda por recursos, principalmente alimentos (MARIMUTHU *et al.*, 2017). Segundo a Organização das Nações Unidas

para Agricultura e Alimentação, será necessário um aumento de 70% na produção agrícola total, associado a um incremento de 20% no consumo de água (FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS - FAO, 2019).

Atualmente, os países cuja economia se baseia predominantemente nas práticas agrícolas e pecuárias são classificados como países subdesenvolvidos. O termo "subdesenvolvido" surgiu após a segunda guerra mundial para explicar a diferença entre os países mais ricos e industrializados, classificados como desenvolvidos, e aqueles mais pobres, geralmente exportadores de matéria-prima, classificados como subdesenvolvidos. Justamente pelo fato de serem mais pobres, esses países costumam ter menor acesso às tecnologias modernas e, por isso, empregam na área agrícola e pecuária técnicas ditas convencionais, tais como os métodos tradicionais de distribuição manual de sementes e aragem, padrão de duas safras por ano, sistemas não científicos de cultivo, além de sofrerem com a irregularidade da disponibilidade de água ao longo do ano. A combinação desses fatores leva a um rendimento inadequado e baixa produtividade (O'GRADY; O'HARE, 2017).

A ampliação do uso de métodos científicos na agricultura pode trazer mudanças na produtividade das culturas, pois tende a aumentar a eficiência das técnicas agrícolas aplicadas na gestão dos sistemas produtivos. Neste sentido, Wolfert *et al.* (2017) destacam o desenvolvimento de um novo modelo, denominado *Smart Farming*, no qual é enfatizado o uso das tecnologias da informação e comunicação (TIC) no processo de gerenciamento da fazenda. Como exemplo, a adoção de modelos e tecnologias computacionais contemporâneas, como a *Internet of Things* (IoT) e *Cloud Computing*, tendem a alavancar esse desenvolvimento e facilitar a introdução de mais robôs e da inteligência artificial na agricultura. Esses conceitos são potencializados pelo fenômeno de *Big Data*, em que, grandes volumes de dados com uma ampla variedade, podem ser periodicamente capturados, analisados e usados para melhorar os processos de tomada de decisão.

Iniciativas que empregam redes de sensores sem fio na coleta de dados com uso de diversos tipos de sensores introduzidos no campo e o envio dos dados para um servidor central principal são, cada vez mais, observadas na literatura, com foco em diferentes culturas. No entanto, apenas monitorar os fatores ambientais não permite aumentar a produtividade das culturas, já que outros fatores têm um papel relevante, como por exemplo, a necessidade de pulverização de inseticidas e pesticidas para evitar a invasão de pragas, a demanda por monitoramento contínuo dos campos no intuito de identificar

ataques de animais e roubos em rebanhos e plantações, etc. (AMANDEEP *et al.*, 2017).

A Pecuária de Precisão ou *Precision Livestock Farming* (PLF) é definida por Wathes *et al.* (2008) como o "gerenciamento da produção de gado por meio da engenharia e suas tecnologias". Ainda, com relação à Pecuária de Precisão, Carvalho *et al.* (2009) trazem que a principal atividade relacionada a esta diz respeito às tecnologias de monitoramento animal em ambiente pastoril. Isso é feito principalmente por meio de sensores capazes de captar dados de localização, peso, alimentação, temperatura, inclinação, umidade, entre outros, referentes ao animal e ao meio onde o mesmo está inserido. Esse tipo de prática proporciona maior entendimento do relacionamento do animal com plantas e solo, além de possibilitar a descoberta de fatores que podem influenciar no bem-estar animal, redução nos impactos ambientais e aumento na produção (LACA, 2009).

O monitoramento periódico das atividades pecuárias, em um sistema de Pecuária de Precisão, gera grandes quantidades de dados. Contudo, a posse de dados por si só não implica em uma gestão qualificada, pois dados brutos não agregam conhecimento aos gestores. Para Primak (2008), a informação é a base para a construção do conhecimento. Portanto, a informação não é conhecimento, mas seu componente. Sendo assim, ainda mais importante que a coleta de dados em si, é dar significado a eles, isto é, ser capaz de extrair informação do conjunto de dados armazenado.

2.2 Tecnologias da Informação e da Comunicação na agricultura

As Tecnologias da Informação e da Comunicação (TIC) têm contribuído para diversas áreas de conhecimento, à medida que permitem o armazenamento e processamento de grandes volumes de dados, automatização de processos e o intercâmbio de informações e de conhecimento. Na agricultura, essa contribuição tem crescido nos últimos anos, inclusive dando origem ao termo: Tecnologias da Informação e da Comunicação na Agricultura (AgroTIC), encontrado na literatura. Massruhá, Leite e Moura (2014) definem AgroTIC como:

> Um conjunto de aplicações específicas para agricultura que utilizam ferramentas baseadas em TIC, tais como sistemas de informação geográfica (SIG), sistemas baseados em conhecimento, sistemas de suporte à decisão e modelos que são incorporados em novas tecnologias empregadas no campo.

Algumas aplicações no campo envolvem agricultura de precisão e o desenvolvimento de automação de rede de sensores para monitoramento de variáveis

diversas, que geram um grande volume de dados, o que vem sendo tratado na literatura como *Big Data*, além de sistemas inteligentes que aplicam técnicas de mineração de dados e inteligência computacional visando identificar padrões e gerar conhecimentos para uso do setor agrícola (MASSRUHÁ; LEITE; MOURA, 2014). Nesse sentido, outro conceito semelhante ao AgroTIC que vem sendo apresentado na literatura é o *Smart Farming* tratado na Seção 2.2.1.

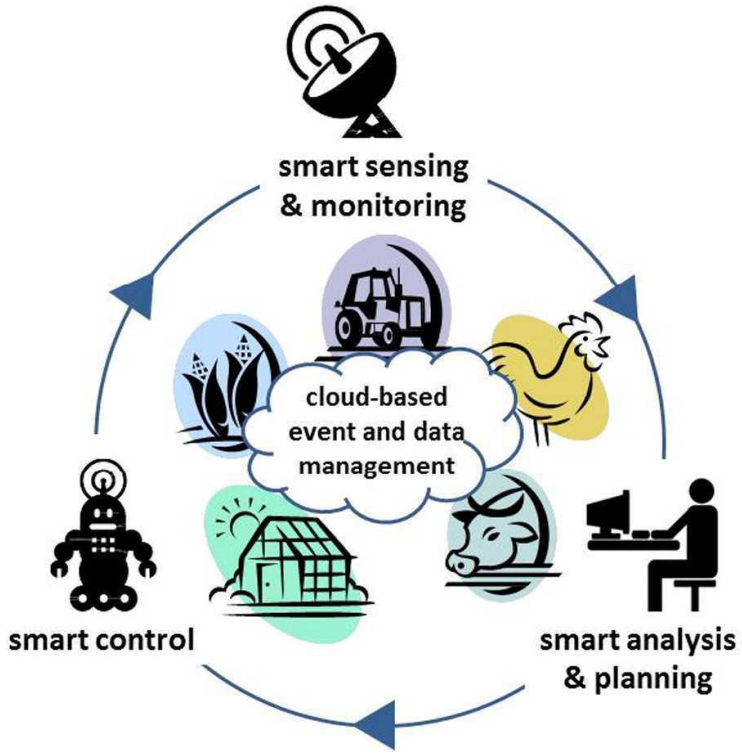
2.2.1 *Smart Farming*

Segundo Pivoto *et al.* (2018), o desenvolvimento tecnológico e o uso crescente de sistemas eletrônicos e transmissão de dados, introduziu mudanças radicais no ambiente de trabalho agrícola nos últimos anos. Essas mudanças exigem informações atualizadas dos sistemas de produção e dos mercados e agentes envolvidos na produção para fornecer informações de tomada de decisão para a produção, bem como para as questões estratégicas e gerenciais envolvidas.

De acordo com Wolfert *et al.* (2017), enquanto a Agricultura de Precisão está levando em conta a variabilidade do campo, o *Smart Farming* vai além, baseando as tarefas de gerenciamento não apenas no local, mas também nos dados, aprimorados pelo conhecimento do contexto e da situação, desencadeados por eventos em tempo real. O *Smart Farming* baseia-se na incorporação de tecnologias de informação e comunicação em máquinas, equipamentos e sensores em sistemas de produção agrícola, permite que um grande volume de dados e informações seja gerado com a inserção progressiva da automação no processo. A agricultura inteligente depende da transmissão de dados e da concentração de dados em sistemas de armazenamento remoto para permitir a combinação e análise de vários dados agrícolas para a tomada de decisões (PIVOTO *et al.*, 2018). A Figura 1 representa o conceito de *Smart Farming* ao longo do ciclo de gerenciamento do sistema, o que significa que os dispositivos inteligentes, conectados à internet, estão controlando o sistema agrícola. Os dispositivos inteligentes aprimoram as ferramentas convencionais, adicionando reconhecimento de contexto autônomo por todos os tipos de sensores e incorpora inteligência ao sistema, tornando-o capaz de executar ações autônomas de forma remota.

O *Smart Farming* herda diversos conceitos de *Internet of Things*, *Cloud Computing* e *Big Data*. As Seções 2.2.2, 2.2.3 e 2.2.5 abordam respectivamente, essas tecnologias.

Figura 1 – O ciclo de gerenciamento do Smart Farming



Fonte: Wolfert *et al.* (2017)

2.2.2 *Internet of Things*

De acordo com Lohan e Singh (2017), a Internet das Coisas (*Internet of Things* - IoT) representa um conceito geral de interconexão de dispositivos que contém sensores, softwares, atuadores e conectividade de rede. A IoT permite que os dispositivos coletem e transfiram informações pela rede, sendo uma combinação de duas palavras, Internet e coisas. As "coisas" na IoT têm a capacidade de enviar e receber informações para uma rede sem a intervenção de seres humanos e contém objetos físicos que possuem um identificador particular com um sistema embarcado.

Recentemente, o número de dispositivos conectados à Internet tem aumentado significativamente. Segundo Saadeh, Almobaideen e Sabri (2017), espera-se que até 2020 mais de 50 bilhões de dispositivos estejam conectados à internet. A definição clássica de Internet, tratada como rede mundial de computadores, não é mais uma descrição clara do que temos hoje. Tipos variados de hardware independentes e embarcados estão se tornando os componentes dominantes, sendo que os computadores não são mais os componentes principais. Portanto, a Internet está se movendo em direção à Internet das coisas.

Para Li, Xu e Zhao (2018), a evolução das redes móveis de quinta geração (5G) é um dos principais impulsores do crescimento de aplicativos baseados em IoT. Segundo relatório da International Data Corporation (IDC), os serviços globais 5G farão com que 70% das empresas invistam em soluções de gerenciamento de conectividade.

A Internet das Coisas promete um novo paradigma tecnológico, conectando qualquer coisa e qualquer um a qualquer hora e em qualquer lugar, usando qualquer caminho e qualquer serviço. A IoT traz uma visão de "mundo inteligente" equipado com tecnologias de detecção e componentes inteligentes. Espera-se que a IoT tenha um impacto significativo sobre indivíduos, empresas e políticas, à medida que os modelos sociais e de negócios forem desafiados e os novos serviços introduzidos. Por outro lado, a IoT não é isenta de desafios e ressalvas, tais como a quantidade de dados gerados e preocupações que envolvem a invasão de privacidade em um mundo totalmente conectado (LU; PAPAGIANNIDIS; ALAMANOS, 2018).

2.2.3 *Cloud Computing*

Segundo Senyo, Addae e Boateng (2018), o fenômeno da computação em nuvem (*Cloud Computing*) tem sua origem nas tecnologias de sistemas paralelos e distribuídos, virtualização, *chips multicore* e tecnologias da Internet. Os recursos que distinguem a *Cloud Computing* das tecnologias relacionadas são o autoatendimento, disponibilização de recursos sob demanda, amplo acesso à rede, o *pool* de recursos, a rápida elasticidade e o serviço medido.

Atualmente não existe uma definição padrão de *Cloud Computing*, mas tanto os acadêmicos quanto os participantes do setor estão dando passos significativos para uma definição padrão (SENYO; ADDAE; BOATENG, 2018). Uma tentativa de definir *Cloud Computing* foi feita por Buyya *et al.* (2009), como um sistema de computação paralelo e distribuído que consiste em uma coleção de computadores interconectados e virtualizados que são provisionados dinamicamente e apresentados como um ou mais recursos de computação unificada com base em acordos de nível de serviço estabelecidos por meio de negociação entre o provedor de serviços e os consumidores .

O Instituto Nacional de Padrões e Tecnologia dos Estados Unidos (NIST), define que *Cloud Computing* é um modelo para permitir acesso onipresente e conveniente através da rede a um conjunto compartilhado de recursos computacionais configuráveis, por exemplo, redes, servidores, armazenamento, aplicativos e serviços, que podem ser

rapidamente provisionados e liberados com o mínimo esforço de gerenciamento ou interação com fornecedores de serviços.

Para Stergiou *et al.* (2018), a Internet das Coisas é uma rede de objetos físicos que unem software, eletrônica, sensores e conectividade sendo que para diversas aplicações os dispositivos possuem restrições de custos, consumo de energia, etc. assim, processamentos intensos e armazenamento em massa, por vezes precisam ser realizados fora dos dispositivos, o que motiva a combinação da tecnologia de *Cloud Computing* com a *Internet of Things*. Além disso, vem sendo discutido um novo paradigma, chamado *Fog Computing*, capaz de reduzir o volume de dados transmitidos na rede e a complexidade computacional necessária na nuvem.

2.2.4 *Fog Computing*

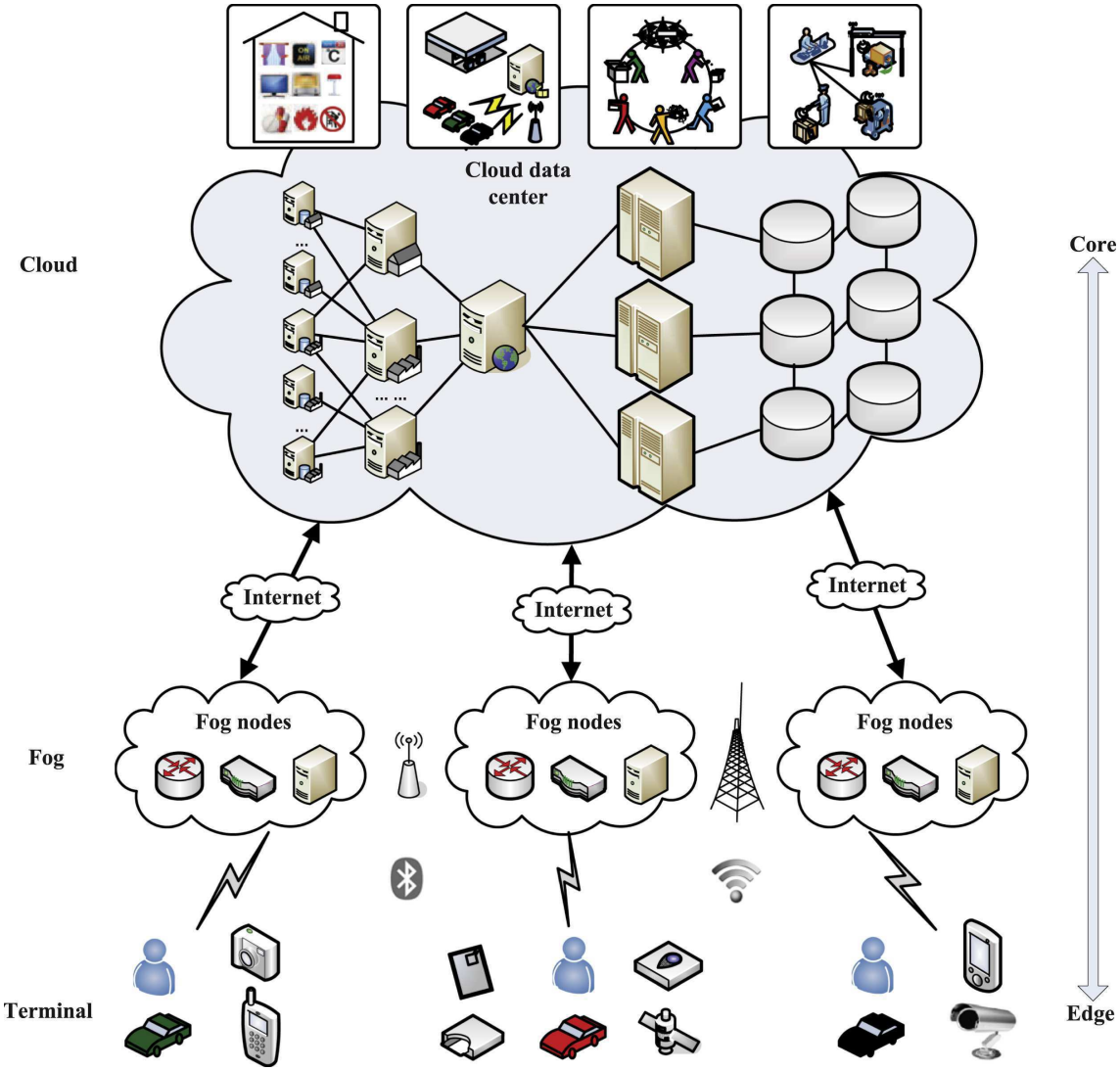
A medida que cresce o número de objetos que se conectam e trocam informações em uma estrutura de IoT, concentrar a predominância do armazenamento e processamento em uma estrutura centralizada como a *Cloud Computing* traz alguns problemas referentes a latência, consumo de energia, largura de banda e segurança, por exemplo, que se tornam limitantes para algumas aplicações (LAURENT *et al.*, 2018).

De acordo com Bonomi *et al.* (2012), o *Fog Computing* é uma extensão da *Cloud Computing*, onde parte dos recursos de armazenamento de dados e poder de processamento são transferidos para dispositivos periféricos, permitindo que os dados sejam pré processados e apenas informações significativas sejam comunicadas à uma estrutura central na nuvem. A *Fog Computing* possui uma topologia hierárquica, fazendo com que os dados dos usuários ou dispositivos finais sejam processados em vários níveis, cada nível realizando filtragem e análise de dados específicas, consolidando os dados repassados ao próximo nível (TSENG; LIN, 2018).

Segundo Hu *et al.* (2017), a arquitetura básica da *Fog Computing* consiste em três camadas, onde parte dos serviços da nuvem são transferido para a borda da rede, introduzindo a camada de *fog* entre os dispositivos finais e a nuvem. A Figura 2 representa a arquitetura hierárquica da *Fog Computing*. A camada terminal é a mais próxima do usuário final e do ambiente físico, ela consiste em vários dispositivos tais como, sensores, telefones celulares, veículos inteligentes, cartões inteligentes, leitores, etc. A camada *fog* está localizada na borda da rede e é composta de um grande número de nodos que geralmente são: roteadores, gateways, comutadores, pontos de acesso, estações base,

servidores de *fog* específicos, etc. que são amplamente distribuídos entre os dispositivos finais e a nuvem. A camada *cloud* consiste em vários servidores de alto desempenho e dispositivos de armazenamento que suportam análise e armazenamento permanentemente de grandes conjuntos de dados.

Figura 2 – Arquitetura hierárquica da Fog Computing



Fonte: Hu *et al.* (2017)

Além dos desafios referentes à interconexão e divisão das tarefas de armazenamento e processamento, à medida que cresce o número de dispositivos e volume de dados surgem também outras questões que são tratadas pelo *Big Data* e apresentadas na Seção 2.2.5.

2.2.5 *Big Data*

Segundo Dasoriya (2017), *Big Data* se refere aos conjuntos de dados extremamente grandes que são coletados periodicamente e analisados computacionalmente. Nesse sentido, surge o *Big Data Analytics*, que utiliza análises e técnicas avançadas para examinar dados brutos em busca de informação que possa gerar conhecimento (GUNTHER *et al.*, 2017). Algumas características do *Big Data* são:

- volume - Enorme quantidade de dados, na escala de Exabytes (10^{18} Bytes), coletados e armazenados de forma distribuída, disponível para processamento e extração de conhecimento.
- variedade - Diferentes tipos de conteúdo podem ser armazenados.
- veracidade - Refere-se à precisão e à validade dos dados. A análise de dados é útil se o conteúdo for válido. Em vez disso, uma grande quantidade de dados que podem não ser válidos levará a falsas interpretações.
- valor - Extração de valor em um tempo razoável. Se demorar mais do que o habitual, não poderá ser usado em tempo real.
- velocidade - Velocidade de geração de dados e a taxa de mudança. O resultado deve ser gerado e extraído rapidamente para que possa ser usado em sistemas de tempo real e nenhum armazenamento seja necessário.

De acordo com Kamilaris, Kartakoullis e Prenafeta-Boldú (2017), o *Big Data* agrícola cria a necessidade de investimentos em infraestruturas para armazenamento e processamento de dados, que precisam operar quase em tempo real para algumas aplicações, como por exemplo, previsão do tempo, monitorização das pragas das culturas e das doenças dos animais. Com base em *Big Data Analytics*, agricultores e organizações relacionadas podem extrair valor econômico de grandes volumes de uma ampla variedade de dados, permitindo a captura, descoberta e análise em alta velocidade.

Para Wolfert *et al.* (2017), as tecnologias de *Big Data* têm papel essencial no desenvolvimento do *Smart Farming* visto que as máquinas são equipadas com vários tipos de sensores que medem dados do ambiente e podem ser usados para realimentar o comportamento das máquinas. Isso ainda poderia ser combinado com outras fontes no *Big Data*, como dados meteorológicos, de mercado ou *benchmarks* de outras fazendas. Estes dados após processados podem auxiliar o processo de tomada de decisão em um sistema denominado *Farm Management Information System*, que é abordado na Seção 2.3.

2.3 *Farm Management Information Systems*

Os *Farm Management Information Systems* (FMIS) são ferramentas eletrônicas para coleta e processamento de dados com o objetivo de fornecer informações de valor potencial para a tomada de decisões gerenciais no contexto de fazendas (FOUNTAS *et al.*, 2015). De acordo com Novkovic *et al.* (2015), a gestão de qualidade de fazendas é uma tarefa importante e desafiadora, sendo três os fatores principais que justificam esta afirmativa:

- ambiente complexo: à medida que não se restringe mais apenas ao abastecimento da sociedade com produtos alimentares baratos e suficientes. Atualmente, existe a expectativa de conformidade com diversas legislações que envolvem requisitos de bem-estar animal, diretrizes mais rigorosas para o uso de agroquímicos visando segurança alimentar e também preocupações ambientais, etc.
- estruturas agrícolas complexas: em geral, a partir da década de 1970, o número total de fazendas diminuiu enquanto a área cultivada não mudou substancialmente. Com isso, as fazendas restantes se tornaram maiores e, à medida que buscaram se beneficiar de economias de escala, elas se tornaram mais difíceis de gerenciar.
- introdução de tecnologias modernas no setor agrícola: prática que contribui com o gerenciamento qualificado de fazendas, à medida que incorpora o uso de *softwares* financeiros e de planejamento de tarefas para o cultivo da terra, além da introdução de tratores e implementos "inteligentes" com GPS para modelagem de paisagem e outras tecnologias de ponta que proporcionam uma análise especial de diversos processos, dentre outros benefícios.

Dado este cenário, Nikkilä, Seilonen e Koskinen (2010) trazem que um sistema que auxilie a gestão da fazenda é necessário para possibilitar o desenvolvimento de agricultura e pecuária de precisão. O sistema deve ser capaz de armazenar dados gerados pelos sensores de implementos agrícolas durante sua operação e utilizar os dados para indicar os planos de operação. Para a geração desses planos, são necessárias quantidades consideráveis de dados e modelos computacionais de descoberta de conhecimento sobre bases de dados. Em geral, o sistema também deve ser capaz de lidar com vários formatos de dados. As Subseções 2.3.1, 2.3.2 e 2.3.3 apresentam conceitos e ferramentas relevantes no sentido da coleta e processamento de dados para tomada de decisões gerenciais.

2.3.1 Persistência de dados

O termo "persistência" no contexto de dados conota um objeto resiliente, concorrentemente compartilhado, o que remete diretamente a um Sistema de Gerenciamento de Banco de Dados (SGBD). A função de um SGBD é permitir o acesso e a atualização simultânea de bancos de dados persistentes, a fim de garantir a persistência dos dados a longo prazo, utilizando-se de várias estratégias de recuperação em caso de falhas na transação, no sistema ou no meio. De acordo com Date (2004), um banco de dados é um sistema computadorizado de armazenamento de dados em forma de registros. Os usuários deste sistema podem executar operações de busca, inserção, modificação e remoção, sobre os registros.

Em meados da década de 1960, todas as aplicações que precisavam guardar dados o faziam por meio de arquivos. Porém, à medida que a quantidade e a complexidade dos dados aumentava, isso começou a se tornar problemático, pois era necessário um grande número de pessoas para desenvolver as aplicações e as manutenções e atualizações se tornavam extremamente custosas.

Bancos de dados se tornaram componentes do cotidiano da sociedade moderna. Na era da informática, da comunicação digital, em que dados podem ser compartilhados de um extremo a outro do planeta em segundos, a informação tornou-se o ativo mais importante das empresas. É comum encontrar-se bancos de dados com centenas de tabelas interrelacionadas com alguns milhões de registros e, por isso, o armazenamento precisa ser feito de forma eficiente.

Os SGBD, são coleções de *softwares* que permitem aos usuários criarem e manterem um banco de dados, facilitando o processo de definição, construção, manipulação e compartilhamento de dados entre vários usuários e aplicações. A construção é o processo de armazenar os dados em alguma mídia apropriada controlada pelo SGBD. A manipulação é feita pelas funções *select*, *insert*, *update* e *delete*, capazes de realizar respectivamente, consultas, inserções, atualizações e exclusões de registros no banco. O compartilhamento permite aos múltiplos usuários e programas acessarem de forma concorrente o banco de dados (ELMASRI, 2003).

De acordo com Date (2004), os SGBD implementam alguns conceitos que facilitam a manipulação e compartilhamento de dados, fazendo com que as aplicações possam ser simplificadas, a medida que podem aproveitar características tais como:

- redundância pode ser reduzida: em sistemas sem bancos de dados cada aplicação

precisa ter os seus próprios arquivos. Esse fato pode levar a uma considerável redundância de dados, e representar considerável desperdício de armazenamento;

- inconsistências podem ser evitadas: se houver redundância não controlada, duas aplicações que utilizam um mesmo arquivo, em determinado momento, podem estar utilizando versões diferentes deste mesmo;
- suporte a transações: uma transação é uma unidade lógica de trabalho que envolve operações de atualização em bancos de dados. Um exemplo típico envolve a transferência de um determinado valor de uma conta bancária A para uma conta bancária B. É claro que, para que essa transação ocorra corretamente, duas atualizações são necessárias. Assim, o suporte a transações garante que ambas sejam realizadas, ou então nenhuma delas, ainda que o sistema venha a falhar (por uma falta de energia por exemplo) em meio ao processo;
- integridade: mesmo quando não há redundâncias, ainda podem haver informações incorretas. Por um erro comum de digitação, o banco poderia mostrar que um empregado trabalhou 400 h em uma semana, o que claramente é impossível. Para esse tipo de problemas é possível implementar restrições de integridade a serem executadas sempre que ocorrer uma operação de atualização;
- segurança: sob a orientação apropriada do administrador, o SGBD pode assegurar que o acesso ao banco se dê através de canais apropriados, e pode definir restrições de segurança a serem verificadas sempre que houver tentativa de acesso a determinados dados. Podem ser definidas, inclusive, restrições diferentes, dependendo do tipo de acesso e do usuário que está tentando acessar;

Existe uma relação entre o compartilhamento e a persistência simultâneos em banco de dados: as atualizações de transação devem persistir, mas, como o banco de dados persistente é ao mesmo tempo acessado e atualizado, o sistema de gerenciamento de banco de dados deve preocupar-se com a coerência dos objetos de dados persistentes. Isso normalmente é obtido por meio de estratégias de controle e recuperação concorrentes (DATE, 2004).

2.3.1.1 Bancos de dados espaço-temporais

No contexto deste trabalho, faz-se necessária ainda a aplicação de uma forma particular de banco de dados devido a natureza espaço-temporal do problema. Segundo Casanova *et al.* (2005), a maioria das aplicações voltadas à geoinformação utilizam

representações estáticas de fenômenos espaciais. Porém, vários fenômenos espaciais são dinâmicos sendo que as representações estáticas não os representam de forma adequada. Deste modo, um dos grandes desafios no desenvolvimento de modelos espaço-temporais, é que estes sejam capazes de representar adequadamente fenômenos que variam tanto no espaço como no tempo.

Atualmente os principais SGBD baseados no modelo relacional suportam dados geométricos para representação de pontos, linhas, polígonos, etc. e as operações referentes a estes elementos, porém, possuem limitações quanto ao desempenho que comprometem sua utilização para determinadas aplicações (QUEIROZ; MONTEIRO; CÂMARA, 2013). A construção de modelos espaço-temporais eficientes envolve duas etapas principais: escolha de conceitos adequados de espaço e de tempo e a construção de representações computacionais apropriadas correspondentes a esses conceitos. Podemos modelar a superfície da terra usando objetos correspondentes ao solo, ou usando campos que indiquem a variação espacial dos elementos da mesma área (CASANOVA *et al.*, 2005).

Para produzir um sistema de informação espaço-temporal que represente o mundo real, precisam ser levadas em conta questões referentes às regras aplicáveis, ao comportamento dos objetos ao longo do tempo, à interpretação da variação do tempo, à natureza das mudanças e à influência dos processos de medida. Acompanhando a necessidade de maior eficiência na representação de elementos espaço-temporais e gerenciamento e processamento de grandes volumes de dados, vêm sendo desenvolvidas novas abordagens que lidam com a espacialização de um grande número de usuários e com o tratamento de dados coletados em sensores de satélites e GPS, por exemplo. Esses sistemas visam aumentar escalabilidade e melhorar o desempenho em aplicações específicas.

2.3.1.2 Qualidade de dados

Desde o final da década de 1960 existem trabalhos motivados a estudar a qualidade de dados. Os estatísticos foram os pioneiros neste assunto ao proporem teorias matemáticas para considerar duplicatas em conjuntos de dados. No início da década de 1980, pesquisadores da área de gestão buscaram controlar os sistemas de manufatura de dados para detectar e eliminar problemas de qualidade de dados. Somente no início dos anos 1990 os cientistas da computação começaram a considerar o problema de definir, medir e melhorar a qualidade dos dados eletrônicos armazenados em bancos de dados,

Data Warehouses e sistemas legados (FAGUNDES; MACEDO; FREUND, 2017).

As questões de qualidade dos dados têm consequências na eficiência e eficácia das organizações e empresas. Relatórios sobre a qualidade dos dados do Data Warehousing Institute estimam que os problemas de qualidade de dados custam às empresas norte-americanas mais de 600 bilhões de dólares por ano. As conclusões do relatório foram baseadas em entrevistas com especialistas do setor, clientes e dados de pesquisas de 647 entrevistados (BATINI; SCANNAPIECO, 2016).

Qualidade de dados pode ser vista como um conjunto de processos que abrangem desde a identificação de problemas com os dados e sua classificação até a correção dos valores e o posterior monitoramento. Esses processos que visam garantir que os dados armazenados sejam corretos, precisos, consistentes, completos, integrados, aderentes às regras de negócio e aos domínios estabelecidos. Qualquer uma dessas facetas pode ser referida como uma variável ou, mais comumente, uma dimensão da qualidade dos dados. Em um sentido mais amplo, a qualidade dos dados também deve ser capaz de indicar os graus em que cada uma das dimensões está de acordo com os requisitos específicos dos usuários de dados em um determinado contexto (FU; EASTON, 2017). Dimensões como precisão, completude, consistência, são fundamentais na definição e medição da qualidade dos dados. Porém, a qualidade dos dados geralmente é avaliada em relação a um determinado requisito, sendo que nos últimos 20 anos, as pesquisas em qualidade de dados basearam-se no princípio fundamental de adequação ao uso (SADIQ; INDULSKA, 2017). Dessa forma, os requisitos de qualidade de dados são determinados de cima para baixo seguindo princípios de uso bem compreendidos e aplicados usando boas práticas de governança de dados. De acordo com Batini *et al.* (2009), as metodologias que visam assegurar a qualidade dos dados seguem três fases principais:

- reconstrução do estado: visa coletar informações contextuais sobre processos e serviços organizacionais, coleta de dados e procedimentos de gestão relacionados, questões de qualidade e custos correspondentes;
- avaliação/medição: mede a qualidade das coletas de dados ao longo de dimensões de qualidade relevantes. O termo medição é usado para abordar a questão da medição do valor de um conjunto de dimensões de qualidade de dados. O termo avaliação é utilizado quando tais medidas são comparadas com valores de referência, a fim de possibilitar um diagnóstico de qualidade.
- melhoria: diz respeito à seleção das etapas, estratégias e técnicas para alcançar novas metas de qualidade de dados.

De modo geral, o gerenciamento de qualidade de dados está focado na avaliação de conjuntos de dados e na aplicação de ações corretivas para garantir que os conjuntos de dados sejam adequados aos objetivos para os quais foram originalmente destinados. Em outras palavras, os princípios de qualidade de dados são respeitados quando os dados são úteis e apropriados para a análise dos processos que eles representam (MERINO *et al.*, 2016). Sendo assim, definidas e respeitadas as dimensões de qualidade de dados, estes passam a ser um ativo para o processo de descoberta de conhecimento cujos princípios são descritos na Seção 2.3.2.

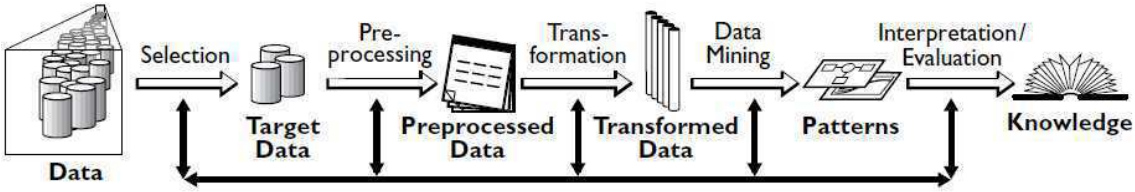
2.3.2 Descoberta de conhecimento em bases de dados

De acordo com Fayyad (1996), Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* - KDD) é um processo não trivial de extração de informações implícitas, previamente desconhecidas e "potencialmente úteis", a partir dos dados armazenados em um banco de dados. Tratando como não trivial, o autor torna clara a necessidade de uso de alguma técnica de processamento de dados. O termo "previamente desconhecidos" aponta que a informação deve ser nova tanto para o sistema quanto para o usuário. E, por fim, "potencialmente úteis" se refere ao fato que a informação deve trazer consigo algum benefício, de forma a possibilitar que o usuário gere conhecimento a partir dela.

O processo de KDD é interativo e iterativo. Interativo, pois o usuário pode intervir e controlar o curso das atividades. Iterativo, por ser uma sequência finita de operações onde o resultado de cada uma é dependente dos resultados das que a precedem (FAYYAD, 1996). Dessa forma, existe flexibilidade para que o usuário retorne etapas no fluxo de desenvolvimento a fim de realizar ajustes nas etapas anteriores que proporcionem melhores resultados As principais etapas do processo de KDD, ilustradas na Figura 3, são:

- domínio da aplicação (*Data*): aquisição de conhecimento prévio sobre os objetivos da aplicação;
- seleção do conjunto de dados de destino (*Selection*): seleção de um conjunto ou subconjunto de variáveis focadas nas descobertas que pretende-se realizar;
- limpeza e pré-processamento (*Pre-processing*): realização de operações básicas, tais como remoção de ruído e *outliers*, escolha de estratégias para lidar com falta

Figura 3 – O processo de KDD



Fonte: Fayyad (1996)

de dados, etc;

- redução e projeção de dados (*Transformation*): definição de recursos para representar os dados, dependendo no objetivo da tarefa, e uso de redução de dimensionalidade ou métodos de transformação para reduzir o número efetivo de variáveis sob análise;
- mineração de dados (*Data Mining*): busca por padrões de interesse em um determinado conjunto de dados de forma a classificar os dados de acordo com regras ou árvores, regressão, agrupamento, modelagem de sequência, dependência ou análise de linha;
- interpretação (*Interpretation/Evaluation*): interpretar os padrões descobertos e, caso necessário, retornar a qualquer uma das etapas anteriores. É importante a visualização dos padrões extraídos para identificar e remover padrões redundantes ou irrelevantes;
- uso do conhecimento (*Knowledge*): incorporar o conhecimento descoberto ao sistema, realizar ações baseadas nele, ou simplesmente documentá-lo e informá-lo às partes interessadas.

A necessidade de dar significado a grandes e complexos conjuntos de dados, enriquecendo-os com informações, tem crescido em todos os campos da tecnologia, negócios e ciência. A capacidade de extrair conhecimento útil desses dados e de agir sobre o conhecimento descoberto está se tornando cada vez mais importante (JOTHI; RASHID; HUSAIN, 2015).

Problemas de predição de séries temporais, onde deseja-se descobrir, com a menor margem de erro possível, valores futuros com base em valores anteriores, geram grande interesse da comunidade científica pois tem diversas aplicações práticas em economia, finanças e hidrologia (TEALAB, 2018). Porém, nos últimos anos tem surgido também aplicações na agricultura e pecuária de precisão, à medida que surgem desafios em torno

da produção em termos de produtividade, impacto ambiental, segurança alimentar e sustentabilidade. Os ecossistemas agrícolas são complexos e multivariados sendo que para serem compreendidos, vários aspectos e fenômenos físicos precisam ser monitorados e analisando continuamente, o que sugere o emprego de novas tecnologias de informação e comunicação (KAMILARIS; PRENAFETA-BOLDÚ, 2018).

Os primeiros modelos para predição de séries temporais foram desenvolvidos nos anos 70, e baseavam-se em modelos matemáticos lineares, onde assume-se que o valor atual da série temporal é uma combinação linear de seus valores passados (SAGHEER; KOTB, 2019). Porém, com o passar dos anos se descobriu que muitas séries temporais seguem um comportamento não linear sendo que a abordagem inicial era insuficiente para representar a sua dinâmica. Assim, a partir dos anos 90, foram apresentados na literatura uma ampla gama de modelos que sugerem diferentes representações matemáticas não lineares presentes nos dados.

Dentre estes, as abordagens baseadas em aprendizado de máquina tem destaque evidenciado, pelo grande número de publicações sobre esse tópico nos últimos anos (TEALAB, 2018). Apesar de possuírem vários pontos em comum, aprendizagem de máquina e mineração de dados possuem diferenças que serão descritas na Seção 2.3.2.1.

2.3.2.1 Mineração de dados x Aprendizado de máquina

Conforme mencionado na Seção 2.3.2, a mineração de dados tem por objetivo encontrar padrões em dados, sendo considerada parte do processo de descoberta de conhecimento. Segundo Mirkin (2011), a mineração de dados possui ênfase em cálculos rápidos para grandes conjuntos de dados para localização de associações e padrões relevantes que antes estavam ocultas.

Para Jothi, Rashid e Husain (2015), existem dois tipos de modelos de mineração de dados: modelo descritivo e modelo preditivo. O modelo descritivo aplica um conjunto de funções na busca por padrões que descrevam os dados. O modelo preditivo geralmente aplica funções de aprendizado para prever valores desconhecidos ou futuros de outras variáveis de interesse. Alguns exemplos de mineração de dados são modelos capazes de realizar agrupamentos conforme similaridade por meio de técnicas de *clustering* ou em termos de atributos ou regras de classificação por meio de técnicas baseadas em árvore de decisão.

O aprendizado de máquina é o campo científico que lida com as técnicas pelas quais as máquinas aprendem com a experiência. Para diversos autores, o

termo "aprendizado de máquina" é equivalente a "inteligência artificial", visto que a capacidade de aprendizado é a principal característica de uma entidade chamada inteligente (KAVAKIOTIS *et al.*, 2017). As tarefas de aprendizado de máquina são tipicamente classificadas em três categorias, sendo elas:

- aprendizado supervisionado: é fornecido um conjunto de treino formado pelas entradas e correspondentes saídas desejadas;
- aprendizado por reforço: para cada entrada apresentada, é produzida uma indicação externa referente a adequação das saídas produzidas pela rede;
- aprendizado não supervisionado: o modelo atualiza seus pesos sem o uso de pares entrada-saídas desejadas e sem indicações sobre a adequação das saídas produzidas.

Sistemas inteligentes buscam a capacidade de adquirir conhecimento ao longo do tempo através da aquisição de experiência e não apenas a extração de informações e transformação em uma estrutura entendível para posterior uso. Uma das principais vantagens das técnicas de aprendizado de máquina é que elas são capazes de resolver autonomamente problemas não-lineares usando conjuntos de dados de múltiplas. O aprendizado de máquina fornece uma estrutura poderosa e flexível para auxiliar a tomada de decisões orientada à dados, através da incorporação de conhecimento especializado ao sistema. Estas características das técnicas de aprendizado de máquina as tornam amplamente utilizadas em muitos domínios, e altamente aplicáveis a agricultura e pecuária de precisão (CHLINGARYAN; SUKKARIEH; WHELAN, 2018). Existem, na literatura, trabalhos que utilizam aprendizado de máquina para realizar a estimativa de biomassa em pastagens dentre os quais se pode destacar o trabalho de Ali *et al.* (2017), que destacam a utilização de Redes Neurais Artificiais e Sistemas de Inferência *fuzzy* Neuro Adaptativas como técnicas potenciais para exploração deste tipo de dados. As técnicas citadas são apresentadas nas Seções 2.3.2.2 e 2.3.2.3.

2.3.2.2 Redes Neurais Artificiais

Nos últimos anos, a utilização de Redes Neurais Artificiais (RNA) tem aumentado. Esse crescimento incentivou o desenvolvimento de diversas aplicações científicas e questões práticas, principalmente no campo dos negócios. Características das redes neurais artificiais, tais como eficiência, robustez e adaptabilidade, fazem delas uma ferramenta para classificação, suporte à decisão, análise financeira (TKAC; VERNER, 2016). As Redes Neurais Artificiais são algoritmos de aprendizado de máquina, cujos

modelos computacionais têm alta capacidade de se adaptar, aprender e generalizar padrões complexos escondidos nos dados. Para isso, as RNA buscam modelar conjuntos de relações entre recursos utilizando estruturas de neurônios artificiais interconectados (MIRKIN, 2011). Uma RNA simula um neurônio biológico onde a informação flui e é processada pelo neurônio e os resultados são geradas através de operações matemáticas. O neurônio tem a capacidade de reagir com base em padrões previamente aprendidos (ALI *et al.*, 2017). As RNA são capazes de resolver problemas não-lineares baseados em composições paralelas.

A definição da arquitetura de uma Rede Neural Artificial é um parâmetro importante na sua concepção, uma vez que ela restringe o tipo de problema que pode ser tratado pela rede. Podem fazer parte da definição da arquitetura os seguintes parâmetros: número de camadas da rede, número de nodos em cada camada, tipo de conexão entre os nodos e topologia da rede (BRAGA; FERREIRA; LUDERMIR, 2007). Quanto ao número de camadas, pode-se ter:

- redes de camada única: só existe um nó entre qualquer entrada e qualquer saída da rede;
- redes de múltiplas camadas: existe mais de um neurônio entre alguma entrada e alguma saída da rede. As camadas intermediárias são denominadas camadas ocultas.

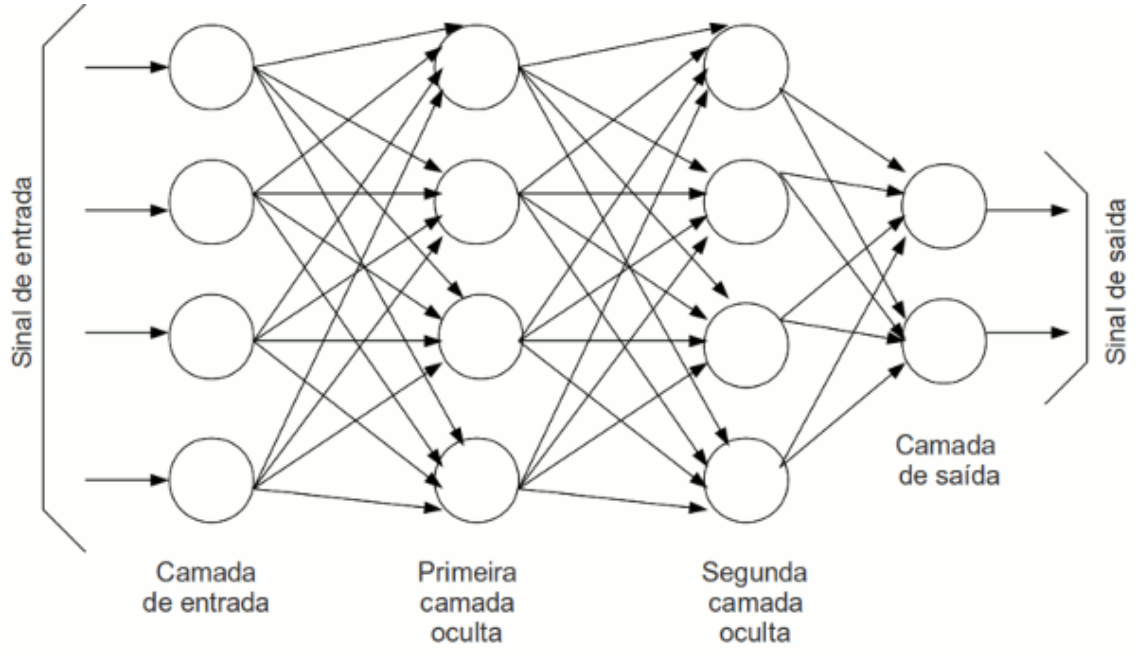
Os nodos podem ter conexões do tipo:

- *feedforward*, ou acíclica: a saída de um neurônio na i-ésima camada da rede não pode ser usada como entrada de nodos em camadas de índice menor ou igual a i;
- *feedback*, ou cíclica: a saída de algum neurônio na *i*-ésima camada da rede é usada como entrada de nodos em camadas de índice menor ou igual a i

A rede é totalmente conectada se para qualquer nodo da i-ésima camada, este está conectado a todos os nodos da camada i+1, do contrário a rede é considerada parcialmente conectada. Um exemplo de arquitetura de RNA de múltiplas camadas, *feedforward* e completamente conectada é apresentado na Figura 4.

As *Long Short-Term Memory* (LSTM) são uma abordagem de RNA proposta por Hochreiter e Schmidhuber (1997), que pode ser considerada uma variação das Redes Neurais Recorrentes capaz de aprender dependências de longo prazo. As Redes Neurais Recorrentes, assim como as LSTM, possuem módulos de memorização em suas estruturas de aprendizado e, por isso, são utilizadas em problemas com características temporais,

Figura 4 – Representação da uma Rede Neural Artificial multilayer perceptron



Fonte: Monolitonimbus (2019)

onde a sequência dos dados é importante.

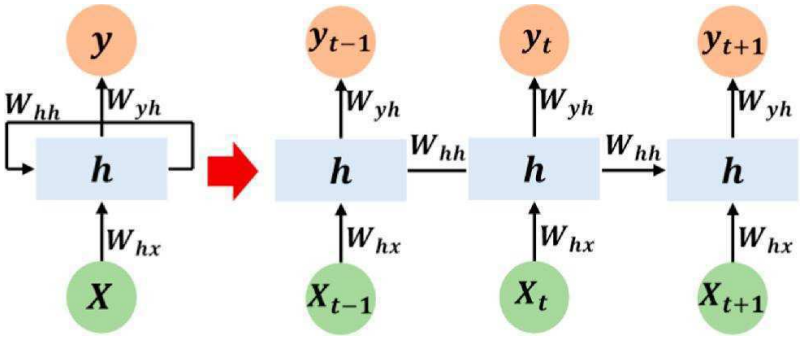
De acordo com Sagheer e Kotb (2019), no passado problemas de previsão de séries temporais eram abordados por métodos estatísticos lineares que buscavam prever comportamento do sistema no futuro, baseando-se em informações do status atual e passado. Porém, nas últimas duas décadas, vários algoritmos de RNA chamaram a atenção e se estabeleceram como concorrentes relevantes para os métodos estatísticos para tarefas de previsão, mostrando maior precisão.

As redes neurais aprendem a correspondência entre entradas e saídas a partir de uma perspectiva estática. No entanto, quando os dados de entrada são uma série temporal, as informações serão perdidas se esses dados forem treinados de maneira independente como entradas e saídas da rede neural (WANG *et al.*, 2019).

A Figura 5 mostra a topologia básica de um rede neural recorrente simples, onde X e Y denotam os dados de entrada e saída; h denota as estruturas presentes na camada oculta; W_{hx}, W_{yh} e W_{hh} denotam as matrizes de peso que descrevem a relação entre X e h, h e Y , e h e h . A saída y_t não é apenas determinada pela entrada xt, mas também pelo último estado oculto h_{t-1}. O estado oculto h_t é o componente chave para manter as dependências dentro da série temporal (WANG *et al.*, 2019).

No entanto, as redes neurais recursivas simples têm apenas um único estado oculto o que as torna sensíveis apenas às dependências de curto prazo. Para capturar

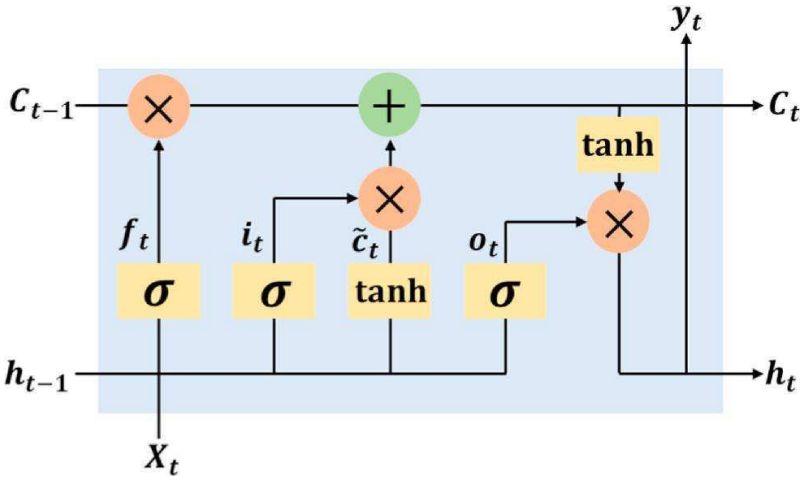
Figura 5 – Topologia básica de um rede neural recorrente simples.



Fonte: Wang *et al.* (2019)

dependências de longo prazo dentro da série temporal, uma unidade LSTM contém dois estados ocultos, que são projetados para manter tanto as informações de curto prazo quanto as de longo prazo. A estrutura interna de uma unidade LSTM é apresentada na Figura 6.

Figura 6 – Estrutura interna de uma unidade LSTM.



Fonte: Wang *et al.* (2019)

O estado oculto contém um mecanismo extra para esquecer estrategicamente informações não relacionadas correspondentes à ocorrência atual. Para reter a informação de longo prazo, três portas de controle são introduzidas na unidade LSTM. Estas são a porta do esquecimento (f_t), a porta de entrada (i_t) e a porta de saída (h_t). As portas de controle essencialmente conectam as camadas, denotadas como σ (WANG *et al.*, 2019).

A primeira porta na unidade LSTM é a porta do esquecimento f_t, que determina quanta informação é mantida do último estado c_{t-1}. O estado de esquecimento no momento t é dado por: $f_t = \sigma(W_f.[h_{t-1}, X_t] + b_f)$; onde $\sigma(.)$ denota a função de ativação

sigmoide; X_t são as entradas para o modelo de regressão, que incluem principalmente dados históricos e fatores externos; f_t, h_{t-1} e b_f representam o vetor da porta de esquecimento no momento t , o vetor de saída no momento $t-1$ e o viés da porta do esquecimento no tempo t , respectivamente; W_f é a matriz de peso da porta do esquecimento; e $[.]$ é o operador de concatenação para vetores (WANG *et al.*, 2019).

A segunda porta é a porta de entrada i_t, que determina quanta informação atual deve ser tratada como entrada para gerar o estado atual c_t. i_t e é calculado por: $i_t = \sigma(W_i.[h_{t-1}, X_t] + b_i)$; onde W_i e b_i denotam a matriz de peso e polarização da porta de entrada, respectivamente. Pode ser visto que i_t tem uma formulação semelhante a f_t. Ambas as portas são determinadas por h_{t-1} e X_t (WANG *et al.*, 2019).

O estado oculto atual c_t é determinado pela adição das partes de informações que eles controlam. A informação de longo prazo é controlada por f_t, e a informação de curto prazo é controlada por i_t: $\tilde{c}_t = tanh(W_c.[h_{t-1}, x_t] + b_c)$, $c_t = f_t * c_{t-1} + i_t * \tilde{c}_t$, onde, $tanh(.)$ denota a função de ativação tangente hiperbólica; W_c e b_c denotam a matriz de pesos e o viés da porta atual, respectivamente; e o operador $*$ representa o produto elementar (WANG *et al.*, 2019).

A última fase constitui em calcular quanta informação pode eventualmente ser tratada como saída. Outra porta de controle é escolhida como porta de saída o_t: $o_t = \sigma(W_o.[h_t, x_t] + b_o)$ (WANG *et al.*, 2019).

A saída final do LSTM é definida por: $h_t = o_t * tanh(c_t)$ (WANG *et al.*, 2019).

As LSTM são um tipo especial de rede neural recorrente, que é adequada para processar e prever eventos importantes com intervalos relativamente longos em séries temporais. Assim, as LSTM alcançaram sucesso considerável e têm sido amplamente utilizadas (SAGHEER; KOTB, 2019).

2.3.2.3 Sistema de Inferência Neuro Fuzzy Adaptativo

Segundo Kayacan, Khanesar e Mendel (2015), sistemas de inferência *fuzzy* empregam regras para modelar aspectos qualitativos do conhecimento e raciocínio humano. Existem diversas aplicações práticas para os modelos nebulosos (*fuzzy*) em sistemas de controle, previsão e inferência. No entanto, existem algumas limitações dessa abordagem, tais como:

- não existem métodos automáticos para transformar o conhecimento ou a experiência humana em regras de um sistema de inferência *fuzzy*;

- há necessidade de métodos eficazes para ajustar as funções de associação de modo a minimizar a medida de erro de saída ou maximizar o índice de desempenho.

Buscando aproveitar as melhores características das RNA e da lógica *fuzzy*, Jang (1993) sugeriu uma nova arquitetura denominada Sistema de Inferência Neuro *Fuzzy* Adaptativo (*Adaptive Neuro-Fuzzy Inference System* - ANFIS), para servir como base para a construção de um conjunto de regras *fuzzy* com funções apropriadas de associação. O ANFIS é um modelo híbrido que integra os aspectos positivos das RNA e da lógica *fuzzy* para construir um modelo robusto que associe as variáveis independentes ou valores de entrada com as variáveis dependentes que são os valores alvo de saída, com erro mínimo de estimativa (ALI *et al.*, 2017).

Um sistema ANFIS é um sistema *fuzzy* que usa um algoritmo de estratégia de aprendizado derivado do comportamento das RNA para encontrar os parâmetros determinados pelos conjuntos *fuzzy*. Este tipo de modelagem permite unir ao processamento linguístico de um sistema de inferência à capacidade de adaptação e aprendizagem das Redes Neurais Artificiais. Uma arquitetura ANFIS típica possui cinco camadas de neurônios. Essa técnica recebe variáveis do mundo real no formato dos conjuntos, sendo que través das iterações em suas camadas, os dados de entrada são transformados em variáveis linguísticas e ao final a variável é convertida novamente para o formato de conjunto (JANG, 1993). As técnicas de exploração aliadas à coleta e armazenamento eficiente de dados constituem a base por trás dos Sistemas de Apoio à Decisão apresentados na Seção 2.3.3.

A utilização de técnicas ANFIS neste trabalho foi descartada devido à escassez de bibliotecas que implementem os recursos necessários para a sua utilização identificado na linguagem Python. Existe a biblioteca ANFIS, porém esta se encontra em uma versão Beta e possui documentação limitada e poucos exemplos disponíveis para consulta (PYPI, 2019).

2.3.3 Sistemas de Apoio à Decisão

Apresentados conceitos e ferramentas de armazenamento e processamento de dados, pode-se construir o conceito de Sistemas de Apoio à Decisão (SAD). Burstein (2008) define SAD como: "um sistema baseado em computador que representa e processa o conhecimento de maneiras que permitem ao usuário tomar decisões mais produtivas,

ágeis, inovadoras e respeitáveis". De acordo com Handoyo e Sensuse (2017), a arquitetura de um SAD possui componentes que proporcionam descoberta de conhecimento e proporcionam benefícios tais como o aumento da produtividade, agilidade e inovação.

Um SAD reúne recursos de persistência de dados e de descoberta de conhecimento em base de dados a uma interface capaz de apresentar de maneira eficiente o conhecimento extraído de forma a auxiliar o seu usuário no processo decisório. Porém, mesmo com o reconhecimento de que as decisões podem ser tomadas com maior rapidez e precisão do que decisões sem a ajuda de um SAD, por vezes usuários em potencial resistem e não aproveitam esse recurso para apoiar a gestão, tornando necessário incentivar o uso de SAD (CHAN *et al.*, 2017).

Nesse sentido, Kukar *et al.* (2018) relatam que diversas soluções comercializadas para a agricultura utilizam sensores para coletar dados e, embora isso resulte em grandes quantidades de dados armazenados, frequentemente os dados são analisados por métodos estatísticos ou técnicas de visualização simplificadas, negligenciando o potencial de aplicação de ferramentas avançadas de análise de dados que se originam dos campos de Mineração de Dados. Por isso, é relevante o desenvolvimento de sistemas que preencham a lacuna entre os sistemas agrícolas e as metodologias de suporte à decisão atuais, fornecendo aos agricultores suporte às decisões de uma maneira acessível e automatizada, proporcionando previsões de cenários simulados e melhor compreensão de padrões e dependências dos dados no qual o sistema estará baseado (KUKAR *et al.*, 2018). No contexto deste trabalho, o processo da coleta de dados já está posto, sendo que ele é baseado em metodologias de amostragem próprias para determinação da massa de forragem, as quais são apresentadas na Seção 2.4.

2.4 Técnicas para estimativa de acúmulo da massa de forragem

Segundo Estrada, Júnior e Regazzi (1991), estimar o peso da forragem ou o volume da massa de forragem faz-se relevante à medida que nos permite calcular a taxa de lotação, estimar a quantidade de forragem consumida e interpretar o rendimento da produtividade animal. Contudo, a estimativa da forragem disponível é frequentemente associada a alto erro experimental, podendo variar bastante entre os métodos e os observadores (CÓSER *et al.*, 2003).

Existem diversas técnicas de estimativa de disponibilidade de forragem, divididas entre métodos diretos ou destrutivos e indiretos ou não-destrutivos. Alguns dos

fatores que devem ser levados em conta na escolha do método para determinação de disponibilidade estão relacionados a aspectos gerais da área de estudo, tais como a uniformidade, a densidade e a altura das plantas, a composição botânica da comunidade vegetal em estudo, bem como a disponibilidade de mão de obra. Em ambas as técnicas, a massa de forragem é dada em quilos por hectare ($kgMS/ha$) (SALMAN; SOARES; CANESIN, 2006).

Na técnica direta, a massa de forragem existente nos vários piquetes é obtida por meio do corte e da pesagem de amostras da planta levando-se em conta o tamanho da área, enquanto que na técnica indireta essa massa de forragem é obtida por estimativa. As Subseções 2.4.1 e 2.4.2 apresentam as principais técnicas utilizadas para coleta de amostras para determinação da massa de forragem disponível na pastagem, considerando-se as vantagens e desvantagens de cada uma, visando fornecer subsídios para a tomada de decisão na hora de escolher o método mais adequado para cada situação (CÓSER; MARTINS; DERESZ, 2002).

2.4.1 Amostragem direta

O "método do quadrado" é a técnica mais aplicada para amostragem de pastagem. Ela utiliza uma moldura, geralmente quadrada, de área conhecida fabricada de madeira ou metal. A área das molduras costuma variar de 0,10 m^2 até 2,0 m^2. O tamanho do quadrado utilizado depende da uniformidade da área a ser amostrada, sendo que as molduras mais comuns são aquelas de 0,5 x 0,5 m (SALMAN; SOARES; CANESIN, 2006).

Segundo Cóser, Martins e Deresz (2002), o método direto geralmente proporciona maior precisão quando comparado com outros métodos. No entanto, para pastagens extensas com grande variabilidade na cobertura o número de amostras necessárias torna, por vezes, a tarefa inviável por necessitar de maior quantidade de mão-de-obra e equipamentos. Essas dificuldades podem fazer com que o pesquisador diminua o número de amostras, tornando a amostragem inadequada e sujeita a baixa precisão. Por outro lado, se o número adequado de amostras for mantido, haverá grande destruição de forragem na área pelo corte de amostras.

Para o método de amostragem direta, o número de amostras necessárias para obtenção de uma estimativa confiável depende do quanto varia a produção de forragem dentro da área a ser amostrada. Porém Salman, Soares e Canesin (2006), sugerem que juntamente com a metodologia direta de amostragem seja utilizado um método indireto,

tal como a avaliação visual, para proporcionar melhores estimativas. A Subseção 2.4.2 apresenta alguns conceitos das metodologias indiretas.

2.4.2 Amostragem indireta

De acordo com Salman, Soares e Canesin (2006), diversos pesquisadores têm buscado o aprimoramento de métodos indiretos de amostragem de pastagem, para uma avaliação mais ágil, menos trabalhosa e com custo reduzido. A produção de forragem varia entre espécies, tipo de hábito de crescimento, altura, estágio de crescimento, entre outros, o que torna a avaliação complexa. Além disso, métodos de avaliação visual são subjetivos e dependem do avaliador, sendo que existe uma tendência de pastagens que possuam altura mais elevada e baixa densidade serem superestimadas, enquanto que as que possuem menor altura e maior densidade, normalmente são subestimadas. Por esses motivos, em geral, esses métodos são considerados menos precisos que o método direto (SALMAN; SOARES; CANESIN, 2006).

O emprego de medidas como a altura da planta e a cobertura do solo pela forrageira, bem como de métodos de estimativas visuais, podem permitir melhor avaliação da produção de forragem em áreas sob pastejo, reduzindo custos, tempo e mão-de-obra. Para isso, o treinamento de avaliadores é de suma importância na recomendação e validação desse método de amostragem (LOPES *et al.*, 2000).

Por isso, pesquisadores têm desenvolvido técnicas de amostragem visando melhorar a eficiência das avaliações. Abramides *et al.* (1982), verificaram que a altura média da vegetação é uma técnica viável para a estimativa da quantidade de forragem, porém, além de medidas de altura, levar em conta a cobertura do solo é importante para melhorar a confiabilidade na estimativa da forragem disponível em gramíneas. Cóser *et al.* (2003) afirma que a altura das plantas da pastagem possui uma razoável correlação com a produção de forragem, sendo esta entre 0,71 e 0,95, entre a altura das plantas e a produção de forragem.

Cóser, Martins e Deresz (2002) sugerem também uma técnica baseada em dupla-amostragem. Neste caso, as áreas são avaliadas a partir da alocação de quadros de referência pré-selecionados de massa conhecida que constituem uma escala. Esse método usa técnicas não-destrutivas e possibilita maior número de estimativas de rendimento e sua precisão é aprimorada em função da calibração dos quadrados de referência.

Para Salman, Soares e Canesin (2006), mesmo que os métodos indiretos

apresentem bons resultados na estimativa da produção, existem casos em que não se consegue resultados satisfatórios com sua utilização. Uma vantagem dos métodos indiretos é que a alta variabilidade encontrada nas amostras pode ser compensada com um maior número de amostras, com o intuito de aumentar a precisão do trabalho, em função da facilidade e rapidez que esses métodos apresentam.

Além disso, existem alternativas capazes de automatizar o processo de estimar a massa de forragem de uma pastagem, uma delas é apresentada por Bremm *et al.* (2015), com o uso de sensores *Normalized Difference Vegetation Index* (NDVI) para realizar uma estimativa de forragem em pastagens naturais do Bioma Pampa. Esse tipo de solução é capaz de minimizar alguns dos problemas do método atual utilizado na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul para coletar as informações sobre a disponibilidade instantânea de forragem. Entretanto, a utilização de sensores de reflectância, como o NDVI, assim como o uso de sonda de capacitância (SERRANO *et al.*, 2011), apesar de facilitar o trabalho de coleta, permitindo obter um maior número de amostras, fornecem uma estimativa da disponibilidade instantânea de forragem, e não equacionam os problemas da necessidade de áreas de exclusão de pastejo, nem da estimativa da taxa de acúmulo de forragem realizada com base na taxa do mês anterior.

2.5 Trabalhos correlatos

Barrett, Laidlaw e Mayne (2005) mostram um modelo de crescimento de forragem construído para apoiar o manejo de pastagens em sistemas de produção leiteira no noroeste da Europa. O trabalho foi baseado no modelo LINGRA proposto por Schapendonk *et al.* (1998) para simular a produtividade das pastagens na Europa. A validação do modelo LINGRA a partir de dados históricos europeus mostrou que as previsões do modelo são suficientemente precisas para torná-lo útil para os processos de tomada de decisão nas fazendas. No entanto, Barrett, Laidlaw e Mayne (2005) ressaltam que ao prever a produção de forragem, a precisão dos resultados depende da precisão dos parâmetros de entrada e, com isso, a saída do modelo fica limitada principalmente pela precisão da previsão do tempo. Assim, tarefa dos modeladores de crescimento de pastagens é construir um modelo com a maior precisão possível, usando parâmetros de entrada realistas mesmo com as limitações conhecidas da previsão do tempo.

Romera *et al.* (2010) apresentam um *software* denominado *Pasture Growth Simulation Using Smalltalk* (PGSUS) desenvolvido com o intuito de prever a massa

de forragem em um piquete individual, por meio de modelo de simulação de pastagem baseado em dados de medições da massa de forragem anteriores e dados meteorológicos diários, incluindo média, temperatura mínima e máxima do ar, radiação solar, chuva e evapotranspiração potencial para treinar o modelo. Testes preliminares feitos em duas fazendas leiteiras, uma do Norte e outra do Sul da Nova Zelândia mostraram índices de correlação maiores que 0,8 para ambas.

O trabalho de Hanrahan *et al.* (2017), comunica o desenvolvido de um sistema chamado *Pasture Base Ireland* (PBI), na Irlanda. Esse sistema visa auxiliar na tomada de decisões em torno da gestão de pastagens através de uma interface web que os produtores acessam semanalmente via smartphone e informam dados de sua propriedade. Estes são cruzados com dados meteorológicos com o objetivo de estimar o crescimento da pastagem. Além disso, os dados são enviados para uma base de dados nacional. Os resultados obtidos dependem da precisão dos dados informados pelo produtor, mas os autores informam um erro relativo de previsão de 15,4%.

Além dos trabalhos científicos descrevendo o desenvolvimento de sistemas para estimativas da massa de forragem, foram encontradas duas soluções já implementadas e disponíveis na internet. O *Pasture Growth Forecaster* é um modelo que calcula o crescimento das pastagens a partir de condições ambientais como temperatura do solo, radiação solar, potencial de evapotranspiração, capacidade de retenção de água no solo e a fertilidade do solo. Assim, o sistema gera estimativas de taxa de crescimento medidos para a fazendas da Nova Zelândia. As estimativas apresentam índices menos realistas se comparadas aos medidos, por exemplo, em gaiolas de pastagem, mas são mais viáveis para estimativas em grande áreas (DAIRYNZ, 2019).

De acordo com seu site oficial, o *Sense-T Pasture Predictor* é uma ferramenta on-line gratuita que visa ajudar os agricultores a tomar melhores decisões no manejo de seus rebanhos, produção e custos, ao fornecer previsões de 30 dias para o crescimento de pastagens e tendências de longo prazo por até 90 dias. O sistema se baseia em uma série de dados, incluindo condições e previsões do tempo atual, eventos de precipitação, registros climáticos passados e umidade do solo em tempo real. As previsões são fornecidas para sete locais nas principais áreas produtoras de carne bovina e laticínios do noroeste da Austrália, mas espera-se que futuramente a ferramenta seja estendida para todo o país (SENSE-T, 2019).

Tanto para o *Sense-T Pasture Predictor* quanto para o *Pasture Growth Forecaster* não se tem informações suficientes para afirmar qual metodologia de predição está sendo

usada pelo sistema. Porém, pode-se observar nos trabalhos de Barrett, Laidlaw e Mayne (2005), Romera *et al.* (2010) e Hanrahan *et al.* (2017) que, em geral, tem sido utilizados modelos estatísticos para modelar o crescimento da pastagem. Os métodos utilizados têm ênfase na utilização de dados meteorológicos e em modelos de crescimento que levam e consideração a fisiologia da planta estudada para inferir um índice para o potencial de crescimento da planta por região. Apesar disso, cada trabalho apresentado tem suas particularidades com relação a dados de entrada e modelagem que permitem ajustes para melhor adaptação ao seu contexto. Não foi encontrado nenhum trabalho que relate a utilização de Aprendizado de Máquina para realizar a predição da massa de forragem, o que evidencia a importância da realização deste trabalho.

3 DESENVOLVIMENTO

Este capítulo apresenta a metodologia adotada visando a identificação, projeto e implementação de um modelo eficiente para realizar a estimativa de massa de forragem em pastagens ao longo do tempo. Inicialmente é apresentado, como referencial, o trabalho de pesquisa que já vinha sendo realizada unidade Pecuária Sul da Empresa Brasileira de Pesquisa Agropecuária. Após, é descrito o desenvolvimento da arquitetura da solução proposta, bem como o detalhamento da construção do banco de dados espacial concebido como modelo persistência para os dados coletados,seguido pelo processo de migração dos dados das planilhas eletrônicas para o banco de dados e o desenvolvimento do modelo da RNA LSTM para predizer a massa de forragem ao longo do tempo.

3.1 Área experimental

O ambiente de estudo é uma área experimental utilizada para mensurar a produção animal e o controle da invasora *Eragrostis plana Nees*, comumente conhecida como capim-annoni, com pastejo contínuo com lotação variável, localizada dentro da fazenda experimental da Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul, em Bagé (RS). Esta parcela constitui-se de uma pastagem natural da Região da Campanha do RS, em relevo suavemente ondulado, com presença significativa da referida gramínea invasora. Além da vegetação campestre, existe no sistema um açude com 2.120 m^2 e uma área florestada com eucalipto com mais de vinte anos, com densidade não uniforme, que ocupa 2,4 ha. Nos períodos em que as áreas são expostas ao pastejo, estas são manejadas tendo como meta manter a carga animal adequada à quantidade de forragem, isto é, com uma quantidade de forragem verde igual a 12% do peso do animal.

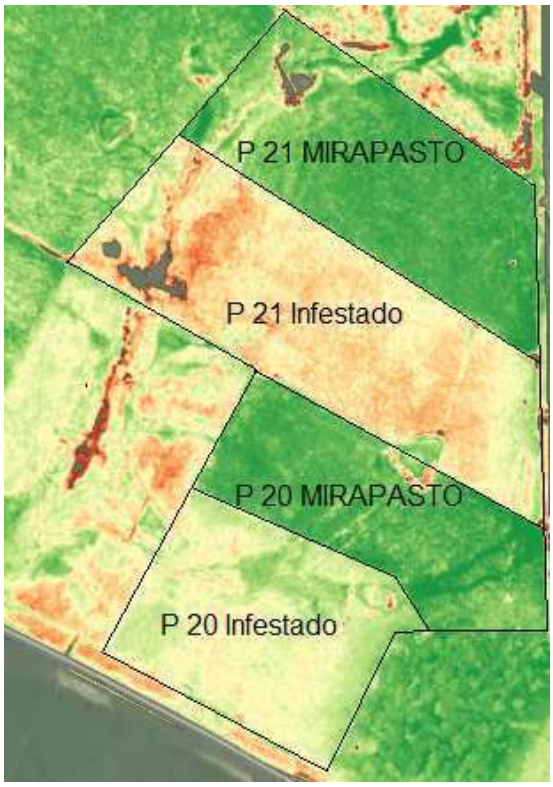
A área de pesquisa foi implantada em dois potreiros denominados 20 e 21 localizados em uma área infestada pelo capim-annoni por mais de 20 anos. Cada um dos potreiros recebe dois tipos de tratamentos. O primeiro, denominado infestado, é mais homogêneo já que apresenta vegetação dominada pelo capim-annoni (*Eragrostis plana Ness*), com relíquias de vegetação nativa campestre, onde predomina *Axonopus afinis* nas partes mais elevadas e *Laersia spp.* nas partes baixas e úmidas, o segundo tratamento, recebe práticas de manejo do Método Integrado para Recuperação de Pastagens, MIRAPASTO (PEREZ, 2015), onde é possível observar uma maior diversidade de espécies da vegetação nativa com predomínio de Poáceas do gêneros

Paspalum, *Axonopus*, *Cynodon*, além de *Asteráceas*, *Ciperaceas* e outras espécies nativas. A área experimental totaliza cerca de 36 ha e possui quatro áreas distintas de estudo com duas repetições de cada tratamento.

O MIRAPASTO faz uso do equipamento Campo Limpo que é um aplicador seletivo de herbicida utilizado na recuperação de pastagens degradadas, pois controla de forma seletiva gramíneas indesejáveis como o capim-annoni e o capim-navalha (*Paspalum virgatum L.*) (EMBRAPA, 2019). Em virtude da diferença de altura que se estabelece entre as plantas forrageiras consumidas pelo gado e as espécies indesejáveis, que assumem uma maior altura, somente as plantas indesejáveis entram em contato com os aplicadores de herbicida da Campo Limpo, com isso ela preserva as espécies forrageiras, dificultando a reinfestação pelas espécies invasoras (PEREZ, 2010).

Essa forma de aplicação direta do herbicida, sem a necessidade de pulverização, aumenta a segurança da aplicação, evitando riscos de deriva do produto e de inalação indevida pelo operador, beneficiando o meio. A Figura 7 mostra uma imagem coletada com um sensor de NDVI onde é possível perceber grande variação deste índice para os diferentes tratamentos testados.

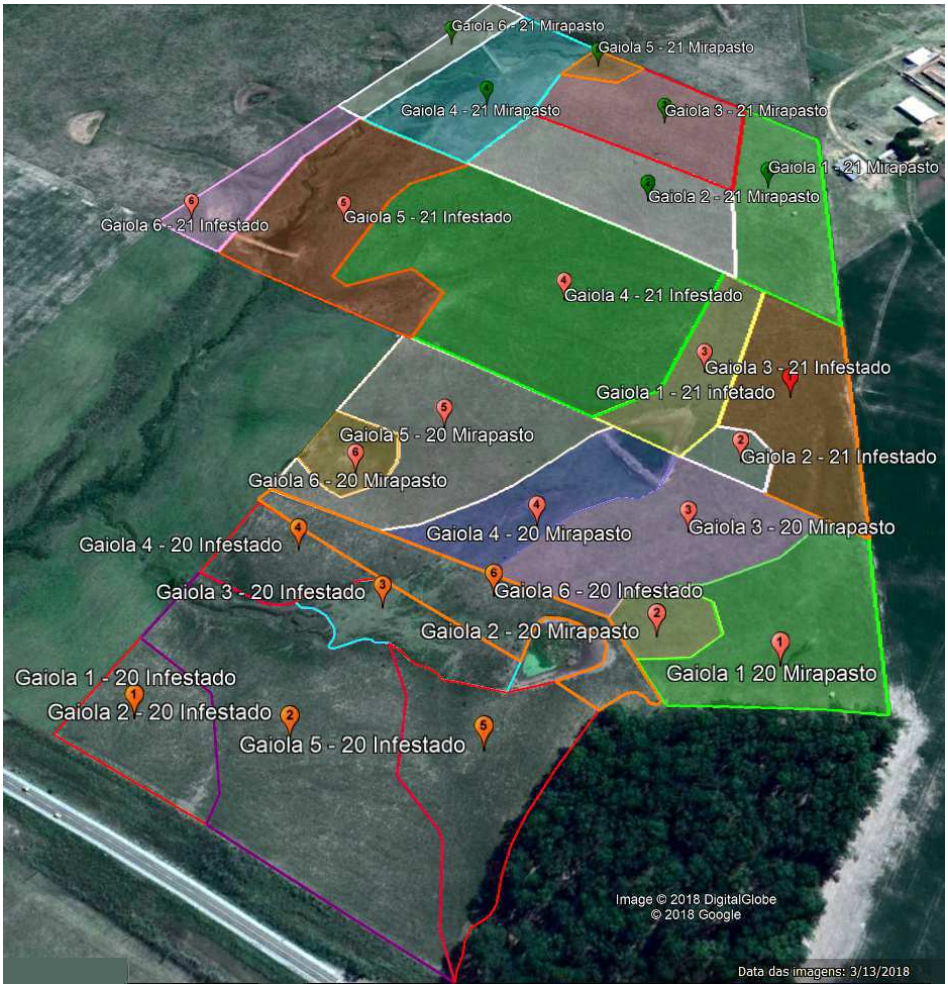
Figura 7 – Imagem capturada com sensor NDVI representando as áreas de estudo e respectivos tratamentos



Fonte: Autor (2018)

Cada uma das quatro áreas foi dividida inicialmente em seis subáreas de acordo com a similaridade da vegetação verificada por análise visual, usada como referência para o método de amostragem direta estratificada da estimativa de massa de forragem. Essa divisão é apresentada na Figura 8. No caso do potreiro 20 com tratamento infestado a divisão das seis subáreas foi feita de acordo com os níveis de infestação pelo capim-annoni como mostra a Figura 9.

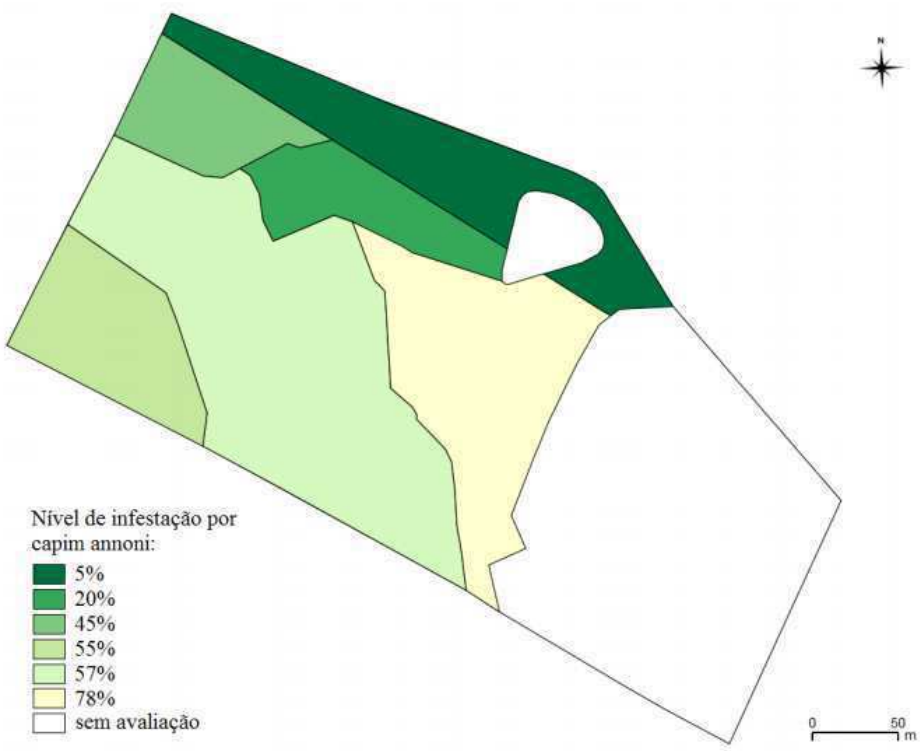
Figura 8 – Zonas amostrais para seis gaiolas



Fonte: Autor (2018)

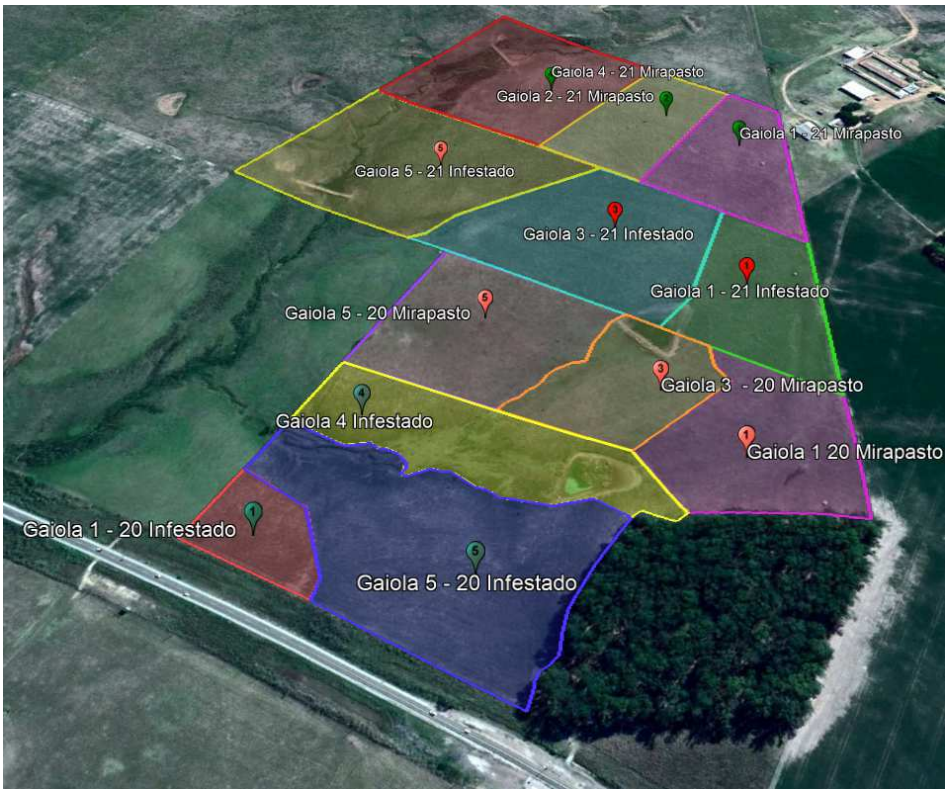
Posteriormente, devido à dificuldade de viabilizar as coletas com tal número de amostras com regularidade, a divisão por similaridade passou a ser feita em três subáreas, como mostra a Figura 10.

Figura 9 – Divisões do potreiro 20 infestado



Fonte: Neves *et al.* (2015)

Figura 10 – Zonas amostrais para três gaiolas



Fonte: Autor (2018)

3.2 Coleta de amostras e armazenamento de dados

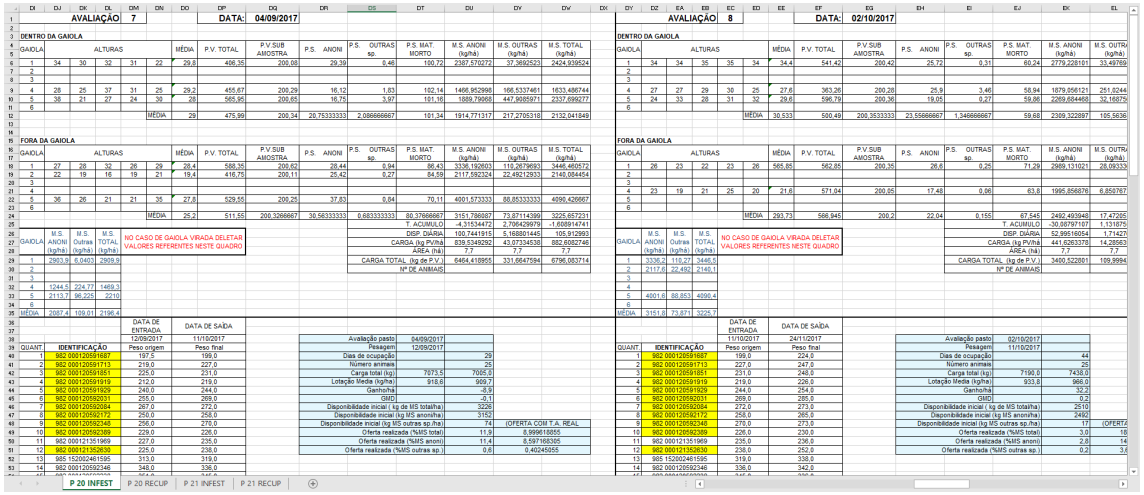
O processo de coleta e processamento de amostras periódicas foi baseado no "método do quadrado" de amostragem direta, apresentado por Cóser, Martins e Deresz (2002), com a diferença de que além da coleta com a moldura foram utilizadas gaiolas de mesma área para isolar pontos da área de estudo, de forma a obter a taxa de acúmulo de forragem, pela diferença entre o valor registrado fora da gaiola e o valor registrado dentro da gaiola no mês posterior. Cada potreiro foi subdividido, virtualmente, em diferentes subáreas com distintos níveis de infestação ou posição topográfica. Dentro de cada subárea, o método direto de amostragem foi realizado de forma orientada, a fim de que o corte representasse a disponibilidade média de forragem da subárea amostrada. Foram utilizadas molduras de 0,5 x 0,5 m para delimitar a área de corte da forragem, bastão graduado para verificar cinco alturas pontuais representativas da área amostrada pela moldura e, dentro desta área, instrumentos para realização do corte da forragem, rente ao chão, na área delimitada. O material foi separado nos componentes verde e seco. O componente verde foi separado em capim-annoni e outras espécies, essas últimas com predominância de gramíneas nativas. por fim, foram utilizadas uma balança para verificação da massa de forragem coletada e uma estufa para realizar a secagem do material verde coletado.

Em cada uma das subáreas foi feita uma análise visual, em busca de dois pontos similares que representassem a média da disponibilidade de forragem na pastagem. No primeiro ponto foi posicionada uma gaiola com o objetivo de isolar a pastagem dos animais para que ela possa crescer livremente e acumular o crescimento da forragem. No segundo foi posicionada uma moldura quadrada de área conhecida onde a forragem foi cortado, seco e pesado. No mês seguinte, esse processo foi repetido amostrando também, através de corte, a quantidade de forragem acumulada dentro da gaiola. Assim, a taxa de acúmulo da pastagem foi estimada com base nos valores do mês anterior, ou seja, pela diferença do total acumulado dentro da gaiola ao final do mês, menos o valor disponível no início do mês (KLINGMAN *et al.*, 1943). No presente estudo, os valores de cada subárea compõem uma média aritmética, a qual é utilizada para estimar a taxa de acúmulo do mês posterior.

Os dados coletados por esse processo foram registrados em planilhas eletrônicas no formato XLS e XLSX como ilustrado na Figura 11. Todas as etapas de processamento dos dados são realizadas na própria planilha por um profissional responsável. Essa

planilha evoluiu ao longo dos anos em que os dados foram registrados, sendo registrados novos dados e de maneiras diferentes à medida que a demanda surgia. Sendo assim, os dados não foram registrados de forma homogênea, sendo necessário um processo para consolidação dos mesmos.

Figura 11 – Planilhas de registro de dados de pastagens



Fonte: Autor (2018)

No processo de coleta de dados foram identificados alguns problemas: a) as amostras coletadas são pontuais e não há garantias de que essas amostras representassem de forma fiel toda a subárea amostrada; b) o processo demanda um período de tempo considerável para a obtenção de resultados e uso de laboratório para processamento de amostras; c) a taxa de acúmulo da forragem quando isolado é diferente de quando está em contato com os animais em pastejo; d) o processo é manual e bastante trabalhoso; e) a média aritmética dos dados das seis subáreas não representa necessariamente a realidade da área total uma vez que as subáreas têm áreas diferentes e foram determinadas de maneira subjetiva; f) a escolha de dois pontos que, teoricamente, representam subáreas, muitas vezes com mais de um hectare dentro potreiro, é subjetiva e sujeita a introduzir alto erro à estimativa devido a multiplicação que é realizada.

Observaram-se algumas anormalidades nos dados, por exemplo, no potreiro 20, com tratamento Infestado, entre 04 de Setembro de 2017 e 02 de Outubro de 2017. Nesse período, onde iniciou a primavera e esperava-se que houvesse crescimento da vegetação, já que as condições climáticas foram favoráveis, com média de temperatura de 17^oC, acúmulo de 251 mm de precipitação, somatório dos valores diários de deficit e excedente hídrico igual a 11 mm e 140 mm, respectivamente, foi registrado no início do período,

fora das gaiolas de exclusão, uma média de 3225 kg/ha de massa de forragem, contra uma média de 2414 kg/ha, dentro das gaiolas, no final do período, evidenciando uma taxa de acúmulo negativa. Anormalidades semelhantes foram observadas com relativa frequência no conjunto de dados, assim, de acordo com os especialistas do domínio (pesquisadores com formação e atuação em manejo de pastagens e agrometeorologia) que participaram do desenvolvimento do trabalho, considera-se que esses valores sejam fruto de erro amostral do método utilizado, possivelmente fruto da escolha inadequada dos pontos representativos da pastagem, já que é improvável que, nessas condições climáticas, a pastagem estando isolada dos animais pelas gaiolas tenha perdido massa. Dessa forma, estima-se que possam existir erros de mais de mil kg/ha por avaliação, associado ao método de amostragem utilizado.

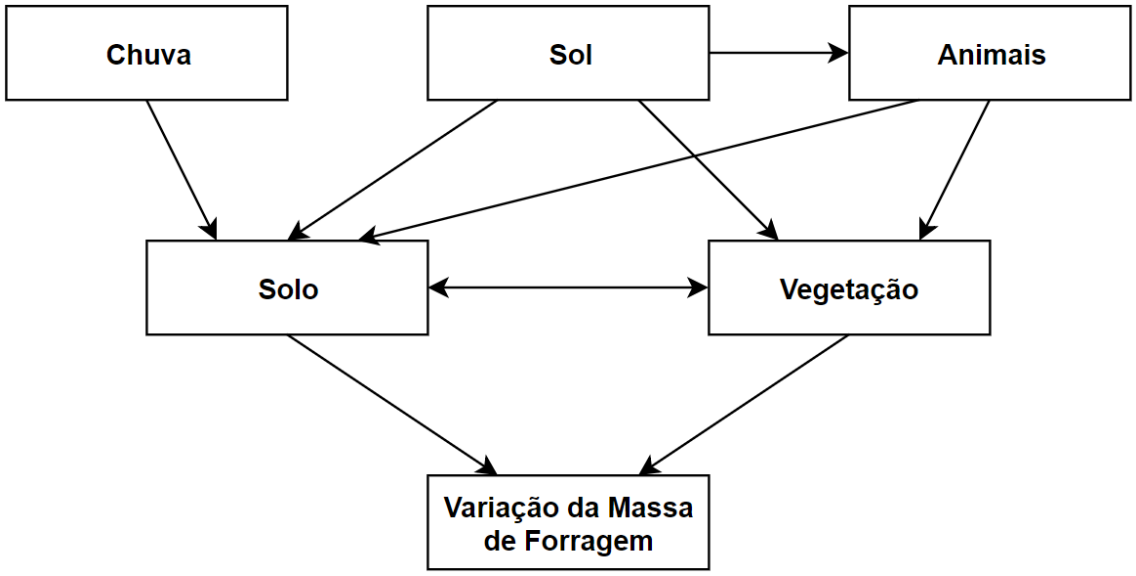
Além disso, quando se assume que o crescimento da pastagem no mês atual é igual ao crescimento da vegetação no mês anterior, tem-se uma estimativa pouco precis a, uma vez que os fatores climáticos, que exercem influência no crescimento da vegetação, variam de um mês para outro (PILLAR, 1995).

Os dados meteorológicos por sua vez, foram obtidos através da estação automática do Instituto Nacional de Meteorologia localizada na Empresa Brasileira de Pesquisa Agropecuária - Pecuária Sul. Segundo o Instituto Nacional de Meteorologia, uma estação meteorológica automática é composta de uma unidade de memória central, ligada a vários sensores dos parâmetros meteorológicos tais como, pressão atmosférica, temperatura e umidade relativa do ar, precipitação, radiação solar, direção e velocidade do vento, etc. que integra os valores observados minuto a minuto e os disponibiliza automaticamente a cada hora, por meio de arquivos no formato CSV que podem ser acessados no site do Instituto Nacional de Meteorologia (INMET, 2019).

3.3 Modelo conceitual

Visando melhorar a estimativa da massa de forragem disponível ao longo do tempo, foram elencadas algumas variáveis consideradas determinantes para o problema. A Figura 12 apresenta essas variáveis e como se relacionam.

Figura 12 – Modelo conceitual



Fonte: Autor (2019)

As variáveis se caracterizam como:

- chuva: interfere nos níveis de umidade disponíveis no sistema, na disponibilidade de nutrientes e na capacidade de absorção das plantas;
- sol: impacta na capacidade fotossintética das plantas, na temperatura e evapotranspiração do sistema;
- animais: quantidade de animais expressa em quilogramas de peso vivo por hectare, que tem relação com o nível de desfolha e pisoteio, o que reflete na estrutura da vegetação e nas características físicas e químicas do solo. Essa variável não foi utilizada no modelo final, pois no método onde utilizam-se gaiolas de exclusão os animais não tem acesso à área amostrada;
- solo: as características do solo são definidas pela sua composição física e química, além da sua interação com fatores meteorológicos e outros agentes. Grande parte do potencial de crescimento das plantas é definido pela composição do solo;
- vegetação: a composição botânica da pastagem vai definir o seu perfil de crescimento e também o seu potencial nutritivo para os animais. O tipo de cobertura que o solo recebe também interfere no perfil de retirada de nutrientes do solo, bem como na sua capacidade de manutenção de água e das suas características físicas e químicas;
- Variação da massa de forragem: a massa de forragem disponível impacta

diretamente no ganho de peso dos animais que ocupam o potreiro e é o objeto de estudo deste trabalho.

Os dados amostrais das pastagens, assim como as variáveis meteorológicas monitoradas pela estação automática do Instituto Nacional de Meteorologia, foram comparadas com o modelo conceitual proposto e usadas como base para orientar os estudos referentes a relação solo-planta-animal necessários para o entendimento das associações e dependências que causam influencia na taxa de acúmulo de forragem. Esse estudo serviu de referência para a caracterização e identificação de um modelo computacional para relacionar as variáveis do problema de forma a realizar uma predição da massa de forragem ao longo do tempo que fosse mais precisa que a realizada atualmente pelo método tradicional. O detalhamento do modelo de RNA desenvolvido para este propósito é apresentado na Seção 3.4.5.

Durante esse estudo, identificou-se também a necessidade do desenvolvimento de uma solução completa de FMIS, que não fosse composta somente de um modelo de aprendizado de máquina para extração de conhecimento, mas sim que tratasse desde as questões iniciais de unificação, consistência, persistência, etc. dos dados, de forma a facilitar o seu uso, através de uma interface que possibilitasse consultas de maneira mais simples, até a aplicação de técnicas de visualização de dados para avaliação dos resultados. O detalhamento da solução proposta é apresentado na Seção 3.4.

3.4 Metodologia

Para atingir o objetivo proposto, foi adotada uma metodologia baseada em pesquisa exploratória e experimental interdisciplinar que, segundo Gil (2008), tem como objetivo proporcionar maior familiaridade com o problema através do aprimoramento de ideias ou a descoberta de intuições, com vistas a torná-lo mais explícito. Seu planejamento é flexível, de modo que possibilite a consideração dos mais variados aspectos relativos ao fato estudado, embora, na maioria dos casos, assume a forma de pesquisa bibliográfica ou de estudo de caso. No presente trabalho foi realizado um estudo de caso por meio de entrevistas com os especialistas do domínio, juntamente com uma revisão bibliográfica para embasar o desenvolvimento do trabalho.

A partir da revisão de literatura, foi feita uma análise sobre os dados armazenados originalmente em planilhas, tendo em perspectiva as questões de pesquisa, com o objetivo

de identificar um modelo para realizar sua persistência em um Sistema Gerenciador de Banco de Dados (SGBD), visando facilitar a posterior aplicação de técnicas de descoberta de conhecimento e aprendizagem de máquina. Após identificado o modelo, foi construída uma base de dados referente a ele no PostgreSQL, o qual possui um complemento denominado PostGIS para operações com dados espaciais, necessário para realizar a espacialização das áreas experimentais (POSTGRESQL, 2019) (POSTGIS, 2019). Para realizar a extração dos dados das planilhas, ajustes e correções de erros, e carga no SGBD, processo conhecido como *Extract, Transform, and Load* (ETL), foram desenvolvidos programas na linguagem Python (PYTHON, 2019).

Concluída essa etapa, estando os dados de 2014 a 2018 persistidos no banco de dados desenvolvido, iniciou-se a definição da conjunto de variáveis representativas do problema para compor a entrada do modelo de aprendizado de máquina escolhido. As variáveis escolhidas são apresentadas na Seção 3.4.5. Devido ao fato de o processo para estimativa de massa de forragem ser realizado a cada aproximadamente 35 dias, foi necessário realizar uma sumarização dos dados meteorológicos que são coletados de hora em hora nas estações automáticas do Instituto Nacional de Meteorologia para que eles pudessem representar da melhor forma possível o período entre as coletas realizadas na Empresa Brasileira de Pesquisa Agropecuária. Como as datas de coletas poderiam ter variação entre os diferentes potreiros, os dados foram sumarizados primeiramente de forma diária e armazenados dessa forma no modelo de persistência. A sumarização no período entre as amostras coletadas nas pastagens foram realizadas individualmente para cada potreiro no pré-processamento.

Além de realizar a sumarização do dados meteorológicos, no pré processamento também foi feita uma sumarização entre as amostras pontuais coletadas em cada gaiola para representar o potreiro. Posteriormente, passaram a ser aplicadas as técnicas de aprendizado de máquina, visando a descoberta de conhecimento sobre o banco de dados. Foram realizados testes com diversas modelagens de RNA com diferentes arquiteturas e ajustes de pré processamento. Sendo que nesta etapa as RNA LSTM se adaptaram melhor às características do problema e, por isso, foi dada ênfase a essa técnica nas etapas seguintes, onde foram definidos o número de camadas oculta e neurônios em cada camada de maneira experimental. Para modelar a RNA, foi utilizada a plataforma Anaconda, a qual se caracteriza como uma distribuição da linguagem Python voltada para a ciência de dados, além da biblioteca de aprendizado de máquina Keras (ANACONDA, 2019) (KERAS, 2019).

O conjunto de dados de entrada foi separado em um conjunto para treinamento e outro para testes, sendo que para os testes foram selecionados dados representativos de um ano completo e os demais formaram o conjunto de treinamento. Devido às características temporais do problema, os dados foram processados de forma sequencial, já que uma medição depende das anteriores. Os resultados foram avaliados por meio da métrica do Erro Quadrático Médio (EQM), dado pela Equação 1, onde $real_i$ são os valores coletados pelo método direto que são usados como referência, $pred_i$ são os valores estimados pelo método proposto neste trabalho e n é o número de mostras usadas para os testes. Os resultados são apresentados na forma de tabelas e figuras comparativos entre os dados de referência e as estimativas realizadas pela RNA ao longo do tempo (RESENDE; DUARTE, 2007). Para a geração das figuras foi utilizada a biblioteca Matpltlib do Python (MATPLOTLIB, 2019).

$$EQM = \sqrt{\sum_{1}^{n} \frac{(real_i - pred_i)^2}{n}} \tag{1}$$

Por fim, buscando quantificar o impacto econômico da adoção do modelo proposto em uma propriedade rural, o mesmo foi comparado em um cenário hipotético, onde não seria feito ajuste de carga, ou mesmo onde esse ajuste seria feito com base na estimativa do mês anterior, método geralmente usado por quem faz estimativas de massa de forragem, frente a um cenário onde se assume a utilização do método proposto. Para tal, utilizou-se a equação apresentada por Maraschin *et al.* (1997), que é dada por:

$$G/ha = -17,9 + 29,2MSO - 1,3MSO^2 \tag{2}$$

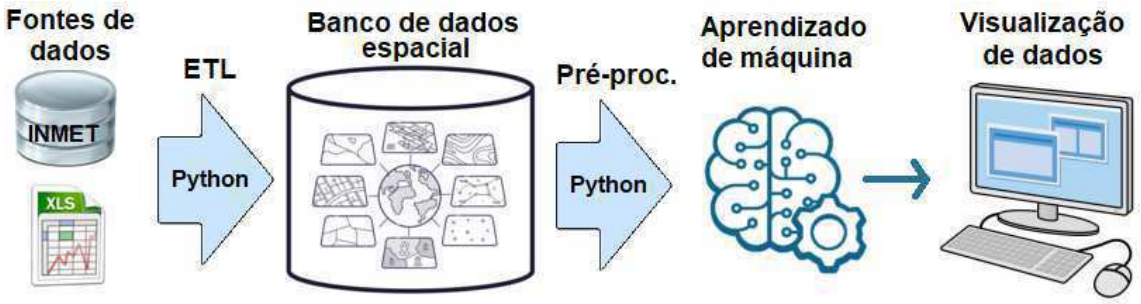
A Equação 2 foi proposta para o campo nativo, no período quente do ano, onde os ganhos de peso vivo dos animais são positivos, o que exclui o inverno, e mostra o ganho de peso vivo dos animais por hectare no período de um ano, em função do percentual de peso vivo animal em relação à Massa Seca Ofertada (MSO). Dessa forma, é possível estimar quanto o método proposto poderia acrescentar nos ganhos de um produtor. O resultado dessa análise é apresentado na Seção 4.5.

Nas Seções 3.4.1, 3.4.2, 3.4.3, 3.4.4 e 3.4.5 são detalhadas todas as etapas da metodologia. Estas tratam respectivamente da arquitetura proposta, processo de ETL, modelo de persistência, pré-processamento e o modelo para estimar a massa de forragem.

3.4.1 Arquitetura proposta

O projeto da arquitetura da solução buscou contemplar os quatro princípios básicos do modelo conceitual de FMIS proposto por Sørensen *et al.* (2010), que tratam de: coleta interna de dados (referentes às pastagens), coleta de informações externas (dados meteorológicos do Instituto Nacional de Meteorologia), geração de relatórios (estimativa baseada em RNA para disponibilidade de forragem) e geração de planos (ajuste da carga animal com base na estimativa de disponibilidade de forragem). Sendo assim, a solução pode ser considerada um módulo para ajuste de carga animal em FMIS (SCHULTE *et al.*, 2019).

Figura 13 – Arquitetura Proposta



Fonte: Autor (2019)

A Figura 13 representa a arquitetura da solução proposta. As fontes de dados representam as planilhas com os dados coletados nos campos experimentais da Empresa Brasileira de Pesquisa Agropecuária e os dados meteorológicos obtidos da estação automática no Instituto Nacional de Meteorologia. O processo de ETL extrai os dados de suas fontes, realiza alguma atividades para verificar a validade dos dados e os carrega em um banco de dados espacial proposto como modelo de persistência. Esses dados passam então por um processo de pré processamento onde foram sumarizados e normalizados para serem usados como entrada no modelo de aprendizado de máquina composto por uma RNA LSTM. Por fim, foram aplicadas técnicas de visualização de dados para gerar figuras referentes aos resultados. Os elementos da arquitetura são detalhados nas Seções 3.4.2, 3.4.3, 3.4.4, 3.4.5 e as figuras são apresentados no Capítulo 4.

3.4.2 Processo de ETL

Antes de ser iniciado o processo de ETL propriamente dito, foi realizada uma atividade para espacializar as amostras, já que no método aplicado na Empresa Brasileira de Pesquisa Agropecuária para coleta de dados de massa de forragem, a coordenada geográfica exata das amostras não era coletada. Esta atividade consistiu em carregar as geometrias com correspondência geográfica das áreas e subáreas de estudo, apresentadas na Seção 3.1, para banco de dados espacial por meio da ferramenta QGIS (QGIS, 2019). Assim, as amostras, quando carregadas no banco, recebem referências para a geometria previamente carregada, a qual cada amostra representada mesma maneira como a representação pretendida pelo método adotado pela Empresa Brasileira de Pesquisa Agropecuária

Para realizar o processo de ETL, foi utilizada a linguagem Python e as suas bibliotecas openpyxl, para leitura das planilhas no formato XLSX, psycopg2 (OPENPYXL, 2019) (PSYCOPG, 2019). Segundo seu site oficial, o Python foi criado por Guido van Rossum em 1991. Os objetivos do projeto da linguagem são produtividade e legibilidade. O Python é uma linguagem livre e multiplataforma, isso significa que os programas escritos em uma plataforma serão executados na maioria das plataformas existentes sem nenhuma modificação (PYTHON, 2019).

O processo de ETL foi implementado como dois algoritmos, sendo que o primeiro foi desenvolvido especificamente para o presente trabalho e faz o mapeamento e validação dos dados presentes nas planilhas que contém as estimativas de massa de forragem coletados por meio do método direto, e em seguida faz a interface com o modelo de persistência e carrega os dados no banco de dados espacial. O segundo algoritmo implementado para o ETL recebeu um documento no formato CSV contendo os dados meteorológicos, que são coletados a cada hora na estação automática mais próxima a área experimental e, através de funções de agregação adequadas, faz a sumarização dos dados de forma diária para então carregá-los no banco de dados espacial.

Durante o processo de construção dos códigos foram verificados alguns problemas na padronização das planilhas. Estes foram corrigidos na própria planilha visto que se fossem feitas adaptações no código este se tornaria específico para aquele caso, fugindo do objetivo inicial de automatizar o processo de ETL para qualquer conjunto de dados que usasse aquele modelo de planilha. Os principais erros corrigidos diretamente na planilha foram o número de linhas e colunas em branco fora do padrão entre o conjunto de dados

de uma coleta e outra, além de diferentes posicionamentos das datas das coletas dentro do modelo.

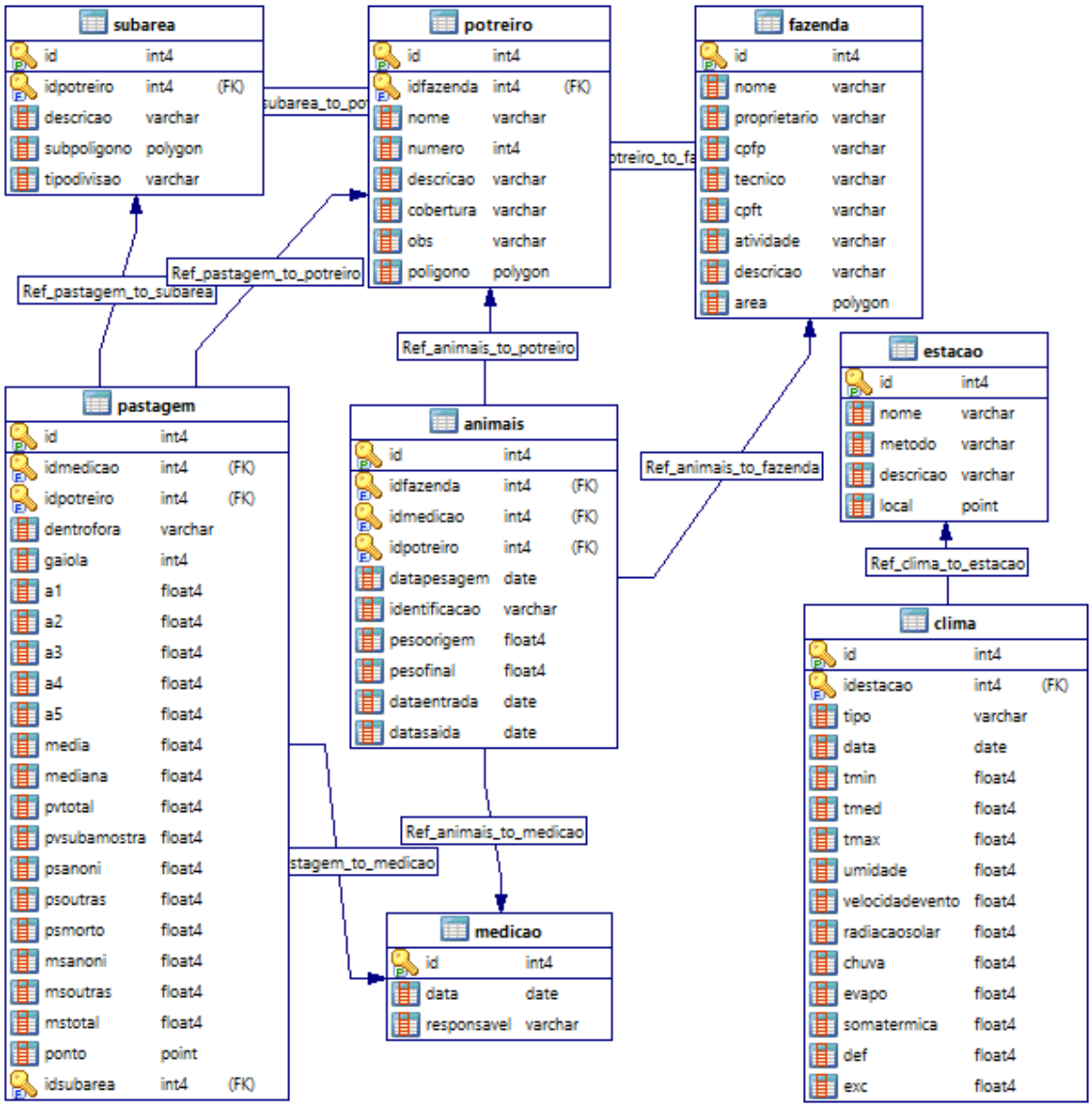
O algoritmo desenvolvido para migrar os dados do ano de 2018 da planilha para o banco é apresentado no Apêndice A. Também foram desenvolvidos códigos para as demais planilhas, as quais seguem uma lógica semelhante mas possuem algumas adaptações referentes a nomenclaturas adotadas nas planilhas e posicionamento de alguns dados, principalmente no modelo que continha os dados mais antigos e seguia um padrão com maiores diferenças.

3.4.3 Modelo de persistência

Como modelo de persistência foi desenvolvida uma base de dados com o PostgreSQL. O PostgreSQL é um Sistema Gerenciador de Banco de Dados (SGBD) completo que está entre os mais usados atualmente. Segundo *PostgreSQL (2019)*, o PostgreSQL começou a ser desenvolvido em 1977. É, portanto, um banco maduro, que possui as principais características desejadas em um banco de dados como, por exemplo, recuperação automática após falhas do sistema, controle de concorrência, suporte a transações, backup *on-line*, tamanho ilimitado de registro (os únicos limites de armazenamento de tipos de dados são impostos pela configuração do hardware), etc. O PostgreSQL é multiplataforma, e oferece baixo custo total de instalação, reduzindo de forma significativa os custos de administração, suporte e licenciamento e, ao mesmo tempo, fornece alta performance, confiabilidade e escalabilidade (POSTGRESQL, 2019). Além disso, este SGBD possui um complemento denominado PostGIS para operações com dados espaciais (POSTGIS, 2019).

O modelo Entidade-Relacionamento (ER) do banco de dados projetado é apresentado na Figura 14. Ele foi desenvolvido de forma a comportar os dados presentes nas planilhas além de prever o acréscimo de novos campos, para que o mesmo possa ser usado como base para um Sistema de Apoio à Decisão que auxilie na gestão de pastagens. Esse banco de dados atende às necessidades do projeto uma vez que facilita o acesso e manipulação dos dados por meio de consultas SQL, possibilitando a realização de testes de forma mais eficiente e ágil.

Figura 14 – Modelo ER do banco de dados proposto como modelo de persistência de dados



Fonte: Autor (2019)

O banco de dados foi estruturado em oito tabelas:

- fazenda: armazena dados de proprietário, técnico responsável e outras características de uma fazenda. Está tabela foi criada supondo que, no futuro, o modelo proposto neste trabalho possa ser ampliado e aplicado em diversas propriedades e/ou campos experimentais;
- potreiro: são registrados os dados referentes as áreas de estudo, tais como, nome, número de identificação, descrição, tipo de vegetação que faz sua cobertura e a geometria correspondente a sua localização geográfica;

- subárea: armazena as geometrias correspondentes as subáreas dos potreiros. No presente estudo permitiu discriminar diferentes situações, tanto para os casos onde a divisão era feita em seis subáreas quanto em três, assim, através das operações oferecidas pelo PostGIS é possível ponderar cada amostra pela área que ela representa do potreiro;
- medição: guarda as datas em que foram realizadas as estimativas pelo método direto;
- pastagem: armazena todos os dados coletados durante a aplicação da metodologia, tais como, alturas em cinco pontos da gaiola, altura média, percentual de annoni, massa verde, massa de forragem total, etc;
- animais: registra todos os bovinos que ocuparam a área bem como as datas e peso de entrada e saída;
- estação: guarda características como descrição e posição geográfica do local onde os dados meteorológicos são coletados, uma vez que com a adição de diferentes propriedades o modelo deve utilizar dados da estação mais próxima;
- clima: armazena os dados meteorológicos diários coletados da estação automática do Instituto Nacional de Meteorologia que se localiza próxima à área de estudo. Essa tabela não se relaciona com as demais pois esses dados precisam ser sumarizados nos intervalos entre as estimativas realizadas pelo método direto, o que é feito no pré-processamento.

3.4.4 Pré-processamento

No pré-processamento foram realizadas três etapas: (i) a sumarização dos dados meteorológicos nos intervalos entre as datas onde foram realizadas coletas de dados amostrais de estimativa de forragem; (ii) a sumarização entre as amostras pontuais coletadas em cada gaiola para representar o potreiro; e (iii) a normalização dos dados. O algoritmo desenvolvido para realizar o processo de ETL é apresentado no Apêndice B, sendo este responsável por acessar o modelo de persistência, realizar as operações de sumarização e normalização dos dados e posteriormente criar um arquivo no formato CSV que é usado como entrada no modelo de RNA LSTM.

A sumarização dos dados meteorológicos foi feita por meio de uma consulta na linguagem SQL que recebe duas datas e aplica às variáveis as suas respectivas funções de

agregação para o intervalo entre as datas. Essa consulta é apresentada no Apêndice B.

A sumarização entre as amostras pontuais coletadas em cada gaiola para representar o potreiro é apresentada no Apêndice B. Esta leva em conta a representatividade da subárea da amostra dentro da área do potreiro através de uma média ponderada apresentada na Equação 3, onde MS_{total} é a massa de forragem total por hectare em kg/ha, MS_{gaiola} é a massa de forragem por hectare na subárea que a gaiola é representativa em kg/ha, $subarea$ é a área que uma determinada gaiola representa em m^2 e $areatotal$ é a área total do potreiro em m^2. Dessa forma, as amostras consideradas nulas ou inválidas foram desprezadas e um único valor foi considerado para representar a quantidade de massa de forragem por hectare em determinado potreiro para uma medição desde que houvesse pelo menos, uma gaiola com estimativa válida. Processo análogo ao realizado para sumarizar a MS_{total} é feito para a altura média da pastagem. Essas estimativas, juntamente com os dados meteorológicos sumarizados, compõem o conjunto de entrada do modelo.

$$MS_{total} = \frac{\sum_1^n MS_{gaiolai} * \frac{subarea_i}{areatotal}}{\sum_1^n \frac{subarea_i}{areatotal}} \tag{3}$$

3.4.5 Modelo para estimar a massa de forragem

Para compor a entrada do modelo de aprendizado de máquina baseado em RNA LSTM foram elencadas inicialmente quatorze variáveis, sendo que, para as variáveis meteorológicas elencadas, foram usados também, em um segundo momento, a variância e desvio padrão associados, como forma de agregar medidas de dispersão da variação ao longo do período que representam:

- número de dias: quantidade de dias referentes ao intervalo da estimativa atual e da anterior;
- altura da forragem média anterior: altura média registrada na estimativa anterior;
- MS total da forragem anterior: massa de forragem total da estimativa anterior;
- altura média: altura média da pastagem na data da estimativa atual;
- percentual de annoni: estimativa do percentual de infestação da vegetação pelo capim-annoni;
- t min: média das temperaturas mínimas registradas no intervalo entre a estimativa

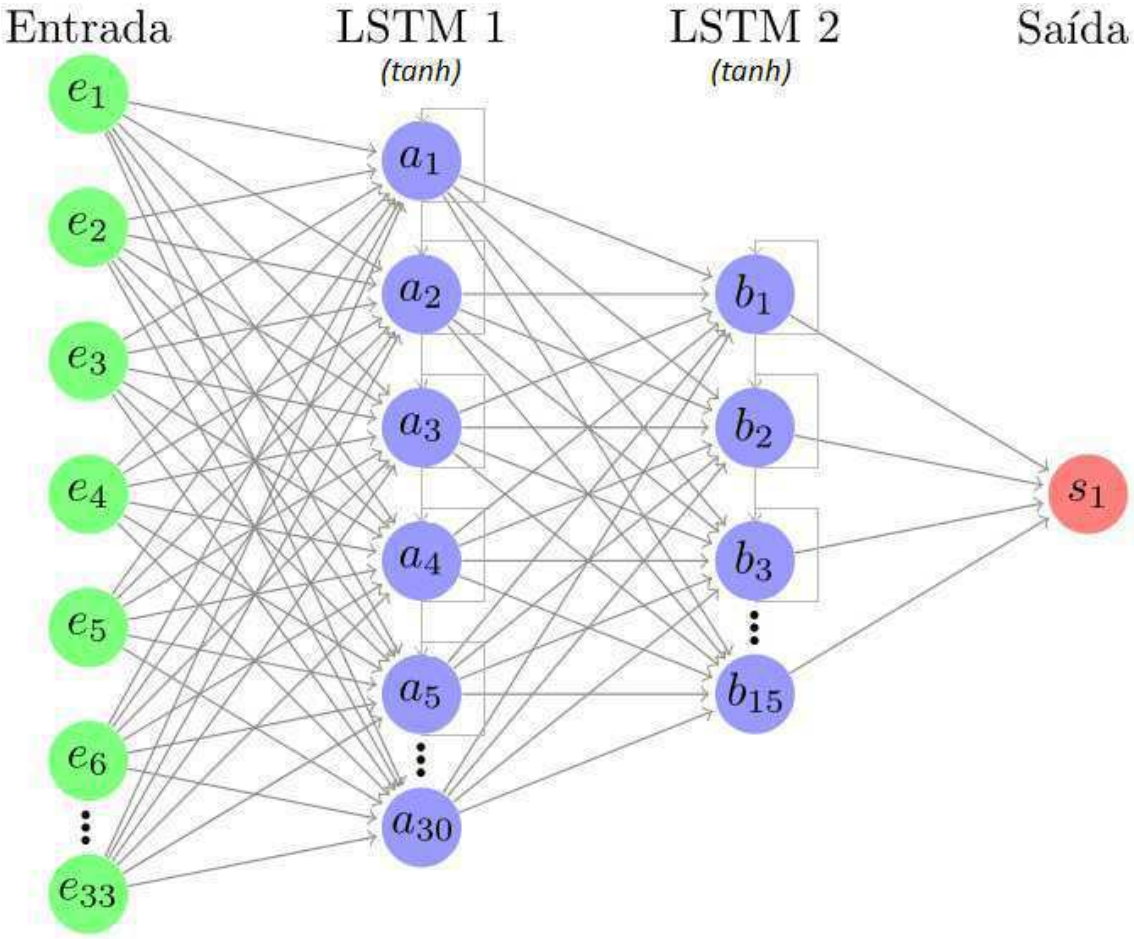
atual e a anterior;

- t med: temperatura média registrada no intervalo entre a estimativa atual e a anterior;
- t max: média das temperaturas máximas registradas no intervalo entre a estimativa atual e a anterior;
- umidade: média da umidade relativa do ar registrada no intervalo entre a estimativa atual e a anterior;
- velocidade do vento: média da velocidade do vento registrada no intervalo entre a estimativa atual e a anterior;
- radiação solar: somatório da radiação solar registrada no intervalo entre a estimativa atual e a anterior;
- chuva: acumulado da precipitação registrada no intervalo entre a estimativa atual e a anterior;
- soma térmica: acumulado da soma térmica dada por: $st = \frac{t_{max}+t_{min}}{2} - t_{base}$, onde t_{max} é a temperatura máxima registrada no dia, t_{min} é a temperatura mínima registrada no dia e t_{base} é a temperatura basal para a vegetação estudada. No contexto deste trabalho foi utilizado t_{base} igual a 7^oC no período mais frio (entre maio e setembro) e 10^oC no período mais quente (entre outubro e abril) devido à diversidade de espécies encontradas (SENTELHAS; PEREIRA; ANGELOCCI, 2000);
- def e exc: déficit e excedente hídrico calculado através do método para cálculo de balanço hídrico proposto por Thornthwaite e Mather (1955).

Assim, a camada de entrada do modelo de RNA LSTM foi composta pelas variáveis apresentadas e mais as medidas de dispersão associadas as variáveis meteorológicas, totalizando 35 unidades. Esta estrutura foi utilizada para realizar uma estimativa de disponibilidade instantânea na pastagem,onde as variáveis meteorológicas fossem conhecidas, podendo-se obter uma estimativa da altura e do percentual de capim-annoni no momento da análise. Porém, em um segundo momento, foi assumido um cenário onde fosse realizada uma predição de estimativa para uma data futura, e por isso as variáveis "altura média" e "percentual de annoni" foram retiradas da estrutura de entrada do modelo, uma vez que elas não são conhecidas previamente, totalizando então 33 unidades. Nesse caso, supôs-se também que os dados meteorológicos fossem substituídos por uma previsão do período que se deseja estimar. Os resultados para ambos cenários são apresentados no Capítulo 4

A Figura 15 apresenta a especificação do modelo de predição proposto para a atividade de previsão de estimativa para períodos futuros, com as 33 variáveis de entrada. Alguns elementos foram omitidos e substituídos por três pontos na vertical para facilitar a visualização.

Figura 15 – Modelo de RNA LSTM proposto para predição de massa de forragem



Fonte: Autor (2019)

A quantidade de camadas ocultas, número de unidades LSTM em cada camada e a função de ativação foram definidos empiricamente, ou seja, foram realizados ensaios, variando-se esses parâmetros, para posterior análise do erro médio. Assim, ao final foram fixadas duas camadas ocultas, a primeira com 30 unidades LSTM e a segunda com 15 unidades. A camada de saída tem a massa de forragem (MStotal) estimada. Foi utilizada a função de ativação padrão *tanh*, visto que produziu os melhores resultados.

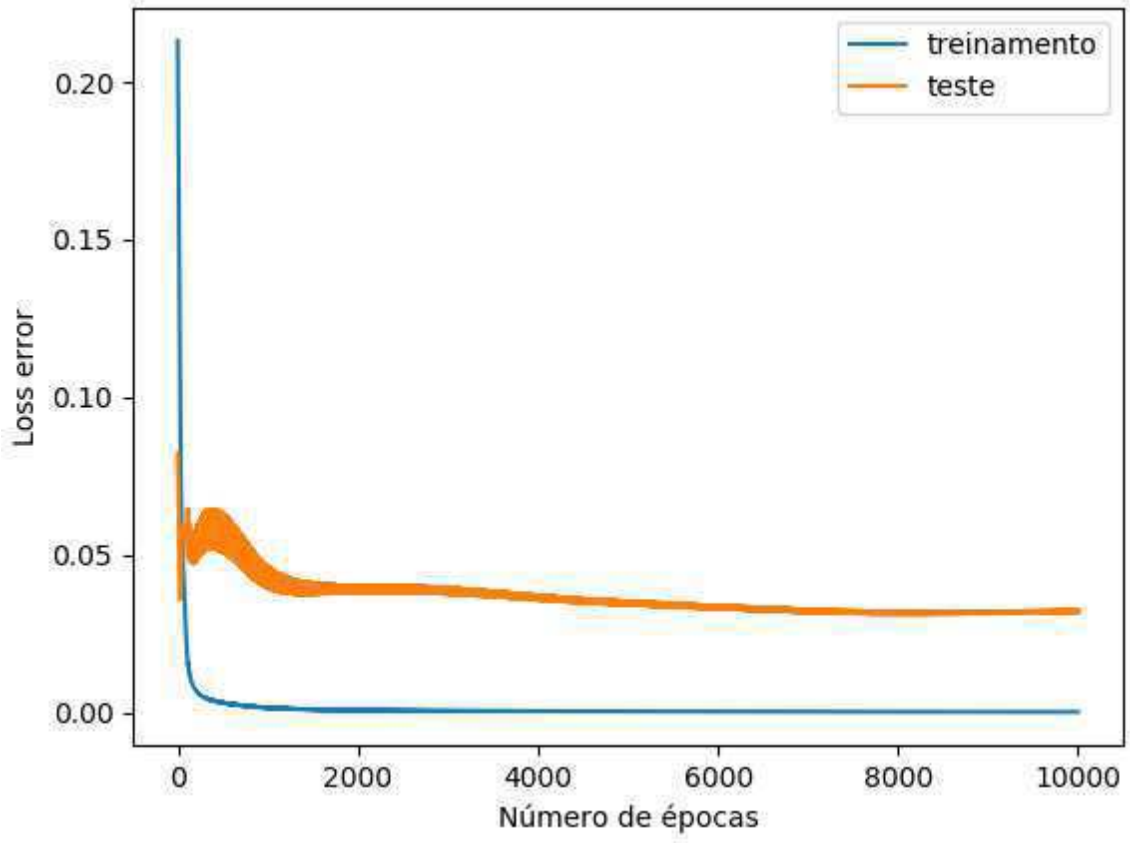
Para o processo de treinamento e testes os dados foram normalizados e inseridos sequencialmente. O conjunto de testes foi definido de forma a ter amostrado no mínimo um ano completo com o objetivo de ter representadas todas as estações completas deste

ano. Os demais dados foram utilizados para o treinamento realizado em 5000 épocas.

Foi dada ênfase na exploração dos dados do potreiro 20, por haver mais amostras válidas disponíveis no período entre janeiro de 2014 e agosto de 2018, fator que se mostrou limitante no processo de treinamento. Para o tratamento Infestado foram coletadas 39 amostras válidas, sendo 27 utilizadas para treino e 12 para testes. Para o tratamento MIRAPASTO foram coletadas 35 amostras válidas, sendo 24 utilizadas para treino e 11 para testes. Cada um dos tratamentos foi treinado e testado separadamente.

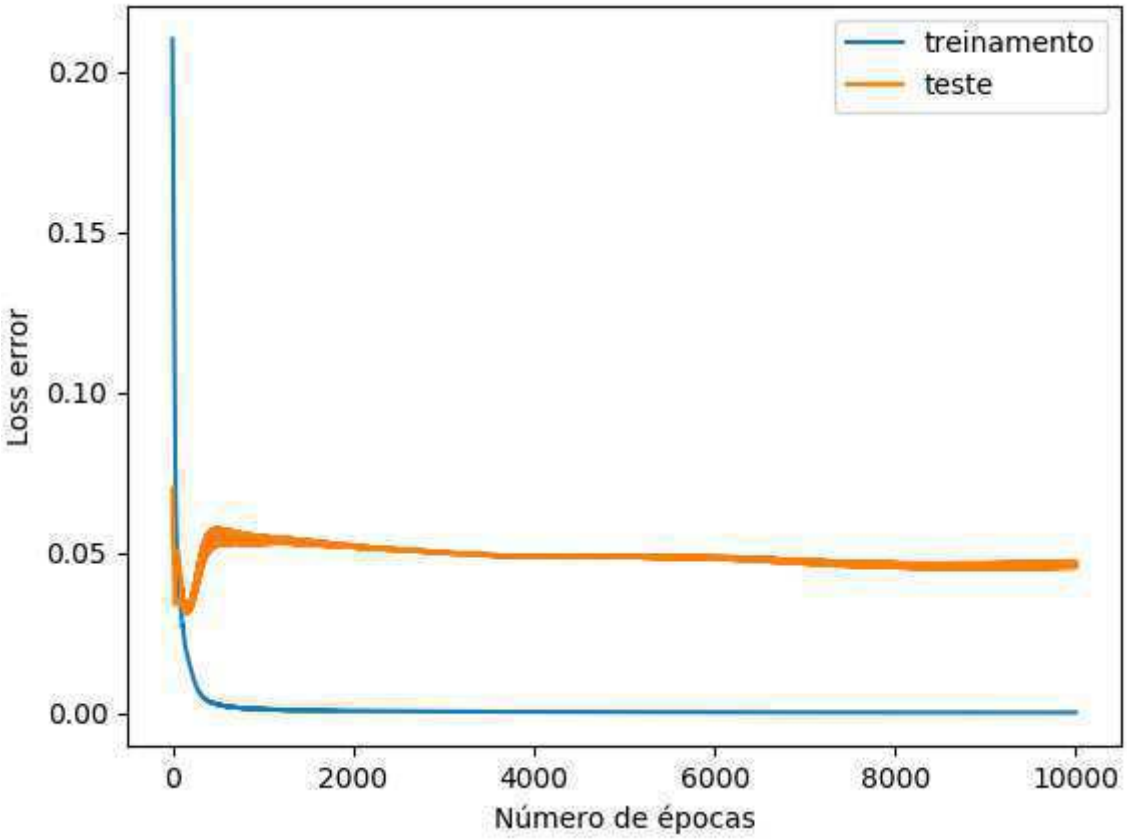
As Figuras 16 e 17 mostram o erro em função do número de épocas de treinamento para o tratamento Infestado e MIRAPASTO respectivamente. Foi observado que o erro apresenta uma queda grande nas primeiras épocas e posteriormente tende a ficar quase estável. Na prática, observou-se uma melhora dos resultados até as 5000 épocas e depois disso a melhora foi pouco significativa e não compensou o tempo extra de treinamento.

Figura 16 – Erro em função do número de épocas para o tratamento Infestado



Fonte: Autor (2019)

Figura 17 – Erro em função do número de épocas para o tratamento MIRAPASTO



Fonte: Autor (2019)

Dado o número reduzido de amostras disponíveis, o processo de treinamento e testes não demandam muitos recursos computacionais. Para um computador com um processador Intel Core i7 7700HQ, placa de vídeo Geforce GTX 1050 Ti 4 GB GDDR5, 16 GB de memória DDR4 2133 MHz, armazenamento primário com SSD M.2 Corsair Force MP500 240 GB, secundário com HD Toshiba MQ02ABD100H de 1TB e sistema operacional Windows 10 Home, o processo de treinamento com 5000 épocas demanda de 10 a 15 segundos.

4 RESULTADOS E DISCUSSÕES

Este capítulo apresenta os resultados alcançados a partir da metodologia adotada, além de discussões sobre os mesmos. As seções tratam da aplicação do modelo proposto em quatro perspectivas: na seção 4.1 são feitas estimativas com base em dados de entrada que supõe uma análise sendo feita na data da estimativa, podendo substituir a aplicação do método direto; na seção 4.2, foram omitidas as variáveis "altura média atual" e "percentual de anonni" para simular um cenário onde deseja-se fazer uma estimativa para uma data futura com base nos dados históricos e previsão meteorológica; na seção 4.3 é feita uma análise do impacto dos valores considerados *outliers*; na seção 4.4, o modelo é testado para estimar a massa de forragem em datas antigas com base em dados mais atuais. Ao final, a seção 4.5 apresenta uma análise de custos para mensurar o impacto da aplicação do modelo proposto na rentabilidade de uma propriedade rural.

4.1 Estimativa instantânea de massa de forragem

O resultado dos testes, utilizando o conjunto de 35 variáveis de entrada apresentadas, convertidos novamente para a escala de origem dos dados, são agrupados na Tabela 1, com erro médio quadrado igual a 1949 kg/ha para o tratamento Infestado (Inf) e 1405 kg/ha para o tratamento MIRAPASTO (MIRA). A Figura 18 apresenta os valores de referência (Real), obtidos pelo método direto de amostragem e os valores preditos (Predito) ao longo do tempo.

Tabela 1 – Valores de referência x estimativas realizadas pela RNA LSTM em kg/ha

Data	Inf Ref	Inf Est	Inf Ajuste	MIRA Ref	MIRA Est	MIRA Ajuste
26/07/2017	2372	3122	+31,6	1710	1661	-2,9
04/09/2017	2130	3771	+77,0	1667	3744	+124,6
02/10/2017	2306	4578	+98,5	1976	2392	+21,1
13/11/2017	8907	7201	-19,2	6004	3348	-44,2
11/12/2017	4937	6641	+34,5	-	-	-
15/01/2018	4259	5635	+32,3	2214	859	-61,2
19/02/2018	2799	3223	+15,1	1776	332	-81,3
23/03/2018	6266	2536	-59,5	1115	3095	+177,6
24/04/2018	5300	5791	+9,3	1932	3104	+60,7
08/06/2018	5495	2282	-58,5	2129	3059	43,7
13/07/2018	3296	2516	-23,7	992	671	-32,4
20/08/2018	4359	2300	-47,2	879	979	+11,4
EQM	-	**1949**		-	**1405**	

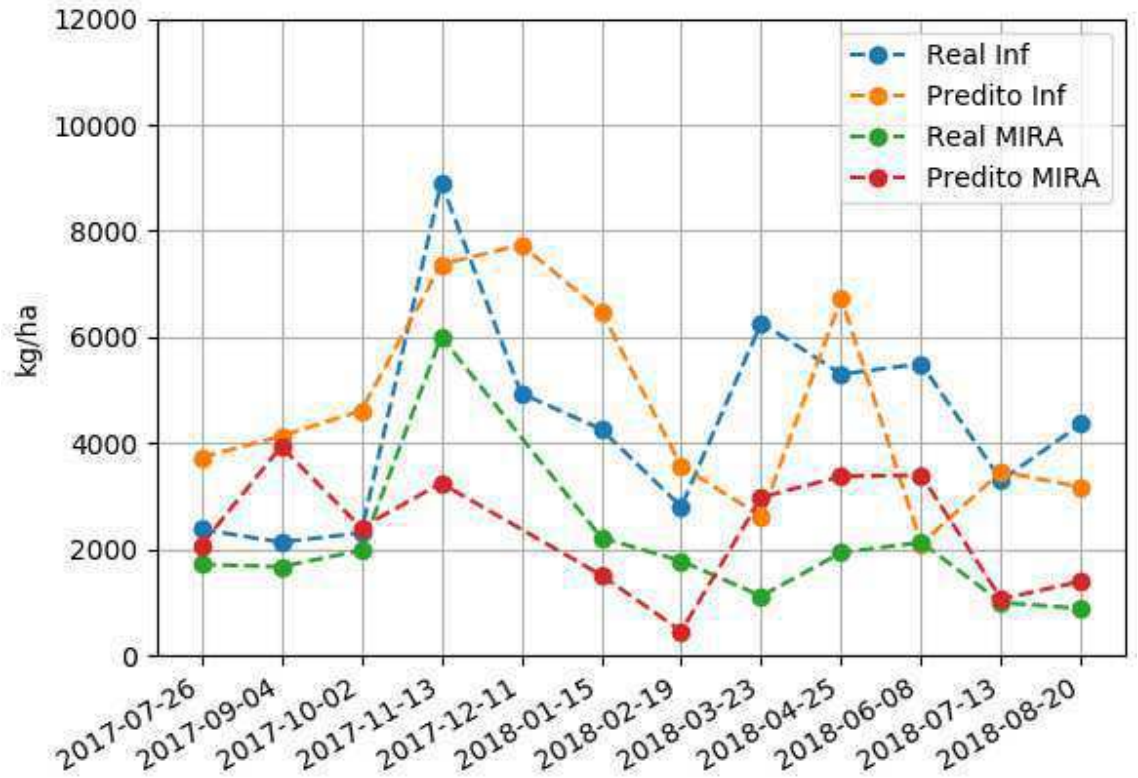
Data: data da avaliação de massa de forragem; **Inf Ref:** valor de referência para o tratamento Infestado; **Inf Est:** valor estimado pelo método proposto para o tratamento Infestado; **Inf Ajuste:** percentual de variação entre o valor de referência e o valor estimado para o tratamento Infestado; **MIRA Ref:** valor de referência para o tratamento MIRAPASTO; **MIRA Est:** valor estimado pelo método proposto para o tratamento MIRAPASTO; **MIRA Ajuste:** percentual de variação entre o valor de referência e o valor estimado para o tratamento MIRAPASTO.

Fonte: Autor (2019)

Apesar de o método do erro médio quadrado indicar um erro relativamente alto, ele é proporcional ao erro associado ao método de amostragem utilizado, que pode ser superior a mil kg/ha. Além disso, o método baseado em RNA LSTM consegue representar as tendências de aumento e diminuição da disponibilidade na maioria dos casos.

No período utilizado para validação do método, no tratamento Infestado, há uma maior ocorrência de erros para mais no período de clima mais quente, enquanto no período frio, o erro tende a ser para menos. Para saber se essa relação é estatisticamente válida, seria necessário analisar os resultados em períodos maiores, porém o número reduzido de amostras impossibilita tal atividade. Já para o tratamento MIRAPASTO, essa relação não fica evidente, sendo que os erros para menos estão mais frequentemente distribuídos ao

Figura 18 – Valores de referência da área experimental x Estimativas realizadas pela RNA LSTM ao longo do tempo



Fonte: Autor (2019)

longo do período, em comparação com o tratamento Infestado.

Também chama atenção a grande variação nos valores de referência de Outubro de 2017 para Novembro de 2017, que tiveram um incremento superior a três vezes em relação ao mês anterior. Tal evento não é desejado, já que quando o ajuste de carga animal é feito de forma adequada, em tese, a disponibilidade de forragem deve ter pouca variação ao longo do tempo. Isso mostra que, nesse caso, houve um alto erro na carga animal alocada no potreiro, ou o método de amostragem não conseguiu representar de forma fiel a disponibilidade de forragem na área.

4.2 Estimativa de disponibilidade futura

Os resultados dos testes, utilizando o conjunto de 33 variáveis de entrada apresentadas, convertidos novamente para a escala de origem dos dados, são sintetizados na Tabela 2, com erro médio quadrado igual a 2201 kg/ha para o tratamento Infestado e 1567 kg/ha para o tratamento MIRAPASTO.

Tabela 2 – Valores de referência x predições realizadas pela RNA LSTM em kg/ha

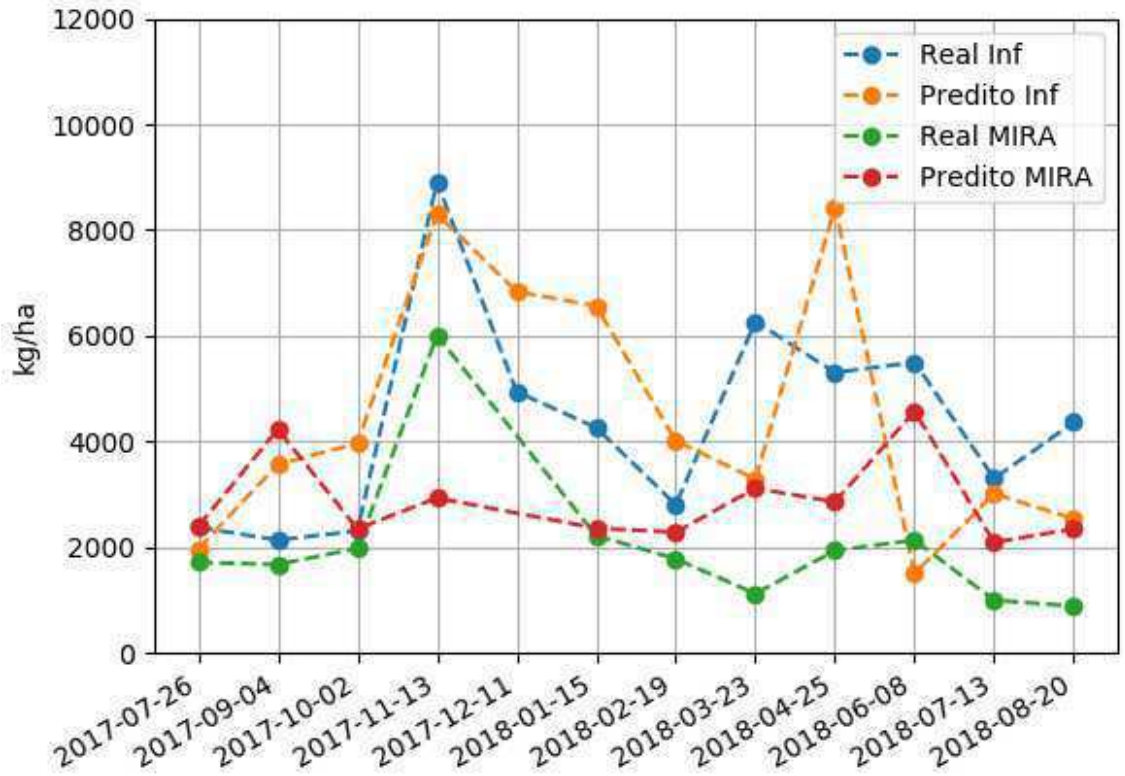
Data	Inf Ref	Inf Pred	Inf Ajuste	MIRA Ref	MIRA Pred	MIRA Ajuste
26/07/2017	2372	2047	-13,7	1710	1666	-2,6
04/09/2017	2130	4054	+90,3	1667	4037	+142,2
02/10/2017	2306	4028	+74,7	1976	2401	+21,5
13/11/2017	8907	8597	-3,5	6004	2658	-55,7
11/12/2017	4937	6938	+40,5	-	-	-
15/01/2018	4259	6683	+56,9	2214	2238	+1,1
19/02/2018	2799	4661	+66,5	1776	1769	-0,4
23/03/2018	6266	3383	-46,0	1115	2950	+164,6
24/04/2018	5300	8703	+64,2	1932	2889	+49,5
08/06/2018	5495	1758	-68,0	2129	4201	+97,3
13/07/2018	3296	2890	-12,3	992	1697	+71,1
20/08/2018	4359	2380	-45,4	879	1852	+110,7
EQM	-	**2201**		-	**1567**	

Data: data da avaliação de massa de forragem; **Inf Ref:** valor de referência para o tratamento Infestado; **Inf Est:** valor estimado pelo método proposto para o tratamento Infestado; **Inf Ajuste:** percentual de variação entre o valor de referência e o valor estimado para o tratamento Infestado; **MIRA Ref:** valor de referência para o tratamento MIRAPASTO; **MIRA Est:** valor estimado pelo método proposto para o tratamento MIRAPASTO; **MIRA Ajuste:** percentual de variação entre o valor de referência e o valor estimado para o tratamento MIRAPASTO.

Fonte: Autor (2019)

A Figura 19 apresenta os valores de referência e os valores preditos dispostos ao longo do tempo para o tratamento Infestado e MIRAPASTO. É possível observar que o método do erro médio quadrado indicou diminuição da acurácia do modelo, mostrando que as variáveis retiradas tem representatividade nos resultados, mas apesar disso, o método baseado em RNA LSTM ainda consegue representar as tendências de aumento e diminuição da disponibilidade na maioria dos casos.

Figura 19 – Valores de referência da área experimental x Predições realizadas pela RNA LSTM ao longo do tempo



Fonte: Autor (2019)

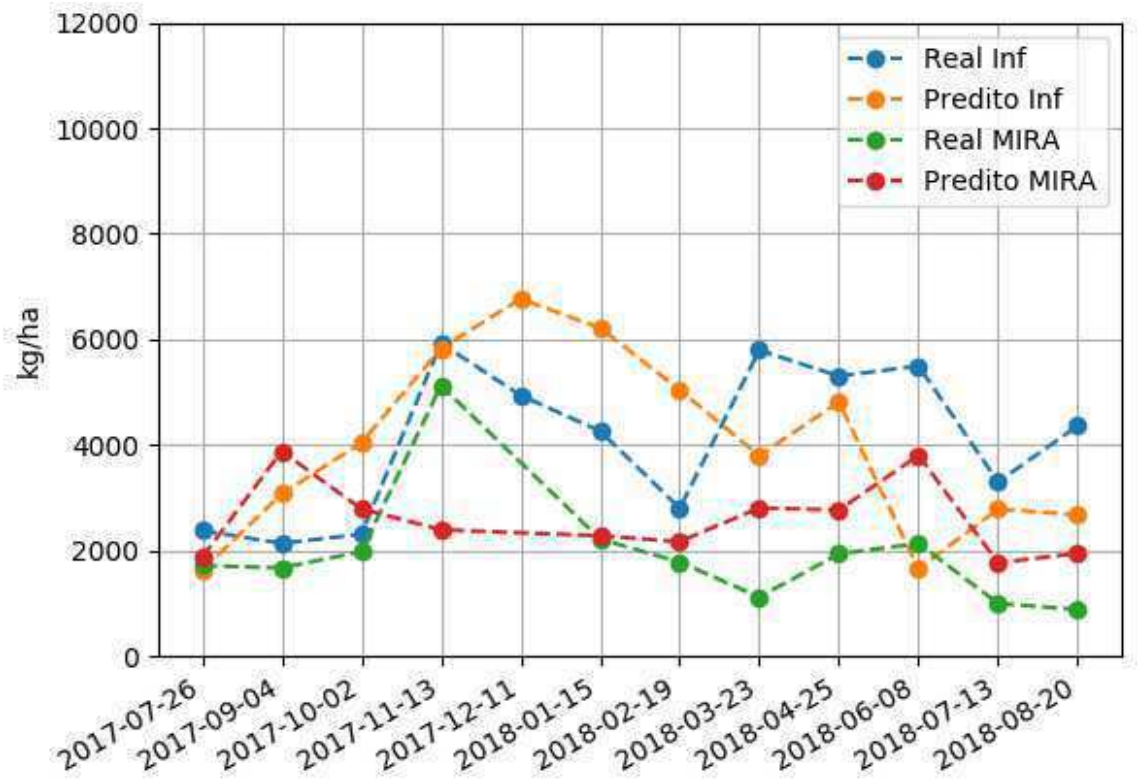
Neste caso, novamente há uma maior ocorrência de erros para mais nos meses de clima mais quente, enquanto no período frio o erro tende a ser para menos, no tratamento Infestado. Em contraste, para o tratamento MIRAPASTO, considerando todo o período avaliado, a maioria dos erros observados são por superestimação.

4.3 Análise de *outliers*

Nas análises realizadas, chamou atenção a alta variação de disponibilidade de massa de forragem entre a estimativa do dia 13 de Novembro de 2017, o mês anterior e o posterior. Por isso, levantou-se a hipótese de haver *outliers* nos dados causados por erros de digitação ou nos cálculos realizados nas planilhas. Para identificar os valores críticos que poderiam conter erros foi utilizado o método de Tukey para detecção de *outliers* (TUKEY, 1977). Porém, em análise realizada junto aos especialistas, descartaram-se ambas as hipóteses iniciais. Porém, concluiu-se que as estimativas poderiam se tratar de superestimação devido a imprecisão do método de coleta adotado.

Optou-se então por conduzir um experimento suavizando os *outliers* identificados pelo método de Tukey de forma a torná-los iguais ao limite máximo identificado como normal pelo método. Os resultados mostram diminuição no erro médio quadrado para o tratamento Infestado de 2201 kg/ha para 1918 kg/ha, enquanto para o tratamento MIRAPASTO o erro passou de 1567 kg/ha para 1322 kg/ha. A Figura 20 mostra os valores de referência suavizados e os valores preditos dispostos ao longo do tempo para o tratamento Infestado e MIRAPASTO.

Figura 20 – Valores de referência da área experimental x predições realizadas pela RNA LSTM ao longo do tempo com suavização dos *outliers*



Fonte: Autor (2019)

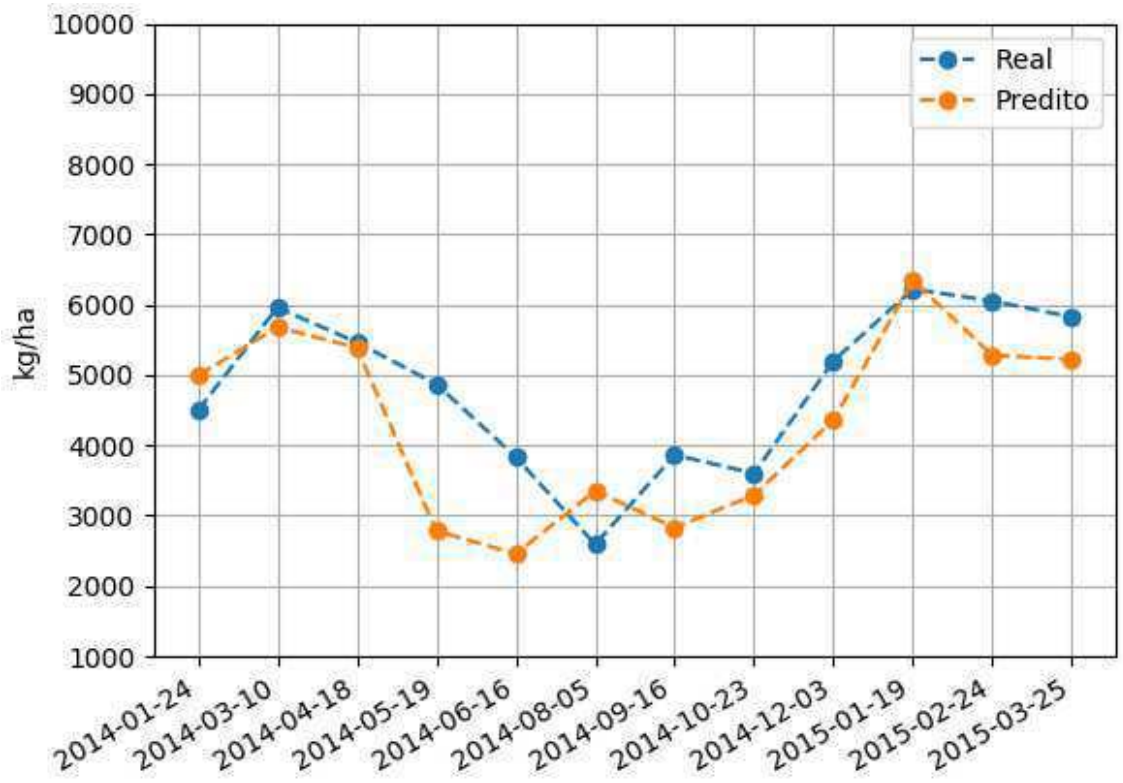
4.4 Estimativa de massa de forragem no passado

Uma vez verificada a acurácia do modelo desenvolvido para a predição da massa de forragem com base em dados de estimativas anteriores e dados meteorológicos, buscou-se verificar se seria possível inferir a disponibilidade de massa de foragem em datas mais antigas com base no treinamento realizado em dados mais recentes.

Para isso, o conjunto de dados suavizados foi introduzido na rede na ordem

temporal inversa sendo que os valores mais recentes constituíram o conjunto de treinamento de mesmo tamanho que os experimentos anteriores e os dados mais antigos foram usados para teste. Os resultados mostram um erro médio quadrado igual a 910 kg/ha para o tratamento Infestado e 1873 kg/ha para o tratamento MIRAPASTO. A Figura 21 apresenta os valores de referência suavizados e os valores preditos dispostos ao longo do tempo para o tratamento Infestado e a Figura 22 apresenta os esses valores para o tratamento MIRAPASTO.

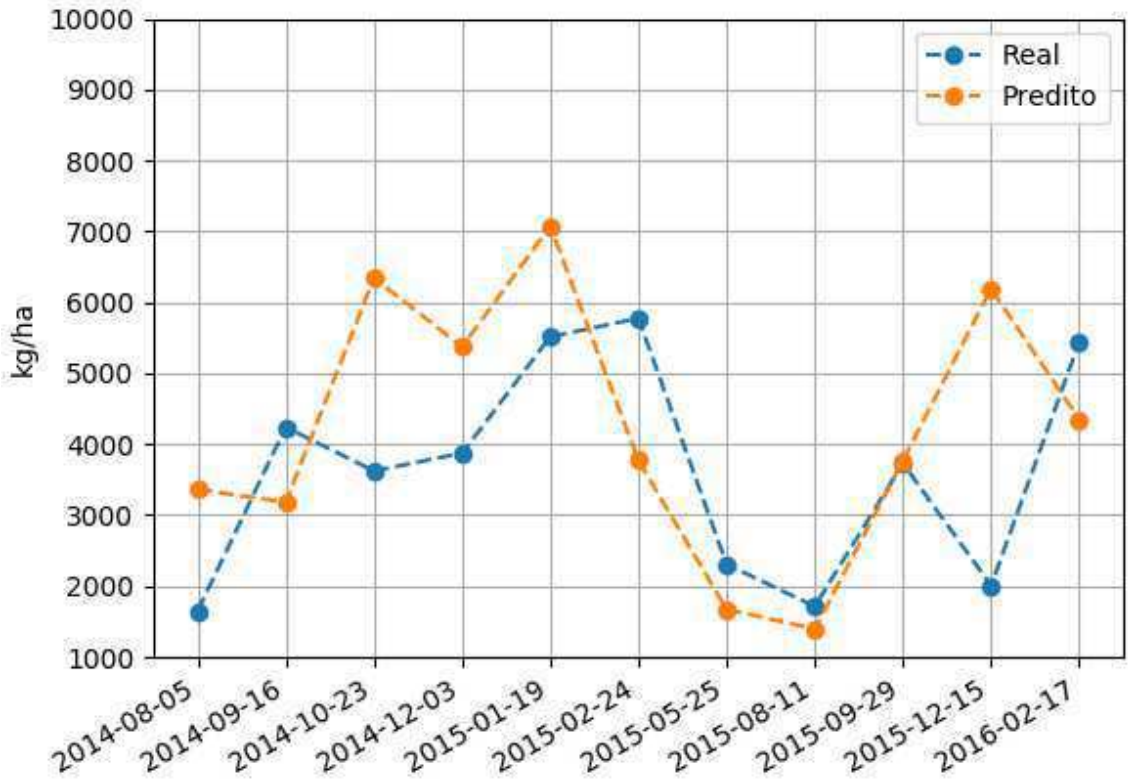
Figura 21 – Valores de referência da área experimental x predição dos valores mais antigos para o tratamento Infestado com suavização dos *outliers*



Fonte: Autor (2019)

As principais limitações para a precisão do método proposto são a quantidade de dados de treinamento, uma vez que idealmente deveria haver um volume maior de dados. Para aumentar o volume dos dados, poderiam ser buscados dados com um intervalo menor entre estimativas, ou que representassem o histórico de um período mais longo. Além disso, sabe-se que o método de amostragem utilizado está sujeito a um erro relativamente alto, principalmente quando se reduzem os pontos representativos dentro de uma área, a medida que cada amostra pontual coletada deve representar uma área com mais de um hectare.

Figura 22 – Valores de referência da área experimental x predição dos valores mais antigos para o tratamento MIRAPASTO com suavização dos *outliers*



Fonte: Autor (2019)

4.5 Análise de viabilidade

Usando como base o potreiro com tratamento Infestado, a Equação 2, e as estimativas realizadas na seção 4.1, a Tabela 3 mostra os valores de referência para a disponibilidade de massa de forragem no período do estudo, acompanhados pelos valores preditos através da metodologia proposta e os valores calculados com a taxa de acúmulo observada no mês anterior. Inicialmente, os valores preditos e os valores do mês anterior foram multiplicados por 11,5%, simulando os cenários nos quais seriam usados como base para um ajuste com o objetivo de otimizar o ganho por hectare (G/ha). Esses valores, multiplicados por 100 e divididos pelo valor de referência correspondente na data, mostram qual percentual de massa de forragem realmente foi ofertado com cada uma das abordagens. As colunas "% OP"e "% OA"apresentam, respectivamente, os valores preditos e os valores calculados com base na taxa de acúmulo do mês anterior.

Em média, foi ofertado aproximadamente 14,2% utilizando os valores preditos pelo método proposto neste trabalho, contra aproximadamente 13,3% utilizando como

base os valores calculados com base na taxa de acúmulo do mês anterior. Ambas estão próximas dos 11,5% buscados. Porém, é possível observar que os valores em "% OP"possuem menor variação e se mantêm mais constantemente próximos ao valor desejado. Isso pode ser comprovado pela variância e desvio padrão, que são de 29,2 e 5,4 respectivamente, para os valores calculados com base na metodologia proposta, contra 128,6 e 11,3 para os valores calculados com base na taxa de acúmulo do mês anterior.

Tabela 3 – Comparativo de potenciais ganhos por hectare (G/ha) com base nos valores preditos pela metodologia proposta e no método tradicional baseado na taxa de acúmulo do mês anterior para o potreiro Infestado

Data	Ref	Pred	Ant	% OP	% OA	G/ha P	GP	G/ha A	GA
04/09/2017	2131	3771	1492	20,4	8,1	37,95	2,91	132,91	10,2
02/10/2017	2306	4578	3158	22,8	15,8	0	0	119,54	13,76
13/11/2017	8908	7201	5240	9,3	6,8	141,2	10,83	120,15	9,22
11/12/2017	4937	6641	11583	15,5	27,0	122,72	11,77	0	0
15/01/2018	4259	5635	2395	15,2	6,5	125,43	12,03	116,56	11,18
19/02/2018	2799	3223	1485	13,2	6,1	140,81	12,35	111,88	9,81
23/03/2018	6266	2536	147	4,7	0,3	89,84	8,12	0	0
25/04/2018	5301	5791	16404	12,6	35,6	143,76	17,33	0	0
Total							**75,34**		**54,15**

Data: data da avaliação de massa de forragem; **Ref:** valor de referência para o tratamento Infestado; **Pred:** valor predito pelo método proposto para o tratamento Infestado; **Ant:** valor calculado pelo método tradicional para o tratamento Infestado; **% OP:** percentual que realmente seria ofertado usando o método proposto; **% OA:** percentual que realmente seria ofertado usando o método tradicional; **G/ha P:** Ganho potencial por ano para o método proposto; **G/ha A:** Ganho potencial por ano para o método tradicional; **GP:** Ganho potencial ponderado pelo número de dias para o método proposto; **GA:** Ganho potencial ponderado pelo número de dias para o método tradicional.

Fonte: Autor (2019)

Utilizando a Equação 2 é possível estimar o G/ha no período para cada um dos cenários. Na Tabela 3, esses valores são apresentados como "G/ha P", para as estimativas de G/ha para o ajuste realizado com base nos valores preditos, e "G/ha A" representa os ganhos estimados ao se realizar o ajuste com base nos valores do mês anterior.

É importante ressaltar que como a Equação 2 representa uma parábola com concavidade virada para baixo e ponto mais alto em 11,5%, quando ofertado um

percentual consideravelmente maior ou menor ao valor ótimo, a equação indica valores de ganhos negativos, o que não é verdade, pelo menos para ofertas superiores, uma vez que os animais não perdem peso no período do ano utilizado na análise quando há oferta de alimento abundante. Dessa forma, por mais que o G/ha não esteja otimizado ele nunca deveria ser negativo, o que mostra uma limitação da Equação 2. Por esse motivo, valores negativos foram desconsiderados e igualados a zero na análise.

Sabendo o número de dias no intervalo entre as amostras e calculando sua representatividade no período de um ano (número de dias / 365), é possível calcular a estimativa de ganho médio por hectare em cada intervalo do estudo. Na Tabela 3, a coluna GP mostra a estimativa de ganho por hectare com o ajuste com base nos valores preditos pela metodologia proposta neste trabalho e a coluna GA mostra essa estimativa para um ajuste realizado com base em uma avaliação do mês anterior. O acumulado dessas estimativas mostra que, em um período de 277 dias, o método proposto poderia proporcionar ao produtor um ganho de 21,18 kg de peso vivo a mais por hectare, dessa forma, em um ano o incremento seria de 27,9 kg/ha.

Considerando um estudo de caso para o estado do RS onde o tamanho médio das propriedade rurais é de 59,39 ha segundo o censo agropecuário do Instituto Brasileiro de Geografia e Estatistica (IBGE) de 2017 e o preço de bovinos gordos pode passar de R$ 5,00 por kg de peso vivo, o produtor poderia ter um incremento nas suas receitas anuais de até R$ 8.288,46, valor que é superior ao preço médio para aquisição de um hectare de terra com pastagem de alto suporte na região da campanha do RS o qual, segundo o Instituto Nacional de Colonização e Reforma Agrária (INCRA), tem sido negociado por R$ 6.879,21 em média (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE, 2019; NÚCLEO DE ESTUDOS EM SISTEMAS DE PRODUÇÃO DE BOVINOS DE CORTE E CADEIA PRODUTIVA - NESPRO - UFRGS, 2019; INSTITUTO NACIONAL DE COLONIZAÇÃO E REFORMA AGRÁRIA - INCRA, 2019). Além disso, considerando que a produção média de peso vivo/ha/ano em sistemas de recria e terminação no RS é da ordem de 70kg por ano (NABINGER *et al.*, 2009), em sistemas que não fazem ajustes com base em metodologias de amostragem de pastagens, o incremento nas receitas pode chegar a R$ 8.691,73 por ano.

A mesmo método foi utilizado para analisar os valores do tratamento MIRAPASTO. Os valores apresentados na Tabela 4 mostram que, em média, foi ofertado aproximadamente 12,5% utilizando os valores preditos pelo método proposto neste trabalho, contra aproximadamente 7,1% utilizando como base os valores calculados com

base na taxa de acúmulo do mês anterior. Novamente, a média de "% OP"se aproximou dos 11,5%, enquanto a média de "% OA"ficou consideravelmente abaixo deste valor, o que provavelmente causaria baixa produtividade dos animais devido a pouca oferta de forragem. Já a diferença nos ganhos é de 6,47 kg/ha, o que representa 8,53 kg/ha em um ano, valor que poderia proporcionar um incremento de até R$ 2.533,28.

Tabela 4 – Comparativo de potenciais ganhos por hectare (G/ha) com base nos valores preditos pela metodologia proposta e no método tradicional baseado na taxa de acúmulo do mês anterior para o potreiro MIRAPASTO

Data	Ref	Pred	Ant	% OP	% OA	G/ha P	GP	G/ha A	GA
04/09/2017	1667	3744	1784	25,8	12,3	0	0	144,56	11,09
02/10/2017	1976	2392	1877	13,9	10,9	136,66	0	145,95	16,79
13/11/2017	6004	3348	2757	6,4	5,3	115,89	20	100,03	17,27
15/01/2018	2214	859	0	4,4	-	86,51	8,3	0	0
19/02/2018	1776	332	0	2,1	-	38,87	3,41	0	0
23/03/2018	2225	3095	376	16,0	1,9	116,54	10,54	33,9	0
25/04/2018	1932	3104	887	18,5	5,3	77,82	9,38	100	0
Total							**51,62**		**45,15**

Data: data da avaliação de massa de forragem; **Ref:** valor de referência para o tratamento Infestado; **Pred:** valor predito pelo método proposto para o tratamento Infestado; **Ant:** valor calculado pelo método tradicional para o tratamento Infestado; **% OP:** percentual que realmente seria ofertado usando o método proposto; **% OA:** percentual que realmente seria ofertado usando o método tradicional; **G/ha P:** Ganho potencial por ano para o método proposto; **G/ha A:** Ganho potencial por ano para o método tradicional; **GP:** Ganho potencial ponderado pelo número de dias para o método proposto; **GA:** Ganho potencial ponderado pelo número de dias para o método tradicional.

Fonte: Autor (2019)

Cabe destacar que a qualidade da pastagem não foi levada em conta nesta análise. Assim, por possuir maior volume de forragem verde, a área infestada por annoni apresentou maiores ganhos calculados, o que pode não ser verdade devido à menor valor nutritivo. Além disso, o método de amostragem direta estratificada demanda treinamento de mão de obra e maior orientação técnica, já que envolve cálculos e avaliações matemáticas para se chegar ao número mais preciso da estimativa de massa de forragem (SALMAN; SOARES; CANESIN, 2006). O custo associado a estas atividades deve ser levado em conta na análise de viabilidade.

5 CONSIDERAÇÕES FINAIS

Após a realização deste trabalho, é possível afirmar que o modelo desenvolvido, baseado em técnicas de aprendizagem de máquina, foi capaz de estimar a massa de forragem com significativa acurácia, tanto a disponibilidade instantânea quanto a do mês subsequente ou anterior, a partir de séries temporais de dados amostrais da pastagem e meteorológicos. Esta afirmação é fundamentada na constatação de que o método apresenta um erro médio de mesma magnitude do erro associado ao método de amostragem utilizado para coletar os dados e, portanto, é mais eficiente que o método tradicional, baseado na taxa de acúmulo do mês anterior. Em função desta comprovação, constitui-se como potencial alternativa inovadora para a realização do ajuste de carga animal em pastagens.

Uma dificuldade encontrada na análise dos resultados foi a ausência, na revisão bibliográfica sistemática realizada, de trabalhos correlatos que descrevessem modelos completos e resultados numéricos passíveis de comparação. Assim o presente trabalho poderá se tornar uma referência para futuros trabalhos que visem estimar a massa de forragem.

Os resultados mostram que o modelo proposto apresenta erro menor para situações onde há maior homogeneidade na massa de forragem amostrada, como é o caso da área infestada por capim-annoni, evidenciado pelo maior incremento na produtividade neste cenário. Além disso, acredita-se que afetam negativamente a precisão do modelo a quantidade ainda pequena de amostras usadas para o treinamento e a imprecisão do método de coleta dos dados. Tratando essas limitações através da realização da coleta durante um período mais longo e em intervalos menores, assim como o aumento do número de pontos de coleta, seria possível melhorar significativamente os resultados deste trabalho.

A construção de um banco de dados espacial como modelo de persistência, unificando dados de diferentes formatos de planilhas e o desenvolvimento de ferramentas computacionais, que facilitam a adição de novos registros ao banco, também representam contribuições relevantes deste trabalho. O banco de dados espacializado permite o cálculo de médias de disponibilidade de forragem, ponderadas pela representatividade da subárea a qual representam na área total do potreiro e o registro da posição exata onde cada amostra foi coletada. Além disso, pode ser utilizado como base para a construção de uma interface para facilitar o registro de novos dados, além de oferecer recursos para a

visualização de gráficos e relatórios, constituindo um FMIS para gestão de pastagens, capaz de auxiliar a tomada de decisão em relação ao ajuste de taxa de lotação animal em um sistema de pecuária de precisão.

Como trabalhos futuros, pode-se buscar a evolução do sistema para que ele seja capaz de receber dados de sistemas pecuários estruturados de diferentes formas, espacializá-los, e buscar periodicamente dados reais de estações meteorológicas próximas e de modelos de previsão meteorológica, tais como os dados de previsão meteorológica diária disponibilizados pelo *National Oceanic and Atmospheric Administration* (NOAA).

Além disso, poderá ser iniciado o desenvolvimento da interface do Sistema de Apoio à Decisão que auxiliará na gestão de pastagens. Posteriormente, pode-se definir uma estrutura de rede visando tornar o sistema capaz de conectar produtores ou experimentos, em diferentes níveis. Para tal, vislumbra-se uma estrutura de *Fog Computing*, supondo que na camada terminal estariam os produtores inserindo dados na interface do SAD, ou mesmo a concepção de um sistema de coleta automática de dados por meio de sensores de imagem em Veículos Aéreos Não Tripulados, por exemplo, além da integração dos dados coletados pelos diversos sensores das estações automáticas do Instituto Nacional de Meteorologia, compondo um cenário de IoT. Na camada de *fog*, podem ser usados SGBDs com extensões espaciais para reunir dados de produtores geograficamente próximos e que tenham composições de pastagem semelhantes, já que as variáveis de clima são similares para ambos, agregando assim, informações regionais ao sistema que podem melhorar a precisão do modelo, ou mesmo fornecer estimativas para usuários que não possuem dados históricos de suas propriedades, incentivando assim o uso do SAD para cada vez mais pessoas. Na camada de *cloud*, os dados podem ser consolidados compondo uma base de dados nacional ou mundial, por exemplo, com alto potencial de exploração para descoberta de conhecimento e aumento da produtividade. Essa prática possibilitaria que o sistema se tornasse mais robusto, a medida que seria alimentado com um grande volume de dados, trazendo a necessidade de incorporação de conceitos de *Big Data* para garantir a qualidade dos dados, além do seu eficiente armazenamento, processamento e análise

Por fim, é possível considerar que a proposta deste trabalho irá impulsionar o desenvolvimento de diversos trabalhos futuros complementares que irão contribuir significativamente com o estado da arte, a medida que incorpora diversos conceitos de *Smart Farming* e tecnologias modernas, aplicadas à agricultura e a pecuária de precisão.

REFERÊNCIAS

ABRAMIDES, P. *et al.* Estimativa da quantidade de forragem em pastagens de capins prostrados tropicais, através da medida da altura média da vegetação. **Zootecnia, Nova Odessa**, v. 20, n. 1, p. 17–41, 1982.

ALI, I. *et al.* Modeling managed grassland biomass estimation by using multitemporal remote sensing data: A machine learning approach. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 10, n. 7, p. 3254–3264, July 2017. ISSN 1939-1404.

AMANDEEP *et al.* Smart farming using iot. In: **2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)**. Vancouver: [s.n.], 2017. p. 278–280.

ANACONDA. **The enterprise data science platform for**. 2019. Disponível em: https://www.anaconda.com/. Acesso em: 01 mar. 2019.

BARRETT, P.; LAIDLAW, A.; MAYNE, C. Grazegro: a european herbage growth model to predict pasture production in perennial ryegrass swards for decision support. **European Journal of Agronomy**, v. 23, n. 1, p. 37 – 56, 2005. ISSN 1161-0301. Disponível em: http://www.sciencedirect.com/science/article/pii/S1161030104000875. Acesso em: 01 mar. 2019.

BATINI, C. *et al.* Methodologies for data quality assessment and improvement. **ACM Comput. Surv.**, ACM, New York, NY, USA, v. 41, n. 3, p. 16:1–16:52, jul. 2009. ISSN 0360-0300. Disponível em: http://doi.acm.org/10.1145/1541880.1541883. Acesso em: 01 mar. 2019.

BATINI, C.; SCANNAPIECO, M. **Data and information quality: dimensions, principles and techniques**. [S.l.]: Springer, 2016.

BONOMI, F. *et al.* Fog computing and its role in the internet of things. In: ACM. **Proceedings of the first edition of the MCC workshop on Mobile cloud computing**. [S.l.], 2012. p. 13–16.

BRAGA, A. de P.; FERREIRA, A. C. P. de L.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. [S.l.]: LTC Editora, 2007.

BREMM, C. *et al.* Estimativa de forragem por sensor remoto ativo de superfície em pastagens naturais do bioma pampa. **XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, João Pessoa, PB, 2015.

BURSTEIN, C. W. H. a. F. **Handbook on Decision Support Systems 1: Basic Themes**. 1. ed. [S.l.]: Springer-Verlag Berlin Heidelberg, 2008. (International Handbooks Information System). ISBN 3540487158,9783540487159,9783540487166.

BUYYA, R. *et al.* Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility. **Future Generation Computer Systems**, v. 25, n. 6, p. 599 – 616, 2009. ISSN 0167-739X. Disponível em: http://www.sciencedirect.com/science/article/pii/S0167739X08001957. Acesso em: 01 mar. 2019.

CARVALHO, P. C. d. F. *et al.* Do bocado ao pastoreio de precisão: compreendendo a interface planta-animal para explorar a multi-funcionalidade das pastagens. **Revista brasileira de zootecnia= Brazilian journal of animal science. Viçosa, MG. Vol. 38, supl. especial (2009), p. 109-122**, 2009.

CASANOVA, M. A. *et al.* **Banco de dados geográficos**. [S.l.]: MundoGEO Curitiba, 2005.

CHAN, S. H. *et al.* Decision support system (dss) use and decision performance: Dss motivation and its antecedents. **Information & Management**, v. 54, n. 7, p. 934 – 947, 2017. ISSN 0378-7206. Disponível em: http://www.sciencedirect.com/science/article/pii/S0378720617300617. Acesso em: 01 mar. 2019.

CHLINGARYAN, A.; SUKKARIEH, S.; WHELAN, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. **Computers and Electronics in Agriculture**, Elsevier, v. 151, p. 61–69, 2018.

CÓSER, A.; MARTINS, C.; DERESZ, F. Metodologias para estimativa da produção de forragem em pastagem de capim-elefante. **Embrapa Minas Gerais. Circular Técnica**, Embrapa Gado de Leite, 2002.

CÓSER, A. C. *et al.* Métodos para estimar a forragem consumível em pastagem de capim-elefante. **Pesquisa Agropecuária Brasileira**, v. 38, n. 7, p. 875–879, 2003.

DAIRYNZ. **Pasture growth forecaster**. 2019. Disponível em: http://pasture-growth-forecaster.dairynz.co.nz/. Acesso em: 01 mar. 2019.

DASORIYA, R. A review of big data analytics over cloud. In: **2017 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)**. [S.l.: s.n.], 2017. p. 1–6.

DATE, C. **Introduction to database systems**. 8ed. ed. [S.l.]: Pearson, 2004.

ELMASRI, S. B. N. R. **Fundamentals of Database Systems**. 4. ed. [S.l.]: Pearson Education, 2003.

EMBRAPA. **Aplicador seletivo de herbicida Campo Limpo**. 2019. Disponível em: https://www.embrapa.br/busca-de-solucoes-tecnologicas/-/produto-servico/558/aplicador-seletivo-de-herbicida-campo-limpo. Acesso em: 01 mar. 2019.

ESTRADA, C.; JÚNIOR, D. N.; REGAZZI, A. Efeito do número e tamanho do quadrado nas estimativas pelo botanal da composição botânica e disponibilidade de matéria seca de pastagens cultivadas. **Revista da Sociedade Brasileira de Zootecnia**, v. 20, n. 5, p. 483–493, 1991.

FAGUNDES, P. B.; MACEDO, D. D. J. de; FREUND, G. P. A produção científica sobre qualidade de dados em big data: um estudo na base de dados web of science. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 16, n. 1, p. 194–210, 2017.

FAYYAD, U. The kdd process for extracting useful knowledge from volumes of data. **Communications of the ACM**, ACM, v. 39, n. 11, p. 27–34, 1996.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS - FAO. **How to feed the world in 2050**. 2019. Disponível em: http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf. Acesso em: 01 mar. 2019.

FOUNTAS, S. *et al.* Farm management information systems: Current situation and future perspectives. **Computers and Electronics in Agriculture**, v. 115, p. 40 – 50, 2015. ISSN 0168-1699. Disponível em: http://www.sciencedirect.com/science/article/pii/S0168169915001337. Acesso em: 01 mar. 2019.

FU, Q.; EASTON, J. M. Understanding data quality: Ensuring data quality by design in the rail industry. In: **2017 IEEE International Conference on Big Data (Big Data)**. [S.l.: s.n.], 2017. p. 3792–3799.

GIL, A. C. **Métodos e técnicas de pesquisa social**. [S.l.]: 6. ed. Ediitora Atlas SA, 2008.

GUNTHER, W. A. *et al.* Debating big data: A literature review on realizing value from big data. **The Journal of Strategic Information Systems**, v. 26, n. 3, p. 191 – 209, 2017. ISSN 0963-8687. Disponível em: http://www.sciencedirect.com/science/article/pii/S0963868717302615. Acesso em: 01 mar. 2019.

HANDOYO, I. T.; SENSUSE, D. I. Knowledge-based systems in decision support context: A literature review. In: **2017 4th International Conference on New Media Studies (CONMEDIA)**. [S.l.: s.n.], 2017. p. 81–86.

HANRAHAN, L. *et al.* Pasturebase ireland: A grassland decision support system and national database. **Computers and Electronics in Agriculture**, v. 136, p. 193 – 201, 2017. ISSN 0168-1699. Disponível em: http://www.sciencedirect.com/science/article/pii/S0168169916304781. Acesso em: 01 mar. 2019.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. Disponível em: https://doi.org/10.1162/neco.1997.9.8.1735. Acesso em: 01 mar. 2019.

HU, P. *et al.* Survey on fog computing: architecture, key technologies, applications and open issues. **Journal of Network and Computer Applications**, v. 98, p. 27 – 42, 2017. ISSN 1084-8045. Disponível em: http://www.sciencedirect.com/science/article/pii/S1084804517302953. Acesso em: 01 mar. 2019.

INMET. **Estações Automáticas**. 2019. Disponível em: http://www.inmet.gov.br/portal/index.php?r=estacoes/estacoesautomaticas. Acesso em: 01 mar. 2019.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Censo agro 2017**. 2019. Disponível em: https://censoagro2017.ibge.gov.br/templates/censo_agro/resultadosagro/pdf/RS.pdf. Acesso em: 01 mar. 2019.

INSTITUTO NACIONAL DE COLONIZAÇÃO E REFORMA AGRÁRIA - INCRA. **Relatório de Análise de Mercado de Terras no Estado do Rio Grande do Sul - RAMT/RS**. 2019. Disponível em: http://www.incra.gov.br/sites/default/files/uploads/relatorios-analise-mercados-terras/sr-11-rio-grande-do-sul/ramt_sr11.pdf. Acesso em: 01 mar. 2019.

JANG, J. S. R. Anfis: adaptive-network-based fuzzy inference system. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 23, n. 3, p. 665–685, May 1993. ISSN 0018-9472.

JAYARAMAN, P. P. *et al.* Internet of things platform for smart farming: Experiences and lessons learnt. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 16, n. 11, p. 1884, 2016.

JESUS, H. A. d.; RESENDE, A. M. P. d.; ZAMBALDE, A. L. **Revisão Sistemática De Engenharia De Software Experimental In Vitro: Uma Análise Preliminar**. Dissertação (Mestrado) — Universidade Federal de Lavras, 2013.

JOTHI, N.; RASHID, N.-A. A.; HUSAIN, W. Data mining in healthcare - a review. **Procedia Computer Science**, v. 72, p. 306 – 313, 2015. ISSN 1877-0509. The Third Information Systems International Conference 2015. Disponível em: http://www.sciencedirect.com/science/article/pii/S1877050915036066. Acesso em: 01 mar. 2019.

KAMILARIS, A.; KARTAKOULLIS, A.; PRENAFETA-BOLDÚ, F. X. A review on the practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, v. 143, p. 23 – 37, 2017. ISSN 0168-1699. Disponível em: http://www.sciencedirect.com/science/article/pii/S0168169917301230. Acesso em: 01 mar. 2019.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. **Computers and Electronics in Agriculture**, Elsevier, v. 147, p. 70–90, 2018.

KAVAKIOTIS, I. *et al.* Machine learning and data mining methods in diabetes research. **Computational and Structural Biotechnology Journal**, v. 15, p. 104 – 116, 2017. ISSN 2001-0370. Disponível em: http://www.sciencedirect.com/science/article/pii/S2001037016300733. Acesso em: 01 mar. 2019.

KAYACAN, E.; KHANESAR, M. A.; MENDEL, J. M. **Fuzzy neural networks for real time control applications : concepts, modeling and algorithms for fast learning**. 1. ed. [S.l.]: Elsevier, 2015. ISBN 0128026871,978-0-12-802687-8,9780128027035,0128027037.

KERAS. **The Python deep learning library**. 2019. Disponível em: https://keras.io/. Acesso em: 01 mar. 2019.

KLINGMAN, D. L. *et al.* The cage method for determining consumption and yield of pasture herbage. **Journal of the American Society of Agronomy**, v. 35, p. 739–746, 1943.

KUKAR, M. *et al.* Agrodss: A decision support system for agriculture and farming. **Computers and Electronics in Agriculture**, 2018. ISSN 0168-1699. Disponível em: http://www.sciencedirect.com/science/article/pii/S0168169917314205. Acesso em: 01 mar. 2019.

LACA, E. A. Precision livestock production: tools and concepts. **Revista brasileira de zootecnia**, SciELO Brasil, v. 38, n. SPE, p. 123–132, 2009.

LAURENT, J. *et al.* Computing in the fog with reconfigurable gateways. In: **2018 IEEE International Symposium on Circuits and Systems (ISCAS)**. [S.l.: s.n.], 2018. p. 1–4.

LI, S.; XU, L. D.; ZHAO, S. 5g internet of things: A survey. **Journal of Industrial Information Integration**, 2018. ISSN 2452-414X. Disponível em: http://www.sciencedirect.com/science/article/pii/S2452414X18300037. Acesso em: 01 mar. 2019.

LOHAN, V.; SINGH, R. P. Research challenges for internet of things: A review. In: **2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)**. [S.l.: s.n.], 2017. p. 109–117.

LOPES, R. dos S. *et al.* Avaliação de métodos para estimação da disponibilidade de forragem em pastagem de capim-elefante1. **Rev. bras. zootec**, v. 29, n. 1, p. 40–47, 2000.

LU, Y.; PAPAGIANNIDIS, S.; ALAMANOS, E. Internet of things: A systematic review of the business literature from the user and organisational perspectives. **Technological Forecasting and Social Change**, 2018. ISSN 0040-1625. Disponível em: http://www.sciencedirect.com/science/article/pii/S0040162518301136. Acesso em: 01 mar. 2019.

MARASCHIN, G. *et al.* Native pasture, forage on offer and animal response. In: **XVIII International Grassland Congress**. [S.l.: s.n.], 1997. p. 26–27.

MARIMUTHU, R. *et al.* Design and development of a persuasive technology method to encourage smart farming. In: **2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)**. [S.l.: s.n.], 2017. p. 165–169.

MASSRUHÁ, S. M. F. S.; LEITE, M. d. A.; MOURA, M. F. Os novos desafios e oportunidades das tecnologias da informação e da comunicação na agricultura (agrotic). **Embrapa Informática Agropecuária-Capítulo em livro científico (ALICE)**, In: MASSRUHÁ, SMFS; LEITE, MA de A.; LUCHIARI JUNIOR, A.; ROMANI, LAS (Ed.). Tecnologias da informação e comunicação e suas relações com a agricultura. Brasília, DF: Embrapa, 2014. Cap. 1., 2014.

MATPLOTLIB. **Python 2D plotting library**. 2019. Disponível em: https://matplotlib.org/. Acesso em: 01 mar. 2019.

MERINO, J. *et al.* A data quality in use model for big data. **Future Generation Computer Systems**, v. 63, p. 123 – 130, 2016. ISSN 0167-739X. Modeling and Management for Big Data Analytics and Visualization. Disponível em: http://www.sciencedirect.com/science/article/pii/S0167739X15003817. Acesso em: 01 mar. 2019.

MIRKIN, B. Data analysis, mathematical statistics, machine learning, data mining: Similarities and differences. In: **2011 International Conference on Advanced Computer Science and Information Systems**. [S.l.: s.n.], 2011. p. 1–8.

MONOLITONIMBUS. **Rede Neural Artificial**. 2019. Disponível em: https://www.monolitonimbus.com.br/wp-content/uploads/2017/05/neuronios_rna.gif. Acesso em: 01 mar. 2019.

NABINGER, C. Manejo e produtividade das pastagens nativas do subtrópico brasileiro. **I Simpósio de Forrageiras e Pastagens**, p. 25–76, 2006.

NABINGER, C. *et al.* Produção animal com base no campo nativo: aplicações de resultados de pesquisa. **Campos Sulinos, conservação e uso sustentável da biodiversidade.(Eds VDP Pillar, SC Müller) pp**, p. 175–198, 2009.

NEVES, M. C. *et al.* Análise exploratória de dados de monitoramento da dinâmica do gado em uma pastagem natural invadida pelo capim-annoni. **XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, João Pessoa-PB, Brasil, 2015.

NIKKILä, R.; SEILONEN, I.; KOSKINEN, K. Software architecture for farm management information systems in precision agriculture. **Computers and Electronics in Agriculture**, v. 70, n. 2, p. 328 – 336, 2010. ISSN 0168-1699. Special issue on Information and Communication Technologies in Bio and Earth Sciences. Disponível em: http://www.sciencedirect.com/science/article/pii/S0168169909001859. Acesso em: 01 mar. 2019.

NOVKOVIC, N. *et al.* Farm management information systems. **Hellenic Association for Information and Communication Technologies in Agriculture Food and Environment (HAICTA)**, p. 705–712, 2015.

NÚCLEO DE ESTUDOS EM SISTEMAS DE PRODUÇÃO DE BOVINOS DE CORTE E CADEIA PRODUTIVA - NESPRO - UFRGS. **Histórico de índices do boi gordo**. 2019. Disponível em: http://www.ufrgs.br/nespro/historico_indices_boi_gordo.php. Acesso em: 01 mar. 2019.

O'GRADY, M. J.; O'HARE, G. M. Modelling the smart farm. **Information Processing in Agriculture**, v. 4, n. 3, p. 179 – 187, 2017. ISSN 2214-3173. Disponível em: http://www.sciencedirect.com/science/article/pii/S2214317316301287. Acesso em: 01 mar. 2019.

OPENPYXL. **A Python library to read/write Excel 2010 xlsx/xlsm files**. 2019. Disponível em: https://openpyxl.readthedocs.io/en/stable/. Acesso em: 01 mar. 2019.

PEREZ, N. Campo limpo: controle de plantas indesejáveis em pastagens. **Embrapa Pecuária Sul-Fôlder/Folheto/Cartilha (INFOTECA-E)**, Bagé: Embrapa Pecuária Sul, 2010. Disponível em: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/68020/1/CL-Ago-2010.pdf. Acesso em: 01 mar. 2019.

PEREZ, N. Método integrado de recuperação de pastagens mirapasto: foco capim-annoni. **Embrapa Pecuária Sul-Fôlder/Folheto/Cartilha (INFOTECA-E)**, Bagé: Embrapa Pecuária Sul, 2015. Disponível em: https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1023496. Acesso em: 01 mar. 2019.

PILLAR, V. Clima e vegetação. **Clima. UFRGS, Departamento de Botânica. Disponível em: http://ecoqua. ecologia. ufrgs. br**, 1995.

PIVOTO, D. *et al.* Scientific development of smart farming technologies and their application in brazil. **Information Processing in Agriculture**, v. 5, n. 1, p. 21 – 32, 2018. ISSN 2214-3173. Disponível em: http://www.sciencedirect.com/science/article/pii/S2214317316301184. Acesso em: 01 mar. 2019.

POSTGIS. **Spatial and geographic objects for PostgreSQL**. 2019. Disponível em: https://postgis.net/. Acesso em: 01 mar. 2019.

POSTGRESQL. **The world's most advanced open source database**. 2019. Disponível em: https://www.postgresql.org/. Acesso em: 01 mar. 2019.

PRIMAK, F. V. **Decisões com bi (business intelligence)**. [S.l.]: Editora Ciência Moderna, 2008.

PSYCOPG. **The most popular PostgreSQL adapter for the Python programming language**. 2019. Disponível em: http://initd.org/psycopg/. Acesso em: 01 mar. 2019.

PYPI. **ANFIS**. 2019. Disponível em: https://pypi.org/project/anfis/. Acesso em: 01 mar. 2019.

PYTHON. **Welcome to Python**. 2019. Disponível em: https://www.python.org/. Acesso em: 01 mar. 2019.

QGIS. **Um sistema de Informação Geográfica livre e aberto**. 2019. Disponível em: https://www.qgis.org/pt_BR/site. Acesso em: 01 mar. 2019.

QUEIROZ, G. R. de; MONTEIRO, A. M. V.; CÂMARA, G. Bancos de dados geográficos e sistemas nosql: onde estamos e para onde vamos. **Revista Brasileira de Cartografia**, n. 65/3, 2013.

RESENDE, M. D. Vilela de; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, Escola de Agronomia e Engenharia de Alimentos, v. 37, n. 3, 2007.

ROMERA, A. *et al.* Use of a pasture growth model to estimate herbage mass at a paddock scale and assist management on dairy farms. **Computers and Electronics in Agriculture**, v. 74, n. 1, p. 66 – 72, 2010. ISSN 0168-1699. Disponível em: http://www.sciencedirect.com/science/article/pii/S0168169910001237. Acesso em: 01 mar. 2019.

SAADEH, H.; ALMOBAIDEEN, W.; SABRI, K. E. Internet of things: A review to support iot architecture's design. In: **2017 2nd International Conference on the Applications of Information Technology in Developing Renewable Energy Processes Systems (IT-DREPS)**. [S.l.: s.n.], 2017. p. 1–7.

SADIQ, S.; INDULSKA, M. Open data: Quality over quantity. **International Journal of Information Management**, v. 37, n. 3, p. 150 – 154, 2017. ISSN 0268-4012. Disponível em: http://www.sciencedirect.com/science/article/pii/S0268401216309021. Acesso em: 01 mar. 2019.

SAGHEER, A.; KOTB, M. Time series forecasting of petroleum production using deep lstm recurrent networks. **Neurocomputing**, v. 323, p. 203 – 213, 2019. ISSN 0925-2312. Disponível em: http://www.sciencedirect.com/science/article/pii/S0925231218311639. Acesso em: 01 mar. 2019.

SALMAN, A.; SOARES, J.; CANESIN, R. Métodos de amostragem para avaliação quantitativa de pastagens. **Embrapa Rondônia. Circular Técnica**, Porto Velho: Embrapa Rondônia., 2006.

SCHAPENDONK, A. *et al.* Lingra, a sink/source model to simulate grassland productivity in europe. **European Journal of Agronomy**, v. 9, n. 2, p. 87 – 100, 1998. ISSN 1161-0301. Disponível em: http://www.sciencedirect.com/science/article/pii/S1161030198000276. Acesso em: 01 mar. 2019.

SCHULTE, L. G. *et al.* Sistema de apoio à decisão para pecuária de precisão: módulo para ajuste de taxa de lotação amparado por predição baseada em aprendizagem de máquina. **XV Simpósio Brasileiro de Sistemas de Informação (SBSI)**, SBC, 2019.

SENSE-T. **Pasture predictor**. 2019. Disponível em: http://dashboard.sense-t.org.au/. Acesso em: 01 mar. 2019.

SENTELHAS, P. C.; PEREIRA, A. R.; ANGELOCCI, L. R. **Meteorologia agrícola**. [S.l.]: ESALQ-Depto de Física e Meteorologia, 2000.

SENYO, P. K.; ADDAE, E.; BOATENG, R. Cloud computing research: A review of research themes, frameworks, methods and future research directions. **International Journal of Information Management**, v. 38, n. 1, p. 128 – 139, 2018. ISSN 0268-4012. Disponível em: http://www.sciencedirect.com/science/article/pii/S0268401217305923. Acesso em: 01 mar. 2019.

SERRANO, J. M. *et al.* Calibration of a capacitance probe for measurement and mapping of dry matter yield in mediterranean pastures. **Precision Agriculture**, Springer, v. 12, n. 6, p. 860–875, 2011.

SøRENSEN, C. *et al.* Conceptual model of a future farm management information system. **Computers and Electronics in Agriculture**, v. 72, n. 1, p. 37 – 47, 2010. ISSN 0168-1699. Disponível em: http://www.sciencedirect.com/science/article/pii/S0168169910000396. Acesso em: 01 mar. 2019.

STERGIOU, C. *et al.* Secure integration of iot and cloud computing. **Future Generation Computer Systems**, v. 78, p. 964 – 975, 2018. ISSN 0167-739X. Disponível em: http://www.sciencedirect.com/science/article/pii/S0167739X1630694X. Acesso em: 01 mar. 2019.

TEALAB, A. Time series forecasting using artificial neural networks methodologies: A systematic review. **Future Computing and Informatics Journal**, v. 3, n. 2, p. 334 – 340, 2018. ISSN 2314-7288. Disponível em: http://www.sciencedirect.com/science/article/pii/S2314728817300715. Acesso em: 01 mar. 2019.

THORNTHWAITE, C.; MATHER, J. **The Water Balance**. Drexel Institute of Technology, Laboratory of Climatology, 1955. (Publications in climatology). Disponível em: https://books.google.com.br/books?id=DTdtcgAACAAJ. Acesso em: 01 mar. 2019.

TKAC, M.; VERNER, R. Artificial neural networks in business: Two decades of research. **Applied Soft Computing**, v. 38, p. 788 – 804, 2016. ISSN 1568-4946. Disponível em: http://www.sciencedirect.com/science/article/pii/S1568494615006122. Acesso em: 01 mar. 2019.

TSENG, C. L.; LIN, F. J. Extending scalability of iot/m2m platforms with fog computing. In: **2018 IEEE 4th World Forum on Internet of Things (WF-IoT)**. [S.l.: s.n.], 2018. p. 825–830.

TUKEY, J. W. **Exploratory data analysis**. [S.l.]: Reading, 1977.

WANG, Y. *et al.* Probabilistic individual load forecasting using pinball loss guided lstm. **Applied Energy**, v. 235, p. 10 – 20, 2019. ISSN 0306-2619. Disponível em: http://www.sciencedirect.com/science/article/pii/S0306261918316465. Acesso em: 01 mar. 2019.

WATHES, C. *et al.* Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? **Computers and Electronics in Agriculture**, Elsevier, v. 64, n. 1, p. 2–10, 2008.

WOLFERT, S. *et al.* Big data in smart farming - a review. **Agricultural Systems**, v. 153, p. 69 – 80, 2017. ISSN 0308-521X. Disponível em: http://www.sciencedirect.com/science/article/pii/S0308521X16303754. Acesso em: 01 mar. 2019.

APÊNDICE A – PROCESSO DE ETL

Este apêndice apresenta o código desenvolvido na linguagem Python para realizar o processo de *Extract, Transform and Load* dos dados da planilha do ano de 2018 para o banco de dados.

```
import openpyxl
from openpyxl import Workbook
from openpyxl import load_workbook
from datetime import datetime
import psycopg2
import re
import statistics

now = datetime.now()

log = open('log.txt', 'a')
log.write("———— ExcecuÃ§Ã£o iniciada em: " + str(now) + "\n\n")

try:
  con = psycopg2.connect("host='localhost' port='5432' dbname='pastagem
      ' user='postgres' password='123456'")
  cur = con.cursor()

except:
  log.write("\nFalha na conexÃ£o com o BD!")

wb = load_workbook(filename='2018_Setembro.xlsx', data_only=True,
    read_only=False)
planilha=['P 20 INFEST','P 20 RECUP', 'P 21 INFEST', 'P 21 RECUP']

for o in planilha:

  pattern= re.compile(r'\d+')
  num= re.findall(pattern, o)

  pattern= re.compile(r'I')
  desc = re.findall(pattern, o)

  if not desc:
```

```
33      desc='R'
    else:
35      desc='I'

37  cur.execute("SELECT \"id\" from \"potreiro\" where \"nome\" like %s;"
        , (o.lower()+'%',))

39  try:
      idpotreiro = cur.fetchone()[0]
41    con.commit()

43  except:
      query =  "INSERT INTO potreiro (nome, numero, descricao) VALUES (%s
          , %s, %s) RETURNING id;"
45    data = (o.lower(), num[0], desc)

47    cur.execute(query, data)

49    idpotreiro = cur.fetchone()[0]
      con.commit()
51
    p=wb[o]
53  y = 0
    for x in range(9,99999999,16):
55    d = p.cell(row=1, column=x)

57    if d.value != None:

59      print (d.value)

61      dt=d.value

63      if type(dt) is not datetime:
            dt = None
65          log.write(o + " - Data invÃ¡lida: " + str(d.value) + " Na
              cÃ©lula : " + str(d) + "\n")

67      cur.execute("SELECT \"id\" from \"medicao\" where \"data\"= %s;",
            (dt,))

69      try:
```

```
        idmedicao = cur.fetchone()[0]
        con.commit()

    except:

        query =  "INSERT INTO medicao (data) VALUES (%s) RETURNING id;"
        data = (dt,)

        cur.execute(query, data)
        idmedicao = cur.fetchone()[0]
        con.commit()

    n = 6
    m = 14

    print("Dentro da Gaiola - " + o)

    matriz1 = []

    tam=-1
    for i in range(n):
        valido=0
        for j in range(m):
            if isinstance(p.cell(row=i+6, column=j+2+y).value, int)
                or isinstance(p.cell(row=i+6, column=j+2+y).value,
                float):
                valido=1

            else:
                p.cell(row=i+6, column=j+2+y).value=None

        if valido==1:
            tam+=1
            matriz1.append([])

            for j in range(m):
                matriz1[tam].append(p.cell(row=i+6, column=j+2+y).value)

            matriz1[tam].append(i+1)
```

```
109         for i in range(len(matriz1)):
                for j in range(len(matriz1[i])):
111                 print(matriz1[i][j], end=" ")
                print ("\n")
113
115         for i in range(len(matriz1)):
              try:
117             mediana= statistics.median(matriz1[i][0:5])

119           except:
                mediana=None
121
              idsubarea=None
123
              if idpotreiro==1 and len(matriz1)==3:
125             idsubarea=25+i

127           if idpotreiro==1 and len(matriz1)==6:
                idsubarea=1+i
129
              if idpotreiro==2 and len(matriz1)==3:
131             idsubarea=28+i

133           if idpotreiro==2 and len(matriz1)==6:
                idsubarea=7+i
135
              if idpotreiro==3 and len(matriz1)==3:
137             idsubarea=31+i

139           if idpotreiro==3 and len(matriz1)==6:
                idsubarea=13+i
141
              if idpotreiro==4 and len(matriz1)==3:
143             idsubarea=34+i

145           if idpotreiro==4 and len(matriz1)==6:
                idsubarea=19+i
147
              query =  "INSERT INTO pastagem (idmedicao, idpotreiro,
                  dentrofora, gaiola, a1, a2, a3, a4, a5, media, mediana,
```

```
        pvtotal, pvsubamostra, psanoni, psoutras, psmorto, msanoni,
         msoutras, mstotal, ponto, idsubarea) VALUES (%s, %s, %s, %
        s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s,
        %s, %s, %s) RETURNING id;"
149     data = (idmedicao , idpotreiro, 'D', matriz1[i][14], matriz1[i
            ][0], matriz1[i][1], matriz1[i][2], matriz1[i][3], matriz1[
            i][4], matriz1[i][5], mediana, matriz1[i][6], matriz1[i
            ][7], matriz1[i][8], matriz1[i][9], matriz1[i][10], matriz1
            [i][11], matriz1[i][12], matriz1[i][13], None, idsubarea )

151     cur.execute(query, data)
        con.commit()
153
    print("Fora da Gaiola - " + o)
155
    matriz2 = []
157
    tam=-1
159 for i in range(n):
        valido=0
161     for j in range(m):
            if isinstance(p.cell(row=i+18, column=j+2+y).value, int) or
                isinstance(p.cell(row=i+18, column=j+2+y).value, float):
163             valido=1
            else:
165             p.cell(row=i+18, column=j+2+y).value=None

167     if valido==1:
            tam+=1
169         matriz2.append([])
            for j in range(m):
171             matriz2[tam].append(p.cell(row=i+18, column=j+2+y).value)
            matriz2[tam].append(i+1)
173
    for i in range(len(matriz2)):
175         for j in range(len(matriz2[i])):
                print(matriz2[i][j], end=" ")
177         print ("\n")

179 for i in range(len(matriz2)):
        try:
```

```
181             mediana= statistics.median(matriz2[i][0:5])

183         except:
                mediana=None
185
            idsubarea=None
187
            if idpotreiro==1 and len(matriz2)==3:
189             idsubarea=25+i

191         if idpotreiro==1 and len(matriz2)==6:
                idsubarea=1+i
193
            if idpotreiro==2 and len(matriz2)==3:
195             idsubarea=28+i

197         if idpotreiro==2 and len(matriz2)==6:
                idsubarea=7+i
199
            if idpotreiro==3 and len(matriz2)==3:
201             idsubarea=31+i

203         if idpotreiro==3 and len(matriz2)==6:
                idsubarea=13+i
205
            if idpotreiro==4 and len(matriz2)==3:
207             idsubarea=34+i

209         if idpotreiro==4 and len(matriz2)==6:
                idsubarea=19+i
211
            query =  "INSERT INTO pastagem (idmedicao, idpotreiro,
                dentrofora, gaiola, a1, a2, a3, a4, a5, media, mediana,
                pvtotal, pvsubamostra, psanoni, psoutras, psmorto, msanoni,
                 msoutras, mstotal, ponto, idsubarea) VALUES (%s, %s, %s, %
                s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s,
                %s, %s, %s) RETURNING id;"
213         data = (idmedicao , idpotreiro, 'F', matriz2[i][14], matriz2[i
                ][0], matriz2[i][1], matriz2[i][2], matriz2[i][3], matriz2[
                i][4], matriz2[i][5], mediana, matriz2[i][6], matriz2[i
                ][7], matriz2[i][8], matriz2[i][9], matriz2[i][10], matriz2
```

```
            [i][11], matriz2[i][12], matriz2[i][13], None, idsubarea )

215         cur.execute(query, data)
            con.commit()
217
        matriz3 = []
219     i=0
        salto=[0,3,5]
221     while 1:
          matriz3.append([])
223       for j in salto:
            matriz3[i].append(p.cell(row=i+38, column=j+2+y).value)
225
          if (matriz3[i][0] is None) and (matriz3[i][1] is None) and (
              matriz3[i][2] is None ):
227         break
          else:
229         if i>=2:
              query =  "INSERT INTO animais (idmedicao, idpotreiro,
                  datapesagem, identificacao, pesoorigem, pesofinal,
                  dataentrada, datasaida) VALUES (%s, %s, %s, %s, %s, %s,
                   %s, %s) RETURNING id;"
231           data = (idmedicao , idpotreiro, None, matriz3[i][0],
                  matriz3[i][1], matriz3[i][2], matriz3[0][1], matriz3
                  [0][2])

233           cur.execute(query, data)
              con.commit()
235
          i+=1
237
        y += 16
239
    else:
241
        log.write(o + " - Parou ao ler: " + str(d.value) + " Na cÃ©lula :
            " + str(d) + "\n")
243     break

245 con.close()
```

```
247 now2 = datetime.now()
    log.write("\n------- ExcecuÃ§Ã£o concluÃda em: " + str(now2))
249 log.write("\n------- Tempo de excecuÃ§Ã£o: " + str(now2-now) + "\n\n\n")
    log.close()
```

APÊNDICE B – PRÉ-PROCESSAMENTO

Este apêndice apresenta o código desenvolvido na linguagem Python para realizar o processo de pré-processamento do dados. Ao final são apresentados os códigos complementares ao pré-processamento, responsáveis pelas consultas no banco de dados espacial, desenvolvidos na linguagem SQL.

```
import psycopg2
import numpy as np
from datetime import datetime
import matplotlib.pyplot as plt

now = datetime.now()

log = open('log.txt', 'a')
log.write("------ ExcecuÃ§Ã£o iniciada em: " + str(now) + "\n\n")

try:
  con = psycopg2.connect("host='localhost' port='5432' dbname='
      pastagemOutlier' user='postgres' password='123456'")
  cur = con.cursor()

except:
  log.write("\nFalha na conexÃ£o com o BD!")

file = open("selectent.sql", 'r')
sql = " ".join(file.readlines())

cur.execute(sql)
con.commit()

try:
  tdo= cur.fetchall()
  ent = []
  ant=None

except:
  print("Erro 3")


```

```
for x in tdo:

  if (ant == None):
    ant=x
    continue

  if (ant[1] != x[1]) and x[0]=='D':
    c=x[1]-ant[1]

    file = open("selectclima.sql", 'r')
    sqlclima = " ".join(file.readlines())

    data = (ant[1].isoformat(), x[1].isoformat())

    cur.execute(sqlclima, data)
    con.commit()

    clima= cur.fetchall()

    #presente
    #ent.append([x[1], (x[1]-ant[1]).days, ant[2], ant[3], clima[0][0],
         clima[0][7], clima[0][14], clima[0][1], clima[0][8], clima
        [0][15], clima[0][2], clima[0][9], clima[0][16], clima[0][3],
        clima[0][10], clima[0][17], clima[0][4], clima[0][11], clima
        [0][18], clima[0][5], clima[0][12], clima[0][19], clima[0][6],
        clima[0][13], clima[0][20], clima[0][21], clima[0][22], clima
        [0][23],clima[0][24], clima[0][25], clima[0][26],clima[0][27],
        clima[0][28], clima[0][29], x[3]])

    #futuro
    ent.append([x[1], (x[1]-ant[1]).days, ant[2], ant[3], clima[0][0],
        clima[0][7], clima[0][14], clima[0][1], clima[0][8], clima
        [0][15], clima[0][2], clima[0][9], clima[0][16], clima[0][3],
        clima[0][10], clima[0][17], clima[0][4], clima[0][11], clima
        [0][18], clima[0][5], clima[0][12], clima[0][19], clima[0][6],
        clima[0][13], clima[0][20], clima[0][21], clima[0][22], clima
        [0][23],clima[0][24], clima[0][25], clima[0][26],clima[0][27],
        clima[0][28], clima[0][29], x[3]])

  ant=x
```

```
import csv

#presente
#header = ['data', 'numerodias', 'alturamedia', 'alturamediaanterior',
    'pcanonni', 'mstotalanterior', 'tmin', 'dptmin', 'vartmin', 'tmed',
     'dptmed', 'vartmed', 'tmax', 'dptmax', 'vartmax', 'umidade', '
    dpumidade', 'varumidade', 'velocidadevento', 'dpvelocidadevento', '
    varvelovidadevento', 'radiacaosolar', 'dpradiacaosolar', '
    varradiacaosolar', 'chuva', 'dpchuva', 'varchuva', 'somatermica', '
    dpsomatermica', 'varsomatermica', 'def', 'dpdef', 'vardef', 'exc',
    'dpexc', 'varexc', 'taxaacumulo']

#futuro
header = ['data', 'numerodias', 'alturamediaanterior', 'mstotalanterior
    ', 'tmin', 'dptmin', 'vartmin', 'tmed', 'dptmed', 'vartmed', 'tmax'
    , 'dptmax', 'vartmax', 'umidade', 'dpumidade', 'varumidade', '
    velocidadevento', 'dpvelocidadevento', 'varvelovidadevento', '
    radiacaosolar', 'dpradiacaosolar', 'varradiacaosolar', 'chuva', '
    dpchuva', 'varchuva', 'somatermica', 'dpsomatermica', '
    varsomatermica', 'def', 'dpdef', 'vardef', 'exc', 'dpexc', 'varexc'
    , 'taxaacumulo']

#reverso
#ent.reverse()

with open('ent.csv', 'w', newline='') as myfile:
    wr = csv.writer(myfile, quoting=csv.QUOTE_ALL, delimiter=';')
    wr.writerow(header)
    wr.writerows(ent)
```

Select para consolidação dos dados climáticos

```
select
cast((avg(tmin)) as NUMERIC(17,2)) as tmin,
cast((avg(tmed)) as NUMERIC(17,2)) as tmed,
cast((avg(tmax)) as NUMERIC(17,2)) as tmax,
```

```
cast((avg(umidade)) as NUMERIC(17,2)) as umidade,
cast((avg(velocidadevento)) as NUMERIC(17,2)) as
   velocidadevento,
cast((sum(radiacaosolar)) as NUMERIC(17,2)) as
   radiacaosolar,
cast((sum(chuva)) as NUMERIC(17,2)) as chuva,
cast((stddev(tmin)) as NUMERIC(17,2)) as dptmin,
cast((stddev(tmed)) as NUMERIC(17,2)) as dptmed,
cast((stddev(tmax)) as NUMERIC(17,2)) as dptmax,
cast((stddev(umidade)) as NUMERIC(17,2)) as dpumidade,
cast((stddev(velocidadevento)) as NUMERIC(17,2)) as
   dpvelocidadevento,
cast((stddev(radiacaosolar)) as NUMERIC(17,2)) as
   dpradiacaosolar,
cast((stddev(chuva)) as NUMERIC(17,2)) as dpchuva,
cast((variance(tmin)) as NUMERIC(17,2)) as vartmin,
cast((variance(tmed)) as NUMERIC(17,2)) as vartmed,
cast((variance(tmax)) as NUMERIC(17,2)) as vartmax,
cast((variance(umidade)) as NUMERIC(17,2)) as varumidade,
cast((variance(velocidadevento)) as NUMERIC(17,2)) as
   varvelocidadevento,
cast((variance(radiacaosolar)) as NUMERIC(17,2)) as
   varradiacaosolar,
cast((variance(chuva)) as NUMERIC(17,2)) as varchuva,
cast((sum(somatermica)) as NUMERIC(17,2)) as somatermica,
cast((stddev(somatermica)) as NUMERIC(17,2)) as
   dpsomatermica,
cast((variance(somatermica)) as NUMERIC(17,2)) as
   varsomatermica,
cast((sum(def)) as NUMERIC(17,2)) as def,
cast((stddev(def)) as NUMERIC(17,2)) as dpdef,
cast((variance(def)) as NUMERIC(17,2)) as vardef,
cast((sum(exc)) as NUMERIC(17,2)) as exc,
cast((stddev(exc)) as NUMERIC(17,2)) as dpexc,
```

```
cast((variance(exc)) as NUMERIC(17,2)) as varexc
from clima
WHERE data between %s and %s;
```

Select para consolidação dos dados amostrais das pastagens

```
select
p.dentrofora,
m.data,
cast((sum(p.media*(ST_Area(s.subpoligono)/ ST_Area(t.
   poligono)))/sum((ST_Area(s.subpoligono)/ ST_Area(t.
   poligono)))) as NUMERIC(7,2)) as media,
cast((sum(p.mstotal*(ST_Area(s.subpoligono)/ ST_Area(t.
   poligono)))/sum((ST_Area(s.subpoligono)/ ST_Area(t.
   poligono)))) as NUMERIC(7,2)) as mstotal,
cast((sum((p.msanoni/(p.msanoni+p.msoutras))*(ST_Area(s.
   subpoligono)/ ST_Area(t.poligono)))/sum((ST_Area(s.
   subpoligono)/ ST_Area(t.poligono)))) as NUMERIC(7,2)) as
    pcanonni
from subarea s, potreiro t, pastagem p, medicao m
where m.data is not NULL and s.idpotreiro=t.id and p.
   idsubarea=s.id and p.idmedicao=m.id and p.mstotal!=0 and
    t.id=2
group by p.dentrofora, m.data
order by m.data;
```

APÊNDICE C – MODELAGEM, TREINAMENTO E TESTE DO MODELO PROPOSTO E VISUALIZAÇÃO DOS RESULTADOS

Este apêndice apresenta o código desenvolvido na linguagem Python para realizar a modelagem, treinamento e teste da RNA LSTM e a visualização dos resultados.

```
import os
from math import sqrt
from numpy import concatenate
from matplotlib import pyplot
from pandas import read_csv
from pandas import DataFrame
from pandas import concat
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import mean_squared_error
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
import numpy as np

#carrega dados
dataset = read_csv('ent.csv', header=0, index_col=0, delimiter=';')
values = dataset.values

#encode inteiros
encoder = LabelEncoder()

#exclui valores nulos
values=values[~np.isnan(values).any(axis=1)]

#converte todos os valores para float
values = values.astype('float32')

#normaliza os dados
scaler = MinMaxScaler(feature_range=(0, 1))
scaled = scaler.fit_transform(values)

scaled = DataFrame(scaled)

```

```
#divide em conjunto de treinamento e teste
values = scaled.values

n_train = 24
#n_train = 27

train = values[:n_train, :]
test = values[n_train:, :]

#divide os dados de entrada da saida
train_X, train_y = train[:, :-1], train[:, -1]
test_X, test_y = test[:, :-1], test[:, -1]

#transforma entrada em 3D [amostras, datas, saidas]
train_X = train_X.reshape((train_X.shape[0], 1, train_X.shape[1]))
test_X = test_X.reshape((test_X.shape[0], 1, test_X.shape[1]))


print(train_X.shape, train_y.shape, test_X.shape, test_y.shape)
print(train_X.shape[1])

#define a estrutura da rede
model = Sequential()

model.add(LSTM(30, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', return_sequences = True))
model.add(LSTM(15, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal'))

model.add(Dense(1, kernel_initializer='normal'))

model.compile(loss='mean_squared_error', optimizer='rmsprop')

#ajusta a rede
history = model.fit(train_X, train_y, epochs=5000, batch_size=72,
    validation_data=(test_X, test_y), verbose=2, shuffle=False)

#imprime historico
pyplot.ylabel('Loss error')
pyplot.xlabel('NÃºmero de Ã©pocas')
pyplot.plot(history.history['loss'], label='treinamento')
```

```
pyplot.plot(history.history['val_loss'], label='teste')
pyplot.legend()
pyplot.show()

#realiza as predicoes
yhat = model.predict(test_X)
test_X = test_X.reshape((test_X.shape[0], test_X.shape[2]))

#invert scaling for forecast
inv_yhat = concatenate((test_X, yhat), axis=1)
inv_yhat = scaler.inverse_transform(inv_yhat)
inv_yhat = inv_yhat[:,-1]

#inverte escala para atual
test_y = test_y.reshape((len(test_y), 1))
inv_y = concatenate((test_X, test_y), axis=1)
inv_y = scaler.inverse_transform(inv_y)
inv_y = inv_y[:,-1]

#calcula o EQM
rmse = sqrt(mean_squared_error(inv_y, inv_yhat))
print('Test RMSE: %.3f' % rmse)

print('Real')
print(inv_y)
print('Predito')
print(inv_yhat)

#imprime a estrutura da rede
print(model.summary())

from keras.utils import plot_model
plot_model(model, to_file='model_plot.png', show_shapes=True,
    show_layer_names=True)

#grafico com o resultado das predicoes
fig, ax = pyplot.subplots()

pyplot.ylabel('kg/ha')

fig.autofmt_xdate()
```

```
114 pyplot.grid(True)
    #pyplot.ylim(bottom=0)
116 #pyplot.ylim(top=12000)

118 pyplot.plot_date(dataset.index.values[(test_X.shape[0]*-1):], inv_y,
        label='Real', linestyle='--', marker='o')
    pyplot.plot_date(dataset.index.values[(test_X.shape[0]*-1):], inv_yhat,
         label='Predito', linestyle='--', marker='o')
120

    #REVERSE GRAPH
122 #pyplot.plot_date(dataset.index.values[(test_X.shape[0]*-1):][::-1],
        inv_y[::-1], label='Real', linestyle='--', marker='o')
    #pyplot.plot_date(dataset.index.values[(test_X.shape[0]*-1):][::-1],
        inv_yhat[::-1], label='Predito', linestyle='--', marker='o')
124

    pyplot.legend()
126 pyplot.show()
```