

UNIVERSIDADE FEDERAL DO PAMPA

ÂNDERSON FISCHOEDER SOARES

**TOUCEIRATECH: UM FARM
MANAGEMENT INFORMATION
SYSTEM PARA PECUÁRIA DE
PRECISÃO BASEADO EM PREDIÇÃO
COM REDES NEURAS RECORRENTES**

**Bagé
2021**

ÂNDERSON FISCHOEDER SOARES

**TOUCEIRATECH: UM FARM
MANAGEMENT INFORMATION
SYSTEM PARA PECUÁRIA DE
PRECISÃO BASEADO EM PREDIÇÃO
COM REDES NEURAIIS RECORRENTES**

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

Orientador: Leonardo Bidese de Pinho
Coorientador: Naylor Bastiani Perez

**Bagé
2021**

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).

S546t Soares, Anderson Fiscoeder

TouceiraTech: Um Farm Management Information System para Pecuária de Precisão Baseado em Predição com Redes Neurais Recorrentes / Anderson Fiscoeder Soares.

144 p.

Dissertação (Mestrado) - Universidade Federal do Pampa, PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA, 2021.

“Orientação: Leonardo Bidese de Pinho; Coorientação: Naylor Bastiani Perez”.

1. Manejo da pastagem. 2. Sistemas de apoio à decisão. 3. Métodos de amostragem. 4. Sensoriamento remoto. 5. Ajuste de lotação. 6. Taxa de acúmulo. I. Título.

ÂNDERSON FISCHOEDER SOARES

**TOUCEIRATECH: UM FARM MANAGEMENT
INFORMATION SYSTEM PARA
PECUÁRIA DE PRECISÃO BASEADO
EM PREDIÇÃO COM REDES NEURAIAS
RECORRENTES**

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: 9 de abril de 2021.

Banca examinadora:

Prof. Dr. Leonardo Bidese de Pinho

Orientador

UNIPAMPA

Profa. Dra. Ana Paula Lüdtke Ferreira

UNIPAMPA

Profa. Dra. Carolina Bremm
DDPA / SEAPDR

Prof. Dr. Gustavo Trentin
EMBRAPA



Assinado eletronicamente por **LEONARDO BIDESE DE PINHO, PROFESSOR DO MAGISTERIO SUPERIOR**, em 16/06/2021, às 11:19, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **ANA PAULA LUDTKE FERREIRA, PROFESSOR DO MAGISTERIO SUPERIOR**, em 16/06/2021, às 12:00, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



A autenticidade deste documento pode ser conferida no site https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0549448** e o código CRC **6CF3A78F**.

AGRADECIMENTO

Agradeço à minha família por todo suporte e apoio nos momentos mais difíceis durante o desenvolvimento, principalmente à minha mãe.

O presente trabalho foi realizado com suporte da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, tendo em vista o vínculo dos autores ao Programa de Pós-Graduação em Computação Aplicada (PPGCAP), resultado da cooperação entre UNIPAMPA e EMBRAPA, incluindo uma bolsa de mestrado. Parte do trabalho foi desenvolvido com o apoio do Programa de Auxílio da Pós-Graduação (PAPG) da UNIPAMPA.

RESUMO

Para aumentar a produtividade na pecuária de corte é fundamental aprimorar o manejo do pasto, o que demanda a mensuração periódica da massa e do acúmulo de forragem e o ajuste da taxa de lotação. Portanto, mostra-se relevante desenvolver ferramentas capazes de auxiliar os pecuaristas neste processo. O presente trabalho, amparado inicialmente por uma metodologia exploratória e, posteriormente, explicativa, propõe, implementa e avalia o TouceiraTech, um protótipo de FMIS para Pecuária de Precisão capaz de coletar, armazenar, pré-processar, predizer e visualizar dados sobre a taxa de acúmulo, necessária para o ajuste da taxa de lotação em uma área de interesse georreferenciada. O TouceiraTech foi projetado de forma interdisciplinar, a partir da expansão de um modelo original de predição de disponibilidade de forragem, baseado em aprendizado de máquina com Redes Neurais Artificiais do tipo LSTM com dados de vegetação campestre obtidos pela amostragem histórica, direta, destrutiva, em experimentos realizados para avaliar a produção animal em diferentes condições de manejo do campo nativo. No modelo de predição são agregados dados meteorológicos, coletados remotamente de estação meteorológica automática próxima à região de interesse e, também, dados de previsão meteorológica. Esses dados são automaticamente pré-processados pelo TouceiraTech para estimação de variáveis específicas, com destaque para uma nova abordagem automatizada para a estimativa da evapotranspiração. Além disso, permite o uso de amostras indiretas não destrutivas da vegetação, com base no NDVI, a partir do processamento de imagens aéreas incorporadas ao seu banco de dados espacial. Os resultados demonstram a eficácia das coletas periódicas automatizadas dos dados meteorológicos necessários para a predição, a partir de bases remotas do INMET, INPE e NOAA. Em especial, comprovam que mudanças na abordagem de treinamento do modelo, de forma estratificada, complementada por ajustes nas suas variáveis de entrada, permitiram a concepção de modelos especializados por tipos de tratamento com acurácia significativamente superior à do modelo original. Adicionalmente, indicam o potencial da incorporação de imagens aéreas georreferenciadas para viabilizar um sistema de suporte à decisão de ajuste de taxa de lotação com base em amostragem indireta de baixo custo operacional, em substituição ao método direto.

Palavras-chave: Manejo da pastagem. Sistemas de apoio à decisão. Métodos de amostragem. Sensoriamento remoto. Ajuste de lotação. Taxa de acúmulo.

ABSTRACT

To increase productivity in beef cattle ranching it is essential to improve pasture management, which demands periodic measurement of the mass and accumulation of forage and adjustment of stocking rate. Therefore, it is important to develop tools capable of helping cattle ranchers in this process. The present work, supported initially by an exploratory methodology and later by an explanatory one, proposes, implements and evaluates TouceiraTech, a prototype of FMIS for Precision Cattle Raising capable of collecting, storing, preprocessing, predicting and visualizing data on forage accumulation rate, necessary for adjusting the stocking rate in a georeferenced area of interest. TouceiraTech was designed in an interdisciplinary way, from the expansion of an original model of forage availability prediction, based on machine learning with LSTM-type Artificial Neural Networks with field vegetation data obtained by historical, direct, destructive sampling in experiments carried out to evaluate animal production in different management conditions of the native field. In the prediction model are aggregated meteorological data, remotely collected from an automatic weather station near the region of interest, and also meteorological forecast data. These data are automatically pre-processed by TouceiraTech to estimate specific variables, with emphasis on a new automated approach to estimate evapotranspiration. In addition, it allows the use of non-destructive indirect sampling of vegetation, based on NDVI, from the processing of aerial images incorporated into its spatial database. The results demonstrate the effectiveness of automated periodic collection of meteorological data needed for prediction, from remote databases of INMET, INPE and NOAA. In particular, they prove that changes in the model training approach, in a stratified way, complemented by adjustments in its input variables, allowed the design of specialized models by treatment types with accuracy significantly higher than that of the original model. Additionally, they indicate the potential of incorporating georeferenced aerial images to enable a decision support system to adjust stocking rates based on indirect sampling with low operational cost, replacing the direct method.

Keywords: Pasture management. Decision support systems. Sampling methods. Remote sensing. Stocking rate. Accumulation rate.

LISTA DE FIGURAS

Figura 1	Sequência de etapas para desenvolvimento	23
Figura 2	Utilização de gaiola	28
Figura 3	Método do quadrado	29
Figura 4	Régua de medição de altura.....	30
Figura 5	Sonda Eletrônica.....	31
Figura 6	Exemplo de arquitetura em uma Smart Farming.....	33
Figura 7	Aspectos fundamentais de Big Data.....	34
Figura 8	Blocos básicos da IoT.....	36
Figura 9	Aspectos fundamentais de Cloud Computing	37
Figura 10	Arquitetura de um sistema baseado em FMIS.....	39
Figura 11	Fases do processo de KDD.....	43
Figura 12	Rede multicamadas.....	44
Figura 13	Arquitetura de uma Rede Neural Recorrente.....	45
Figura 14	Arquitetura LSTM	46
Figura 15	Componentes básicos da fase de aquisição de dados em sensoriamento remoto	46
Figura 16	Arquitetura proposta para o TouceiraTech.....	51
Figura 17	Arquivo no formato .csv gerado pelo <i>script</i> de coleta.....	53
Figura 18	Imagem original obtida através do Sentinel-2 da área de estudo principal	54
Figura 19	Recorte da área total de estudo principal.....	55
Figura 20	Método de registro das amostras diretas - (a) Planilha sem padronização e (b) Planilha padronizada	57
Figura 21	Cabeçalho da tabela de dados disponibilizado pelo INMET.....	57
Figura 22	Cabeçalho da tabela clima existente no banco de dados	57
Figura 23	Diagrama Entidade-Relacionamento do modelo original	59
Figura 24	Diagrama Entidade-Relacionamento do modelo atual	60
Figura 25	Comparação dos valores de Evapotranspiração.....	68
Figura 26	Tabela com dados meteorológicos disponibilizada no website do INMET ...	69
Figura 27	Tabela com dados meteorológicos obtida através da execução do <i>script</i> de coleta	69
Figura 28	Valores de temperatura disponibilizadas pelo NCAR e CPTEC	70
Figura 29	Interface gráfica do website do rda e variáveis meteorológicas selecionáveis	71
Figura 30	Interface gráfica do website do rda, onde é necessário indicar os parâmetros de busca.....	71
Figura 31	Dados de previsão meteorológica obtidos com uso do website do rda e com <i>script</i> de coleta para o dia 16 de fevereiro de 2021 na hora 00.....	71
Figura 32	Resultado da busca realizada no repositório Open Access Hub do Copernicus	72
Figura 33	Resultado da busca realizada pelo <i>script</i> de coleta de imagens do satélite Sentinel-2	72
Figura 34	Ajuste de épocas	74
Figura 35	Reprodução do experimento de Schulte (2019) com o Potreiro 20 Infestado	74
Figura 36	Sensibilidade ao tamanho do conjunto de treinamento	75
Figura 37	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de massa de forragem, para cada subárea	76
Figura 38	valores reais e preditos de massa de forragem, para melhor e pior caso	77
Figura 39	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de massa de forragem.....	77

Figura 40	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de massa de forragem, para cada subárea	79
Figura 41	valores reais e preditos de massa de forragem, para melhor e pior caso	79
Figura 42	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de massa de forragem.....	80
Figura 43	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de massa de forragem, para cada subárea	81
Figura 44	valores reais e preditos de massa de forragem, para melhor e pior caso	82
Figura 45	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de massa de forragem.....	83
Figura 46	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de massa de forragem, para cada subárea	84
Figura 47	valores reais e preditos de massa de forragem, para melhor e pior caso	84
Figura 48	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de massa de forragem.....	85
Figura 49	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso	86
Figura 50	valores reais e preditos de taxa de acúmulo, para melhor e pior caso	87
Figura 51	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo.....	88
Figura 52	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso	89
Figura 53	valores reais e preditos de taxa de acúmulo, para melhor e pior caso	89
Figura 54	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo.....	90
Figura 55	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso	91
Figura 56	Valores reais e preditos de taxa de acúmulo, para melhor e pior caso.....	92
Figura 57	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo.....	92
Figura 58	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso	93
Figura 59	valores reais e preditos de taxa de acúmulo, para melhor e pior caso	94
Figura 60	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo.....	94
Figura 61	Valores reais <i>versus</i> preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso	95
Figura 62	valores reais e preditos de taxa de acúmulo, para melhor e pior caso	96
Figura 63	Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo.....	96
Figura 64	Valores para estimativa futura de taxa de acúmulo, reais <i>versus</i> preditos pela RNN LSTM para cada uma das subáreas de estudo, para treinamento estratificado por tratamento (T)	97
Figura 65	Resultado do cálculo de NDVI para as diferentes subáreas (potreiros)	99
Figura 66	Resultado do cálculo de NDVI para a área total de interesse.....	100
Figura 67	NDVI calculado com o software QGIS para a área total de estudo	101
Figura 68	Valores de taxa de acúmulo preditos pela RNN LSTM, utilizando NDVI, para o presente para cada uma das subáreas de estudo	102
Figura 69	Valores de taxa de acúmulo preditos pela RNN LSTM para o futuro, utilizando valores de NDVI, para cada uma das subáreas de estudo	103
Figura 70	Imagem original e o resultado do algoritmo de classificação K-means	104

Figura 71 Mapa temático com valores de taxa de acúmulo preditos pela RNN LSTM para o futuro, utilizando dados históricos de pastagem, para cada uma das subáreas de estudo apresentados no mapa de vegetação	105
Figura 72 Mapa temático com valores de taxa de acúmulo preditos pela RNN LSTM para o futuro, utilizando NDVI, para cada uma das subáreas de estudo apresentados no mapa de vegetação.....	106
Figura 73 Seis primeiras imagens presentes na GIF animada criada para visualização do comportamento do NDVI ao longo do tempo	107

LISTA DE TABELAS

Tabela 1	Diferenças entre Agricultura Tradicional e Smart Farming	34
Tabela 2	Trabalhos correlatos.....	48
Tabela 3	Desvios padrão dos dados de previsão meteorológica	70
Tabela 4	Valores de previsão meteorológica e dados reais da estação automática do INMET para o ano de 2019 (°C).....	108
Tabela 5	Correlação entre valores de NDVI e reais para cada uma das gaiolas utilizadas nos poteiros 20 e 21	112
Tabela 6	Funcionalidades e aspectos dos sistemas	113

LISTA DE ABREVIATURAS E SIGLAS

BDOO	Banco de Dados Orientados a Objetos
BDR	Banco de Dados Relacionais
CPTEC	Centro de Previsão de Tempo e Estudos Climáticos
KDD	Descoberta de Conhecimento em Banco de Dados
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FMIS	<i>Farm Management Information Systems</i>
GPS	<i>Global Positioning System</i>
GIF	<i>Graphics Interchange Format</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
INMET	Instituto Nacional de Meteorologia
IA	Inteligência Artificial
IoT	Internet das Coisas
LSTM	<i>Long Short-Term Memory</i>
MCDM	Metodologia Multicritério de Tomada de Decisão
MCDA	Metodologias Multicritério em Apoio à Decisão
MD	Mineração de Dados
NCAR	<i>National Center for Atmospheric Research</i>
NDVI	Índices de Vegetação por Diferença Normalizada
RDA	<i>Research Data Archive</i>
RNA	Rede Neural Artificial
RNN	Rede Neural Recorrente
SAD	Sistema de Apoio a Decisão
SAVI	Índices de Vegetação Ajustado para o Solo
SIG	Sistema de Informações Geográficas

SGBD Sistemas de Gerenciamento de Banco de Dados

TIC Tecnologias da Informação e da Comunicação

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Motivação.....	15
1.1.1 Perspectiva da agropecuária	15
1.1.2 Perspectiva da computação	17
1.2 Temática interdisciplinar	18
1.3 Questões de pesquisa	20
1.4 Objetivos	20
1.5 Metodologia de pesquisa	21
1.6 Estrutura do texto	24
1.7 Considerações finais.....	24
2 REFERENCIAL TEÓRICO	25
2.1 Desafios de gestão da pastagem	25
2.1.1 Métodos diretos	28
2.1.2 Métodos indiretos.....	29
2.2 Tecnologias da Informação e Comunicação aplicáveis à Agropecuária	32
2.2.1 Smart Farming	32
2.2.2 Big Data.....	34
2.2.3 Internet das Coisas.....	35
2.2.4 Cloud Computing.....	37
2.3 Farm Management Information Systems	38
2.3.1 Sistemas de Apoio a Decisão	39
2.3.2 Bancos de Dados Espaço-Temporais	40
2.3.3 Mineração de Dados	41
2.3.4 Aprendizado de Máquina	42
2.3.5 Sensoriamento Remoto.....	45
2.4 Trabalhos correlatos	47
2.5 Considerações finais.....	50
3 SOLUÇÃO PROPOSTA	51
3.1 Fontes de dados	52
3.2 Atualização periódica	55
3.3 Descrição das entradas	56
3.4 Processo de ETL.....	56
3.5 Banco de dados espacial	58
3.6 Pré-processamento	61
3.7 Aprendizado de máquina	61
3.8 Visualização dos resultados.....	62
3.9 Considerações finais.....	62
4 METODOLOGIA EXPERIMENTAL	63
4.1 Material e Métodos	63
4.2 Delineamento dos Experimentos	65
4.3 Considerações finais.....	67
5 RESULTADOS	68
5.1 Estimativa Automática de Evapotranspiração.....	68
5.2 Coleta de Dados Meteorológicos e de Previsão	68
5.3 Módulo de Predição	73
5.3.1 Resultados do Modelo Original com Conjunto A	73
5.4 Resultados do Modelo Original com propostas de ajustes.....	78
5.4.1 Resultados do Modelo Ajustado com Conjunto A	85

5.4.2 Resultados do Modelo Ajustado com Conjunto B	87
5.4.3 Resultados do Modelo Ajustado com a união dos Conjuntos A e B.....	90
5.4.4 Resultados do Modelo Ajustado do Conjunto A para Predição Futura.....	97
5.5 Avaliação do NDVI e teste com K-means.....	98
5.6 Visualização com Mapa Temático e GIF animada.....	104
5.7 Considerações Finais	106
6 DISCUSSÃO	110
6.1 Considerações Finais	113
7 CONCLUSÃO	115
7.1 Trabalhos Futuros.....	116
REFERÊNCIAS.....	120
APÊNDICE A – SCRIPT DE SUMARIZAÇÃO DOS DADOS DE PASTAGEM.	128
APÊNDICE B – SCRIPT DO MÓDULO DE PREDIÇÃO DO MODELO AJUSTADO UTILIZANDO O CONJUNTO DE DADOS A, PARA O POTREIRO P20 INFESTADO, UTILIZANDO TREINAMENTO ES- TRATIFICADO POR TRATAMENTO.....	137

1 INTRODUÇÃO

Este capítulo tem como objetivo introduzir a proposta do presente projeto de pesquisa, desenvolvimento e inovação, caracterizado no contexto de tecnologias para a produção agropecuária. A seção 1.1 apresenta a motivação para o estudo, seguida da delimitação do tema interdisciplinar (seção 1.2), da explicitação das questões de pesquisa (seção 1.3) e dos objetivos (seção 1.4). Posteriormente, são resumidamente descritos os referenciais metodológicos adotados na concepção do projeto (seção 1.5), seguidos da apresentação da estrutura do restante do trabalho (seção 1.6) e pelas considerações finais (seção 1.7).

1.1 Motivação

A motivação para o desenvolvimento desta pesquisa aplicada partiu de um problema identificado no contexto da agropecuária, mais especificamente da bovinocultura de corte, associado à oportunidade de desenvolvimento de tecnologias computacionais aplicáveis à produção agropecuária.

1.1.1 Perspectiva da agropecuária

Segundo o IBGE (2020), o Brasil possuía um efetivo de aproximadamente 214,8 milhões de cabeças de gado em 2019. No mesmo ano, no estado do Rio Grande do Sul estavam concentradas 11,9 milhões (IBGE, 2020). Para que os pecuaristas, produtores rurais especializados em sistemas pecuários, maximizem seu lucro, de forma sustentável, é fundamental fazer a gestão dos seus sistemas de pastejo (*grazing systems*), representados pela “combinação integrada entre os componentes animal, planta, solo, e fatores ambientais, mais o método de pastejo, com o objetivo de se atingir metas específicas” (PEDREIRA, 2002).

De acordo com Salman, Soares e Canesin (2006), no que se refere aos custos de produção animal, de 50% a 80% estão associados ao gasto com alimentação. Além da redução das despesas de custeio com alimentação, o correto uso das pastagens está diretamente ligado com a produtividade, visto que em um contexto de pecuária a produtividade é representada pela quantidade de peso vivo (PV) produzido anualmente em

cada hectare¹ de pastagem (FERNANDES et al., 2004). Neste contexto, há a necessidade de se evitar a degradação das pastagens e também intensificar a sua produtividade, visto que em pastagens degradadas a produtividade de carne fica aproximadamente em 2 kg de PV.ha⁻¹ e em pastagens consideradas em bom estado pode-se atingir em torno de 16 kg de PV.ha⁻¹ (KICHEL; MIRANDA; ZIMMER, 1999). Esta variação na produtividade² é impactada diretamente pela oferta de forragem (*forage allowance* - OF), isto é, pela relação quantitativa e instantânea entre forragem e animal, mais precisamente pela relação estabelecida entre a matéria seca (MS) de forragem por unidade de área e o número de unidades animais em um determinado instante (PEDREIRA, 2002). Portanto, “as relações quantitativas entre o componente 'forragem' (kg MS/ha) e o componente 'animal' (kg PV/ha) têm um impacto marcante sobre a produção e a produtividade” (PEDREIRA, 2002), de modo que conclui-se que a gestão da MS é essencial em um sistema de pastejo.

O alicerce para a pecuária sustentável é o adequado manejo das pastagens, já que, com o devido manejo, é possível garantir a qualidade e a quantidade necessária para a alimentação dos animais, sem a degradação das pastagens. O desafio para a sustentabilidade de uma pastagem é equilibrar os interesses dos animais, ou seja, uma alimentação adequada, com as necessidades da pastagem (MELADO, 2007). Portanto, além de fazer a correta mensuração da taxa de lotação (*stocking rate*), que corresponde à “relação entre o número de animais ou de unidades animais (UA) e a área da unidade de manejo por eles ocupada, durante um período específico de tempo” (PEDREIRA, 2002), também é necessário disponibilizar pastagens com valores nutritivos adequados (SILVEIRA et al., 2005). É importante levar em conta esses dois fatores devido à capacidade dos bovinos selecionarem sua própria dieta durante o pastejo. A produção animal em sistemas extensivos é uma função da relação solo-planta-animal e até mesmo o homem, devido a essa relação é importante compreender como o pastejo afeta o solo, a superfície hidrológica, etc. (SANTOS et al., 2002). Os animais por sua vez afetam o sistema de produção devido as suas características genéticas, sua idade, sua condição produtiva, entre outras (NABINGER, 1997). As pastagens nativas correspondem a um grande recurso que vem sendo utilizado na pecuária extensiva (MOOJEN; MARASCHIN, 2002). Contudo, como nos estados do Sul do Brasil as estações frias são muito rigorosas, a disponibilidade de pastagem nativa reduz nestes períodos, com isso se faz necessária a introdução de pastagens de estação fria, como aveia e azevém. Com a implantação de forrageiras cultivadas em períodos específicos do ano, é possível garantir boa disponibilidade de pastagem para o

¹Unidade de medida para superfícies agrárias correspondente a 10.000 metros quadrados.

²Também expresso na literatura como kg de PV/ha.

rebanho durante todo ano. A partir disso, com a implantação de novas tecnologias é possível realizar a preservação das pastagens naturais, por meio do melhoramento das mesmas ou pela introdução de espécies de estação fria, que tenham crescimento no período de outono-inverno (CÓRDOVA et al., 2014).

Conhecendo a oferta de forragem, é possível estimar o ganho de peso dos animais e com isso a produtividade do rebanho (ANDRADE et al., 2005). Neste contexto, o principal problema passa a ser o processo de tomada de decisões, sob responsabilidade do produtor, relacionado à taxa de lotação, tendo como referência dois fatores distintos: a quantidade de forragem necessária para otimizar o ganho médio diário (GMD) de massa dos animais é em torno de 13,5%, porém, se for considerado o ganho de massa por hectare ($G \cdot ha^{-1}$), essa estimativa gira em torno de 11,5% (NABINGER, 2006).

1.1.2 Perspectiva da computação

O aumento da utilização da TI pode causar impactos positivos na agropecuária, destacando-se a substituição da força de trabalho e o aumento no controle das atividades, contribuindo para a redução da incerteza e dos riscos. Dessa forma, é possível esperar que sua incorporação no dia a dia da propriedade aumente o ingresso aos mercados, por facilitar o acesso à informação. (MACHADO; NANTES, 2011).

Nas últimas décadas, diversas Tecnologias da Informação e da Comunicação (TIC) passaram a ser aplicadas ao agronegócio. Neste contexto, diferentes TIC estão sendo também integradas com o objetivo de aumentar os recursos e resultados obtidos, tendo como exemplo, as tecnologias de informação unindo-se com as tecnologias de controle, as tecnologias de monitoramento com as tecnologias de telecomunicações (CAVALHEIRO et al., 2018). Segundo Bambini e Romani (2014), existem diversos benefícios para a adoção de TIC voltadas ao agronegócio, como redução dos custos de comunicação entre os agentes da agropecuária, redução dos custos de acesso a serviços e informação, redução de riscos relacionados a eventos climáticos, pragas e desastres naturais, entre outros.

No contexto de pecuária, as AgroTIC, as quais serão definidas na seção 2.2, podem ser usadas para auxiliar no armazenamento de informações acerca de condições sanitárias, nutricionais e genéticas dos animais (MACHADO, 2002 *apud* MACHADO; NANTES, 2008, p. 1). Contudo, desde 1997 se destaca que “inúmeras são as aplicações do uso das

tecnologias da informação na bovinocultura, sendo a mais importante delas o desenvolvimento de sistemas computacionais que buscam auxiliar produtores e profissionais na tomada de decisão” (LOPES, 1997 *apud* CEOLIN et al., 2008, p. 64).

Portanto, percebe-se que é relevante estudar o desenvolvimento de novas Agrotécnicas, a partir da integração de tecnologias presentes no estado da arte, com foco na gestão de sistemas de pastejo, capazes de auxiliar na tomada de decisão e, conseqüentemente, na obtenção de resultados positivos na sua produtividade (CEOLIN et al., 2008).

1.2 Temática interdisciplinar

A Empresa Brasileira de Pesquisa Agropecuária (Embrapa), através da sua unidade descentralizada chamada de Embrapa Pecuária Sul, desenvolve pesquisa nos Campos Sul-brasileiros, compreendidos pelos Estados do Rio Grande do Sul, Santa Catarina e Paraná. Dentre as áreas nas quais a unidade atua estão a bovinocultura de corte, de leite e ovinos. Na unidade é realizado estudo da disponibilidade de forragem, utilizando dados históricos de coleta através de um método direto de amostragem estratificada. Para realizar a coleta de pontos específicos dos campos experimentais são utilizadas gaiolas que isolam áreas de pastagem. Mensalmente a pastagem isolada é cortada, seca e pesada. A taxa de lotação de animais nos campos experimentais é mensurada através da estimativa de crescimento da pastagem, que por sua vez é calculada através da massa de forragem da amostra coletada. O ajuste da taxa de lotação busca manter uma oferta de forragem igual a 12% do peso vivo do animal. A estimativa realizada atualmente é imprecisa, não apenas pelo erro amostral decorrente do método de medição, mas sobretudo devido ao fato de não considerar fatores importantes para o crescimento das plantas, tais como variáveis climáticas e variações meteorológicas.

Para que se possa realizar a estimativa de produtividade de forragem e também do desempenho dos animais, existem técnicas diretas e indiretas para estimar a massa de forragem (MF). A MF é definida como “a quantidade de matéria seca (MS) presente instantaneamente acima do nível do solo por unidade de área” (HODGSON, 1979 *apud* ARRUDA et al., 2011, p. 1). Dentre as técnicas diretas está o corte da forragem, porém é um método destrutivo que gera muito trabalho para coleta das amostras e, com isso, limita o número pontos de coleta na pastagem para que se tenha uma quantidade substancial de dados. Existem também os métodos indiretos para a estimativa de MF, dentre eles estão o disco medidor (DM) que estima a massa através da altura e densidade da pastagem

com a utilização de uma placa de metal com peso conhecido que desliza em um eixo central que possui uma escala graduada. O bastão graduado (BG) tem um funcionamento similar ao disco medidor, porém o pasto não é comprimido, já que a leitura é feita assim que o dispositivo toca a planta. Também existe o medidor de capacitância (MC), que funciona através de um sinal produzido pelo oscilador de circuito elétrico (CAUDURO et al., 2006).

A partir da necessidade de aprimorar o método realizado, Schulte (2019) iniciou a criação de um modelo baseado em aprendizado de máquina para a estimativa de forragem na qual utiliza os dados históricos de cortes realizados na Embrapa Pecuária Sul, considerando fatores meteorológicos no crescimento das pastagens. Os dados meteorológicos são obtidos da estação meteorológica de superfície automática, do Instituto Nacional de Meteorologia (INMET). Os dados de corte, tanto quanto os referentes ao clima, foram inseridos em um banco de dados para que, através de aprendizagem de máquina, fosse possível prever a quantidade de massa de forragem disponível em determinada data.

A técnica de aprendizado de máquina utilizada por Schulte (2019) é um tipo especial de Rede Neural Recorrente (Recurrent Neural Networks - RNN), denominada Long Short-Term Memory (LSTM), a qual pode ser usada como base para um Sistema de Apoio a Decisão (SAD) para os produtores rurais. Ou seja, em tese permitiria a concepção de um Farm Management Information System (FMIS) voltado para pecuária de precisão, especializado para ajuste de taxa de lotação por meio da predição da oferta de forragem. Sistema este que deveria ser composto por módulos capazes de realizar a coleta, o processamento, o armazenamento e a visualização dos dados espacializados referentes às áreas de interesse. Contudo, o modelo baseado em RNN LSTM desenvolvido por Schulte (2019) se ampara apenas em dados meteorológicos históricos. Com isso, mostra-se relevante o aprimoramento do modelo para que não só sejam utilizados dados históricos, mas também dados de previsões meteorológicas, bem como automatizar o processo de coleta destes a partir de bases de dados remotas. Além disso, também se apresentou como importante agregar ao modelo uma coleta de dados sobre a pastagem de maneira automatizada, através de imagens, visto o custo operacional e suscetibilidade a erro inerentes aos processos de amostragem destrutiva periódica da pastagem. Para este fim, seria possível utilizar imagens de satélite para, através de índices, estimar dados referentes a pastagem como, por exemplo, a altura e/ou densidade. Por fim, nota-se a relevância de estudar alternativas para a apresentação, dos dados espaço-temporais correspondentes à evolução da pastagem, de forma intuitiva para produtores rurais.

1.3 Questões de pesquisa

O presente trabalho visa responder três principais questões de pesquisa:

- É possível propor, implementar e avaliar um protótipo de prova de conceito de FMIS partindo de princípios de pecuária de precisão, capaz de coletar, armazenar, processar e visualizar dados das pastagem - massa de forragem (MF) instantânea - necessários para o ajuste de taxa de lotação?
- É possível introduzir no modelo já existente dados de previsão meteorológica que aumentem a eficácia na predição da MF para o período de 30 dias subsequente?
- É possível, através do processamento de imagens de satélite, obter, com acurácia suficiente, amostras indiretas referentes à MF que possam ser usadas como entrada no modelo de RNN desenvolvido, substituindo parte do processo de amostragem direta?

1.4 Objetivos

De forma geral, este trabalho objetivou, como parte do processo de desenvolvimento de um protótipo de prova de conceito, aprimorar e aplicar o método de predição desenvolvido por Schulte (2019), incrementando dados de coleta de amostras indireta por meio de imagens de satélite. Também visava incorporar ao modelo de banco de dados espacial proposto os dados de previsão meteorológica de determinadas regiões de interesse. Ao fim do projeto também era esperado que o resultado da predição fosse apresentado de maneira visual, através de mapas da vegetação usando animações no formato Graphics Interchange Format (GIF).

Para atingir o objetivo geral do trabalho, foram definidos objetivos específicos, divididos entre essenciais e desejáveis. Os objetivos essenciais são aqueles considerados como fundamentais e, portanto, prioritários, elencados como segue:

- Dominar o problema;
- Identificar o estado da arte;
- Compreender os dados meteorológicos;
- Incorporar imagens aéreas ao banco de dados espacial;
- Automatizar a coleta de dados meteorológicos e de imagens de satélite;
- Flexibilizar a inserção de dados sobre a forragem com amostragem direta e indireta;
- Apresentar graficamente os resultados do modelo; e
- Desenvolver o protótipo funcional do FMIS.

Por outro lado, os objetivos desejáveis, especificados abaixo, são aqueles caracterizados como importantes mas secundários, passíveis de serem contemplados no trabalho após o atingimento dos essenciais:

- Refinar o modelo RNN LSTM;
- Estimar a confiabilidade de métodos indiretos já existentes; e
- Coletar dados de sensores de ultrassom para estimar altura de pastagens.

1.5 Metodologia de pesquisa

Para se obter resultados efetivos, há necessidade de definir a metodologia de pesquisa a ser adotada no projeto e, a partir desta definição, planejar uma sequência de atividades, orientadas pelos objetivos do trabalho, capazes de subsidiar respostas às questões de pesquisa, preliminares inicialmente, mas posteriormente conclusivas. Conforme Prodanov e Freitas (2013), toda pesquisa científica:

[...] percorre, desde a formulação do problema até a apresentação dos resultados, a seguinte sequência de fases:

- a) preparação da pesquisa: seleção, definição e delimitação do tópico ou problema a ser investigado; planejamento de aspectos logísticos para a realização da pesquisa; formulação de hipóteses e construção de variáveis;
 - b) trabalho de campo (coleta de dados);
 - c) processamento dos dados (sistematização e classificação dos dados);
 - d) análise e interpretação dos dados;
 - e) elaboração do relatório da pesquisa.
- (PRODANOV; FREITAS, 2013, p. 48)

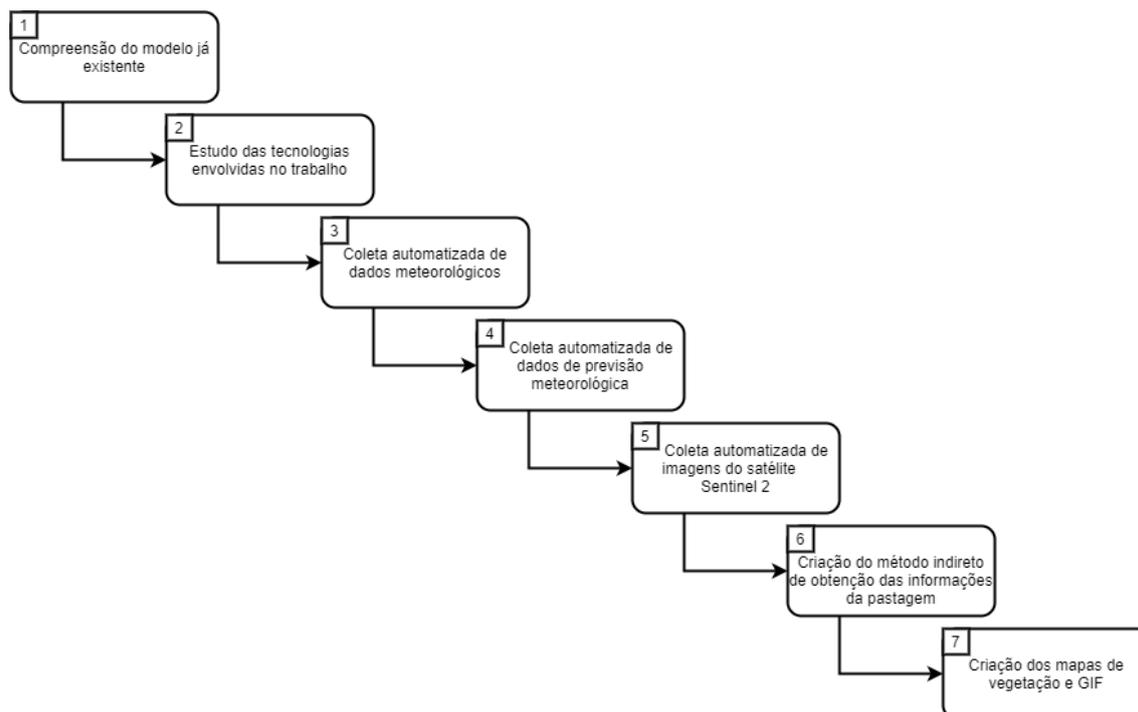
A pesquisa realizada neste trabalho percorreu os cinco itens mencionados anteriormente, visto que foi delimitado um problema de pesquisa, coletados dados de pastagem, meteorológicos e também de imagens, os dados foram processados para que fosse possível passar como entrada para o modelo de predição, após isso também foram feitas análises dos dados gerados pelo modelo de predição.

Além disso, o viés científico da pesquisa se caracteriza quando:

- a) discute ideias e fatos relevantes relacionados a determinado assunto, a partir de um marco teórico bem-fundamentado;
 - b) o assunto tratado é reconhecível e claro, tanto para o autor quanto para os leitores;
 - c) tem alguma utilidade, seja para a ciência, seja para a comunidade;
 - d) demonstra, por parte do autor, o domínio do assunto escolhido e a capacidade de sistematização, recriação e crítica do material coletado;
 - e) diz algo que ainda não foi dito;
 - f) indica com clareza os procedimentos utilizados, especialmente as hipóteses (que devem ser específicas, plausíveis, relacionadas com uma teoria e conter referências empíricas) com que trabalhamos na pesquisa;
 - g) fornece elementos que permitam verificar, para aceitar ou contestar, as conclusões a que chegou;
 - h) documenta com rigor os dados fornecidos, de modo a permitir a clara identificação das fontes utilizadas;
 - i) a comunicação dos dados é organizada de modo lógico, seja dedutiva, seja indutivamente;
 - j) é redigido de modo gramaticalmente correta, estilisticamente agradável, fraseologicamente claro e terminologicamente preciso.
- (PRODANOV; FREITAS, 2013, p. 48-49)

Este trabalho se amparou em uma revisão sistemática da literatura e possui relevância para a comunidade, visto que a pecuária é importante para a realidade econômica da região e do país, apresenta claramente os métodos usados, bem como os elementos necessários à verificação das conclusões. Portanto, constata-se que possui as principais características apresentadas por Prodanov e Freitas (2013). Além disso, mostrou-se relevante dividir o projeto em duas partes, com objetivos e procedimentos distintos. A primeira, exploratória, buscou, por meio de um planejamento flexível, a aquisição dos conhecimentos e habilidades necessárias sobre o domínio do problema e sobre as alternativas computacionais pertencentes ao estado da arte aplicáveis para a sua solução. Para a sua execução, definiu-se a necessidade da execução de procedimentos característicos de pesquisa documental, bibliográfica e experimental. Já a segunda, explicativa, mais complexa e capaz de aprofundar o entendimento sobre a realidade, almejou “explicar os porquês das coisas e suas causas, por meio do registro, da análise, da classificação e da interpretação dos fenômenos observados” (PRODANOV; FREITAS, 2013).

Figura 1 – Sequência de etapas para desenvolvimento



Fonte: Autor (2021)

Mais especificamente, a figura 1 apresenta as etapas delimitadas para o desenvolvimento do trabalho. A etapa inicial, necessária para compreensão detalhada do modelo de previsão de disponibilidade de forragem usado como base, foi seguida por um estudo de tecnologias passíveis de serem empregadas na concepção da solução. As etapas três, quatro, cinco, seis e sete envolveram a especificação, proposta, implementação e experimentação de novas funcionalidades do sistema, incluindo a coleta automatizada dos dados meteorológicos, previsão meteorológica e de imagens de satélite que servem como entrada para o modelo de previsão. A etapa seis consistiu na criação do método indireto de obtenção de informações de pastagens, utilizando o Índice de Vegetação com Diferença Normalizada (NDVI) para cada uma das áreas de estudo. Por fim, a etapa sete correspondeu ao módulo de visualização dos resultados, apresentados para o usuário em forma de gráficos de linha, mapas de vegetação e GIF animada do comportamento temporal do NDVI na área de interesse.

1.6 Estrutura do texto

O restante do trabalho está dividido em seis capítulos. No capítulo 2 é apresentada a fundamentação teórica, referente ao domínio do problema no contexto da agropecuária, seguida de uma breve descrição das tecnologias da informação e comunicação identificadas como relevantes para o estudo, encerrando-se com uma discussão sobre trabalhos correlatos identificados a partir da revisão da literatura. O capítulo 3 ilustra o modelo de FMIS proposto, incluindo uma descrição detalhada dos seus módulos, complementada por aspectos relevantes da implementação do protótipo de prova de conceito. No capítulo 4 é apresentada a avaliação experimental, contemplando a metodologia e o delineamento dos experimentos. O capítulo 5 expõe os resultados obtidos a partir da execução dos experimentos planejados para demonstrar o funcionamento e o refinamento dos módulos implementados. Na sequência, o capítulo 6 concentra as principais discussões sobre o processo de desenvolvimento e os resultados obtidos, tendo em perspectiva as questões de pesquisa e os objetivos específicos. Por fim, no capítulo 7, são elencadas as conclusões da pesquisa exploratória-experimental realizada neste projeto, seguidas pela apresentação de trabalhos futuros, elencando possíveis funcionalidades a serem incorporadas ao sistema e também explicitando possíveis formas de implementação.

1.7 Considerações finais

Com esta pesquisa pretendeu-se identificar não só o impacto do clima e dos animais no crescimento das pastagens, mas também quais impactos serão refletidos no modelo de previsão com o incremento de dados de previsão meteorológica. Também visou estabelecer um método indireto confiável comparando-se ao método direto que já vem sendo executado. Além disso, do ponto de vista de desenvolvimento tecnológico e de inovação, almejou-se entregar um protótipo funcional de FMIS que poderá evoluir para uma solução aplicável em pesquisas na área de manejo de pastagem e no sistema produtivo, por meio de processos de transferência de tecnologia futuros. Na sequência, são apresentadas as referências que servem de base para a execução do trabalho.

2 REFERENCIAL TEÓRICO

Como o presente trabalho é interdisciplinar, demanda um levantamento bibliográfico não apenas na área de Pecuária de Precisão (Precision Livestock Farming - PLF), com foco no domínio do problema, mas também para identificação de alternativas pertencentes ao estado da arte na área de computação, sobretudo em relação às Tecnologias da Informação e Comunicação (TIC) aplicáveis na solução do problema. Entre as temáticas principais estudadas, na área de PLF cabe descrever aspectos relevantes observados sobre gestão da pastagem, em particular no que se refere à avaliação e predição da quantidade forragem, dependente de métodos de amostragem diretos e/ou indiretos. Na área de TIC, diversas alternativas aplicáveis à agropecuária (AgroTIC) se apresentam como promissoras, com destaque para os avanços nas áreas de Smart Farming, Big Data, Internet of Things e Cloud Computing. Em especial, os Farm Management Information Systems se destacam como forma potencial para operacionalizar soluções especializadas para problemas específicos da agropecuária, os quais, para serem efetivos, demandam escolhas adequadas não apenas para os seus componentes responsáveis pela coleta, armazenamento, processamento e visualização de dados, as quais, na proposta deste trabalho, incorporam características de Sistemas de Apoio a Decisão, de Banco de Dados Espaço-Temporais e de Aprendizado de Máquina. Por fim, cabe analisar o sistema proposto neste trabalho em contraste com outros encontrados na literatura, os quais possuem características em comum.

2.1 Desafios de gestão da pastagem

Para o Conselho de Pesquisa de Bovinos de Corte (Beef Cattle Research Council - BCRC), agência canadense liderada pela indústria que financia pesquisas com carne (*beef*), gado (*cattle*) e forragem (*forage*), a pastagem (*pasture*) “é um recurso crítico na indústria de gado” (BCRC, 2019, tradução nossa). Segundo o BCRC, inovações na gestão da pastagem permitem que os produtores gerenciem de forma combinada o ecossistema da pastagem com o seu uso para a produção de alimento. Além disso, destaca a importância do plano de gestão:

An effective management plan requires clear understanding of forage production, realistic production goals, effective grazing strategies and timely response to forage availability and environmental changes¹. (BCRC, 2019).

Portanto, é nítido que a gestão da pastagem exerce papel determinante em um sistema de pastejo (*grazing system*), caracterizado pela “combinação integrada entre os componentes animal, planta, solo, e fatores ambientais, mais o método de pastejo, com o objetivo de se atingir metas específicas.” (PEDREIRA, 2002, p. 5).

O aumento na eficiência na criação de bovinos em pastagens demanda o gerenciamento da quantidade de forragem produzida e seu valor nutritivo, bem como da interação entre os animais e as forrageiras. [...] Para este sistema complexo e desafiador, cada vez há maior demanda por tecnologias que permitam melhorar a sua gestão. Em especial, o manejo nutricional traz inúmeras oportunidades de emprego de tecnologias que coletam e gerem informações para auxiliarem na tomada de decisão. Decisões mais bem informadas permitem uso mais ajustado de recursos, o que garante um melhor resultado econômico e, concomitantemente, menor impacto ambiental. (MEDEIROS et al., 2014, p. 1)

A avaliação de massa de forragem (MF) deve ser precisa para um adequado manejo da pastagem. A MF é um importante parâmetro para a eficácia no gerenciamento para tomada de decisão na produção animal.

To choose a particular method to estimate herbage mass production in a pasture, issues such as the growth of the forage species, equipment costs and labor, and the number of samples required for an accurate measurement must be considered². (EDVAN et al., 2016, p. 37)

A estimativa de massa de forragem é fundamental, pois também nos permite calcular a taxa de lotação, estimar a quantidade de forragem consumida e interpretar o rendimento da produtividade animal (ESTRADA; JÚNIOR; REGAZZI, 1991). As variáveis meteorológicas, tais como temperatura, precipitação e luminosidade, são determinantes para a produtividade das forragens e afetam diretamente características de grande importância, como o acúmulo de massa e a estacionalidade de produção dessas plantas (TONATO et al., 2011).

¹Um plano de gestão efetivo requer entendimento claro sobre a produção de forragem, objetivos de produção realísticos, estratégias de pastejo efetivas e respostas atempadas à disponibilidade de forragem e mudanças ambientais.

²Para escolher um método particular para estimar a produção de massa de forragem em uma pastagem, questões como o crescimento das espécies forrageiras, custos de equipamentos e trabalho, e número de amostras requeridas para uma medida precisa devem ser consideradas.

Para a determinação de massa de forragem em áreas específicas, são coletadas amostras da pastagem no local. A utilização de procedimentos de amostragem gera a dificuldade de estabelecer a quantidade mínima de amostras necessárias para representar toda a variabilidade existente na área de estudo. O número mínimo de amostras também deve levar em conta também a não remoção de quantidades excessivas de pastagem (CUNHA, 2002).

São diversas as técnicas para a estimativa de massa de forragem. As técnicas foram sendo criadas, visto a diversidade de fatores nas quais podem afetar o resultado do estudo. Alguns desses fatores são: uniformidade e densidade da forragem, hábito de crescimento, perfilhamento e altura da planta, etc. (CARVALHO, 2008).

As técnicas podem ser classificadas como diretas e indiretas para estimar a massa de forragem. O método direto geralmente proporciona maior precisão quando comparado com outros métodos. Porém, a relação entre quantidade de amostras e destruição da pastagem deve ser muito bem mensurada, visto que pastagens extensas com grande variabilidade na cobertura exigem maior quantidade de mão de obra e equipamentos. Esse fato faz com que a quantidade de amostradas seja reduzida, o que acarreta em uma baixa precisão nos resultados (CÓSER; MARTINS; DERESZ, 2002). Os métodos indiretos são assim chamados pelo fato de não destruírem a pastagem. Estes métodos foram desenvolvidos, principalmente, pela alta demanda de tempo associada a coleta de amostras em um método direto (CUNHA, 2002). Hodgson et al. (1981) definiram que para a criação dos métodos indiretos, alguns aspectos foram utilizados para atender as necessidades. São eles:

- Redução da quantidade de trabalho, equipamentos, tempo e/ou recursos, gerando redução no custo de amostragem;
- Possibilidade de utilização em áreas com animais ou locais de difícil acesso;
- Possibilidade de utilização em áreas menores;
- Permitir a classificação de tratamentos com grandes diferenças comparativas;
- Promover um guia para estimar a massa de forragem em sistemas de produção animal.

2.1.1 Métodos diretos

O método direto baseia-se no corte e remoção da pastagem, podendo ser em uma área amostral ou área total onde está sendo realizada a avaliação. A seguir serão expostos alguns dos métodos diretos.

- (a) Método do corte:** O método do corte, ou método de ceifa, é muito usado em sistemas de pastejo contínuo com uso de gaiolas. Segundo Gardner (1986), para se obter a melhor estimativa de massa de forragem, esta deve ser cortada e pesada. A altura de corte ao nível do solo deve ser padronizada, com objetivo de minimizar erros. Entretanto, cortes drásticos repetidas vezes podem prejudicar a rebrota das forrageiras (JONES, 2000). Utilizando essa técnica, colhem-se amostras e a partir delas, estima-se a quantidade de forragem disponível por unidade de área. A figura 2 mostra uma gaiola utilizada em um ponto específico na área avaliada.

Figura 2 – Utilização de gaiola



Fonte: Adaptado de Machado e Kichel (2004)

- (b) Método do quadrado:** Este método utiliza uma moldura de área conhecida fabricada de madeira ou metal de forma quadrada ou retangular, a mais comum é a forma quadrada. Segundo Salman, Soares e Canesin (2006), as áreas do quadrado variam entre 0,10 m² até 2,0 m². Porém, é ressaltado que o tamanho depende da uniformidade da área a ser amostrada. Para vegetações baixas, rasteiras e densas o mais adequado é fazer cortes rentes ao solo. Para capins de crescimento ereto deve-se considerar a altura de pastejo dos animais. Em razão da pequena área de

amostragem torna-se necessária a tomada da medição em 10 a 20 pontos por hectare da pastagem para aumentar a confiabilidade da amostragem (WHITE; HODGSON, 1999). A figura 3 mostra uma moldura utilizada para realização do procedimento de corte.

Figura 3 – Método do quadrado



Fonte: Adaptado de Garcia (2017)

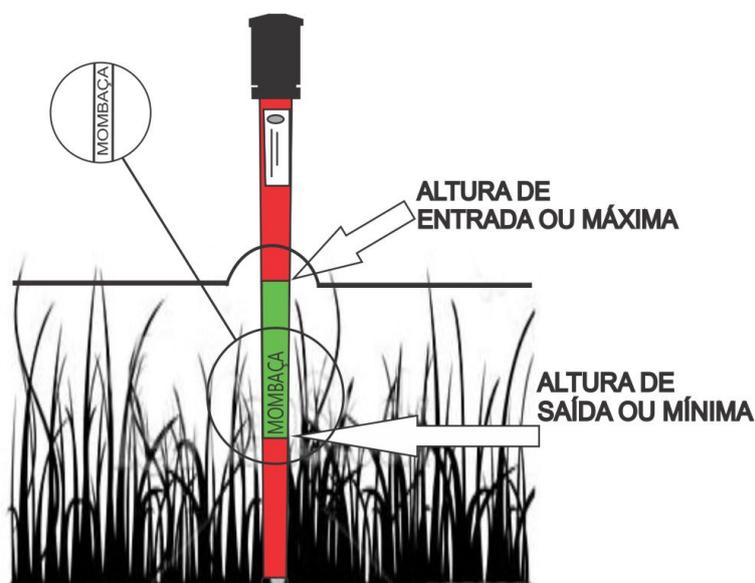
2.1.2 Métodos indiretos

Como a utilização de métodos diretos demanda muita mão de obra e são inviabilizados em grandes extensões de pastagem, foram desenvolvidos métodos que visam tornar operacional o número de amostras, sem inviabilizar o processo de avaliação. Os chamados métodos indiretos, ou não destrutivos, minimizam a remoção da pastagem e diminuem o tempo de coleta e processamento dos dados (CUNHA, 2002). A seguir serão apresentadas algumas das técnicas indiretas utilizadas para coleta de dados.

- (a) **Altura do dossel:** É um procedimento que calibra a altura como indicador de massa de forragem. A altura pode ser tomada com uma régua em um grande número de pontos da área sob avaliação sendo um método simples e não requer equipamento sofisticado (PEDREIRA, 2002). A partir da altura é definido quando os animais devem entrar e sair de determinado piquete consultando-se uma tabela que indica a melhor altura para entrada e saída dos animais. Melhores resultados serão obtidos se a densidade do dossel for uniforme e constante ao longo de todo seu

perfil. Porém, são raros os casos em que isso ocorre, com isso, frequentemente a massa de forragem é superestimada, quanto mais alto for o dossel (SALMAN; SOARES; CANESIN, 2006). A figura 4 apresenta uma ilustração da régua de manejo de pastagens.

Figura 4 – Régua de medição de altura



Capim mombaça na altura de entrada dos animais na pastagem.

Fonte: Adaptado de Tecnologia (2019)

(b) Sonda Eletrônica: Para realizar medições este equipamento baseia-se em diferenças de capacitância. A quantidade de carga elétrica existente em dois condutores isolados é descrita pela capacitância (CUNHA, 2002). Com base na capacidade elétrica da forragem, considerando que a mesma tem uma alta capacitância e a do ar é considerada baixa, as medidas são tomadas. A capacitância da massa de forragem depende da espécie de planta e da sua concentração de umidade, tornando necessário efetuar muitas calibrações (PEDREIRA, 2002). As calibrações devem ser realizadas não só para o tipo de forrageira, mas também para as estações do ano (CARVALHO, 2008). A figura 5 mostra um equipamento de medição de capacitância.

(c) Estimativas Visuais: Esse método tem como princípio a avaliação da massa de forragem com relação a uma medida anterior da mesma área. Isso possibilita a inferência sobre informações da pastagem (SALMAN; SOARES; CANESIN, 2006). Essa técnica tem por vantagem o baixo custo operacional, o grande número de

Figura 5 – Sonda Eletrônica



Fonte: Adaptado de Pro (2016)

amostras que se pode coletar, demanda pouco tempo de trabalho, em comparação aos outros métodos, porém, pode ser muito impreciso, devido as variações sazonais na altura da pastagem, densidade e teor de umidade do pasto. Vale ressaltar também a necessidade de treinamento prévio (CARVALHO, 2008).

(d) NDVI: As pastagens obtém significativa parcela de cobertura do solo do globo terrestre. Associado a isso, existem várias constelações de satélites que disponibilizam seus dados gratuitamente e que auxiliam nas análises para monitoramento dos recursos de pastagens (PUNALEKAR et al., 2018). Segundo Damasceno et al. (2020), a utilização do Índice de Vegetação por Diferença Normalizada (NDVI) apresenta bons resultados, caso o objetivo seja obter informações de safra, como produtividade potencial. Também relata que, em comparação com outros índices, o NDVI é o que menos sofre influência das variações atmosféricas. A avaliação dos índices de vegetação permite perceber características, como estado nutricional das plantas. Algumas das possíveis aplicações na utilização do NDVI estão a estimativa de produtividade, biomassa, teor de proteína, teor de fertilizante de nitrogênio e controle de ervas daninhas. Uma outra questão importante de se ressaltar é que a

precisão em prever a biomassa e o rendimento da cultura usando os índices de vegetação depende de vários fatores, como o clima da região, cultura, estágio fenológico de avaliação (COELHO; ROSALEN; FARIA, 2018).

2.2 Tecnologias da Informação e Comunicação aplicáveis à Agropecuária

Modificações ocasionadas nos processos de desenvolvimento, e suas consequências na democracia e cidadania, convergem para uma sociedade caracterizada pela importância crescente dos recursos tecnológicos e pelo avanço das Tecnologias de Informação e Comunicação (TIC) com impacto nas relações sociais, empresariais e nas instituições. TIC tem contribuído ao longo dos anos de maneira impactante para as diversas áreas de conhecimento, permitindo armazenar e processar grandes volumes de dados, automatizar processos e também o intercâmbio de informações e de conhecimento (MASSRUHÁ et al., 2014).

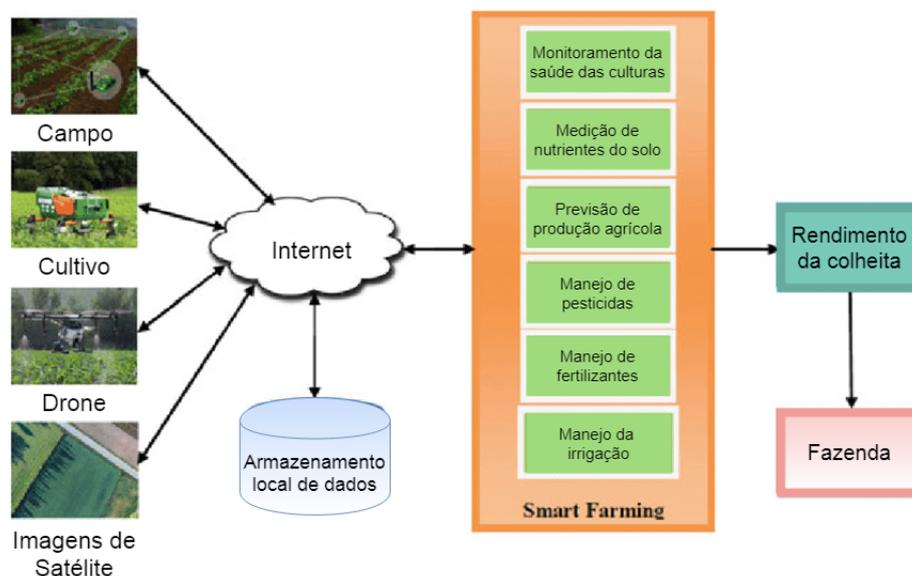
Com o passar dos anos surgiram as TIC voltadas para o contexto de agropecuária, denominadas AgroTIC, sendo definidas por Massruhá, Leite e Moura (2014) como: “a combinação de hardware, software e instrumentos de produção que permitam coleta, armazenamento, troca, processamento e manejo da informação e do conhecimento”. Agricultura de precisão envolvendo a aplicação de inteligência embarcada, automação e rede de sensores locais pra monitoramento da produção e controle de processos como mapeamento de solos, monitoramento de doenças e de variáveis meteorológicas e irrigação inteligente, atividades de sensoriamento remoto, mineração de dados e *big data*, modelos de inteligência computacional e simulação para emissão de alertas e suporte a decisão agropecuária, são algumas das aplicações das AgroTIC (BAMBINI; ROMANI, 2014).

2.2.1 Smart Farming

À medida que máquinas inteligentes e sensores surgem em fazendas e dados de fazenda crescem em quantidade, os processos agrícolas tornam-se cada vez mais orientados a dados e ativados por dados (WOLFERT et al., 2017). Smart Farming visa capacitar os agricultores de hoje com as ferramentas de decisão e as tecnologias de automação que integram perfeitamente produtos, conhecimento e serviços para melhor produtividade, qualidade e lucro (JHURIA; KUMAR; BORSE, 2013).

Segundo Wolfert, Goense e Sørensen (2014) a Agricultura de Precisão apenas leva em consideração a variabilidade no campo, enquanto Smart Farming baseia 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. Também envolve o uso das Tecnologias da Informação e da Comunicação (TIC), em particular, conceitos de Internet das Coisas (IoT) e análises de Big Data, para aumentar a produção e rentabilidade da fazenda, por meio do monitoramento eletrônico das culturas, do solo, de fertilização e meio ambiente e condições de irrigação. Esses dados de monitoramento podem ser analisados para identificar potenciais culturas e variedades específicas que irão atender melhor as metas de produtividade da propriedade (JAYARAMAN et al., 2016). A figura 6 apresenta uma possível arquitetura em uma Smart Farming.

Figura 6 – Exemplo de arquitetura em uma Smart Farming



Fonte: Adaptado de Rajasekaran e Anandamurugan (2019)

Smart Farming explora as limitações da agricultura tradicional por meio de monitoramento em tempo real, minimização da mão-de-obra, tempo, estimativa precisa da água necessária para irrigação e protege a colheita de desastres como doenças (RAJASEKARAN; ANANDAMURUGAN, 2019). A Tabela 1 apresenta as principais diferenças existentes entre o método de Agricultura Tradicional e Smart Farming.

Tabela 1 – Diferenças entre Agricultura Tradicional e Smart Farming

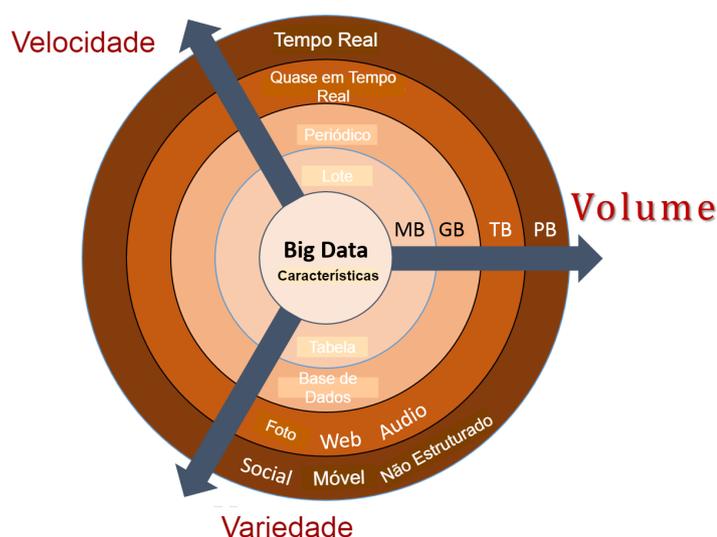
Exigências agrícolas	Agricultura tradicional	Smart Farming
Necessidade de pessoal	Sim	Não
Utilização de água	Alto	Baixo
Monitoramento de campo	Alto	Baixo
Aquisição de dados	Baixo	Alto
Custo	Alto	Baixo
Rendimento	Baixo	Alto
Tempo	Elevado	Reduzido

Fonte: Adaptado de Rajasekaran e Anandamurugan (2019)

2.2.2 Big Data

O conceito de Big Data se refere aos enormes conjuntos de dados que são coletados periodicamente e analisados computacionalmente (DASORIYA, 2017). Este fenômeno é caracterizado por três aspectos: Volume de dados, velocidade de aquisição e grande variedade de tipos de dados. Essas características são ilustradas na figura 7.

Figura 7 – Aspectos fundamentais de Big Data



Fonte: Adaptado de Heavilin (2018)

Devido à apropriação da sociedade atual das tecnologias digitais, algumas que até pouco tempo atrás eram usadas somente por grandes corporações ou governo, como por exemplo o Global Positioning System (GPS), existe uma crescente no volume de dados,

na velocidade em que podem ser gerados e inúmeros tipos de dados (JUNIOR, 2012).

Na área do agronegócio, há necessidade de grande capacidade de armazenamento e processamento de dados em sistemas que precisam operar em tempo real para algumas aplicações, como o monitoramento de animais, gerando a necessidade de investimentos em infraestrutura (KAMILARIS; KARTAKOULLIS; PRENAFETA-BOLDÚ, 2017). Com a utilização do conceito de Smart Farming, praticamente todos os atuadores no processo produtivo geram dados, tais como sensores que medem dados relacionados ao ambiente, que coletam dados meteorológicos e até mesmo recebem dados de outros sensores. Assim, as tecnologias de Big Data se fazem essenciais para que o produtor possa realizar a tomada de decisão da melhor maneira possível em cenários compostos por grande volume de dados (WOLFERT et al., 2017).

2.2.3 Internet das Coisas

A Internet das Coisas (Internet of Things - IoT) diz respeito ao relacionamento real entre objetos, humanos e animais, no qual estes passaram a ser portadores de dispositivos computacionais capazes de se conectar e comunicar entre si. Nesse sentido, não há necessidade de especialistas realizando comandos e os objetos tendem a assumir o controle das ações do dia a dia (SANTAELLA et al., 2013). Segundo Santos et al. (2016), IoT começou a ter destaque devido aos avanços de áreas como microeletrônica, sistemas embarcados, comunicação móvel, sensoriamento e computação ubíqua.

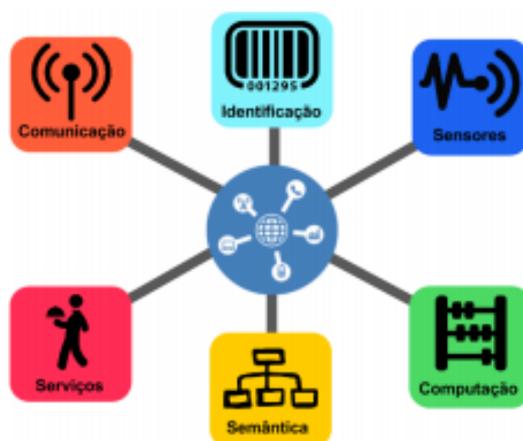
Devido à enorme diversidade de tecnologias de *hardware* e *software*, o paradigma de IoT possuía vários desafios a serem superados. O primeiro deles corresponde à heterogeneidade dos diversos ambientes que são capazes de se conectar, o que faz com que as aplicações suportem a comunicação de diferentes tipos de componentes que pertençam à solução. Com isso surge a figura de uma plataforma de *middleware* que provê a interoperabilidade e gerência entre os componentes. Um outro desafio diz respeito à enorme necessidade de escalabilidade dos ambientes de IoT, visto que uma solução deve ser capaz de suportar inúmeros dispositivos e também a enorme quantidade de dados gerados (PIRES et al., 2015). Esse desafio se justifica pois é estimado que até 2020 o número de dispositivos conectados a Internet ultrapasse 50 bilhões (SAADEH; ALMOBAIDEEN; SABRI, 2017).

Santaella et al. (2013) relatam que para que fosse possível a grande difusão de IoT três aspectos precisaram ser tratados. O primeiro deles era a maneira eficiente de

identificação dos objetos, com relação ao custo, como por exemplo, RFID e QRCode. Outro aspecto foi a evolução do sensoriamento para que os sistemas fossem capazes de ser autônomos com relação a coleta de dados e também fossem capazes de se comunicar com objetos que estivessem no extremo da rede. Por fim, a diminuição de tamanho dos dispositivos fez com que dispositivos cada vez menores fossem capazes de se conectar.

A gama de novas aplicações é diretamente proporcional a crescente de dispositivos capazes de conexão com a Internet. Conseqüentemente, criando-se novas aplicações, novos desafios surgem, visto que os dados coletados podem apresentar imperfeições, pode haver inconsistências, serem de diferentes tipos, assim, deve haver um cuidado no tratamento dos dados por parte das aplicações e algoritmos e estas devem estabelecer níveis de confiança sobre os dados obtidos (SANTOS et al., 2016).

Figura 8 – Blocos básicos da IoT



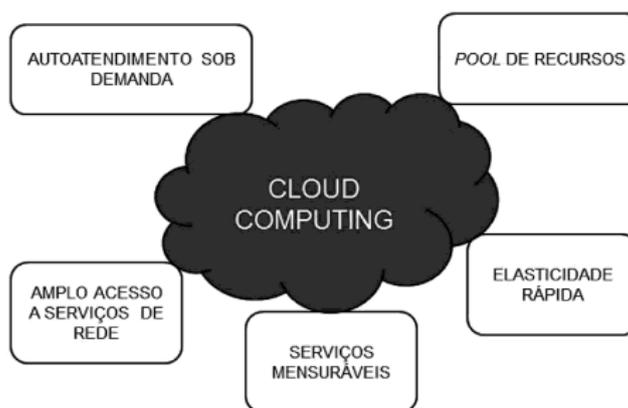
Fonte: Adaptado de Santos et al. (2016)

A figura 8 apresenta os blocos básicos da IoT. O primeiro bloco corresponde na identificação de maneira única dos objetos para conectá-los à Internet. O segundo bloco diz respeito aos sensores que coletam, armazenam e/ou encaminham dados para centros de armazenamento. No bloco referente a computação está a unidade de processamento dos dados. O bloco de semântica refere-se a habilidade dos objetos em adquirir conhecimento a partir dos dados obtidos. No penúltimo bloco estão os diversos serviços que a IoT pode fornecer, como por exemplo, serviços de identificação, serviços de agregação de dados, etc. Por fim o bloco de comunicação diz respeito aos protocolos que são utilizados para conectar os diversos objetos (ALVES; CAVALCANTI, 2018).

2.2.4 Cloud Computing

De acordo com Mell e Grance (2011), Cloud Computing é um modelo para permitir acesso ubíquo, conveniente e de rede sob demanda a um conjunto compartilhado de recursos de computação configuráveis que podem ser rapidamente provisionados e liberados com o mínimo esforço de gerenciamento ou interação do provedor de serviços. Este modelo deve seguir cinco aspectos fundamentais, ilustrados na figura 9:

Figura 9 – Aspectos fundamentais de Cloud Computing



Fonte: Adaptado de Veras (2012)

- **Serviço sob demanda:** Um consumidor pode solicitar recursos de computação unilateralmente, automaticamente, conforme necessário, sem a necessidade de interação humana com cada provedor de serviços.
- **Amplo acesso aos serviços:** Os recursos estão disponíveis na rede e podem ser acessados por dispositivos heterogêneos.
- **Agrupamento de recursos:** Os recursos de computação do provedor são agrupados para atender vários consumidores usando um modelo de multilocatário, com diferentes recursos físicos e virtuais dinamicamente atribuídos e reatribuídos de acordo com a demanda do consumidor, sem que haja o conhecimento da localização dos recursos utilizados.
- **Elasticidade:** A necessidade por recursos pode ser atendida rapidamente a partir da demanda do consumidor.
- **Monitoramento de serviços:** O uso de recursos pode ser monitorado, controlado e relatado, fornecendo transparência ao provedor e ao consumidor do serviço utilizado.

Existem três modelos de serviço em Cloud Computing, são eles: Software como um serviço, plataforma como um serviço e infraestrutura como um serviço. Quatro modelos de implantação são possíveis: Nuvem privada, nuvem comunidade, nuvem pública e nuvem híbrida (VERDI et al., 2010). Esses modelos subsidiam novos modelos de negócios envolvendo serviços como armazenamento, organização, validação, processamento e análise de grandes volumes de dados.

Devido ao grande volume de dados que são gerados por soluções, em qualquer contexto de negócio, se faz necessário o uso de Cloud Computing para processamentos intensos e armazenamento em massa, já que muitas vezes se torna inviável realizá-los no próprio dispositivo de processamento/armazenamento (STERGIOU et al., 2018).

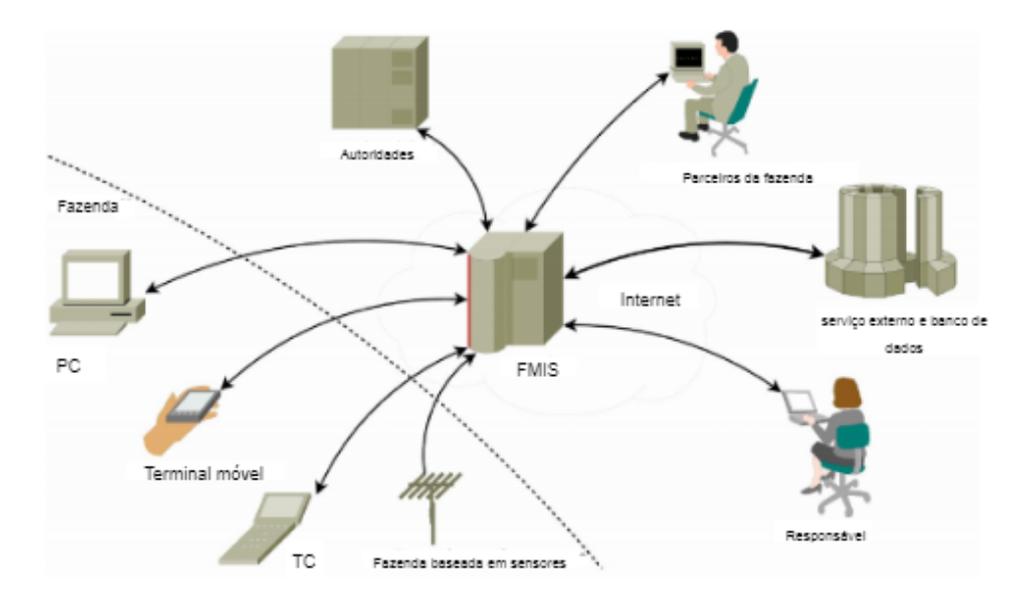
2.3 Farm Management Information Systems

Um Farm Management Information System (FMIS) é definido como um sistema planejado para a coleta, processamento, armazenamento e disseminação de dados na forma de informações necessárias para executar as funções operacionais da fazenda (SØRENSEN et al., 2010). Segundo Boehlje, Eidman e Errington (1985), os FMIS são definidos como ferramentas eletrônicas para coleta e processamento de dados com o objetivo de fornecer informações de valor potencial na tomada de decisões gerenciais. Murakami et al. (2007) afirma que os componentes essenciais do FMIS incluem projetos específicos orientados ao agricultor, interfaces de usuário dedicadas, funções automatizadas de processamento de dados, conhecimento especializado e preferências do usuário, comunicação de dados padronizada e escalabilidade.

As razões pelas quais uma gestão sofisticada da fazenda é uma tarefa tão importante e desafiadora são certamente diversas, no entanto, um ambiente complexo, estruturas agrícolas complexas e a introdução de tecnologias modernas no setor agrícola são os três fatores principais de acordo com Sørensen et al. (2010). O ambiente tornou-se complexo devido as novas preocupações em relação ao processo produtivo de uma propriedade, tais como: novas e mais rigorosas diretrizes para o uso de agroquímicos, segurança alimentar, requisitos de bem-estar animal e preocupações ambientais. As estruturas agrícolas tornaram-se mais complexas devido ao aumento das extensões das mesmas, porém, sem redução substancial das áreas cultivadas, conseqüentemente a complexidade da gestão aumentou. O terceiro fator está na adoção de novas tecnologias ao ambiente agrícola, visto que muitos produtores introduziram *softwares* para coleta, armazenamento e processa-

mentos dos dados, máquinas acionadas por Global Positioning System (GPS) e aplicações que utilizam Sistema de Informações Geográficas (SIG).

Figura 10 – Arquitetura de um sistema baseado em FMIS



Fonte: Adaptado de Salami e Ahmadi (2010)

Através da figura 10 é possível verificar a complexidade de um sistema agrícola atualmente, visto a quantidade de tecnologias que podem ser incorporadas ao sistema. A arquitetura apresentada tem como ponto central o FMIS, ao qual todas as outras partes estão conectadas. As subseções 2.3.1, 2.3.2, 2.3.3 e 2.3.4 apresentam conceitos e ferramentas relevantes no sentido da coleta, armazenamento e processamento de dados para tomada de decisão aplicáveis em um FMIS.

2.3.1 Sistemas de Apoio a Decisão

Um sistema baseado em computador que representa e processa o conhecimento de maneiras a auxiliar o usuário a tomar decisões mais produtivas, ágeis, inovadoras e respeitáveis, é considerado um Sistema de Apoio a Decisão (SAD) (FJERMESTAD, 2009). A importância de um SAD no contexto da cadeia produtiva, se justifica através da melhor compreensão e, conseqüentemente, melhor tomada de decisão para a resolução de problemas específicos (FERNANDES; PESSOA, 2005).

No contexto de SAD existem estudos nos quais apontam dois focos em termos de metodologia multicritério: a Metodologia Multicritério em Apoio à Decisão (MCDA)

que visa gerar conhecimento aos decisores e a Metodologia Multicritério de Tomada de Decisão (MCDM) que visa encontrar uma solução ótima a todos os decisores envolvidos (AREND, 2010).

Um SAD reúne recursos de persistência de dados e de Descoberta de Conhecimento em Banco de Dados (Knowledge Discovery in Databases - KDD) a uma interface capaz de apresentar de maneira eficiente o conhecimento extraído, com o objetivo de auxiliar o seu usuário no processo de tomada de decisão. Porém, segundo Chan et al. (2017) existe uma resistência dos potenciais usuários acerca da utilização de um SAD, mesmo estando cientes de que decisões podem ser tomadas com maior rapidez e precisão.

A crescente complexidade dos problemas na agropecuária, como por exemplo, a queda da produção, erosão do solo, declínio no preço de mercado devido a competição internacional, aumento dos custos de fertilizantes e de resistência às pragas pelas culturas, barreiras econômicas e fitossanitárias, entre outras, justificam o aumento no desenvolvimento dos SAD (FERNANDES; PESSOA, 2005).

2.3.2 Bancos de Dados Espaço-Temporais

Banco de Dados pode ser definido como a organização e armazenagem de informações sobre um domínio específico, ou seja, é o agrupamento de dados que tratam do mesmo assunto, e que precisam ser armazenados para segurança ou conferência futura. (TANAKA; CAMARGO; GOTARDO, 2012).

Os dados eram mantidos aleatoriamente em arquivos, geralmente como partes integrantes da aplicação até meados da década de 60. Posterior a essa época começaram a surgir os primeiros Sistemas de Gerenciamento de Banco de Dados (SGBD) provendo armazenamento dos dados de forma independente da aplicação, contudo, sem mecanismos de acesso eficientes (BOSCARIOLI et al., 2006). Um SGBD é uma coleção de programas que permite aos usuários criar e manter um banco de dados. Em outras palavras é um sistema de *software* de propósito geral que facilita os processos de definição, construção, manipulação e compartilhamento de bancos de dados entre vários usuários e aplicações.

A definição de um banco de dados consiste na especificação dos tipos de dados, restrições e estruturas. A construção é o processo de armazenar os dados em alguma mídia na qual seja capaz de ser controlada pelo SGBD. A manipulação dos dados é feita a partir de funções de pesquisa, inserção, atualização e exclusão, denominadas como *select*, *insert*, *update* e *delete*, respectivamente. Por fim o compartilhamento permite a inúmeros

usuários e aplicações, acessar de forma concorrente os dados armazenados (ELMASRI et al., 2005).

Existem Bancos de Dados Relacionais (BDR) nos quais uma coleção de tabelas, todas com nomes únicos, compõe a base de dados, podendo estar relacionada a uma ou mais tabelas. Conceitos como integridade referencial de dados e chaves primárias estão presentes e garantem que um conjunto de informações possa ser representado de maneira consistente. Os chamados Bancos de Dados Orientados a Objetos (BDOO) possuem três pilares principais: herança, polimorfismo e encapsulamento. Este modelo apresenta maior flexibilidade na manipulação de seu conteúdo e por meio de identificadores de objetos manipula os dados de forma consistente (BOSCARIOLI et al., 2006).

No contexto do trabalho, faz-se necessário a utilização de um banco de dados espaço-temporal. Os SGBD-Relacionais a partir de meados da década passada se tornaram capazes de lidar de maneira ampla e eficiente com dados geográficos. Atualmente os SGBD-Relacionais abrangem o modelo geométrico que contém representações para pontos, linhas, polígonos, coleções homogêneas e heterogêneas de geometrias, com isso é possível introduzir em determinada coluna da tabela informações que representam geometrias (QUEIROZ; MONTEIRO; CÂMARA, 2013).

Segundo Casanova et al. (2005), os modelos espaço-temporais reúnem dois aspectos distintos: a escolha de conceitos adequados do espaço e do tempo e a construção de representações computacionais apropriadas correspondentes a esses conceitos. A superfície pode ser modelada a partir de geo-objetos que podem ser estáticos, mudar de lugar ou até mesmo estar em constante movimento. Estes geo-campos podem representar uma parcela de solo ou indicar a variação espacial da vegetação da mesma área.

2.3.3 Mineração de Dados

Com o advento de novas tecnologias, como por exemplo, banco de dados, Data Warehouses, Bibliotecas Virtuais, etc e também com o preço do hardware cada vez mais baixo, o volume de dados produzido é cada vez maior. Porém a problemática deixou de ser em relação ao armazenamento dos dados e sim o que fazer com os dados produzidos. (CAMILO; SILVA, 2009).

Na busca por informações importantes existentes nos dados, a Mineração de Dados é uma técnica muito eficaz para extrair conhecimento a partir de um grande volume de dados. A partir disso é possível estabelecer relações e padrões entre os dados, com

isso, é possível auxiliar na tomada de decisão (GALVÃO; MARIN, 2009).

Para Amo (2004) a Mineração de Dados é apenas uma etapa do processo de KDD. Sendo essa a etapa responsável por extrair padrões de interesse utilizando técnicas inteligentes. A Seção 2.3.4 apresenta todas as etapas presentes no processo de KDD.

Braga (2005) estabelece um modelo genérico de projeto de Mineração de Dados. Dentre as etapas está a aquisição e avaliação dos dados que é considerada um dos pilares de todo projeto e se compromete cumprir os seguintes passos:

- Adquirir os dados
- Formatar os dados
- Criar ambiente e ferramentas
- Validar aquisição e formatação
- Criar amostras (aleatórias) de trabalho
- Particionar os dados

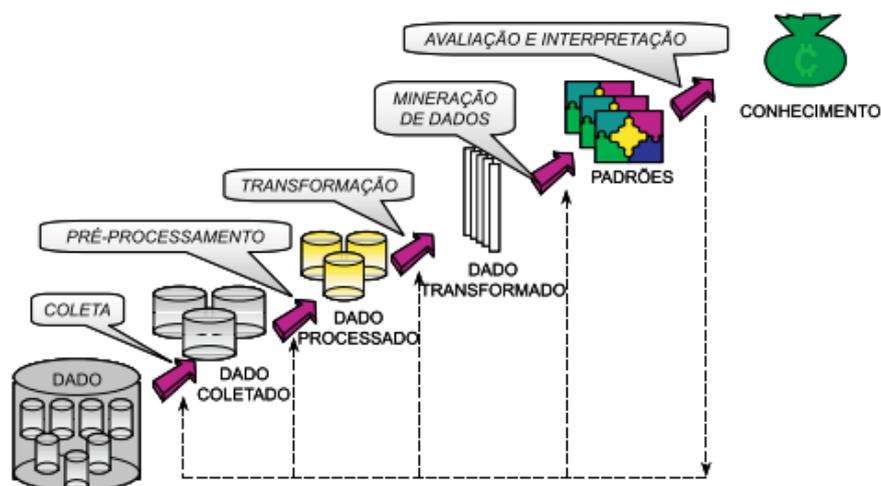
2.3.4 Aprendizado de Máquina

O aprendizado de máquina pode ser definido como uma subárea da Inteligência Artificial (IA), que pesquisa métodos computacionais que visam a aquisição automática de novos conhecimentos, habilidades e formas de organizar o conhecimento já adquirido (BATISTA, 2003). Um sistema de aprendizado de máquina envolve a definição de seu paradigma de aprendizado e também a estratégia de aprendizado. O paradigma de aprendizado define a forma como o conhecimento adquirido será representado, já a estratégia definirá como o aprendizado ocorre (LEE, 2000). Características como a capacidade de resolver autonomamente problemas não lineares e o fornecimento de uma estrutura poderosa e flexível para auxiliar a tomada de decisões orientada a dados, através da incorporação de conhecimento especializado, as tornam amplamente utilizadas em vários contextos, e altamente aplicáveis a agricultura e pecuária de precisão (CHLINGARYAN; SUKKARIEH; WHELAN, 2018). As técnicas de aprendizado de máquina existentes são aprendizado supervisionado e não supervisionado. Em supervisionado é fornecido um conjunto de dados de treinamento formado pelas entradas e saídas desejadas. Já em não supervisionado apenas os atributos de entrada estão disponíveis, ou seja, não contém pares de entradas-saídas desejadas e sem indicações sobre a adequação das saídas produzidas (SOUTO et al., 2003).

Dentro do contexto de aprendizado de máquina, está a técnica de KDD. Devido ao crescimento exponencial no tamanho das bases de dados, tornou-se inviável o trabalho de busca por padrões por especialistas humanos. Com isso, há necessidade por técnicas nas quais seja possível auxiliar os especialistas, de maneira inteligente e autônoma, na procura por informações desconhecidas nos dados. Diversas fases são incorporadas no processo de KDD, sendo elas: identificação e entendimento do problema, identificação de dados relevantes, coleta de dados, pré-processamento, transformação de dados, mineração de dados e avaliação e interpretação de resultados (BATISTA, 2003).

Na figura 11 são ilustradas as fases do processo de KDD, onde na primeira fase (dados) são atribuídos, o processo de identificação e entendimento do problema, ou seja, saber dentre as necessidades do usuário quais podem ser resolvidas utilizando algum método de Mineração de Dados (MD) e também a identificação de dados relevantes. Na segunda fase (coleta), é feita a coleta dos atributos que serão analisados. Na fase de pré-processamento busca-se aprimorar os dados coletados. Na quarta fase (transformação) pode ser necessário transformar a forma como os dados estão representados. A fase de MD envolve a decisão de quais algoritmos serão aplicados aos dados. Caso o objetivo seja criar um modelo preditivo, decidir qual algoritmo é ótimo para o problema não é uma tarefa trivial. Por fim na última fase (avaliação e interpretação) são estabelecidas métricas para determinar se o classificador atingiu os objetivos, como por exemplo, taxa de erro, tempo de CPU e complexidade do modelo (BATISTA, 2003).

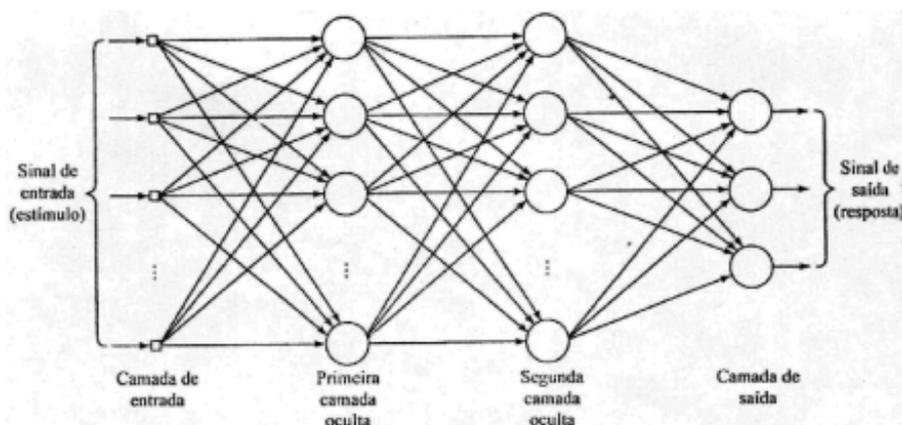
Figura 11 – Fases do processo de KDD



Fonte: Adaptado de Batista (2003)

A crescente utilização das Redes Neurais Artificiais (RNA) incentivou o desenvolvimento de diversas aplicações científicas e questões práticas, principalmente no campo dos negócios. As RNA 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 gerados através de operações matemáticas, geralmente não lineares (ALI et al., 2016). O neurônio tem a capacidade de reagir com base em padrões previamente aprendidos. O aprendizado em rede neural ocorre por meio da apresentação de um conjunto de padrões à rede. Estes padrões são utilizados por um algoritmo de treinamento para, iterativamente, ajustar os pesos das conexões (SOUTO et al., 2003). A figura 12 apresenta um exemplo de RNA multicamadas.

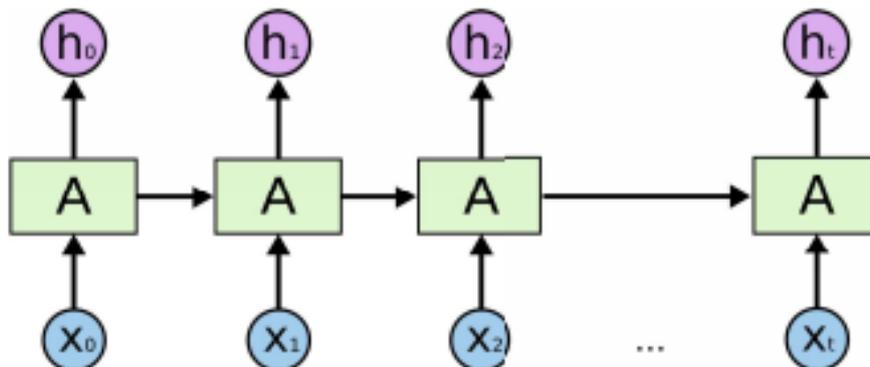
Figura 12 – Rede multicamadas



Fonte: Adaptado de Vassalli (2018)

Dentre os tipos de RNA, existe as RNN. Esse tipo de arquitetura foi projetada para reconhecer padrões em sequências de dados, como texto, genomas, caligrafia, palavra falada ou dados de séries numéricas que emanam de sensores, bolsas de valores e agências governamentais. Esses algoritmos consideram tempo e sequência, eles têm uma dimensão temporal, podendo ser utilizadas em aplicações de previsão e aplicação de séries temporais (ACADEMY, 2021). A figura 13 apresente a arquitetura de uma RNN, onde X_0 , X_1 , X_2 e X_t representam as entradas e h_0 , h_1 , h_2 e h_t as saídas. A camada A por ser recorrente, além de ter como entrada X_t também recebe h_{t-1} . Através da imagem é possível constatar que as RNN apresentam um problema de não conseguir aprender com recorrências muito longas, por esse fato foi criada a RNN do tipo LSTM (MOURA et al., 2018).

Figura 13 – Arquitetura de uma Rede Neural Recorrente



Fonte: Adaptado de Moura et al. (2018)

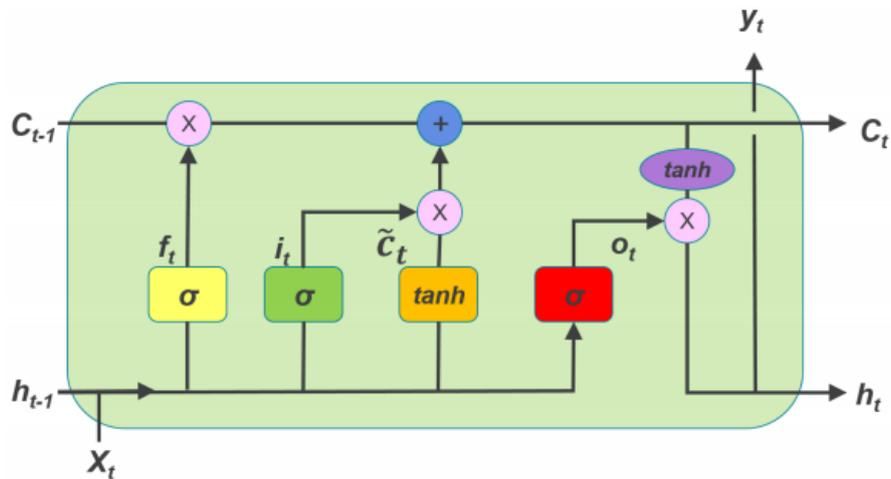
As redes neurais LSTM possuem o objetivo de aprender padrões complexos em estruturas com dependência temporal e vários estágios de processamento. Podem ser consideradas uma variação das redes neurais recorrentes, capazes de aprender dependências a longo prazo (SANTOS, 2019). A criação do modelo de RNA LSTM se estabeleceu para resolver problemas de previsão de séries temporais, visto que as redes neurais até então só eram capazes de estabelecer correspondência entre suas entradas e saídas através de uma perspectiva estática (WANG et al., 2019). Segundo Sagheer e Kotb (2019), os modelos de redes neurais se mostram como uma grande alternativa no processo de previsão do comportamento do sistema no futuro, comparado com métodos estatísticos lineares, apresentando uma maior precisão.

A figura 14 apresenta a arquitetura de uma rede neural do tipo LSTM, onde X_t é o vetor de entrada, H_{t-1} é a saída da célula anterior, C_{t-1} é a memória da célula anterior, H_t consiste na memória de célula anterior e C_t é a memória da célula atual. f_t , i_t , \tilde{C}_t e o_t são redes neurais simples que decidem quais informações serão adicionadas no estado da célula (ISMAIL; WOOD; BRAVO, 2018).

2.3.5 Sensoriamento Remoto

O termo sensoriamento remoto foi criado em 1960 e é uma tecnologia capaz de se obter dados da superfície terrestre, através da captação de energia emitida ou refletida pela superfície por meio de sensores acoplados em equipamentos em altitude, esses dados coletados são posteriormente processados em terra (SLATER, 1980). É atribuído o termo sensoriamento devido ao fato de obter-se dados de maneira com que o sensor está

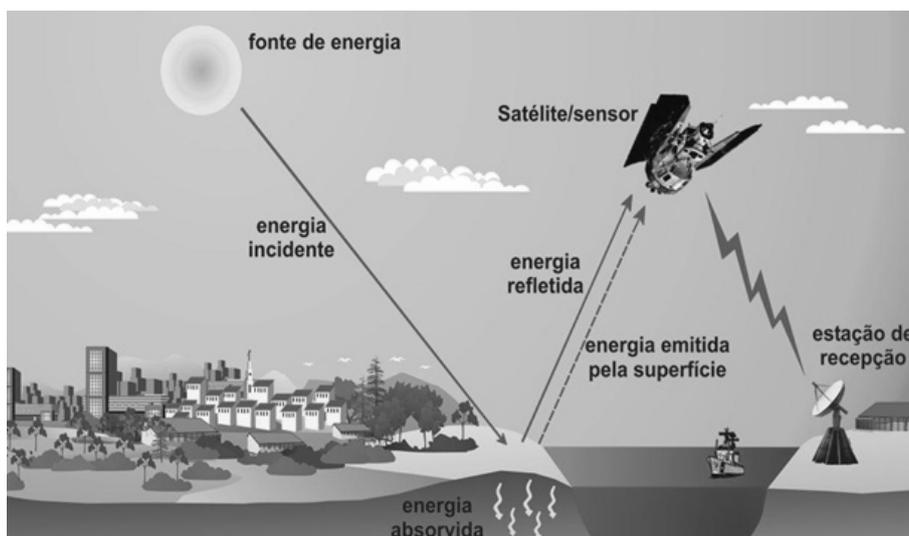
Figura 14 – Arquitetura LSTM



Fonte: Adaptado de Ismail, Wood e Bravo (2018)

acoplado em plataformas terrestres, aéreas ou orbitais. O termo remoto é devido ao fato de o objeto de interesse não estar em contato físico com o sensor, como ilustra a figura 15. Segundo Florenzano (2007), as etapas de processamento, análise e interpretação do dados também fazem parte do sensoriamento remoto.

Figura 15 – Componentes básicos da fase de aquisição de dados em sensoriamento remoto



Fonte: Adaptado de Freitas, Basso e Ferri (2015)

A obtenção de imagens por câmeras é feita a partir de energia eletromagnética que se propaga em forma de ondas com a velocidade da luz, em torno de 300.000Km/s. Estas ondas são medidas em frequência, cuja unidade é Hertz (Hz), a unidade de com-

primento de onda é metro (m). A frequência (F) de onda é o número de vezes em que a onda se repete por unidade de tempo e o comprimento de onda (λ) é medido a partir dos picos de ondas sucessivas (MENESES et al., 2012). Quanto maior for a velocidade de propagação (C) da onda maior será a sua frequência já que estas são diretamente proporcionais e quanto menor for a velocidade maior será o comprimento de onda (MENESES; ALMEIDA, 2012). Belward e Skøien (2015) define que por meio do uso de imagens de sensores remotos é possível realizar mapeamentos, avaliações e monitoramento da cobertura terrestre.

As missões de observação da Terra do programa Copernicus, chamadas de Sentinel-1, Sentinel-2, Sentinel-3, Sentinel-4, Sentinel-5, Sentinel-5P e Sentinel-6, tem por objetivo garantir a continuidade da aquisição de dados, visto que estas substituíram as missões mais antigas de observação. Cada uma das missões possui uma constelação de dois satélites e foca em aspectos diferentes na obtenção de dados. O Sentinel-1 tem por objetivo o monitoramento terrestre e oceânico, Sentinel-2 monitoramento terrestre, Sentinel-3 observação marinha e estudos sobre a topografia da superfície do mar, Sentinel-4 e -5 monitoramento da qualidade do ar, Sentinel-5P é o precursor do Sentinel-5 e foi lançada para preencher a lacuna entre a retirada do satélite Envisat e da missão Aura da NASA. A missão Sentinel-6 complementa a Sentinel-3, visto que analisa a altura global da superfície do mar. O presente trabalho utiliza dados obtidos através da missão Sentinel-2.

Através do instrumento acoplado nos satélites Sentinel-2, chamado MSI, é possível a captura de diferentes tipos de bandas, dependendo da resolução da imagem de interesse. Para a banda utilizada no trabalho (10m) as bandas disponibilizadas são a b2 (azul), b3 (verde), b4 (vermelho) e b8 (infravermelho próximo). Através da composição das bandas são feitos então os produtos disponibilizados. Para a geração do NDVI é feita uma combinação entre as banda b4 e b8. Já para a geração de imagem na escala RGB, é feita a combinação das bandas b3, b3 e b4,

2.4 Trabalhos correlatos

Para encontrar trabalhos correlatos foi realizada uma revisão sistemática da literatura, seguindo os passos definidos por Neiva (2016). O primeiro passo foi a definição de palavras-chave, através da análise das questões de pesquisa, esse processo foi incremental ao longo das leituras realizadas em trabalhos encontrados. Algumas palavras definidas

são: manejo de pastagens, pecuária de precisão, fmis, modelo de estimativa de massa de forragem, ajuste na taxa de lotação.

Foram definidas bases de busca, tais como: IEEE Xplore, Scopus, Scielo, ScienceDirect. Depois de encontradas as bases, foi criada uma *string* de busca e realizados testes através da análise dos artigos encontrados a partir da *string*, isso permite o refinamento da mesma. Com a *string* definida, ela foi adaptada para a sintaxe de cada base de busca definida. Feita essa adaptação, foi realizado o armazenamento de cada uma delas, junto com a quantidade de artigos encontrados e a data de busca.

Fez-se também o armazenamento do conjunto de resultados obtidos nas bases em um arquivo como uma lista de referências e utilizado na ferramenta de gerenciamento de referências Jabref. No próximo passo foram definidos critérios de inclusão e exclusão de artigos, estes critérios foram definidos através de idioma, área.

A seleção de artigos em um primeiro momento foi feita com a leitura do título e resumo. Após feita a exclusão de artigos que não se enquadram na pesquisa, foi feita a etapa de leitura da introdução e conclusão dos artigos que restaram. No último passo foi realizada a leitura dos artigos que passaram pela etapa anterior, além disso, foram definidos critérios de qualidade para estes artigos.

Tabela 2 – Trabalhos correlatos

Trabalho	Região	Pastagem	Modelo
Cruz et al. (2011)	Sudeste	Capim-marandu	Modelo baseado em regressão linear simples e múltipla.
Costa et al. (2011)	Sul	Azévem	Utiliza modelos de Loomis e Williams, Sinclair e Doorimbo e Kassam.
Fonseca, Formaggio e Ponzoni (2007)	Sul	Nativa	Modelo agrometeorológico-espectral utilizando imagens do sensor ETM+/Landsat 7
Schulte (2019)	Sul	Nativa	Modelo baseado em aprendizagem de máquina, especificamente Redes Neurais Artificiais Long Short-Term Memory.
Tonato et al. (2012)	Sudeste e Centro-Oeste	Brachiaria e Cynodon Panicum	Modelos utilizando sistema Invernada e modelo CROPGRO.

A partir da revisão sistemática foram encontrados os trabalhos apresentados na tabela 2. O trabalho de Cruz et al. (2011) propõe a predição da taxa média de acúmulo de matéria seca utilizando dados agrometeorológicos, através de modelos empíricos de regressão linear simples e múltipla.

O trabalho proposto por Costa et al. (2011) realiza cálculos do rendimento potencial comparando três modelos matemáticos: de Loomis e Williams, que utiliza apenas a radiação solar incidente como parâmetro de entrada, implicando na predição do maior rendimento total de forragem, de Sinclair, que utiliza parâmetros relacionados com a estrutura da pastagem, como Índice de Área Foliar (IAF) e coeficiente de extinção luminosa e de Doorenbos e Kassam, que leva em conta aspectos fisiológicos relativos à cultura, como, IAF e perdas por respiração e também ao ambiente, como por exemplo radiação incidente, temperatura, fotoperíodo e nebulosidade.

O trabalho de Fonseca, Formaggio e Ponzoni (2007) realiza a estimativa de disponibilidade de forragem pela parametrização do submodelo espectral do modelo JONG. Foram coletadas amostras em campo na mesma data em que o satélite ETM+/Landsat 7 passava pela área experimental. Antes de realizar a parametrização foram realizadas operações nas imagens com o intuito de reduzir os efeitos atmosféricos sobre os valores de reflectância e uniformizar as imagens. Foram utilizadas as variáveis mostradas a seguir para parametrizar o submodelo espectral: i) reflectância das bandas individuais do sensor ETM+/Landsat (B3, B4, B5, B7); ii) os Índices de Vegetação por Diferença Normalizada (NDVI) e Ajustado para o Solo (SAVI); iii) as frações da reflectância obtidas pelo modelo linear de mistura espectral: vegetação (Veg), solo (Sl) e sombra (Sb); iv) as variáveis oriundas da transformação Tasseled Cap: Brightness (B), Greenness (G) e Wetness (W).

Schulte (2019) propôs um modelo de predição de disponibilidade de forragem baseado em aprendizado de máquina, especificamente RNN LSTM. Esse modelo leva em conta dados históricos sobre corte de forragem através de um método direto e também dados meteorológicos coletados pela estação automática presente na Embrapa - Pecuária Sul.

O trabalho de Tonato et al. (2012) utiliza o sistema Invernada para criar um modelo de crescimento levando em consideração dados meteorológicos. Também utilizou-se CROPGRO que é um modelo mecanístico que prediz a produção e composição de culturas utilizando informações de planta, clima, manejo e solo, e também há simulações de balanço hídrico e de nitrogênio, matéria orgânica e dinâmica de resíduos do solo e danos por praga/doenças.

2.5 Considerações finais

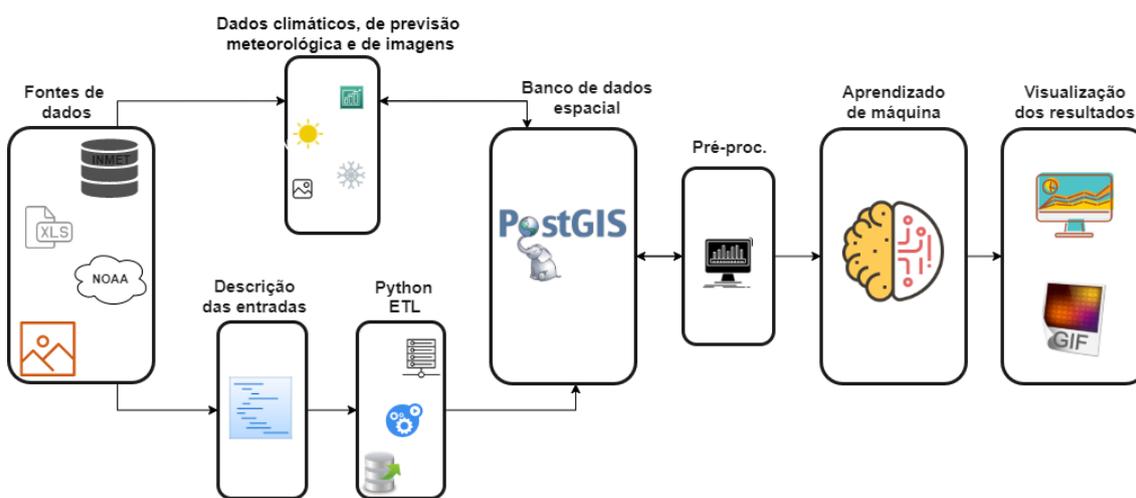
Apresentadas as tecnologias necessárias para o sistema TouceiraTech, pode-se notar que este faz uso de conceitos presentes em Smart Farming, Big Data, IoT e Cloud Computing, visto que possibilita a utilização de grandes volumes de dados gerados por uma fazenda. O conceito de Cloud Computing está na generalização do uso do TouceiraTech, ou seja, diversos pecuaristas, em áreas diferentes, utilizando o sistema para que os auxiliem na tomada de decisão em suas propriedades.

O TouceiraTech foi planejado como um projeto de FMIS, dado que realiza coleta, utilizando conceitos de MD, processamento, através de conceitos de aprendizado de máquina, ETL, descrição de entradas, etc., o armazenamento se dá através da utilização de banco de dados espaço-temporais e a disseminação dos dados através da disponibilização visual dos resultados de predição. O diferencial do TouceiraTech comparado aos trabalhos encontrados na literatura, se dá pelo fato de ser um sistema que utiliza como fonte de dados do pasto um método direto. Embora, no futuro, poderá fazer uso de dados oriundos de métodos indiretos. Realiza coletas, processamento, armazenamento e disseminação dos dados de forma automática. O capítulo 3 apresenta o sistema baseado em FMIS para a predição da disponibilidade de forragem, assim como, a arquitetura proposta, a descrição detalhada e implementação dos módulos presentes no sistema.

3 SOLUÇÃO PROPOSTA

Este capítulo detalha o projeto de um sistema baseado em FMIS, criado com base nos conceitos definidos por Sørensen et al. (2010), o qual usa princípios de pecuária de precisão para a predição do crescimento da vegetação em pastagens por meio da aplicação de técnicas de redes neurais recorrentes.

Figura 16 – Arquitetura proposta para o TouceiraTech



Fonte: Adaptado de Schulte (2019)

A figura 16 apresenta uma visão geral do modelo da arquitetura do protótipo de FMIS proposto, denominado TouceiraTech, sendo que os retângulos representam cada módulo do sistema. Através da arquitetura proposta, o TouceiraTech é capaz de realizar a coleta automatizada de dados históricos meteorológicos, de previsão meteorológica e de imagens de satélite. Permite agregar diferentes formatos de fontes de dados sobre pastagem e, com isso, realizar estudos de áreas de interesse em diferentes escalas (área total ou potreiro), considerando a diversidade natural que ocorre dentro das áreas de manejo. O banco de dados espacial torna possível realizar o processamento de imagens e utilizá-las como fonte de dados sobre a pastagem. A partir do pré-processamento dos dados, ou seja, operações e cálculos sobre os dados armazenados no sistema sobre áreas de interesse, é possível alimentar um modelo de predição especialmente ajustado para prever, com acurácia suficiente, a evolução de indicadores de disponibilidade de forragem. Com base nos dados de predição, são criados mapas de vegetação representando o quanto a pastagem irá variar até determinada data e também uma GIF animada ilustrando a variação temporal do NDVI em um determinado período.

A seguir, os módulos são apresentados em ordem: coleta (seções 3.1, 3.2, 3.3 e 3.4, tratando respectivamente das Fontes de dados, Atualização periódica, Metadados e ETL), armazenamento (seção 3.5, descrevendo o banco de dados espacial), processamento (seções 3.6 e 3.7, explicando o pré-processamento e o modelo de predição) e visualização (seção 3.8, detalhando as interfaces com gráficos e mapas de vegetação estáticos e animados).

3.1 Fontes de dados

No primeiro módulo são feitas as coletas primárias de dados necessários para os módulos seguintes. Atualmente são feitas as coletas automáticas nas bases de dados do INMET, NOAA e Copernicus Open Access Hub que disponibiliza imagens da missão Sentinel-2. Os arquivos XLS, referentes aos dados sobre a pastagem, são disponibilizados pela Embrapa Pecuária Sul. A coleta de imagens servirá, após feitas as operações e aplicado o NDVI, como uma alternativa de aquisição de dados sobre a pastagem. Uma das vantagens já observadas é o aumento na quantidade de amostras em comparação com o método de corte de pastagem realizado, visto que, o corte é realizado uma vez em cada mês e o satélite passa na mesma área três vezes ao mês, ou seja, é possível a aquisição de três imagens, em um intervalo de 10 dias, da mesma área de estudo.

Para a coleta automática dos dados da estação meteorológica presente na Embrapa Pecuária Sul foi utilizado um *script* em Python, disponibilizado pelo INMET. As informações são obtidas e gravadas em um arquivo no formato JavaScript Object Notation (JSON). Primeiramente esse arquivo é convertido para o formato Comma-separated Values (CSV), após isso são realizadas as operações sobre os dados, e, por fim os valores são armazenados no banco de dados. A figura 17 mostra o arquivo no formato CSV.

Para a coleta dos dados de previsão meteorológica foi utilizado um *script* chamado *rdams-client*, disponibilizado pela Research Data Archive (RDA), na qual é necessário possuir uma conta gratuita no *website*. As previsões feitas pelo National Center for Atmospheric Research (NCAR) utilizam o modelo americano chamado Global Forecast System (GFS) e é executado quatro vezes por dia (00 UTC, 06 UTC, 12 UTC e 18 UTC).

Através de um arquivo de configuração é possível atribuir quais parâmetros, ou seja, quais variáveis estão sendo solicitadas para *download*, a altura da superfície na qual essas variáveis são observadas, o período que se a previsão dos dados, de qual *dataset* se pretende buscar os dados e também informações de coordenadas do ponto de observação.

Figura 17 – Arquivo no formato .csv gerado pelo *script* de coleta

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	CD_ESTACCHUVA	DC_NOME	DT_MEDICA	HR_MEDIC	PRE_INS	PRE_MAX	PRE_MIN	PTO_INS	PTO_MAX	PTO_MIN	RAD_GLO	TEM_INS	TEM_MAX	TEM_MIN	UF	UMD_INS	UMD_MA	UMD_MIN	VEN_DIR	VEN_R	
2	A827	0 BAGÉ	2021-01-10	0	984.2	984.2	983.2	19.7	20	19.6	-3.539	24.1	24.8	24.1	RS	77	77	73	82	82	1
3	A827	0 BAGÉ	2021-01-11	100	985.1	985.2	984.2	19.8	20	19.6	-3.222	23.3	24.1	23.3	RS	81	81	77	132	132	
4	A827	0 BAGÉ	2021-01-12	200	985.2	985.2	985	20	20.1	19.8	-2.991	22.9	23.6	22.9	RS	84	84	81	140	140	
5	A827	0 BAGÉ	2021-01-13	300	985.4	985.5	985.1	20	20.2	20	-3.263	22.3	22.9	22.3	RS	87	87	84	126	126	
6	A827	0 BAGÉ	2021-01-14	400	984.6	985.4	984.6	20.3	20.4	19.9	-3.539	21.9	22.4	21.7	RS	91	91	87	115	115	
7	A827	0 BAGÉ	2021-01-15	500	984.3	984.6	984.3	20	20.5	20	-3.433	22.2	22.3	21.8	RS	88	91	88	122	122	
8	A827	0 BAGÉ	2021-01-16	600	984.1	984.3	984.1	19.9	20.2	19.8	-2.999	21.5	22.2	21.4	RS	91	91	88	119	119	
9	A827	0 BAGÉ	2021-01-17	700	984.4	984.4	984	20	20.2	19.9	-3.32	21.4	21.6	21.4	RS	92	92	91	136	136	
10	A827	0 BAGÉ	2021-01-18	800	984.7	984.7	984.4	19.6	20.4	19.6	-1.685	20.8	21.7	20.8	RS	93	93	92	131	131	
11	A827	0 BAGÉ	2021-01-19	900	985.2	985.2	984.7	21.5	21.5	19.5	176.825	22.9	22.9	20.6	RS	92	94	92	86	86	
12	A827	0 BAGÉ	2021-01-20	1000	985.7	985.8	985.2	21.2	21.5	20.8	966.521	25.3	25.4	22.9	RS	78	92	78	74	74	
13	A827	0 BAGÉ	2021-01-21	1100	985.7	985.8	985.7	20.7	21.7	20.6	1.752.455	26.8	26.9	25	RS	70	79	69	40	40	
14	A827	0 BAGÉ	2021-01-22	1200	985.7	985.8	985.6	20.2	21.2	19.8	2.344.904	27.9	28.7	26.7	RS	63	70	61	6	6	
15	A827	0 BAGÉ	2021-01-23	1300	985.4	985.7	985.4	20.1	21.4	19.5	3098.29	29.9	30.3	27.9	RS	56	64	55	346	346	
16	A827	0 BAGÉ	2021-01-24	1400	985.1	985.5	985.1	20.6	20.8	19.3	3.528.031	31.4	31.7	29.3	RS	53	58	50	281	281	
17	A827	0 BAGÉ	2021-01-25	1500	984.5	985.1	984.5	20.4	20.6	19.1	2.967.611	31.9	32.2	30.4	RS	51	54	47	354	354	
18	A827	0 BAGÉ	2021-01-26	1600	983.9	984.5	983.9	18.3	20.9	18.3	2883.67	31.8	33.5	30.9	RS	45	51	43	3	3	
19	A827	0 BAGÉ	2021-01-27	1700	983.2	983.9	983.2	17.9	19.8	17.6	3184.85	34.2	34.7	31.7	RS	38	47	37	36	36	
20	A827	0 BAGÉ	2021-01-28	1800	982.6	983.2	982.6	17.4	18.8	16	2.651.631	34.5	34.8	32.8	RS	36	41	33	48	48	
21	A827	0 BAGÉ	2021-01-29	1900	981.9	982.6	981.9	15.6	18.1	15.5	2679.91	34.3	35.2	33.7	RS	33	38	32	104	104	
22	A827	0 BAGÉ	2021-01-30	2000	981.1	981.9	981.1	15	16.1	14.9	1.717.931	34.1	34.9	33.6	RS	32	35	31	21	21	
23	A827	0 BAGÉ	2021-01-31	2100	981	981.1	981	16.8	16.9	14.9	698.397	33.2	34.1	32.2	RS	38	38	32	36	36	

Fonte: Autor (2021)

A execução do programa se dá por linha de comando, apenas indicando o nome do arquivo de configuração. A saída gerada pode ser tanto um arquivo em formato netCDF quanto em formato CSV. Vale ressaltar que o período no qual são disponibilizados dados de previsão é, no máximo 15 dias, contados a partir da data de requisição.

A aquisição das imagens de satélite, realizada de maneira automatizada, utiliza um *script* que através de um pacote Python, chamado Sentinel-2Sat, faz requisições ao repositório Copernicus Open Access Hub, informando login e senha. Além das informações de usuário, também são fornecidas informações da área, através de um arquivo GEOJSON. Nesse arquivo constam coordenadas que informam os limites de um polígono, com isso é possível obter imagens em que o polígono estará inserido. Também é possível parametrizar e otimizar a busca por imagens, ou seja, delimitar o intervalo de tempo em que se deseja obter as imagens, a porcentagem de nuvens na imagem, qual produto se deseja obter, de qual satélite se deseja obter a imagem, etc. As imagens são obtidas no formato JPG2000 e posteriormente convertidas para o formato GeoTiff para uma maior facilidade nas operações necessárias. Esse formato é utilizado pois guarda nos seus metadados correspondência geográfica do local de aquisição da imagem. Vale ressaltar que o cadastro no repositório é feito de maneira gratuita. A figura 18 mostra a imagem obtida através do *script* de coleta de imagens de satélite.

Como as imagens obtidas abrangem uma área muito maior do que a de estudo do trabalho, foi necessário realizar o recorte das mesmas. Para realizar efetivamente o recorte foi necessário a criação de um arquivo *shapefile*, similar ao arquivo *geojson*, que armazena as coordenadas que limitam um polígono da área de interesse, esse arquivo foi gerado no software livre com código-fonte aberto chamado QGIS. Este é um software de SIG que

Figura 18 – Imagem original obtida através do Sentinel-2 da área de estudo principal



Fonte: Adaptado de Sentinel-2 (2021)

permite visualização, edição e análise de dados georreferenciados. Esse arquivo gerado deve estar no mesmo sistema de coordenadas da imagem que se deseja recortar. Após isso um *script* utiliza o pacote rioxarray e realiza o recorte através da função `clip()`. O resultado gerado é salvo no mesmo formato GeoTiff da imagem original e também armazena a correspondência geográfica. Considerando que a área de estudo contém divisões, foram criados arquivos *shapefile* para cada uma dessas divisões, chamados de poteiros, com o objetivo de gerar o NDVI para cada uma delas e poder comparar os valores de corte de pastagem com os valores de NDVI médio de cada uma dessas subáreas. O cálculo do NDVI é realizado através da equação 1, sendo que as bandas das imagens, no Sentinel-2, o NIR e o RED correspondem às bandas b8 e b4, respectivamente. O *script* para o cálculo do NDVI utiliza o pacote em Python chamado earthpy, comumente utilizado em trabalhos encontrados na literatura. Esse pacote simplifica algumas operações, como o NDVI, em dados raster e também vetoriais. A figura 19 apresenta o recorte feito na imagem original, contendo a área total de estudo.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

Onde *NIR* representa a banda do infravermelho próximo e *RED* a banda do vermelho visível. E o *NDVI* pode assumir valores no intervalo [-1, 1].

Figura 19 – Recorte da área total de estudo principal



Fonte: Autor (2021)

3.2 Atualização periódica

O módulo responsável pela atualização periódica de dados climáticos, de previsão meteorológica e de imagens se caracteriza por *scripts* agendados para execução diária. A cada dia, busca atualizações nas bases remotas.

Para cada uma das áreas de interesse cadastradas no sistema, coleta os dados meteorológicos do dia anterior e a nova previsão meteorológica, substituindo dados de previsão armazenados no dia anterior. Como a previsão atual contém os dados diários para os próximos quinze dias, na prática o módulo realiza duas etapas: i) substitui no banco de dados as variáveis meteorológicas do último dia, anteriormente preenchidas com dados de previsão, por dados obtidos da estação meteorológica mais próxima; e ii) atualiza os dados de previsão dos próximos 14 dias e insere a previsão do 15o dia. Por exemplo, quando o *script* for executado às 00h do dia X, coletará e armazenará os dados da estação referentes ao dia X-1 e coletará e armazenará os dados, substituindo as previsões anteriores, de X a X+14. De forma análoga, um *scripts* diário é agendado para consultar a base aberta da Copernicus, usando o método descrito anteriormente, para extrair novas imagens para a área de interesse.

3.3 Descrição das entradas

O módulo de descrição de entrada foi planejado com a finalidade de auxiliar o processo de ETL na interpretação (sintaxe e semântica) dos dados de pastagem de fontes variadas, reorganizando-os em um formato intermediário, evitando a necessidade de criação de diferentes *scripts* para tratar de diferentes formatos de fontes primárias de dados (sobretudo planilhas de diferentes formatos) no módulo seguinte. Portanto, o seu objetivo é facilitar a utilização do sistema para estudo de áreas diversas, manejadas por outros especialistas de diferentes instituições e também por pecuaristas, com diferentes maneiras de registro dos dados históricos sobre suas áreas. A ausência deste módulo exigiu que os *scripts* originais de ETL (SCHULTE, 2019), concebidos para o modelo de dados de pastagem adotado na Embrapa Pecuária Sul para registro dos dados experimentais de áreas de pastagem nativa sujeitas a um método de controle de plantas invasoras, fossem replicados e adaptados para serem capazes de lidar com outro modelo. Modelo este que correspondente aos dados de outras áreas experimentais de pastagem nativa da mesma unidade da Embrapa manejadas com diferentes técnicas de melhoramento.

3.4 Processo de ETL

O módulo de ETL é capaz de ler os arquivos XLS e automaticamente inserir no banco de dados. Esse módulo é responsável também por realizar operações nos dados, ou seja, armazenar apenas informações necessárias para o modelo de predição. Para realizar a inserção dos dados sobre a pastagem que, originalmente são inseridos em planilhas no formato XLSX, foi necessário padronizá-las para que o *script* fosse capaz de ler corretamente o arquivo e capturar os valores corretos de cada coluna da tabela clima existentes no banco de dados. Também foi necessário realizar ajustes nas células em que eram realizadas operações, como por exemplo, no cálculo de médias de massa seca de *annoni*. A figura 20 apresenta a planilha não padronizada para a correta leitura do programa (a) e também a planilha após realizado o processo de padronização (b).

A figura 21 apresenta o cabeçalho da tabela dos dados disponibilizados pelo INMET, ou seja, no momento em que é realizada a coleta automática dos dados, estes são obtidos no mesmo formato em que é apresentado na tabela.

A figura 22 apresenta a tabela de dados meteorológicos existente no banco de dados, com isso, para o modelo proposto não se faz necessárias informações de determi-

- ST: O cálculo de somatermica foi feito através da Equação 2;

$$ST = \frac{t_{\max} + t_{\min}}{2} - t_{\text{base}} \quad (2)$$

Onde t_{base} foi estabelecida como sendo 10°C.

- def e exc: O cálculo de deficit e excedente hídrico foi realizado através da equação proposta por Thornthwaite e Mather (THORNTHWAITE, 1957).

Para os dados de previsão meteorológica foi criada uma nova tabela no banco de dados, utilizando o mesmo padrão da tabela clima, com o intuito de que sejam passados para a RNA os mesmos dados históricos de clima. As manipulações necessárias foram as mesmas apresentadas anteriormente.

A inserção dos dados de imagens foi feita através de um *script* que faz uso do executável, chamado raster2pgsql, disponibilizado pela extensão PostGIS¹ do SGBD PostgreSQL. Essa extensão agrega suporte a objetos geográficos, permitindo que consultas de localização sejam executadas no SQL, além de adicionar funções, operadores e aprimoramentos de índice que se aplicam aos dados geográficos. Para a utilização da função raster2pgsql alguns parâmetros precisaram ser definidos para a correta inserção, tais como: atribuição de um sistema de projeção ao *raster*, cortar *raster* em pedaços a serem inseridos por linha da tabela, indicar o nome da tabela, indicar usuário, entre outros.

3.5 Banco de dados espacial

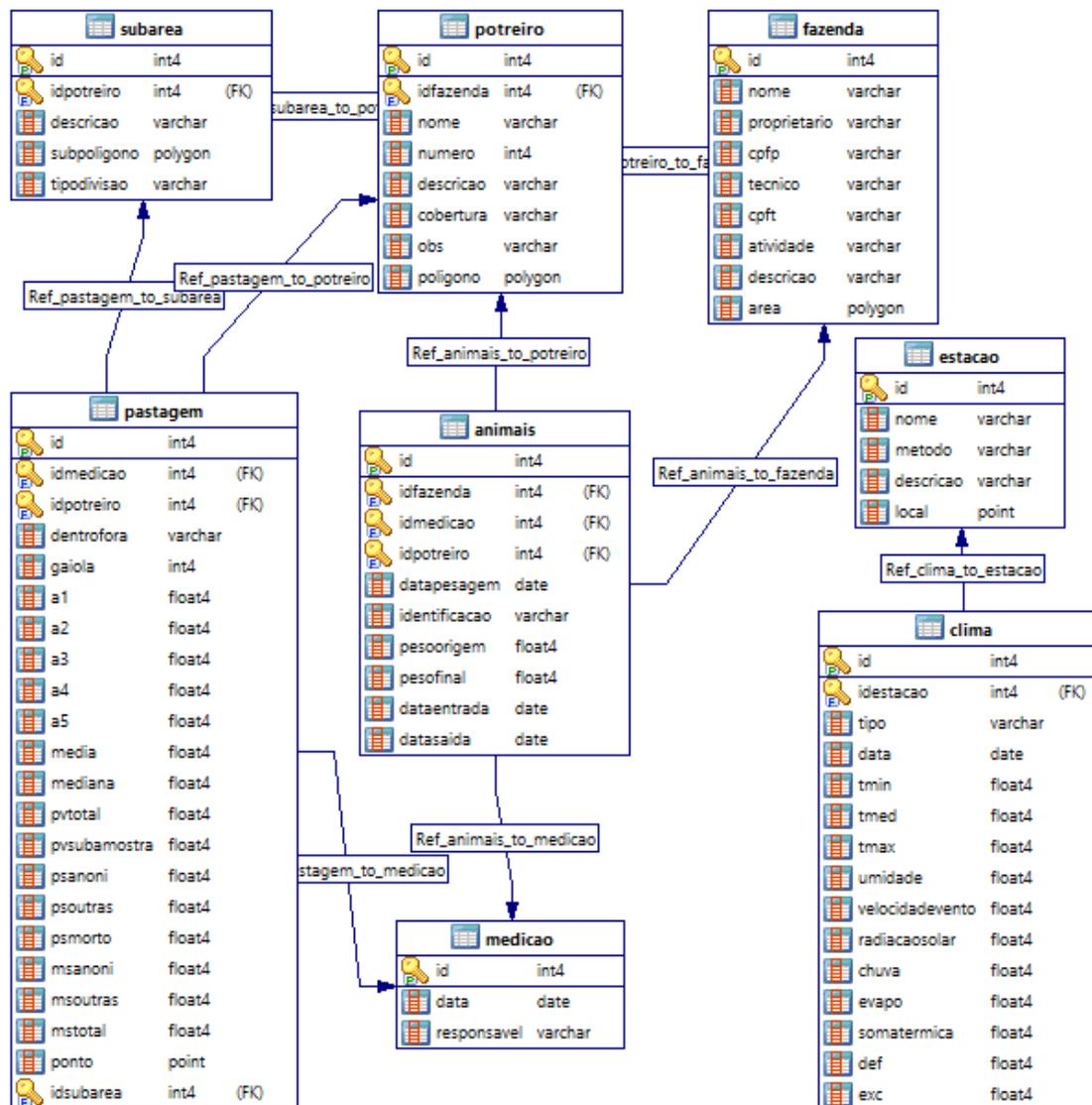
A etapa de armazenamento de dados, no presente trabalho definido como um módulo que utiliza o SGBD PostgreSQL que é um sistema de banco de dados objeto-relacional. Optou-se por utilizar esse SGBD pelo fato de possuir extensões capazes de armazenar objetos com correspondência geográfica.

Foi necessário realizar algumas modificações no modelo original. A figura 23 apresenta o diagrama de entidade-relacionamento do modelo criado por Schulte (2019).

Para que fosse possível criar um método indireto de coleta de dados referentes a pastagem, foi necessário o armazenamento de imagens no banco de dados espacial, com isso, foi criada uma tabela onde foram salvas as imagens. A figura 24 apresenta o diagrama de entidade-relacionamento do modelo atual do banco de dados.

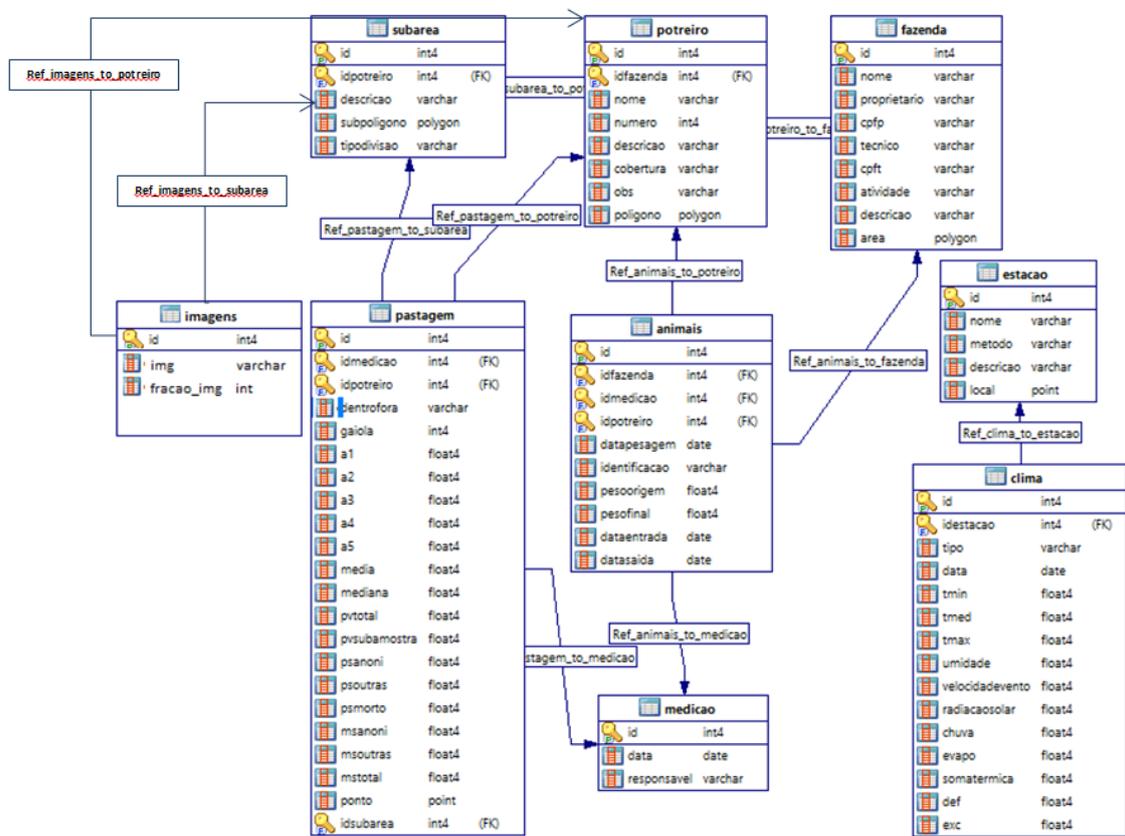
¹<https://postgis.net/>

Figura 23 – Diagrama Entidade-Relacionamento do modelo original



Fonte: Adaptado de Schulte (2019)

Figura 24 – Diagrama Entidade-Relacionamento do modelo atual



Fonte: Autor (2021)

3.6 Pré-processamento

O módulo de pré-processamento dos dados corresponde a um *script* em Python, o qual faz a busca e operações nos dados de pastagem e meteorológicos no intervalo entre os cortes. Este inclui a realização de três tarefas, sendo elas:

- **sumarização dos dados meteorológicos:** É 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.
- **sumarização entre as amostras pontuais coletadas em cada gaiola:** A sumarização é feita através da média ponderada apresentada na equação 3.

$$MS_{total} = \frac{\sum_{i=1}^n MS_{gaiola_i} * subarea_i}{areatotal} \quad (3)$$

Onde MS_{total} é a massa de forragem total por hectare em $kg \cdot ha^{-1}$, MS_{gaiola_i} é a massa de forragem por hectare na subárea em que a gaiola i é representativa em $kg \cdot ha^{-1}$, $subarea_i$ é a área que a gaiola representa i em m^2 , $areatotal$ é a área total do potreiro em m^2 e n é a quantidade de gaiolas existentes no potreiro.

- **normalização dos dados:** Consiste na análise dos dados presentes no banco de dados com o intuito de remover redundâncias, como valores repetidos.

3.7 Aprendizado de máquina

O módulo de predição utiliza o modelo proposto por Schulte (2019), composto por uma RNA LSTM com duas unidades LSTM. São passadas 36 entradas de treinamento e 12 de teste no estudo do tratamento Infestado e 32 entradas de treinamento e 12 de teste para o tratamento MIRAPASTO. A primeira unidade LSTM possui trinta neurônios e a segunda possui 15 neurônios e ambas possuem a função de ativação tangente hiperbólica. Através do modelo, é possível realizar a predição para o presente, ou seja, utilizando como uma alternativa ao método de corte direto, e também é possível realizar uma estimativa futura com dados históricos de corte de pastagem e dados meteorológicos.

3.8 Visualização dos resultados

O módulo de visualização dos dados corresponde ao procedimento de disseminação dos dados presente em um projeto de FMIS. Os resultados do modelo de predição são apresentados através de gráficos de linha, mapas de vegetação e também GIF animada. No capítulo 5 são apresentadas todas as maneiras de visualização dos resultados.

O objetivo da apresentação através dos gráficos de linha é mostrar para o usuário os valores reais e preditos, com isso, poderá verificar a acurácia do método. Também através dos gráficos de linha é possível mostrar os valores reais e um valor predito para 15 dias, após o último valor real, ou seja, o resultado do modelo de predição com a utilização de dados de previsão meteorológica, com o intuito de poder verificar a curva de crescimento ou decrescimento do valor predito.

Os mapas de vegetação apresentam de forma visual e através de cores como estará a disponibilidade de forragem para a data futura, após 15 dias do último dado de corte. O mapa utiliza três cores, vermelho, amarelo e verde, sendo que a cor vermelha representa que naquela data não haverá uma boa disponibilidade de forragem, com isso, é necessário retirar animais, a cor amarela representa que a disponibilidade de forragem se manteve estável comparada ao mês anterior, com isso pode-se manter o número de animais e a cor verde representa que houve um aumento na disponibilidade de forragem, podendo assim, inserir animais na área.

Com a GIF animada é possível ver o comportamento da área na perspectiva do NDVI, ou seja, havendo as imagens de um intervalo de tempo em que se deseja estudar, é possível calcular o NDVI e assim, gerar a GIF animada.

3.9 Considerações finais

De forma geral, o protótipo inicial do TouceiraTech apresenta funcionalidades capazes de comprovar o seu potencial uso, realizando coletas automáticas, armazenando e processando dados para que seja possível prever a estimativa de massa de forragem. Contudo, alguns refinamentos ainda são necessários como, por exemplo, a generalização de entradas de dados sobre pastagens. No capítulo 4 é apresentada a metodologia experimental adotada no trabalho e no capítulo 5 os resultados obtidos com a integralização de todos os módulos do sistema e utilização das suas funcionalidades.

4 METODOLOGIA EXPERIMENTAL

Neste capítulo são apresentados os materiais e métodos adotados para a avaliação dos módulos do sistema TouceiraTech, com destaque para o delineamento dos experimentos planejados e realizados tendo em perspectiva as questões de pesquisa e os objetivos do trabalho.

4.1 Material e Métodos

Para o seu funcionamento, o sistema demanda a definição de áreas de referência, com séries de dados históricos de manejo da pastagem nativa associadas a dados meteorológicos. Neste trabalho, foram usados dados sobre duas áreas submetidas a pastejo contínuo com carga variável, com oferta de forragem visando máximo ganho (12%), correspondentes a dois experimentos de longo prazo desenvolvidos na EMBRAPA Pecuária Sul.

O primeiro conjunto de dados, assim como no trabalho de Schulte (2019), tem origem nos experimentos desenvolvidos em quatro poteiros (dupla repetição), com o intuito de avaliar a eficácia do Método Integrado de Recuperação de Pastagens - MIRAPASTO, para controle do capim-annoni em pastagens nativas (NEVES et al., 2020). Os tratamentos foram: pastagem nativa submetida à recuperação da degradação pela infestação do capim-annoni (MIRAPASTO) e pastagem natural infestada pelo capim-annoni (Infestado). Este conjunto principal (doravante identificado por “Conjunto A”) se caracteriza como uma série de dados temporais correspondente ao período entre 2014 e 2020, no qual seis amostras sistemáticas (contendo a MF e múltiplas medidas aleatórias de altura média em uma área de 0,5 x 0,5 metros quadrados, sendo três de dentro e três de fora das gaiolas de exclusão) são obtidas em cada poteiro, com intervalos de aproximadamente 30 dias.

O segundo corresponde a um outro experimento, realizado em nove poteiros (tripla repetição), com o objetivo de verificar o efeito de três níveis de intensificação, envolvendo a utilização de fertilização e da sobressemeadura de espécies hibernais, em pastagens nativa (COSTA, 2019). Os tratamentos foram: Campo Nativo (CN), Campo Nativo Fertilizado (CNF) e Campo Nativo Melhorado (CNM). Este conjunto complementar (doravante identificado por “Conjunto B”) forma uma série temporal que contempla quatro períodos de um ano, coletados entre os meses de agosto a julho (2007/08, 2012/13,

2013/14 e 2014/15), onde oito amostras aleatórias (contendo a MS total e múltiplas medidas aleatórias de altura média em uma área de 0,5 x 0,5 metros quadrados, sendo quatro dentro e quatro de fora das gaiolas de exclusão) são coletadas em cada potreiro, também com intervalos aproximados de 30 dias.

Após o cadastro das áreas de interesse, na qual referem-se as unidades de manejo de pastejo, denominadas potreiros, os quais podem estar subdivididos em áreas menores com diferentes perfis de acúmulo de forragem. Ressaltando que o sistema permite tanto a definição de áreas de estudo quando de pastagem (potreiro), foram inseridos no banco do sistema os dados sobre a pastagem nestas áreas, os dados meteorológicos obtidos do Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) do INMET, correspondentes à estação meteorológica mais próxima e também os dados de previsão meteorológica.

Para realizar as inserções necessárias sobre as pastagens foi utilizado um *script* Python, adaptado de Schulte (2019), o qual é capaz de ler as planilhas que foram disponibilizadas pelos pesquisadores da Embrapa - Naylor Perez e Cristina Genro - e automaticamente armazenar os registros no banco de dados.

Os demais dados foram inseridos no banco por meio das novas funcionalidades propostas e implementadas no protótipo do sistema como contribuições deste trabalho:

- *script* em Python capaz de realizar a coleta automática de dados, da base do INMET, correspondente a uma determinada estação meteorológica de observação de superfície automática, sendo utilizada a estação localizada na própria unidade da EMBRAPA, a menos de 3 km de distância das áreas de interesse. Após feita a requisição dos dados no período desejado, estes foram disponibilizados em formato JSON e convertidos para CSV;
- *script* em Python, para acessar a base de dados do NCAR e coletar as previsões correspondentes aos dados que são disponibilizados pelo CPTEC/INMET;
- *script* em Python para automatizar o cálculo de evapotranspiração, o qual utilizada uma biblioteca capaz de calcular os valores de evapotranspiração através do método de Penman-Monteith;
- *script* para a busca automática de imagens da área de manejo na base do Sentinel-2 e inserção destas no banco de dados espacial utilizado pelo modelo por meio da ferramenta raster2pgsql do PostGIS. Imagens matriciais necessárias ao método indireto de coleta de amostras sobre a pastagem, como uma alternativa frente ao método de corte de forragem;

- *script* responsável pela estimativa da massa de forragem por meio da correlação entre o NDVI calculado e a série histórica de dados obtidos pelo método direto;
- alteração no *script* de pré-processamento, automatizando a execução da abordagem proposta por Schulte (2019) para suavização de *outliers* identificados pelo método de Tukey, tornando-os iguais ao limite máximo identificado como normal pelo método.

Em relação ao ambiente computacional, cabe destacar que, no contexto atual, a execução do modelo de predição não apresenta demanda significativa de recursos, tendo em vista a quantidade reduzida de amostras disponíveis tanto para treinamento quanto para testes. Os testes foram realizados em um computador do tipo notebook, com microprocessador multicore Intel Core i7-6500U com frequência de 2,5 GHz e dois núcleos com tecnologia Simultaneous multithreading (SMT), placa de vídeo Geforce 940M 4 GB DDR3, 8 GB de memória DDR3 e sistema operacional Windows 10.

4.2 Delineamento dos Experimentos

A partir do conjunto de dados e dos módulos desenvolvidos para o sistema TouceiraTech, foram delineados experimentos para:

- Demonstrar a funcionalidade de estimativa automatizada de evapotranspiração.
- Aferir os módulos de coletas de dados em bases remotas: (i) séries históricas de dados meteorológicos disponibilizadas pelo CPTEC/INPE; (ii) previsões meteorológicas disponibilizadas pelo NCAR; (iii) imagens do Sentinel-2;
- Testar o modelo de predição de MF (dentro da gaiola) de Schulte (2019), denominado “Modelo Original” para os novos dados do Conjunto A com diferentes tipos de treinamento: (i) estratificado por potreiro - um modelo específico por potreiro, mesma abordagem de Schulte (2019); (ii) estratificado por tratamento - nova abordagem, resultando em um modelo específico por tipo de tratamento; e (iii) não estratificado - nova abordagem, resultando um modelo genérico aplicável a qualquer tratamento;
- Avaliar ajustes nas variáveis de entrada do modelo para melhor capturar: (i) o efeito estacional, substituindo a data da amostra por mês e ano, (ii) o acúmulo de forragem observado em relação ao mês anterior, introduzindo a taxa de acúmulo (TA) e (iii) o efeito dos animais, pela inclusão da carga animal (CA);

- Verificar o “Modelo Ajustado” de predição de TA para os dados do Conjunto B com os diferentes tipos de tratamento;
- Avaliar a capacidade de generalização do “Modelo Ajustado” de predição de TA combinando os dados dos Conjuntos A e B, com treinamento estratificado por tratamentos similares: MIRAPASTO + CNM/CNF e INFESTADO + CN;
- Exercitar o “Modelo Ajustado” de predição de TA com treinamento estratificado por tratamentos similares com dados de previsão meteorológica com horizonte de quinze dias;
- Aferir o módulo NDVI, responsável pela estimativa da MF por meio da correlação entre MF real e NDVI, visando a substituição do método direto de amostragem;
- Testar o “Modelo Novo” de predição de TA que usa a correlação entre MF e NDVI em substituição ao método direto;
- Demonstrar o módulo de visualização da evolução da MF em uma área de manejo, delimitada por um polígono espacializado, por meio de mapas temáticos animados, como ferramenta de apoio a decisão de manejo.

O primeiro passo para que seja possível realizar os testes corresponde ao cálculo da evapotranspiração a partir de um pacote em Python chamado PyETo (<https://github.com/woodcrafty/PyETo>). Com este pacote foi possível fazer o cálculo utilizando três métodos: o de Penman-Monteith, o de Hargreaves ou o de Thornthwaite. Para o trabalho foi utilizado o método de Penman-Monteith.

Nos testes do modelo de predição, são considerados dois cenários: (i) predição para o presente, isto é, para a predição da MF de uma série temporal tendo como base uma série histórica de dados meteorológicos reais como entrada no modelo; e (ii) predição para o futuro, ou seja, para a predição com base em dados de previsões meteorológicas. Para ampliar a avaliação da acurácia dos modelos de predição de MF, além da apresentação dos resultados do Root Mean Squared Error (RMSE), ou Raiz do Erro Quadrático Médio (REQM) em Português, de acordo com a equação 4, calculou-se também o coeficiente de determinação (R^2), calculado através de regressão linear.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (pred_j - real_j)^2} \quad (4)$$

4.3 Considerações finais

A partir das definições metodológicas para o experimentos, conforme exposto, no capítulo 5 são apresentados os resultados observados.

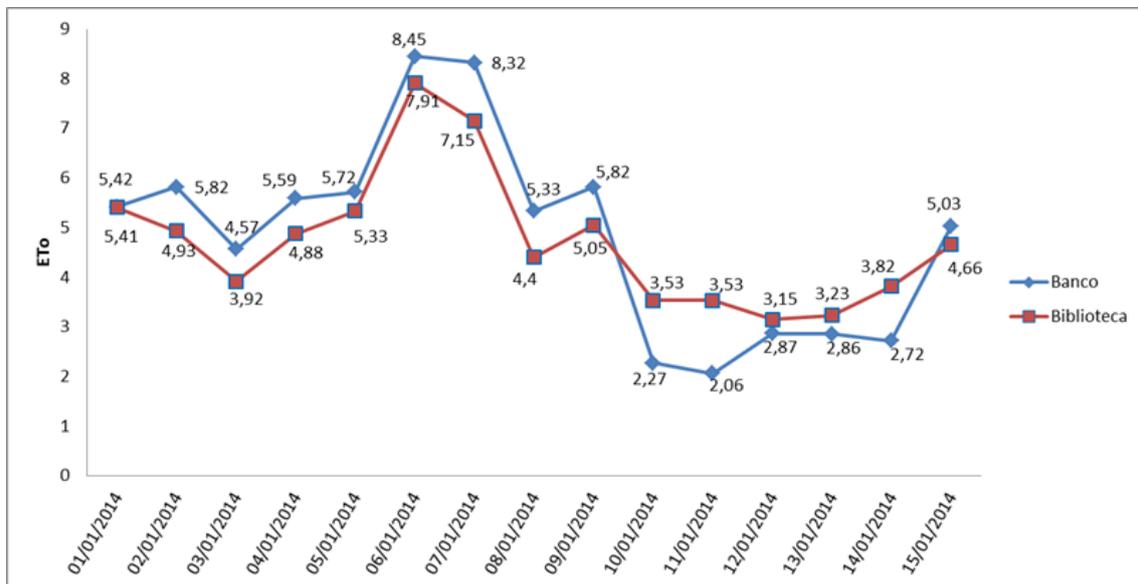
5 RESULTADOS

Este capítulo apresenta os resultados obtidos a partir dos experimentos planejados com o intuito de demonstrar o funcionamento dos módulos do sistema TouceiraTech.

5.1 Estimativa Automática de Evapotranspiração

Para validar os resultados foram calculados os valores de evapotranspiração em datas iguais às que já estavam presentes no banco de dados, obtidos a partir da análise de um agrometeorologista. A figura 25 mostra a diferença entre os valores presentes no banco de dados (SCHULTE, 2019) e os valores calculados pela biblioteca no período de 15 dias. Vale ressaltar que a biblioteca necessita de seis parâmetros existentes no banco, são eles: temperatura média, temperatura máxima, temperatura mínima, umidade, velocidade do vento e radiação.

Figura 25 – Comparação dos valores de Evapotranspiração



Fonte: Autor (2021)

5.2 Coleta de Dados Meteorológicos e de Previsão

Para realizar a coleta de dados meteorológicos é utilizado um *script* em python que realiza a requisição dos dados ao repositório de dados do INMET. O arquivo é dis-

ponibilizado em formato .json, porém, no mesmo *script* já é realizada a transformação do arquivo para o formato .csv. Para verificar que a execução é feita de maneira correta, foi feita a validação comparando os dados obtidos através do website do INMET, mostrados na figura 26, com os dados obtidos através do *script*, mostrados na figura 27. Com isso, é possível constatar que a coleta automatizada busca as mesmas informações que seriam obtidas utilizando a interface do website.

Figura 26 – Tabela com dados meteorológicos disponibilizada no website do INMET

Data	Hora UTC	Temperatura (°C)			Umidade (%)			Pto. Orvalho (°C)			Pressão (hPa)			Vento (m/s)			Radiação KJ/m²	Chuva mm
		Inst.	Máx.	Mín.	Inst.	Máx.	Mín.	Inst.	Máx.	Mín.	Inst.	Máx.	Mín.	Vel.	Dir.	Raj.		
10/01/2021	0000	24.1	24.8	24.1	77.0	77.0	73.0	19.7	20.0	19.6	984.2	984.2	983.2	3.9	82.0	10.3	0.0	
10/01/2021	0100	23.3	24.1	23.3	81.0	81.0	77.0	19.8	20.0	19.6	985.1	985.2	984.2	3.2	132.0	8.4	0.0	
10/01/2021	0200	22.9	23.6	22.9	84.0	84.0	81.0	20.0	20.1	19.8	985.2	985.2	985.0	3.9	140.0	5.8	0.0	
10/01/2021	0300	22.3	22.9	22.3	87.0	87.0	84.0	20.0	20.2	20.0	985.4	985.5	985.1	2.1	126.0	5.5	0.0	
10/01/2021	0400	21.9	22.4	21.7	91.0	91.0	87.0	20.3	20.4	19.9	984.6	985.4	984.6	2.3	115.0	3.5	0.0	
10/01/2021	0500	22.2	22.3	21.8	88.0	91.0	88.0	20.0	20.5	20.0	984.3	984.6	984.3	2.6	122.0	4.2	0.0	
10/01/2021	0600	21.5	22.2	21.4	91.0	91.0	88.0	19.9	20.2	19.8	984.1	984.3	984.1	1.7	119.0	3.7	0.0	
10/01/2021	0700	21.4	21.6	21.4	92.0	92.0	91.0	20.0	20.2	19.9	984.4	984.4	984.0	1.4	136.0	2.1	0.0	
10/01/2021	0800	20.8	21.7	20.8	93.0	93.0	92.0	19.6	20.4	19.6	984.7	984.7	984.4	1.9	131.0	2.4	0.0	
10/01/2021	0900	22.9	22.9	20.6	92.0	94.0	92.0	21.5	21.5	19.5	985.2	985.2	984.7	1.1	86.0	2.4	176.82	
10/01/2021	1000	25.3	25.4	22.9	78.0	92.0	78.0	21.2	21.5	20.8	985.7	985.8	985.2	3.1	74.0	6.0	966.52	
10/01/2021	1100	26.8	26.9	25.0	70.0	79.0	69.0	20.7	21.7	20.6	985.7	985.8	985.7	4.1	40.0	7.2	1752.45	
10/01/2021	1200	27.9	28.7	26.7	63.0	70.0	61.0	20.2	21.2	19.8	985.7	985.8	985.6	3.5	6.0	6.5	2344.90	

Fonte: Autor (2021)

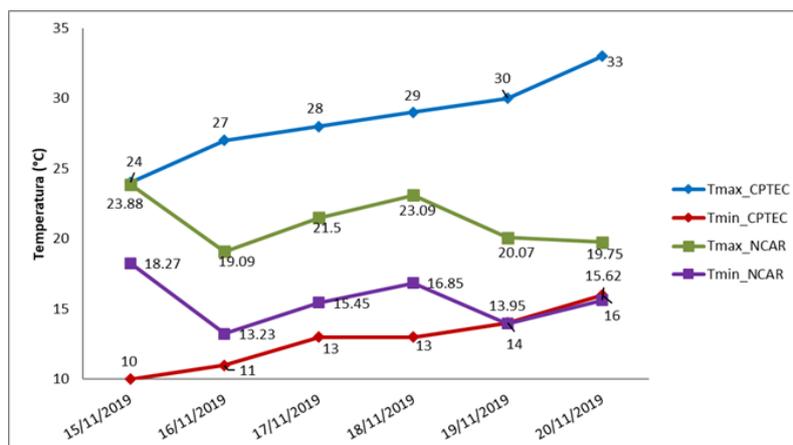
Figura 27 – Tabela com dados meteorológicos obtida através da execução do *script* de coleta

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	CD_ESTAC	CHUVA	DC_NOME	DT_MEDICA	HR_MEDIX	PRE_INS	PRE_MAX	PRE_MIN	PTO_INS	PTO_MAX	PTO_MIN	RAD_GLO	TEM_INS	TEM_MAX	TEM_MIN	UF	UMD_INS	UMD_MAJ	UMD_MIN	VEN_DIR	VEN_RAJ	VEN_VEL	VL_LATITUD	VL_LONGITUDE
2	A827	0	BAGÁf9c	2021-01-10	0	984.2	984.2	983.2	19.7	20	19.6	-3.539	24.1	24.8	24.1	RS	77	77	73	82	10.3	3.9	-31.347.801	-54.013.292
3	A827	0	BAGÁf9c	2021-01-10	100	985.1	985.2	984.2	19.8	20	19.6	-3.222	23.3	24.1	23.3	RS	81	81	77	132	8.4	3.2	-31.347.801	-54.013.292
4	A827	0	BAGÁf9c	2021-01-10	200	985.2	985.2	985	20	20.1	19.8	-2.991	22.9	23.6	22.9	RS	84	84	81	140	5.8	3.9	-31.347.801	-54.013.292
5	A827	0	BAGÁf9c	2021-01-10	300	985.4	985.5	985.1	20	20.2	20	-3.263	22.3	22.9	22.3	RS	87	87	84	126	5.5	2.1	-31.347.801	-54.013.292
6	A827	0	BAGÁf9c	2021-01-10	400	984.6	985.4	984.6	20.3	20.4	19.9	-3.539	21.9	22.4	21.7	RS	91	91	87	115	3.5	2.3	-31.347.801	-54.013.292
7	A827	0	BAGÁf9c	2021-01-10	500	984.3	984.6	984.3	20	20.5	20	-3.433	22.2	22.3	21.8	RS	88	91	88	122	4.2	2.6	-31.347.801	-54.013.292
8	A827	0	BAGÁf9c	2021-01-10	600	984.1	984.3	984.1	19.9	20.2	19.8	-2.999	21.5	22.2	21.4	RS	91	91	88	119	3.7	1.7	-31.347.801	-54.013.292
9	A827	0	BAGÁf9c	2021-01-10	700	984.4	984.4	984	20	20.2	19.9	-3.32	21.4	21.6	21.4	RS	92	92	91	136	2.1	1.4	-31.347.801	-54.013.292
10	A827	0	BAGÁf9c	2021-01-10	800	984.7	984.7	984.4	19.6	20.4	19.6	-1.685	20.8	21.7	20.8	RS	93	93	92	131	2.4	1.9	-31.347.801	-54.013.292
11	A827	0	BAGÁf9c	2021-01-10	900	985.2	985.2	984.7	21.5	21.5	19.5	176.825	22.9	22.9	20.6	RS	92	94	92	86	2.4	1.1	-31.347.801	-54.013.292
12	A827	0	BAGÁf9c	2021-01-10	1000	985.7	985.8	985.2	21.2	21.5	20.8	966.521	25.3	25.4	22.9	RS	78	92	78	74	6	3.1	-31.347.801	-54.013.292
13	A827	0	BAGÁf9c	2021-01-10	1100	985.7	985.8	985.7	20.7	21.7	20.6	1.752.455	26.8	26.9	25	RS	70	79	69	40	7.2	4.1	-31.347.801	-54.013.292
14	A827	0	BAGÁf9c	2021-01-10	1200	985.7	985.8	985.6	20.2	21.2	19.8	2.344.904	27.9	28.7	26.7	RS	63	70	61	6	6.5	3.5	-31.347.801	-54.013.292

Fonte: Autor (2021)

Foram realizados experimentos preliminares com o objetivo de validar as coletas automatizada de fontes de dados de previsão meteorológica, sendo uma delas local e outra de fora do país. Foram coletados os valores de temperatura máxima e mínima disponibilizadas pelo CPTEC, as quais foram comparadas com os valores disponibilizados pelo NCAR. Os valores coletados são apresentados na figura 28, onde é possível notar que os valores de temperatura máxima e mínima disponibilizados pelo CPTEC estão em um intervalo maior ao comparar com os dados disponibilizados pelo NCAR. A tabela 3 mostra o desvio padrão para cada variável disponível pelos centros.

Figura 28 – Valores de temperatura disponibilizadas pelo NCAR e CPTEC



Fonte: Autor (2021)

Tabela 3 – Desvios padrão dos dados de previsão meteorológica

	Temperatura máxima	Temperatura mínima
CPTEC	3,01	2,13
NCAR (NOAA)	1,93	1,84

Fonte: Autor (2021)

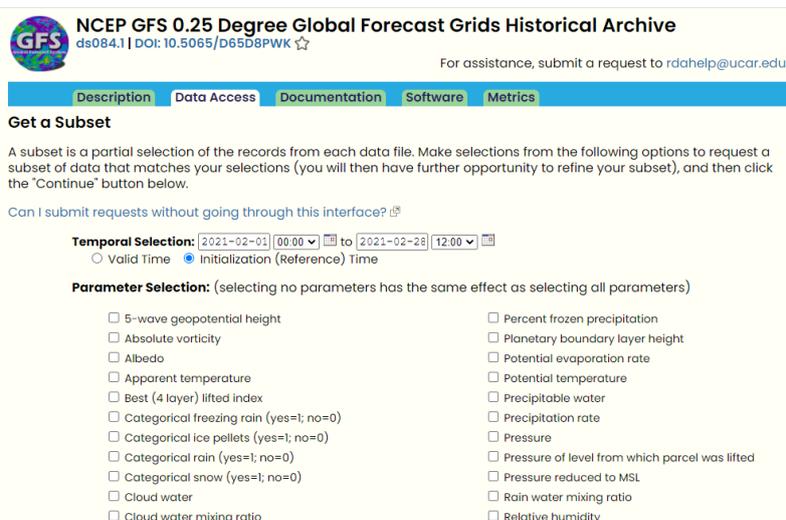
A de coleta de dados de previsão meteorológica foi automatizada e para validar os resultados obtidos, foi comparada a busca realizada através da interface gráfica do website do repositório rda com os resultados obtidas através do *script* de coleta. A figura 29 apresenta a interface gráfica do website, onde é possível escolher as variáveis meteorológicas a serem buscadas, além do período da busca.

Após realizada essa etapa, é necessário indicar alguns parâmetros de busca, como por exemplo, o formato do arquivo em que estarão os dados de previsão meteorológica e a área de interesse. A interface onde é necessário realizar a escolha desses parâmetros é mostrada na figura 30.

A figura 31(a) mostra a tabela obtida com a utilização da interface gráfica do website do rda e a figura 31(b) ilustra a tabela obtida com a utilização do *script* de coleta de dados de previsão meteorológica. Vale ressaltar que a busca foi realizada utilizando os mesmos parâmetros e que o valor de temperatura apresentado nas tabelas está na unidade Kelvin.

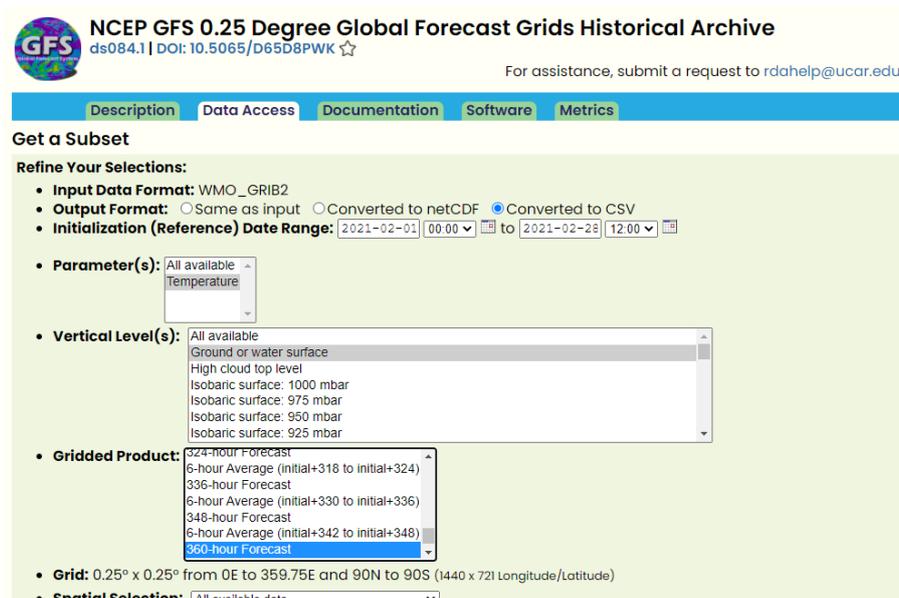
Também foi realizada a validação do *script* de coleta de imagens de satélite. A figura 32 mostra o resultado da busca realizada no repositório de imagens Open Access Hub do Copernicus. Com a utilização do *script* é possível obter os mesmos resultados,

Figura 29 – Interface gráfica do website do rda e variáveis meteorológicas selecionáveis



Fonte: Adaptado de RDA (2021)

Figura 30 – Interface gráfica do website do rda, onde é necessário indicar os parâmetros de busca



Fonte: Adaptado de RDA (2021)

Figura 31 – Dados de previsão meteorológica obtidos com uso do website do rda e com script de coleta para o dia 16 de fevereiro de 2021 na hora 00

(a) Website do rda

	A	B	C	D	E	F	G	H
1	#Paramet	Longitude	Latitude	ValidDate	ValidTime	Level	Value	ParameterValue
2	2t	-54	-31.5	2021-02-16	00:00:00	2	293,5017	
3	#name	lon	lat	date	time	lev	value	
4								

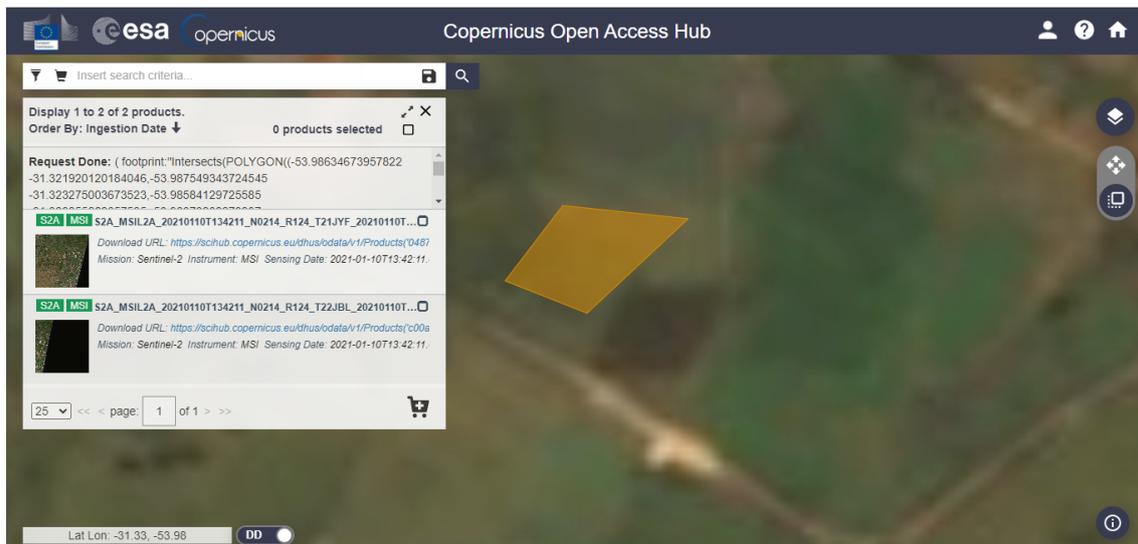
(b) script de coleta automática

	A	B	C	D	E	F	G	H
1	#Paramet	Longitude	Latitude	ValidDate	ValidTime	Level	Value	ParameterValue
2	2t	-54	-31.5	2021-02-16	00:00:00	2	293,5017	
3	#name	lon	lat	date	time	lev	value	
4								

Fonte: Autor (2021)

mostrados na figura 33. A linha vermelha presente na figura tem por objetivo mostrar que também foram encontrados dois produtos para a busca. Vale ressaltar que os parâmetros de filtragem de busca foram iguais para as duas consultas.

Figura 32 – Resultado da busca realizada no repositório Open Access Hub do Copernicus



Fonte: Adaptado de Copernicus Open Access Hub (2021)

Figura 33 – Resultado da busca realizada pelo *script* de coleta de imagens do satélite Sentinel-2

```
(base) C:\Users\Anderson Fischeoeder\Desktop\img_sentinel\download>python sentinel.py
OrderedDict([('f059abc4-a7c4-41cf-9432-05c66e0176f7', {'title': 'SZA_MSIL2A_20210107T133231_N0214_R081_T22JBL_20210107T155304', 'link': 'https://scihub.copernicus.eu/apihub/odata/v1/Products('f059abc4-a7c4-41cf-9432-05c66e0176f7')/$value', 'link_alternative': 'https://scihub.copernicus.eu/apihub/odata/v1/Products('f059abc4-a7c4-41cf-9432-05c66e0176f7')/$value', 'link_icon': 'https://scihub.copernicus.eu/apihub/odata/v1/Products('f059abc4-a7c4-41cf-9432-05c66e0176f7')/Products('Quicklook')/$value', 'summary': 'Date: 2021-01-07T13:32:31.024Z, Instrument: MSI, Satellite: Sentinel-2, Size: 1.06 GB', 'ondemand': 'false', 'beginposition': 'datetime.datetime(2021, 1, 7, 13, 32, 31, 240000)', 'endposition': 'datetime.datetime(2021, 1, 7, 18, 46, 34, 270000)', 'ingestiondate': 'datetime.datetime(2021, 1, 7, 18, 46, 34, 270000)', 'orbitnumber': 28964, 'relativeorbitnumber': 81, 'vegetationpercentage': 55.752432, 'notvegetatedpercentage': 7.704354, 'waterpercentage': 0.098198, 'unclassifiedpercentage': 2.320456, 'mediumprobcloudpercentage': 6.481757, 'highprobcloudpercentage': 20.174007, 'snowicepercentage': 0.05599799999999999, 'cloudcoverpercentage': 33.772887, 'levelididentifier': 'S2A_OPER_MSI_L1C_TL_VGS2_20210107T152041_A028964_T22JBL_N02_00', 'gmloffprint': '<gml:Polygon srsName='http://www.opengis.net/gml/srs/epsg.xml#4326' xmlns:gml='http://www.opengis.net/gml'>\n      <gml:coordinates>-30.694733440748557, -54.13199 -30.71725330253914, -52.98651 -31.707438890022775, -53.067324 -31.683953701756977, -54.104764 -30.694733440748557, -54.13199 />\n    </gml:LinearRing>\n    </gml:LinearRing>\n    </gml:outerBoundaryIs>\n  </gml:Polygon>', 'footprint': 'MULTIPOLYGON (((-53.067324 -31.707438890022775, -52.98651 -30.71725330253914, -54.12199 -30.694733440748557, -54.104764 -31.683953701756977, -53.067324 -31.707438890022775)))', 'format': 'SAFE', 'processingbaseline': '02_14', 'platformname': 'Sentinel-2', 'filename': 'S2A_MSIL2A_20210107T133231_N0214_R081_T22JBL_20210107T155304_SAFE', 'instrumentname': 'Multi-Spectral Instrument', 'instrumentshortname': 'MSI', 'size': '1.06 GB', 's2dataakeid': 'S2A_20210107T133231_028964_N02_14', 'producttype': 'S2MSI2A', 'platformidentifier': '2015-028A', 'orbitdirection': 'DESCENDING', 'platformserialidentifier': 'Sentinel-2A', 'processinglevel': 'Level-2A', 'identifier': 'S2A_MSIL2A_20210107T133231_N0214_R081_T22JBL_20210107T155304', 'uid': 'f059abc4-a7c4-41cf-9432-05c66e0176f7'}), ('2f6cbb1f-3e97-4cda-b638-0850e108ea2', {'title': 'SZA_MSIL2A_20210107T133231_N0214_R081_T21JYF_20210107T155304', 'link': 'https://scihub.copernicus.eu/apihub/odata/v1/Products('2f6cbb1f-3e97-4cda-b638-0850e108ea2')/$value', 'link_alternative': 'https://scihub.copernicus.eu/apihub/odata/v1/Products('2f6cbb1f-3e97-4cda-b638-0850e108ea2')/$value', 'link_icon': 'https://scihub.copernicus.eu/apihub/odata/v1/Products('2f6cbb1f-3e97-4cda-b638-0850e108ea2')/Products('Quicklook')/$value', 'summary': 'Date: 2021-01-07T13:32:31.024Z, Instrument: MSI, Satellite: Sentinel-2, Size: 818.93 MB', 'ondemand': 'false', 'beginposition': 'datetime.datetime(2021, 1, 7, 13, 32, 31, 240000)', 'endposition': 'datetime.datetime(2021, 1, 7, 13, 32, 31, 240000)', 'ingestiondate': 'datetime.datetime(2021, 1, 7, 18, 43, 43, 121000)', 'orbitnumber': 28964, 'relativeorbitnumber': 81, 'vegetationpercentage': 64.204895, 'notvegetatedpercentage': 13.558188, 'waterpercentage': 0.54963, 'unclassifiedpercentage': 1.263591, 'mediumprobcloudpercentage': 4.079884, 'highprobcloudpercentage': 9.168599, 'snowicepercentage': 0.0, 'cloudcoverpercentage': 19.941709, 'levelididentifier': 'S2A_OPER_MSI_L1C_TL_VGS2_20210107T152041_A028964_T21JYF_N02_00', 'gmloffprint': '<gml:Polygon srsName='http://www.opengis.net/gml/srs/epsg.xml#4326' xmlns:gml='http://www.opengis.net/gml'>\n      <gml:coordinates>-31.702100138305713, -54.715515 -31.584233946013272, -54.480726 -31.43720494369236, -54.637177 -31.209340842287585, -54.59378 -31.143460069780428, -54.550415 -30.99660274105446, -54.50714 -30.849796588116416, -54.463867 -30.705682187725003, -54.421875 -30.69224562053003, -53.76645 -31.681364358642234, -53.732605 -31.702100138305713, -54.715515 />\n    </gml:LinearRing>\n    </gml:LinearRing>\n    </gml:outerBoundaryIs>\n  </gml:Polygon>', 'footprint': 'MULTIPOLYGON (((-54.715515 -31.702100138305713, -53.732605 -31.681364358642234, -53.76645 -30.69224562053003, -54.421875 -30.705682187725003, -54.438667 -30.849796588116416, -54.50714 -30.99660274105446, -54.550415 -31.143460069780428, -54.59378 -31.209340842287585, -54.637177 -31.43720494369236, -54.680725 -31.584233946013272, -54.715515 -31.702100138305713)))', 'format': 'SAFE', 'processingbaseline': '02_14', 'platformname': 'Sentinel-2', 'filename': 'S2A_MSIL2A_20210107T133231_N0214_R081_T21JYF_20210107T155304_SAFE', 'instrumentname': 'Multi-Spectral Instrument', 'instrumentshortname': 'MSI', 'size': '818.93 MB', 's2dataakeid': 'S2A_20210107T133231_028964_N02_14', 'producttype': 'S2MSI2A', 'platformidentifier': '2015-028A', 'orbitdirection': 'DESCENDING', 'platformserialidentifier': 'Sentinel-2A', 'processinglevel': 'Level-2A', 'identifier': 'S2A_MSIL2A_20210107T133231_N0214_R081_T21JYF_20210107T155304', 'uid': '2f6cbb1f-3e97-4cda-b638-0850e108ea2'})])
```

Fonte: Autor (2021)

5.3 Módulo de Predição

Nesta seção são apresentados os resultados obtidos com o módulo de predição, os quais embasam diferentes conclusões a respeito do modelo original e de versões ajustadas do modelo, com diferentes conjuntos de dados históricos meteorológicos e de pastagem, bem como permitem demonstrar o seu comportamento com dados de previsão meteorológica.

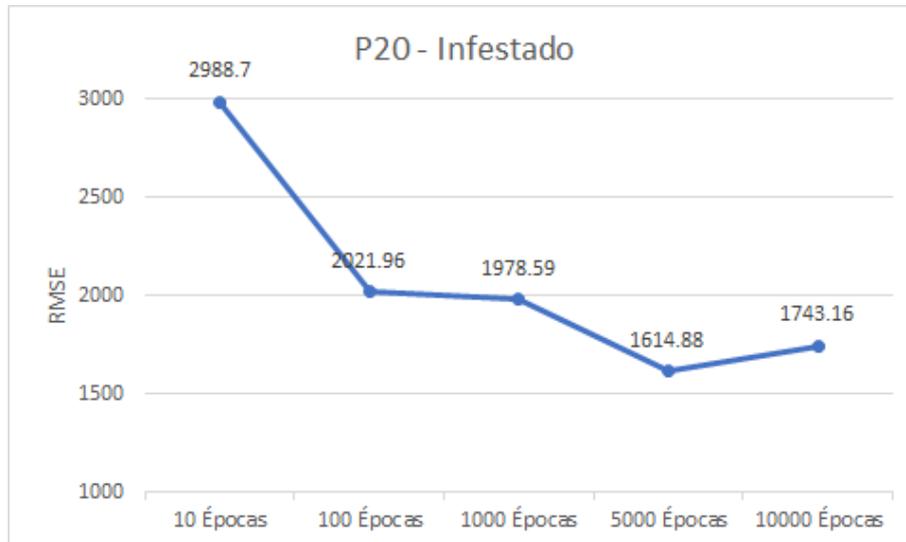
5.3.1 Resultados do Modelo Original com Conjunto A

Para a realização dos testes com o Conjunto A, referentes às subáreas do experimento com manejo de planta invasora, foram utilizadas 50 amostras válidas do potreiro 20 com tratamento Infestado e 46 amostras válidas do potreiro 20 com tratamento MIRAPASTO. Para o potreiro 21 foram utilizadas 43 amostras válidas com tratamento Infestado e 37 amostras válidas com tratamento MIRAPASTO. A variação do número de amostras ocorre pelo fato da exclusão de amostras que estavam com algum tipo de avaria, como por exemplo, a gaiola estar virada. O conjunto de amostras foi dividido em treinamento e teste, para o potreiro 20 Infestado foram utilizadas 38 para treinamento e 12 para teste, para o potreiro 20 MIRAPASTO ficaram 34 para treinamento e 12 para teste. O conjunto de treinamento e teste para o potreiro 21 Infestado ficou dividido em 31 e 12, respectivamente. Por fim para o potreiro 21 MIRAPASTO foram utilizadas 25 amostras para treinamento e 12 para teste. A divisão foi feita dessa maneira com o propósito de utilizar um intervalo de dados de 12 meses para teste, com isso, podendo observar o crescimento em todas as estações do ano. Esses conjuntos de amostras são passadas como entrada para a RNN. É importante ressaltar que foi realizada a suavização dos valores de massa de forragem, isto é, valores que estavam acima do terceiro quartil de todo conjunto de dados foram igualados ao valor do terceiro quartil.

Antes de iniciar os testes com o conjunto completo dos dados, foram realizados testes preliminares. A primeira série de experimentos visou o ajuste da quantidade de épocas, que representa o número de repetições em que os dados serão transmitidos pela rede, para o treinamento do modelo (figura 34), por meio da qual confirmou-se que 5000 épocas é uma quantidade adequada.

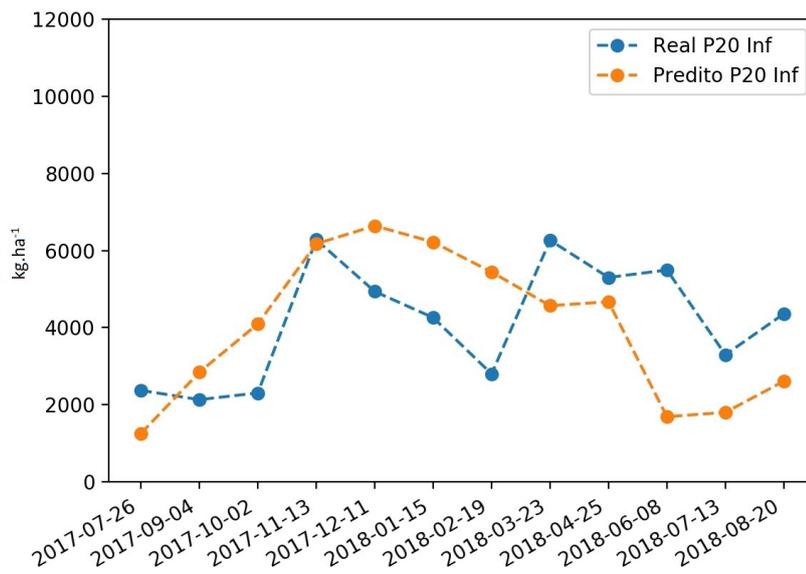
Na sequência foi realizado um experimento para validação com o mesmo conjunto de treinamento/teste usado por Schulte (2019) (27 amostras de treinamento e 12 de teste,

Figura 34 – Ajuste de épocas



Fonte: Autor (2021)

Figura 35 – Reprodução do experimento de Schulte (2019) com o Potreiro 20 Infestado



Fonte: Autor (2021)

correspondentes ao período entre janeiro de 2014 e agosto de 2018), utilizando o potreiro com mais amostras (20 Infestado), obtendo resultado com uma diferença insignificante no RMSE (figura 35).

Figura 36 – Sensibilidade ao tamanho do conjunto de treinamento



Fonte: Autor (2021)

Após, foram realizados experimentos para avaliar o efeito da relação existente entre a quantidade de amostras de treinamento e teste (figura 36), partindo da proporção usada por Schulte (2019) (69%/31%) até a proporção máxima (76%/24%, 38 de treinamento/12 de teste, estendendo o período até outubro de 2019). Para esta análise de sensibilidade do modelo, optou-se por manter fixo o conjunto de testes (agosto de 2018 a outubro de 2019), variando a quantidade de amostras de treinamento contíguas ao período do conjunto de testes, até chegar à última amostra, de janeiro de 2014.

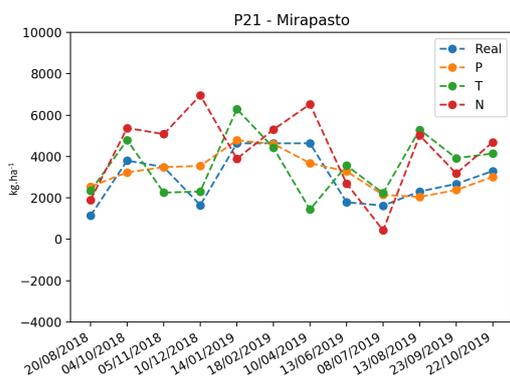
Feitos os testes preliminares, partiu-se então para os experimentos visando observar a acurácia do modelo de predição. A figura 37 ilustra graficamente a estimativa de massa de forragem instantânea ao longo do tempo usando as abordagens dos três tipos de treinamento. Foram feitas 10 repetições para cada potreiro, utilizando todos os tipos de treinamento. Para constatar o melhor e pior caso, foram verificados os RMSE de cada tipo de treinamento em cada uma das repetições, o menor valor de RMSE obtido foi considerado o melhor caso e o maior valor de RMSE obtido foi considerado o pior caso. Para o potreiro P21 MIRAPASTO (melhor caso) o RMSE médio foi de 975,12 kg.ha⁻¹ para o treinamento estratificado por potreiro, para o treinamento estratificado por tratamento foi de 1627,19 kg.ha⁻¹ e para o treinamento não estratificado foi de 1764,76 kg.ha⁻¹ e para o potreiro P20 Infestado (pior caso) o RMSE médio para o treinamento estratificado

por potreiro foi de 1882,62 kg.ha⁻¹, para o treinamento estratificado por tratamento foi de 2135,98 kg.ha⁻¹ e para o treinamento não estratificado foi de 1644,51 kg.ha⁻¹.

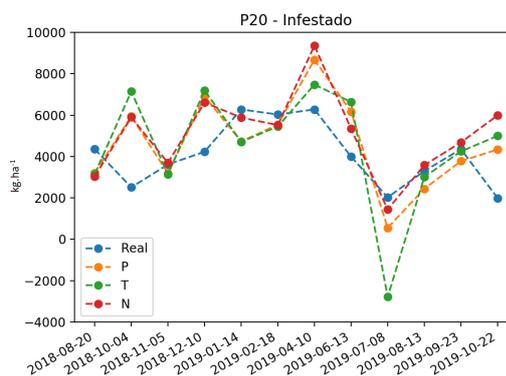
A figura 37(a) mostra os resultados obtidos, para o melhor caso, a partir dos três tipos de treinamento utilizados. Através do gráfico é possível verificar que a curva dos valores preditos que manteve um comportamento mais próximo da curva dos valores reais foi do treinamento estratificado por potreiro (P), obtendo o menor RMSE de 893,28 kg.ha⁻¹. A figura 37(b) apresenta o comportamentos das curvas dos valores reais e preditos, para o pior caso, utilizando os três tipos de treinamento. É possível verificar que a curva dos valores preditos que se manteve mais diferenciada da curva dos valores reais foi a do treinamento estratificado por tratamento (T), com o maior RMSE de 2508,52 kg.ha⁻¹.

Figura 37 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de massa de forragem, para cada subárea

(a) Melhor Caso (P)



(b) Pior Caso (T)



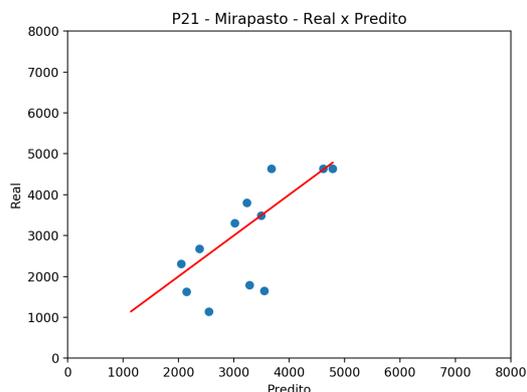
Fonte: Autor (2021)

A figura 38 apresenta os gráficos onde é possível analisar a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso, com seus respectivos treinamentos. A figura 38(a) apresenta a correlação para o potreiro P21 MIRAPASTO (melhor caso), através da regressão linear foi possível calcular um R^2 linear de 0,51 e um coeficiente angular de 1,05°. Para o potreiro P20 Infestado (pior caso) a correlação entre valores reais e preditos pelo treinamento estratificado por tratamento (T) é apresentado na figura 38(b), com um R^2 linear de 0,18 e um coeficiente angular de 0,22°.

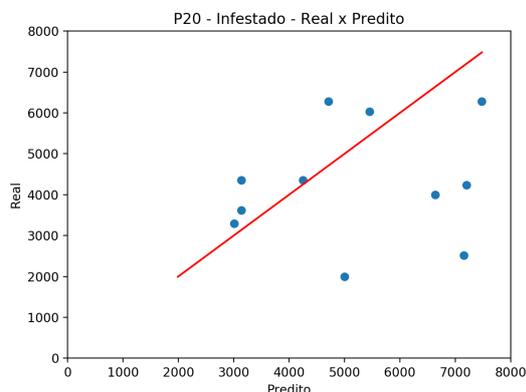
Através da figura 39 é possível verificar a distribuição dos valores reais, preditos e também da diferença entre o valor real e o predito para cada uma das datas, chamado de

Figura 38 – valores reais e preditos de massa de forragem, para melhor e pior caso

(a) Melhor Caso (P)



(b) Pior Caso (T)

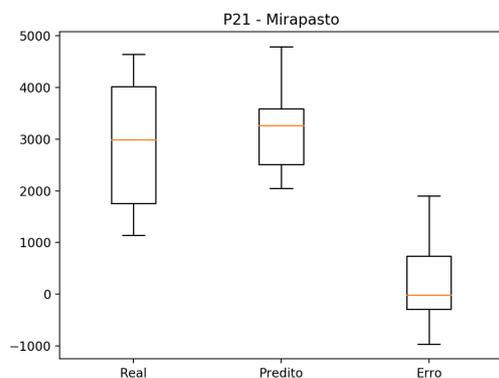


Fonte: Autor (2021)

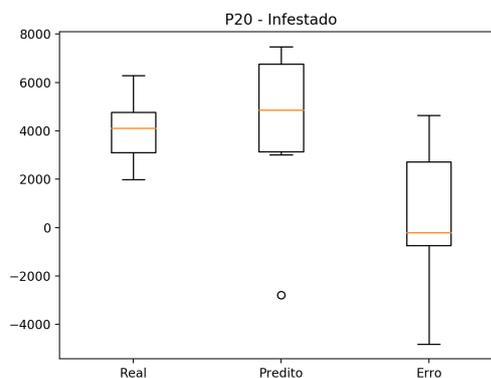
erro. A figura 39(a) apresenta a distribuição para o melhor caso a partir do treinamento estratificado por potreiro (P), podendo verificar que a dispersão dos valores reais é maior do que dos valores preditos, havendo semelhança entre seus valores máximos. A figura 39(b) mostra que, para o potreiro P20 Infestado (pior caso) a partir do treinamento estratificado por tratamento (T), a dispersão dos valores preditos é maior em relação a dispersão dos valores reais, ocorrendo um *outlier* no conjunto de valores preditos, que expressa a ocorrência de um valor discrepante se comparado aos outros. A partir do valor mínimo expresso pelo conjunto de dados de erro é possível constatar a ocorrência de uma predição bem menor do que se comparado ao valor real.

Figura 39 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de massa de forragem

(a) Melhor Caso (P)



(b) Pior Caso (T)



Fonte: Autor (2021)

5.4 Resultados do Modelo Original com propostas de ajustes

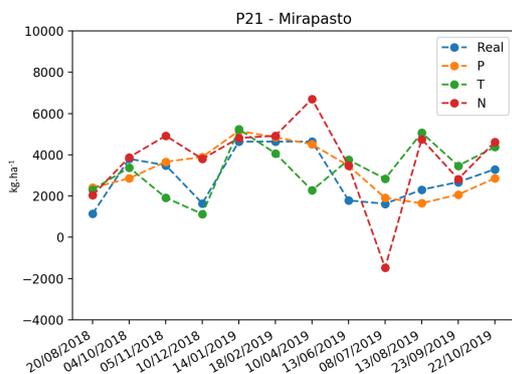
Para realizar refinamentos no modelo original, com o objetivo de aumentar a acurácia, foram pensadas alterações e inclusões nas variáveis de entrada, bem como a substituição da predição da disponibilidade de massa de forragem pela taxa de acúmulo. Primeiramente foi feita a substituição da informação de data da amostra pelas informações de mês e ano correspondentes, com o objetivo de tornar mais evidente a informação da estação do ano em que a amostra foi coletada. A quantidade de amostras utilizadas nos testes para o modelo antigo sem alterações, foram as mesmas utilizadas nos testes para o modelo antigo com alterações, houve apenas a substituição das informações ou incremento de novas informações, como taxa de acúmulo e carga animal. Também foi utilizada a mesma sequência de testes, ou seja, estratificado por potreiro (P), estratificado por tratamento (T) e não estratificado (N).

A figura 40 ilustra graficamente a estimativa de massa de forragem instantânea ao longo do tempo usando as abordagens dos três tipos de treinamento. Foram feitas 10 repetições para cada potreiro, utilizando todos os tipos de treinamento. Para constatar o melhor e pior caso, foram verificados os RMSE de cada tipo de treinamento em cada uma das repetições, o menor valor de RMSE obtido foi considerado o melhor caso e o maior valor de RMSE obtido foi considerado o pior caso. Para o potreiro P21 MIRAPASTO (melhor caso) o RMSE médio foi de 1047,73 kg.ha⁻¹ para o treinamento estratificado por potreiro, para o treinamento estratificado por tratamento foi de 1626,13 kg.ha⁻¹ e para o treinamento não estratificado foi de 1643,8 kg.ha⁻¹ e para o potreiro P21 Infestado (pior caso) o RMSE médio para o treinamento estratificado por potreiro foi de 1722,95 kg.ha⁻¹, para o treinamento estratificado por tratamento foi de 2490,42 kg.ha⁻¹ e para o treinamento não estratificado foi de 2290,27 kg.ha⁻¹.

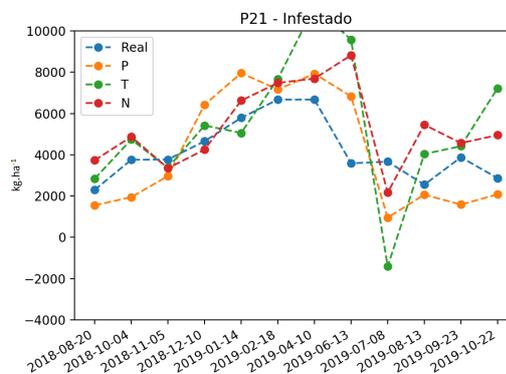
A figura 40(a) mostra os resultados obtidos, para o melhor caso, a partir dos três tipos de treinamento utilizados. Através do gráfico é possível verificar que a curva dos valores preditos que manteve um comportamento mais próximo da curva dos valores reais foi do treinamento estratificado por potreiro (P), obtendo o menor RMSE de 990,05 kg.ha⁻¹. A figura 40(b) apresenta o comportamentos das curvas dos valores reais e preditos, para o pior caso, utilizando os três tipos de treinamento. É possível verificar que a curva dos valores preditos que se manteve mais diferenciada da curva dos valores reais foi a do treinamento estratificado por tratamento (T), com o maior RMSE de 2987,24 kg.ha⁻¹.

Figura 40 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de massa de forragem, para cada subárea

(a) Melhor Caso (P)



(b) Pior Caso (T)

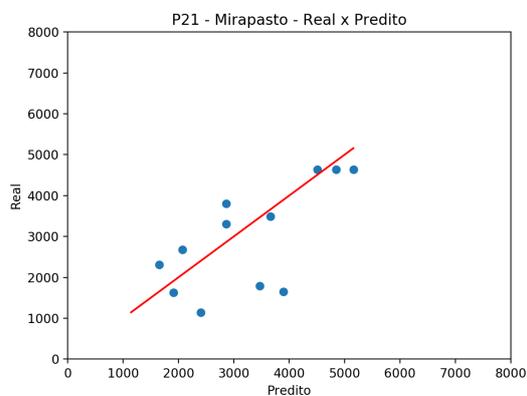


Fonte: Autor (2021)

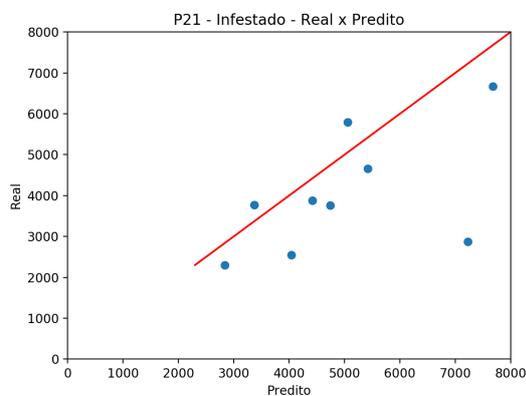
A figura 41 apresenta os gráficos onde é possível analisar a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso, com seus respectivos treinamentos. A figura 41(a) apresenta a correlação para o potreiro P21 MIRAPASTO (melhor caso), através da regressão linear foi possível calcular um R^2 linear de 0,46 e um coeficiente angular de $0,74^\circ$. Para o potreiro P21 Infestado (pior caso) a correlação entre valores reais e preditos pelo treinamento estratificado por tratamento (T) é apresentado na figura 41(b), com um R^2 linear de 0,24 e um coeficiente angular de $0,22^\circ$.

Figura 41 – valores reais e preditos de massa de forragem, para melhor e pior caso

(a) Melhor Caso (P)



(b) Pior Caso (T)



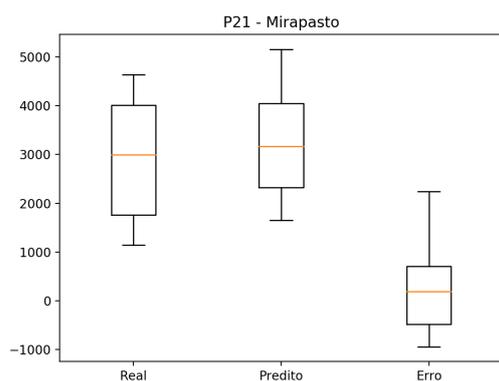
Fonte: Autor (2021)

Através da figura 42 é possível verificar a distribuição dos valores reais, preditos

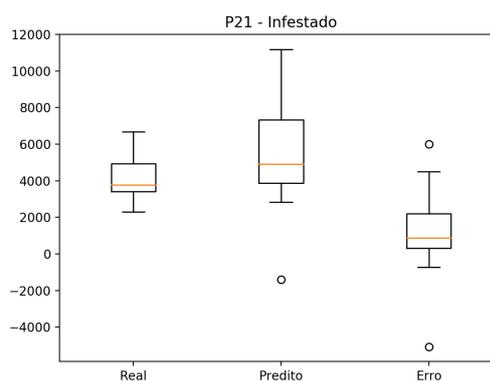
e também de erro. A figura 42(a) apresenta a distribuição para o melhor caso a partir do treinamento estratificado por potreiro (P), podendo verificar que a dispersão dos valores reais é maior do que dos valores preditos, havendo semelhança entre os valores dos terceiros quartis. A figura 42(b) mostra que, para o potreiro P21 Infestado (pior caso) a partir do treinamento estratificado por tratamento (T), a dispersão dos valores preditos é maior em relação a dispersão dos valores reais, havendo diferença entre os valores máximos dos conjuntos de valores reais e preditos. A partir do conjunto de valores de erro, analisando o *outlier* inferior, é possível constatar que ocorreu uma predição consideravelmente menor do que o valor real, já analisando o *outlier* superior é possível constatar que ocorreu uma predição consideravelmente maior do que o valor real.

Figura 42 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de massa de forragem

(a) Melhor Caso (P)



(b) Pior Caso (T)



Fonte: Autor (2021)

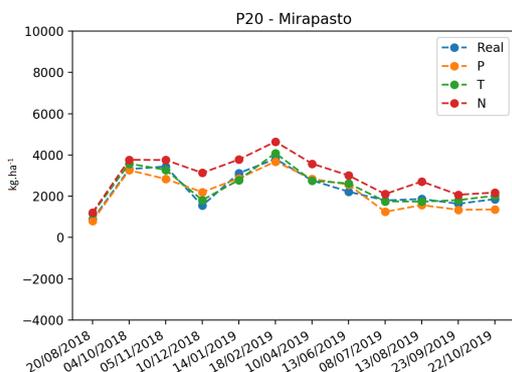
Com o intuito de melhorar a acurácia do modelo, foi inserida, como entrada, para o modelo de predição a informação de taxa de acúmulo. A figura 43 ilustra graficamente a estimativa de massa de forragem instantânea ao longo do tempo usando as abordagens dos três tipos de treinamento. Foram feitas 10 repetições para cada potreiro, utilizando todos os tipos de treinamento. Para constatar o melhor e pior caso, foram verificados os RMSE de cada tipo de treinamento em cada uma das repetições, o menor valor de RMSE obtido foi considerado o melhor caso e o maior valor de RMSE obtido foi considerado o pior caso. Para o potreiro P20 MIRAPASTO (melhor caso) o RMSE médio foi de 381,62 kg.ha⁻¹ para o treinamento estratificado por potreiro, para o treinamento estratificado por tratamento foi de 295,91 kg.ha⁻¹ e para o treinamento não estratificado foi de 783,39 kg.ha⁻¹ e para o potreiro P21 Infestado (pior caso) o RMSE médio para o

treinamento estratificado por potreiro foi de 1232,15 kg.ha⁻¹, para o treinamento estratificado por tratamento foi de 1460,28 kg.ha⁻¹ e para o treinamento não estratificado foi de 1256,89 kg.ha⁻¹.

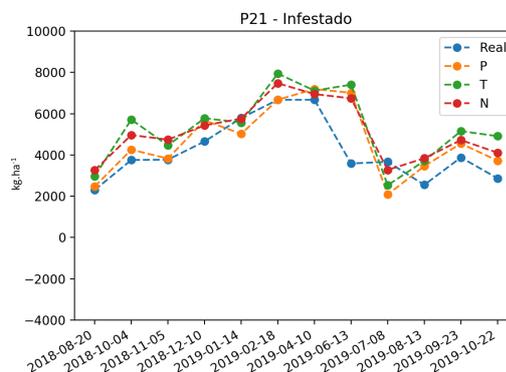
A figura 43(a) mostra os resultados obtidos, para o melhor caso, a partir dos três tipos de treinamento utilizados. Através do gráfico é possível verificar que a curva dos valores preditos que manteve um comportamento mais próximo da curva dos valores reais foi do treinamento estratificado por tratamento (T), com o menor RMSE de 239,32 kg.ha⁻¹. A figura 43(b) apresenta o comportamentos das curvas dos valores reais e preditos, para o pior caso, utilizando os três tipos de treinamento. É possível verificar que a curva dos valores preditos que se manteve mais diferenciada da curva dos valores reais foi a do treinamento estratificado por tratamento (T), com o maior RMSE de 1604,34 kg.ha⁻¹.

Figura 43 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de massa de forragem, para cada subárea

(a) Melhor Caso (T)



(b) Pior Caso (T)



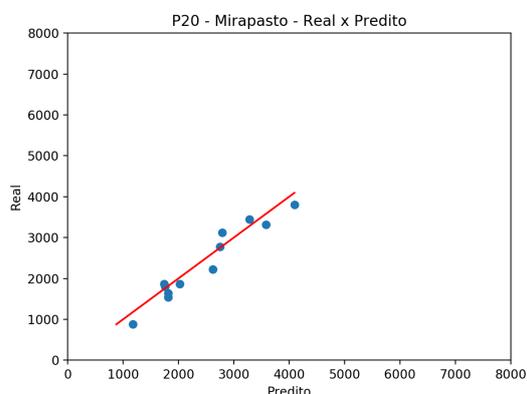
Fonte: Autor (2021)

A figura 44 apresenta os gráficos dispersão que mostram a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso, com seus respectivos treinamentos. A figura 44(a) apresenta a correlação para o potreiro P20 MIRAPASTO (melhor caso), através da regressão linear foi possível calcular um R^2 linear de 0,93 e um coeficiente angular de 1°. Para o potreiro P21 Infestado (pior caso) a correlação entre valores reais e preditos pelo treinamento estratificado por tratamento (T) é apresentado na figura 44(b), com um R^2 linear de 0,5 e um coeficiente angular de 0,62°.

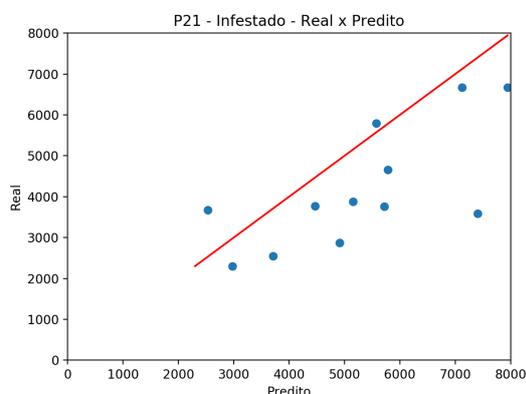
Através da figura 45 é possível verificar a distribuição dos valores reais, preditos

Figura 44 – valores reais e preditos de massa de forragem, para melhor e pior caso

(a) Melhor Caso (T)



(b) Pior Caso (T)



Fonte: Autor (2021)

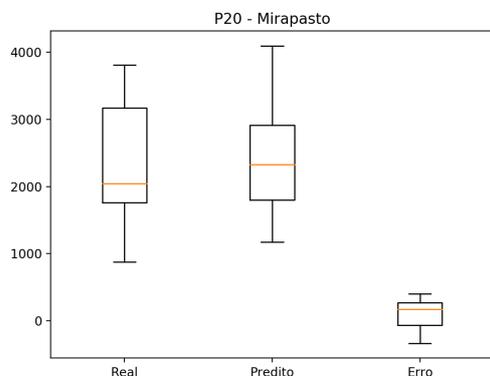
e de erro. A figura 45(a) apresenta a distribuição para o melhor caso a partir do treinamento estratificado por tratamento (T), podendo verificar que a dispersão dos valores reais é maior do que dos valores preditos, havendo uma dispersão baixa entre os valores de erro, o que mostra que os valores preditos se mantiveram próximos dos valores reais. A figura 45(b) mostra que, para o potreiro P21 Infestado (pior caso) a partir do treinamento estratificado por tratamento (T), a dispersão dos valores preditos é maior em relação a dispersão dos valores reais, havendo diferença entre os valores máximos dos conjuntos de valores reais e preditos. A partir do conjunto de valores de erro, analisando o *outlier* inferior, é possível constatar que ocorreu uma predição consideravelmente menor do que o valor real, já analisando o *outlier* superior é possível constatar que ocorreu uma predição consideravelmente maior do que o valor real.

Com o intuito de melhorar a acurácia do modelo e analisar seu comportamento com relação ao acréscimo de informações de entrada, foi inserida para o modelo de predição a informação de carga animal. A figura 46 ilustra graficamente a estimativa de massa de forragem instantânea ao longo do tempo usando as abordagens dos três tipos de treinamento. Foram feitas 10 repetições para cada potreiro, utilizando todos os tipos de treinamento. Para constatar o melhor e pior caso, foram verificados os RMSE de cada tipo de treinamento em cada uma das repetições, o menor valor de RMSE obtido foi considerado o melhor caso e o maior valor de RMSE obtido foi considerado o pior caso.

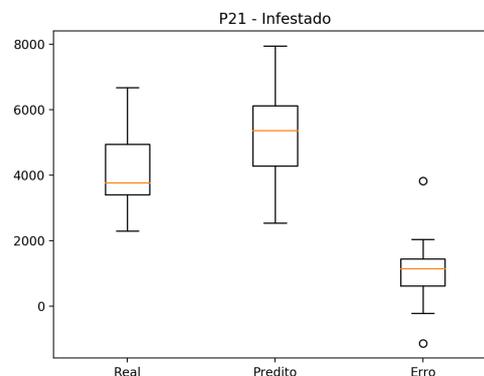
Para o potreiro P20 MIRAPASTO (melhor caso) o RMSE médio foi de 390,14 $\text{kg}\cdot\text{ha}^{-1}$ para o treinamento estratificado por potreiro, para o treinamento estratificado por tratamento foi de 271,26 $\text{kg}\cdot\text{ha}^{-1}$ e para o treinamento não estratificado foi de 782,19

Figura 45 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de massa de forragem

(a) Melhor Caso (T)



(b) Pior Caso (T)



Fonte: Autor (2021)

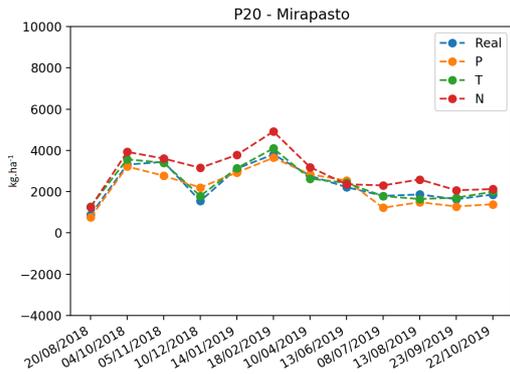
kg.ha⁻¹ e para o potreiro P21 Infestado (pior caso) o RMSE médio para o treinamento estratificado por potreiro (P) foi de 1469,07 kg.ha⁻¹, para o treinamento estratificado por tratamento foi de 1409,56 kg.ha⁻¹ e para o treinamento não estratificado foi de 1257,85 kg.ha⁻¹.

A figura 46(a) mostra os resultados obtidos, para o melhor caso, a partir dos três tipos de treinamento utilizados. Através do gráfico é possível verificar que a curva dos valores preditos que manteve um comportamento mais próximo da curva dos valores reais foi do treinamento estratificado por tratamento (T), com o menor RMSE de 205,36 kg.ha⁻¹. A figura 46(b) apresenta o comportamentos das curvas dos valores reais e preditos, para o pior caso, utilizando os três tipos de treinamento. É possível verificar que a curva dos valores preditos que se manteve mais diferenciada da curva dos valores reais foi a do treinamento estratificado por potreiro (P), com o maior RMSE de 1595,58 kg.ha⁻¹.

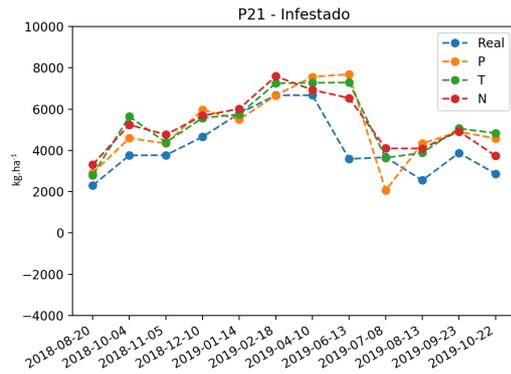
A figura 47 apresenta os gráficos dispersão que mostram a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso, com seus respectivos treinamentos. A figura 47(a) apresenta a correlação para o potreiro P20 MIRAPASTO (melhor caso), através da regressão linear foi possível calcular um R² linear de 0,95 e um coeficiente angular de 0,98°. Para o potreiro P21 Infestado (pior caso) a correlação entre valores reais e preditos pelo treinamento estratificado por potreiro (P) é apresentado na figura 47(b), com um R² linear de 0,4 e um coeficiente angular de 0,55°.

Figura 46 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de massa de forragem, para cada subárea

(a) Melhor Caso (T)



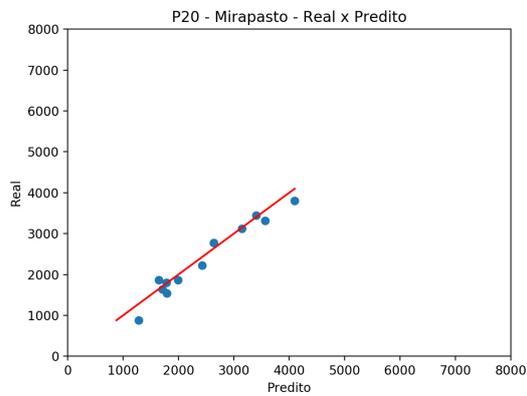
(b) Pior Caso (P)



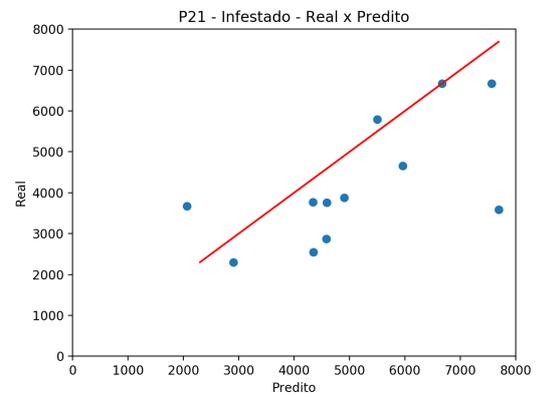
Fonte: Autor (2021)

Figura 47 – valores reais e preditos de massa de forragem, para melhor e pior caso

(a) Melhor Caso (T)



(b) Pior Caso (P)

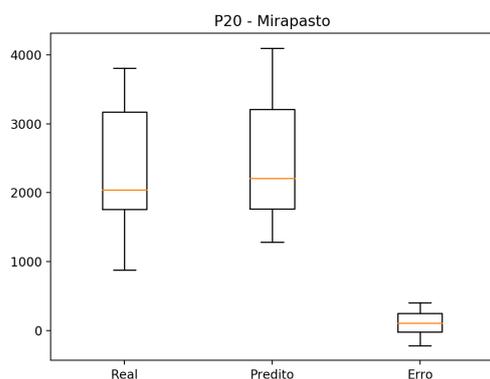


Fonte: Autor (2021)

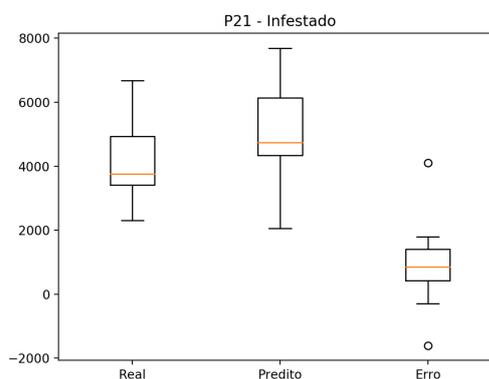
Através da figura 48 é possível verificar a distribuição dos valores reais, preditos e de erro. A figura 48(a) apresenta a distribuição para o melhor caso a partir do treinamento estratificado por tratamento (T), podendo verificar que a dispersão dos conjuntos de dados é similar e que a dispersão do conjunto de dados de erro é baixa, podendo ser verificada pelo tamanho da caixa. A figura 48(b) mostra que, para o potreiro P21 Infestado (pior caso) a partir do treinamento estratificado por potreiro (P), há diferença entre os valores máximos, dos primeiros e dos terceiros quartis dos conjuntos de valores reais e preditos. A partir do conjunto de valores de erro, analisando o *outlier* inferior, é possível constatar que ocorreu uma predição consideravelmente menor do que o valor real, já analisando o *outlier* superior é possível constatar que ocorreu uma predição consideravelmente maior do que o valor real.

Figura 48 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de massa de forragem

(a) Melhor Caso (T)



(b) Pior Caso (P)



Fonte: Autor (2021)

5.4.1 Resultados do Modelo Ajustado com Conjunto A

Para a realização do treinamento estratificado por potreiro, foram utilizadas as mesmas quantidades de amostras, quantidade de repetições de testes e análise do RMSE para o melhor e pior caso realizados na subseção 5.3.1.

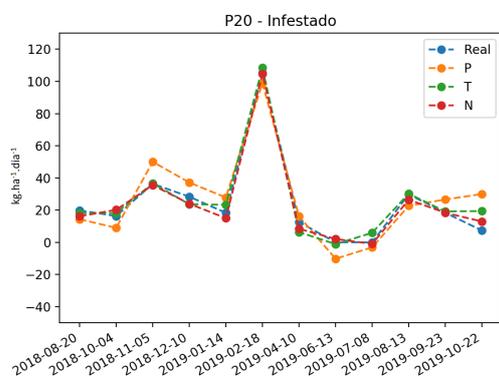
A figura 49 ilustra a estimativa de taxa de acúmulo instantânea ao longo do tempo usando as abordagens dos três tipos de treinamento. Para o potreiro P20 Infestado (melhor caso) o RMSE médio foi de $11,46 \text{ kg} \cdot \text{ha}^{-1} \cdot \text{dia}^{-1}$ para o treinamento estratificado por

potreiro, para o treinamento estratificado por tratamento foi de $5,65 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$ e para o treinamento não estratificado foi de $3,69 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$ e para o potreiro P21 MIRAPASTO (pior caso) o RMSE médio para o treinamento estratificado por potreiro foi de $15,13 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$, para o treinamento estratificado por tratamento foi de $8,02 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$ e para o treinamento não estratificado foi de $22,39 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$.

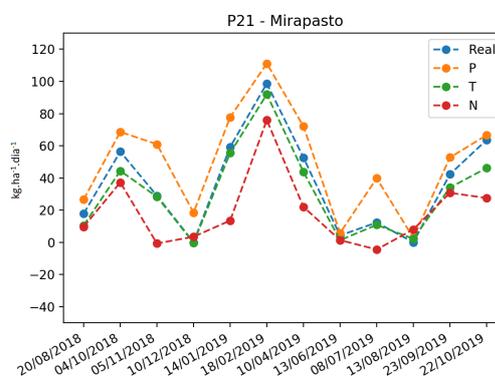
A figura 49(a) mostra os resultados obtidos, para o melhor caso, a partir dos três tipos de treinamento utilizados. Através do gráfico é possível verificar que a curva dos valores preditos que manteve um comportamento mais próximo da curva dos valores reais foi do treinamento não estratificado (N), obtendo o menor RMSE de $3,21 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$. A figura 49(b) apresenta o comportamentos das curvas dos valores reais e preditos, para o pior caso, utilizando os três tipos de treinamento. É possível verificar que a curva dos valores preditos que se manteve mais distante da curva dos valores reais foi a do treinamento não estratificado (N), com o maior RMSE de $23,56 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$.

Figura 49 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso

(a) Melhor Caso (N)



(b) Pior Caso (N)

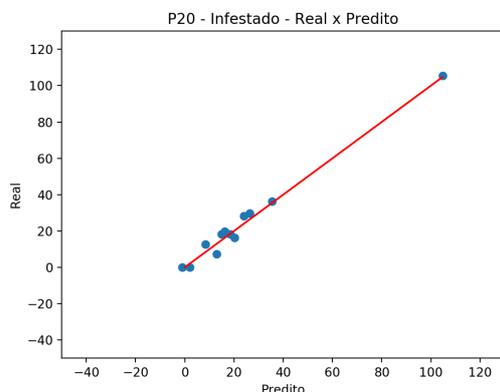


Fonte: Autor (2021)

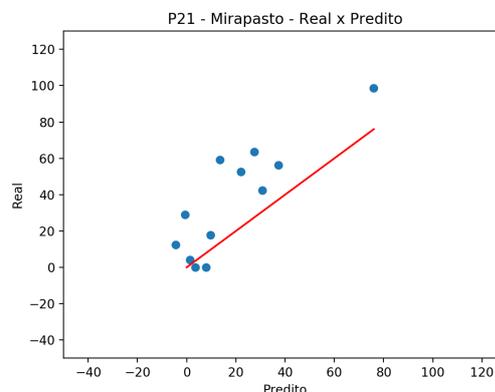
A figura 50 apresenta os gráficos onde é possível analisar a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso, com seus respectivos treinamentos. A figura 50(a) apresenta a correlação para o potreiro P20 Infestado (melhor caso), através da regressão linear foi possível calcular um R^2 linear de 0,98 e um coeficiente angular de 1° . Para o potreiro P21 MIRAPASTO (pior caso) a correlação entre valores reais e preditos pelo treinamento não estratificado (N) é apresentado na figura 50(b), com um R^2 linear de 0,74 e um coeficiente angular de $1,18^\circ$.

Figura 50 – valores reais e preditos de taxa de acúmulo, para melhor e pior caso

(a) Melhor Caso (N)



(b) Pior Caso (N)



Fonte: Autor (2021)

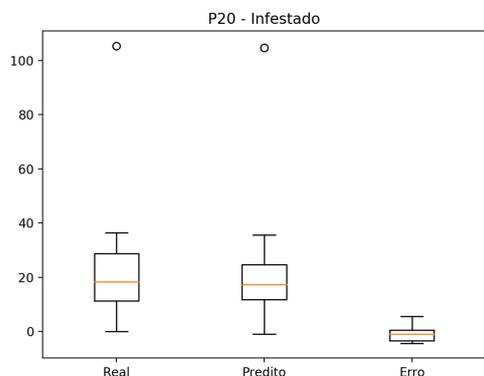
Através da figura 51 é possível verificar a distribuição dos valores reais, preditos e também da diferença entre o valor real e o predito para cada uma das datas, chamado de erro. A figura 51(a) apresenta a distribuição para o melhor caso a partir do treinamento não estratificado (N), podendo verificar que a dispersão dos valores reais é maior do que dos valores preditos, havendo semelhança entre seus valores máximos e também a ocorrência de um *outlier* superior em cada conjunto de valores. A figura 51(b) mostra que, para o potreiro P20 Infestado (pior caso) a partir do treinamento não estratificado (N), a dispersão dos valores reais é maior em relação a dispersão dos valores preditos, ocorrendo um *outlier* no conjunto de valores preditos e uma diferença considerável entre os valores máximos de cada conjunto. A partir do valor mínimo expresso pelo conjunto de dados de erro é possível constatar a ocorrência de uma predição bem menor do que se comparado ao valor real.

5.4.2 Resultados do Modelo Ajustado com Conjunto B

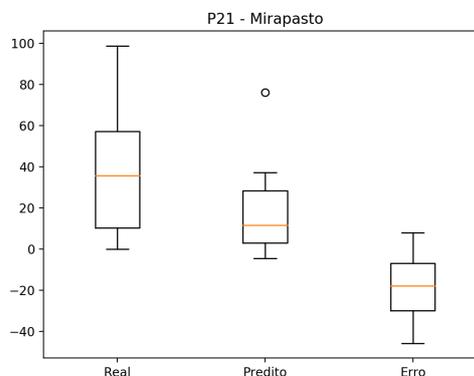
Com o objetivo de flexibilizar as entradas referentes a dados de pastagem, foi utilizado o Conjunto B, no qual a área total está dividida em nove potreiros, com os seguintes manejos: em P1, P2 e P7 é aplicado o tratamento Campo Nativo Melhorado (CNM), em P3, P6 e P8 o tratamento Campo Nativo (CN) e em P4, P5 e P9 o tratamento Campo Nativo Fertilizado (CNF). Para realizar a predição instantânea, para todos os potreiros, foram utilizados 39 dados para treinamento (76,5%) e 12 dados para teste (23,5%).

Figura 51 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo

(a) Melhor Caso (N)



(b) Pior Caso (N)



Fonte: Autor (2021)

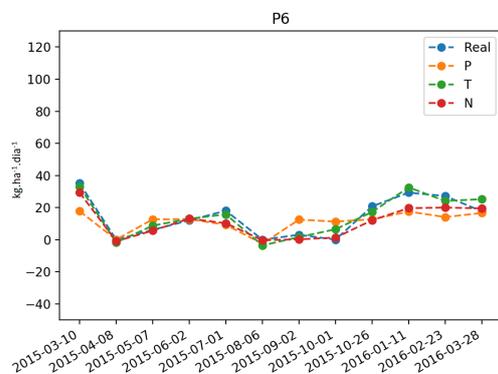
A figura 52 ilustra a estimativa de taxa de acúmulo instantânea ao longo do tempo usando as abordagens dos três tipos de treinamento. Para o potreiro P6 (melhor caso) o RMSE médio foi de $9,27 \text{ kg.ha}^{-1}.\text{dia}^{-1}$ para o treinamento estratificado por potreiro, para o treinamento estratificado por tratamento foi de $4,43 \text{ kg.ha}^{-1}.\text{dia}^{-1}$ e para o treinamento não estratificado foi de $5,01 \text{ kg.ha}^{-1}.\text{dia}^{-1}$ e para o potreiro P2 (pior caso) o RMSE médio para o treinamento estratificado por potreiro foi de $11,09 \text{ kg.ha}^{-1}.\text{dia}^{-1}$, para o treinamento estratificado por tratamento foi de $14 \text{ kg.ha}^{-1}.\text{dia}^{-1}$ e para o treinamento não estratificado foi de $15,3 \text{ kg.ha}^{-1}.\text{dia}^{-1}$.

A figura 52(a) mostra os resultados obtidos, para o melhor caso, a partir dos três tipos de treinamento utilizados. Através do gráfico é possível verificar que a curva dos valores preditos que manteve um comportamento mais próximo da curva dos valores reais foi do treinamento estratificado por tratamento (T), obtendo o menor RMSE de $3,72 \text{ kg.ha}^{-1}.\text{dia}^{-1}$. A figura 52(b) apresenta o comportamentos das curvas dos valores reais e preditos, para o pior caso, utilizando os três tipos de treinamento. É possível verificar que a curva dos valores preditos que se manteve mais distante da curva dos valores reais foi a do treinamento não estratificado (N), com o maior RMSE de $16,3 \text{ kg.ha}^{-1}.\text{dia}^{-1}$.

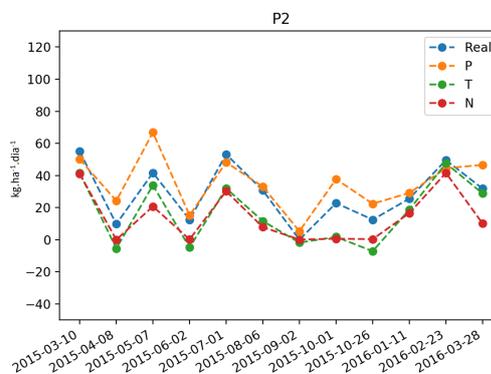
A figura 53 apresenta os gráficos onde é possível analisar a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso, com seus respectivos treinamentos. A figura 53(a) apresenta a correlação para o potreiro P6 (melhor caso), através da regressão linear foi possível calcular um R^2 linear de 0,9 e um coeficiente angular de $0,94^\circ$. Para o potreiro P2 (pior caso) a correla-

Figura 52 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso

(a) Melhor Caso (T)



(b) Pior Caso (N)

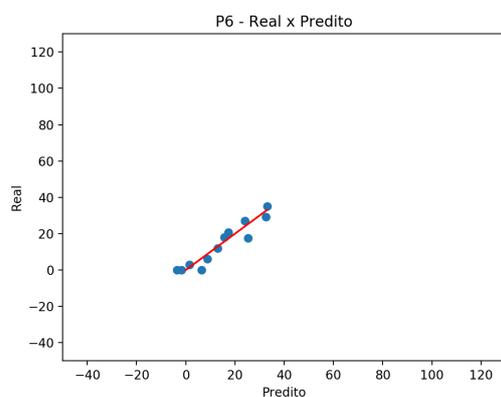


Fonte: Autor (2021)

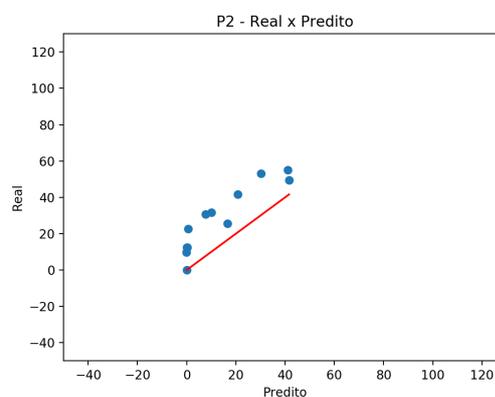
ção entre valores reais e preditos pelo treinamento não estratificado (N) é apresentado na figura 53(b), com um R^2 linear de 0,83 e um coeficiente angular de $1,04^\circ$.

Figura 53 – valores reais e preditos de taxa de acúmulo, para melhor e pior caso

(a) Melhor Caso (T)



(b) Pior Caso (N)



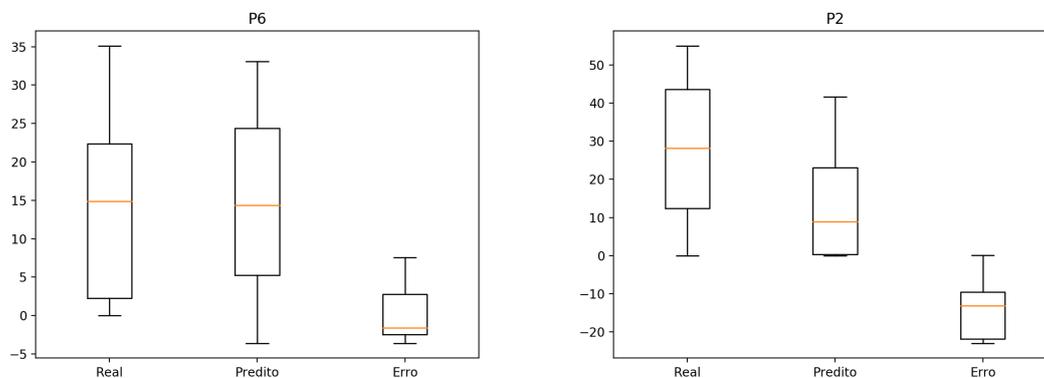
Fonte: Autor (2021)

Através da figura 54 é possível verificar a distribuição dos valores reais, preditos e erro. A figura 54(a) apresenta a distribuição para o melhor caso a partir do treinamento estratificado por tratamento (T), podendo verificar que há semelhança entre seus valores máximos. A figura 54(b) mostra que, para o potreiro P2 (pior caso) a partir do treinamento não estratificado (N), a dispersão dos valores reais é maior em relação a dispersão dos valores preditos, havendo diferença considerável entre os valores máximos.

Figura 54 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo

(a) Melhor Caso (T)

(b) Pior Caso (N)



Fonte: Autor (2021)

5.4.3 Resultados do Modelo Ajustado com a união dos Conjuntos A e B

Para poder analisar a capacidade do modelo em realizar o aprendizado através de fontes variadas de dados de pastagem, foi feita a união entre os conjuntos de dados A e B para o melhor caso observado, ou seja, treinamento estratificado por tratamento. Através da união dos conjuntos de dados de treinamento dos poteiros com tratamentos semelhantes, para os tratamentos Infestado/CN o conjunto total de amostras de treinamento foi de 186 amostras e para os tratamentos MIRAPASTO/CNM e MIRAPASTO/CNF foi de 180 amostras. Foram feitas 10 repetições utilizando o treinamento estratificado por tratamento (T), o menor RMSE foi considerado o melhor caso e o maior RMSE foi considerado o pior caso.

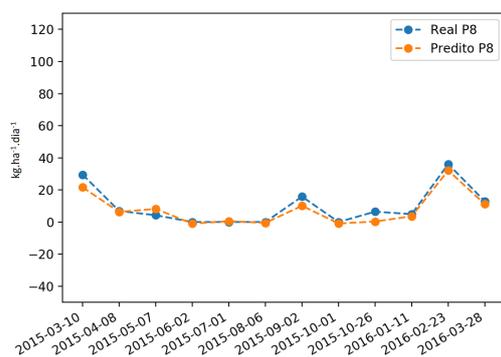
Realizando os testes a partir da união Infestado/CN foram obtidos o melhor e pior caso para os poteiros que recebem ou tratamento Infestado ou tratamento CN. A figura 55 ilustra a estimativa de taxa de acúmulo instantânea ao longo do tempo usando as abordagens dos três tipos de treinamento. Para o poteiro P8 (melhor caso) o RMSE médio para o treinamento estratificado por tratamento foi de $4,79 \text{ kg.ha}^{-1}.\text{dia}^{-1}$ e para o poteiro P21 Infestado (pior caso) o RMSE médio para o treinamento estratificado por tratamento foi de $21,5 \text{ kg.ha}^{-1}.\text{dia}^{-1}$.

A figura 55(a) mostra os resultados obtidos para o melhor caso. Através do gráfico é possível verificar que a curva dos valores preditos manteve um comportamento similar a curva dos valores reais com menor RMSE obtido de $3,79 \text{ kg.ha}^{-1}.\text{dia}^{-1}$. A figura 55(b)

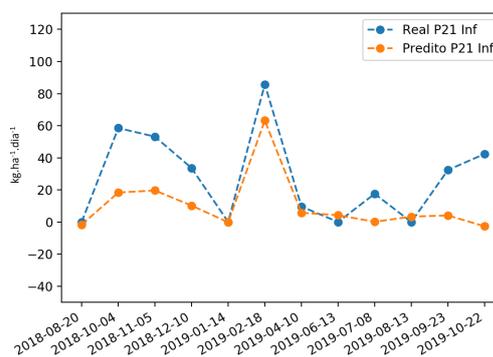
apresenta o comportamento das curvas dos valores reais e preditos, para o pior caso, é possível verificar que a curva dos valores preditos manteve um comportamento similar em metade dos pontos se comparada a curva dos valores reais, com o maior RMSE de $24,12 \text{ kg} \cdot \text{ha}^{-1} \cdot \text{dia}^{-1}$.

Figura 55 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso

(a) Melhor Caso (T)



(b) Pior Caso (T)



Fonte: Autor (2021)

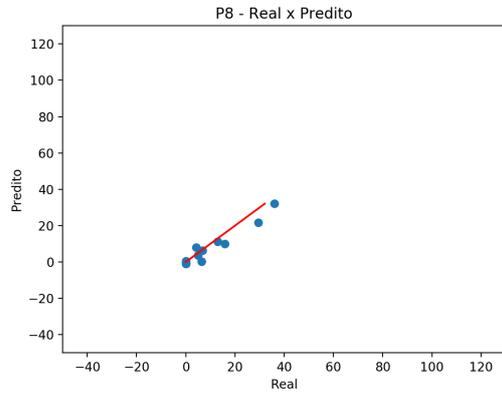
A figura 56 apresenta os gráficos onde é possível analisar a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso. A figura 56(a) apresenta a correlação para o potreiro P8 (melhor caso), através da regressão linear foi possível calcular um R^2 linear de 0,93 e um coeficiente angular de $1,13^\circ$. Para o potreiro P21 Infestado (pior caso) a correlação entre valores reais e preditos é apresentado na figura 56(b), com um R^2 linear de 0,64 e um coeficiente angular de $1,24^\circ$.

Através da figura 57 é possível verificar a distribuição dos valores reais, preditos e erro. A figura 57(a) apresenta a distribuição para o melhor caso, podendo constatar que os valores dos primeiros quartis são consideravelmente semelhantes entre os conjuntos de valores reais e preditos. A figura 57(b) mostra que, para o potreiro P21 Infestado (pior caso), a dispersão dos valores reais é maior em relação a dispersão dos valores preditos, havendo diferença considerável entre os valores máximos.

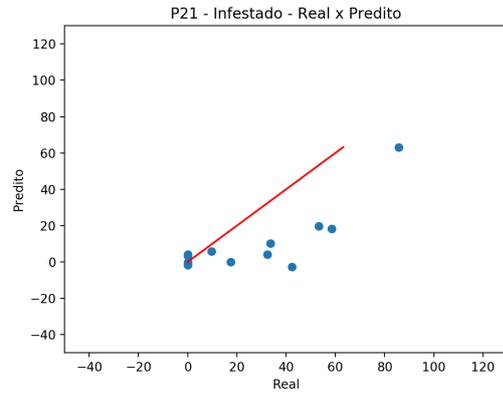
Realizando os testes a partir da união MIRAPASTO/CNM foram obtidos o melhor e pior caso para os potreiros que recebem ou tratamento MIRAPASTO ou tratamento CNM. A figura 58 ilustra a estimativa de taxa de acúmulo instantânea ao longo do tempo usando as abordagens dos três tipos de treinamento. Para o potreiro P7 (melhor caso) o

Figura 56 – Valores reais e preditos de taxa de acúmulo, para melhor e pior caso

(a) Melhor Caso (T)



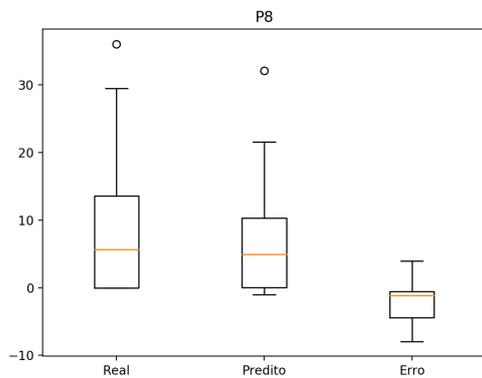
(b) Pior Caso (T)



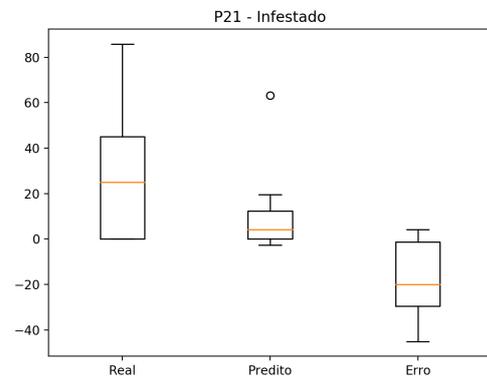
Fonte: Autor (2021)

Figura 57 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo

(a) Melhor Caso (T)



(b) Pior Caso (T)



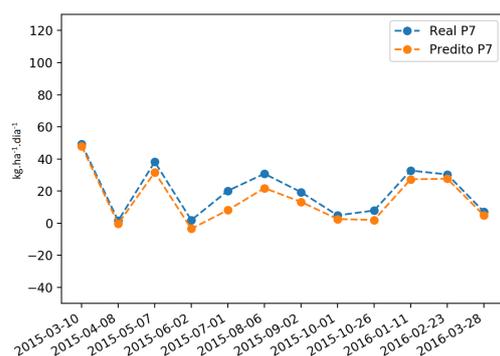
Fonte: Autor (2021)

RMSE médio para o treinamento estratificado por tratamento foi de $6,75 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$ e para o potreiro P21 MIRAPASTO (pior caso) o RMSE médio para o treinamento estratificado por tratamento foi de $15,31 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$.

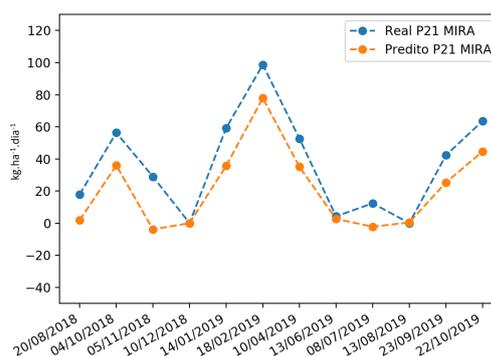
A figura 58(a) mostra os resultados obtidos para o melhor caso. Através do gráfico é possível verificar que a curva dos valores preditos manteve um comportamento similar a curva dos valores reais com menor RMSE obtido de $5,92 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$. A figura 58(b) apresenta o comportamentos das curvas dos valores reais e preditos, para o pior caso, é possível verificar que a curva dos valores preditos manteve um comportamento similar na maioria dos pontos se comparada a curva dos valores reais, porém, havendo uma maior diferença nos valores de cada ponto, com o maior RMSE obtido de $18,04 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$.

Figura 58 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso

(a) Melhor Caso (T)



(b) Pior Caso (T)



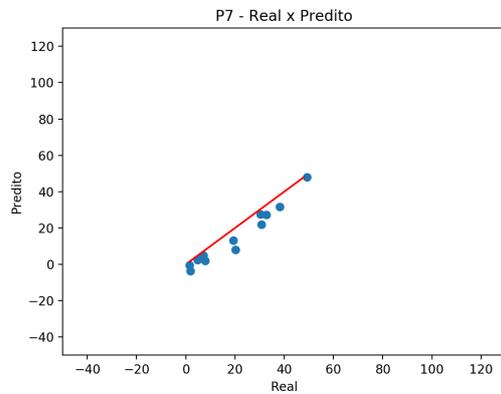
Fonte: Autor (2021)

A figura 59 apresenta os gráficos onde é possível analisar a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso. A figura 59(a) apresenta a correlação para o potreiro P7 (melhor caso), através da regressão linear foi possível calcular um R^2 linear de 0,95 e um coeficiente angular de $0,98^\circ$. Para o potreiro P21 MIRAPASTO (pior caso) a correlação entre valores reais e preditos é apresentado na figura 59(b), com um R^2 linear de 0,9 e um coeficiente angular de $1,15^\circ$.

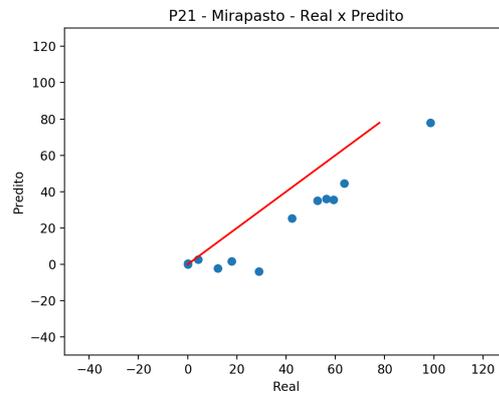
Através da figura 60 é possível verificar a distribuição dos valores reais, preditos e erro. A figura 60(a) apresenta a distribuição para o melhor caso, podendo constatar que há similaridade entre os valores máximos dos conjuntos de valores reais e preditos. Através do *outlier* presente no conjunto de valores de erro é possível constatar que houve uma

Figura 59 – valores reais e preditos de taxa de acúmulo, para melhor e pior caso

(a) Melhor Caso (T)



(b) Pior Caso (T)

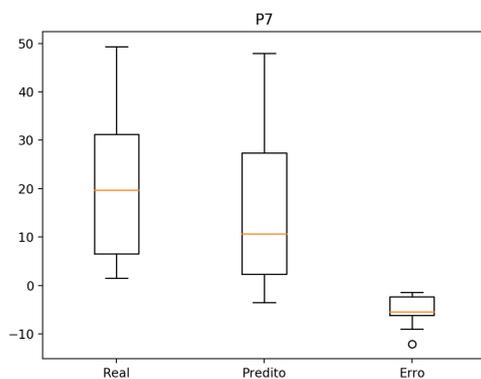


Fonte: Autor (2021)

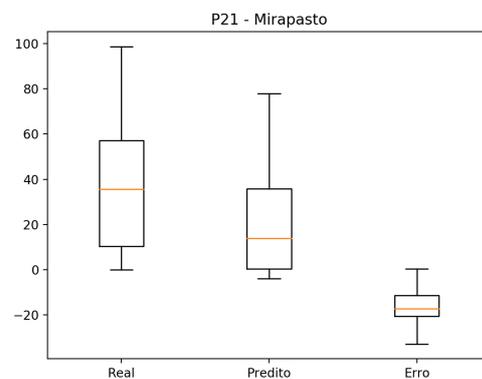
predição consideravelmente menor do que o valor real. A figura 60(b) mostra que, para o potreiro P21 MIRAPASTO (pior caso), a dispersão dos valores reais é maior em relação a dispersão dos valores preditos, havendo diferença considerável entre os valores máximos.

Figura 60 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo

(a) Melhor Caso (T)



(b) Pior Caso (T)



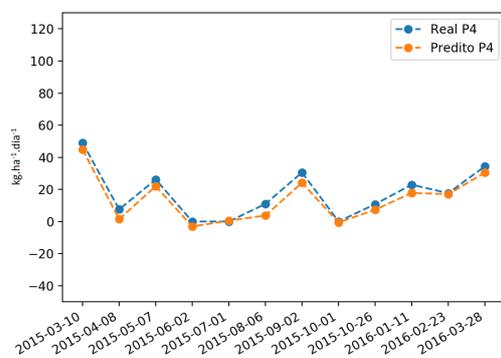
Fonte: Autor (2021)

Realizando os testes a partir da união MIRAPASTO/CNF foram obtidos o melhor e pior caso para os potreiros com esses tipos de tratamento. A figura 61 ilustra a estimativa de taxa de acúmulo instantânea ao longo do tempo usando as três abordagens de treinamento. Para o potreiro P4 (melhor caso) o RMSE médio para o treinamento estratificado por tratamento foi de $6,12 \text{ kg} \cdot \text{ha}^{-1} \cdot \text{dia}^{-1}$ e para o potreiro P21 MIRAPASTO (pior caso) o RMSE médio para o treinamento estratificado por tratamento foi de $20,34 \text{ kg} \cdot \text{ha}^{-1} \cdot \text{dia}^{-1}$.

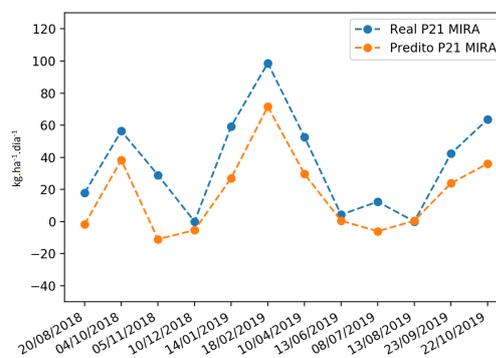
A figura 61(a) mostra os resultados obtidos para o melhor caso. Através do gráfico é possível verificar que a curva dos valores preditos manteve um comportamento similar a curva dos valores reais com menor RMSE obtido de $4,36 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$. A figura 61(b) apresenta o comportamentos das curvas dos valores reais e preditos, para o pior caso, é possível verificar que a curva dos valores preditos manteve um comportamento similar na maioria dos pontos se comparada a curva dos valores reais, porém, havendo uma maior diferença nos valores de cada ponto, com o maior RMSE obtido de $22,53 \text{ kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$.

Figura 61 – Valores reais *versus* preditos pela RNN LSTM da estimativa instantânea de taxa de acúmulo, para o melhor e pior caso

(a) Melhor Caso (T)



(b) Pior Caso (T)



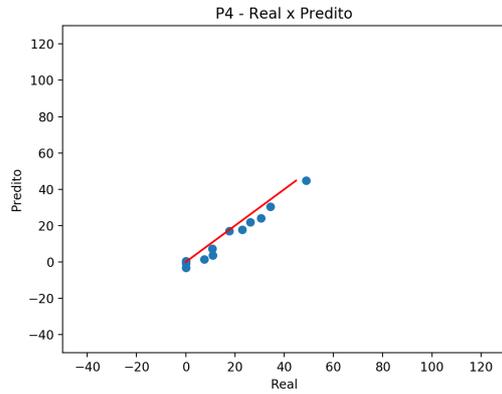
Fonte: Autor (2021)

A figura 62 apresenta os gráficos onde é possível analisar a correlação entre os valores reais e preditos, permitindo verificar se há linearidade entre os valores para o melhor e pior caso. A figura 62(a) apresenta a correlação para o potreiro P4 (melhor caso), através da regressão linear foi possível calcular um R^2 linear de 0,97 e um coeficiente angular de $1,03^\circ$. Para o potreiro P21 MIRAPASTO (pior caso) a correlação entre valores reais e preditos é apresentado na figura 62(b), com um R^2 linear de 0,86 e um coeficiente angular de $1,15^\circ$.

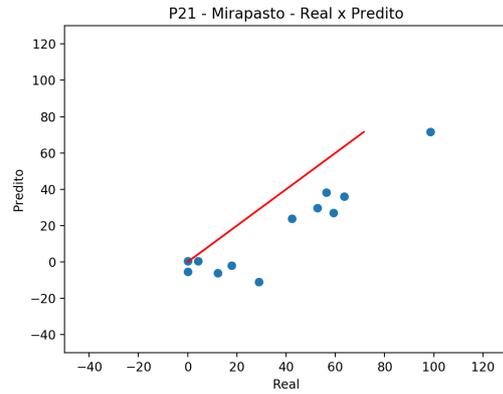
Através da figura 63 é possível verificar a distribuição dos valores reais, preditos e erro. A figura 63(a) apresenta a distribuição para o melhor caso, podendo constatar que há similaridade entre os valores mínimos e máximos dos conjuntos de valores reais e preditos. A figura 63(b) mostra que, para o potreiro P21 MIRAPASTO (pior caso), a dispersão dos valores reais é maior em relação a dispersão dos valores preditos, havendo diferença considerável entre os valores máximos.

Figura 62 – valores reais e preditos de taxa de acúmulo, para melhor e pior caso

(a) Melhor Caso (T)



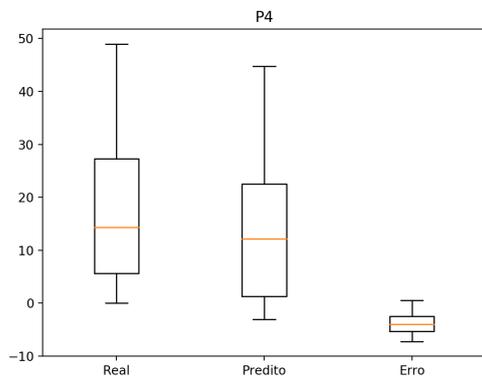
(b) Pior Caso (T)



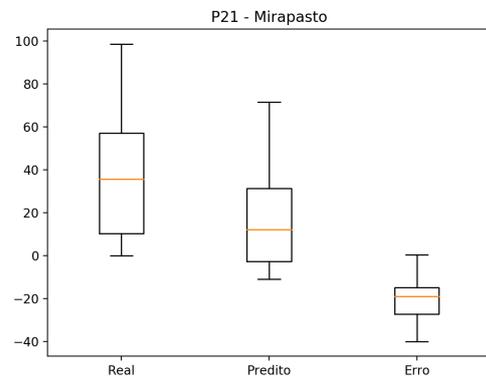
Fonte: Autor (2021)

Figura 63 – Gráficos de caixa, do melhor e pior caso, que mostram os valores máximos e mínimos reais, preditos e de erro, de taxa de acúmulo

(a) Melhor Caso (T)



(b) Pior Caso (T)



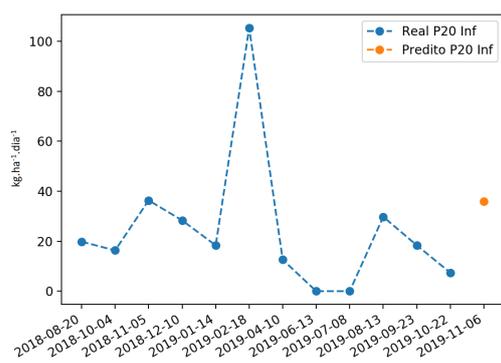
Fonte: Autor (2021)

5.4.4 Resultados do Modelo Ajustado do Conjunto A para Predição Futura

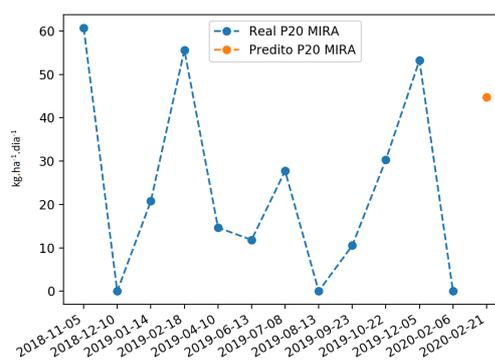
Para a realização dos experimentos para predição futura de taxa de acúmulo utilizando o conjunto de dados A, foi definido o treinamento estratificado por tratamento, utilizando para cada potreiro o conjunto de treinamento do respectivo tratamento. Por exemplo, para o potreiro P20 Infestado, o conjunto de dados de treinamento do tratamento Infestado (P20 + P21). O total de amostras para o tratamento Infestado foi de 69 e para o tratamento MIRAPASTO foi de 63.

Figura 64 – Valores para estimativa futura de taxa de acúmulo, reais *versus* preditos pela RNN LSTM para cada uma das subáreas de estudo, para treinamento estratificado por tratamento (T)

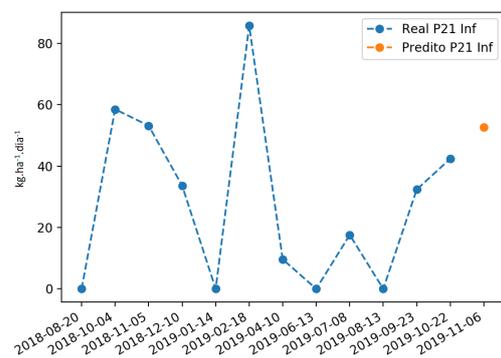
(a) Potreiro 20 Infestado



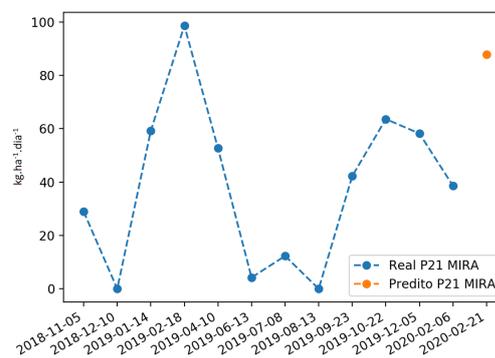
(b) Potreiro 20 MIRAPASTO



(c) Potreiro 21 Infestado



(d) Potreiro 21 MIRAPASTO



Fonte: Autor (2021)

Para que a rede obtivesse melhores resultados, foi utilizado no conjunto de dados de teste o valor de taxa de acúmulo real do mês em que se deseja realizar a previsão, porém, do ano anterior. A figura 64 apresenta a curva dos valores reais e o valor predito

para a rede após 15 dias da última coleta de amostras de pastagem. A figura 64(a) mostra que para todos os poteiros a predição gerou um resultado dentro do intervalo observado nos valores reais apresentados no gráfico.

5.5 Avaliação do NDVI e teste com K-means

Com a aquisição das imagens de satélite, foi possível a criação do método indireto de obtenção de informações sobre a pastagem. Para a execução do método indireto foi necessário seguir os seguintes passos:

- Aquisição das imagens de satélite;
- Criação dos arquivos *shapefile* para cada uma das áreas de interesse;
- Recorte das imagens de satélite;
- Cálculo do NDVI;
- Armazenamento das informações do NDVI, data das imagens e informações das áreas no banco de dados.

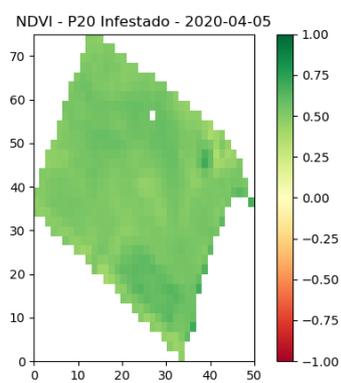
A partir dessa sequência, pode-se realizar os testes e obter resultados. Primeiramente foram calculados os NDVI de cada subárea, como o modelo de predição para dados de corte, utiliza as subáreas das gaiolas, foram calculados os NDVI de cada uma das subáreas onde estão localizadas as gaiolas. Após isso, foram feitas operações estatísticas e obtidos os valores de NDVI médio, desvio padrão e variância para cada uma das subáreas. Por fim, esses valores foram inseridos no banco de dados. Vale ressaltar que o *script* que gera as entradas para o modelo de predição para o NDVI, também leva em conta os fatores climáticos, os tamanhos das subáreas e da área total, assim como é feito quando se utiliza dados de corte da pastagem.

A figura 65 ilustra o resultado para cada uma das subáreas. Ou seja, 65(a), 65(b), 65(c) e 65(d) mostram os NDVI calculados para o Potreiro 20 Infestado, Potreiro 20 MIRAPASTO, Potreiro 21 Infestado e Potreiro 21 MIRAPASTO, respectivamente.

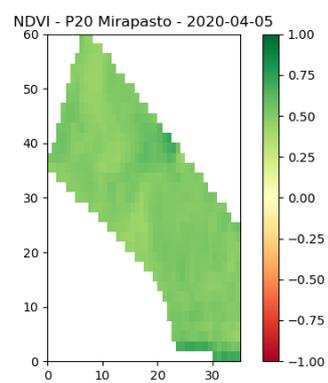
Com o objetivo de validar os resultados do *script* proposto para o cálculo do NDVI, foram comparados os resultados obtidos com a sua utilização (figura 66) e os resultados obtidos com a utilização do software QGIS (figura 67), visto que esse é um dos softwares SIG mais utilizados no mercado, já que é gratuito e de código-fonte aberto, além de disponibilizar um grande número de funcionalidades para a realização de operações com imagens georreferenciadas. A figura 67 apresenta o resultado do NDVI cal-

Figura 65 – Resultado do cálculo de NDVI para as diferentes subáreas (potreiros)

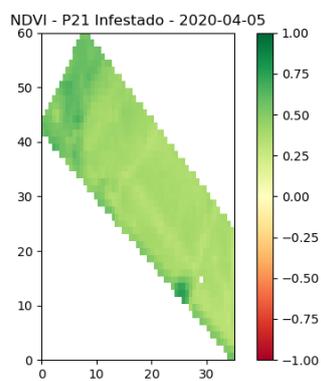
(a) NDVI para o Potreiro 20 Infestado



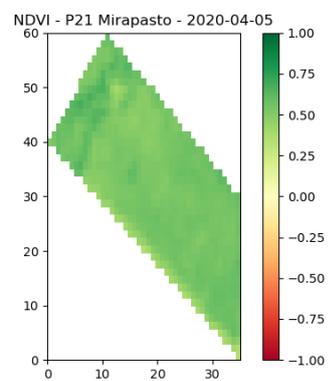
(b) NDVI para o Potreiro 20 MIRAPASTO



(c) NDVI para o Potreiro 21 Infestado



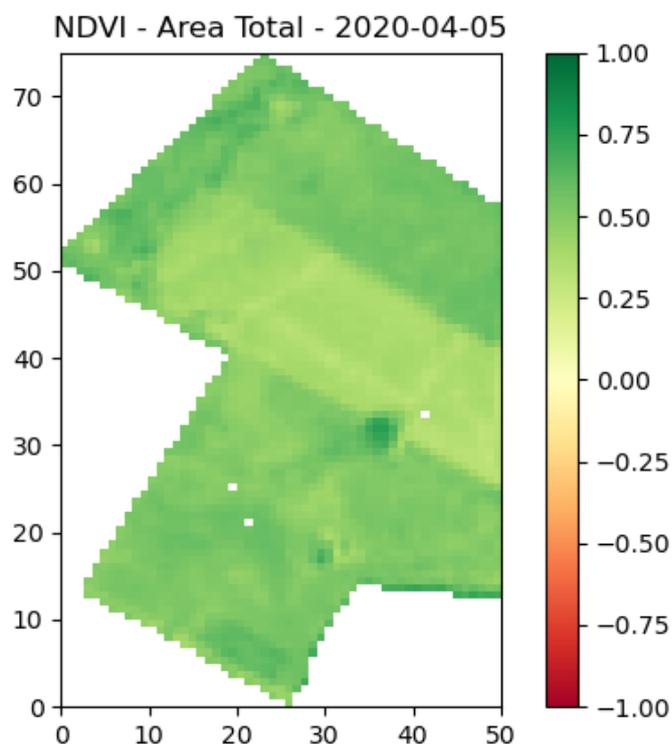
(d) NDVI para o Potreiro 21 MIRAPASTO



Fonte: Autor (2021)

culado pelo software. É possível notar que em ambos os resultados o comportamento do NDVI é igual, porém, havendo uma pequena diferença na escala de cores entre o *script* e o software. Vale ressaltar que as imagens utilizadas para calcular o NDVI apresentado na figura 66 e na figura 67 foram as mesmas.

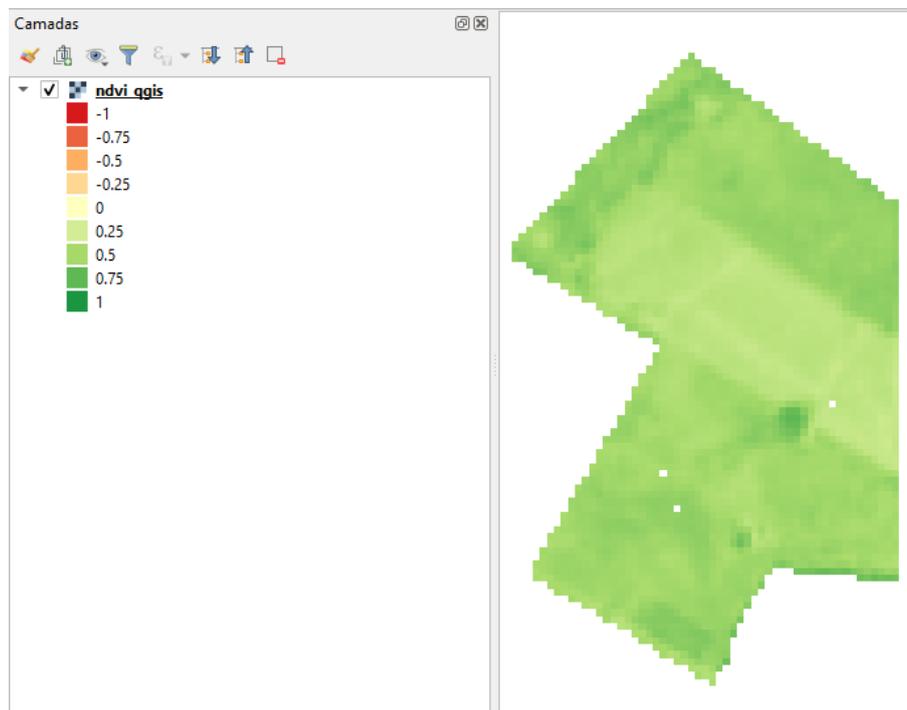
Figura 66 – Resultado do cálculo de NDVI para a área total de interesse



Fonte: Autor (2021)

Após realizados os cálculos do NDVI para cada uma das subáreas de interesse, é possível gerar as entradas tanto para realização da predição de taxa de acúmulo, tanto para o presente quanto para o futuro, substituindo o valor de MF de fora da gaiola pelo valor de NDVI da data mais próxima. A figura 68 apresenta os resultados de predição para o presente para todas as subáreas de estudo, gerando um RMSE de 10,15 $\text{kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$, para o potreiro P20 Infestado, Para o potreiro P20 MIRAPASTO o RMSE foi de 33,34 $\text{kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$ e para os potreiro P21 Infestado e P21 MIRAPASTO o RMSE foi de 6,25 $\text{kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$ e 19,28 $\text{kg}\cdot\text{ha}^{-1}\cdot\text{dia}^{-1}$, respectivamente. É possível notar que o comportamento das curvas é inverso tanto para o potreiro P20 Infestado (figura 68(a)), quanto para o potreiro P20 MIRAPASTO (figura 68(b)). Já para o potreiro P21 Infestado (figura 68(c)) o comportamento das restas é igual e para o potreiro P21 MIRAPASTO (figura 68(d)) é

Figura 67 – NDVI calculado com o software QGIS para a área total de estudo



Fonte: Autor (2021)

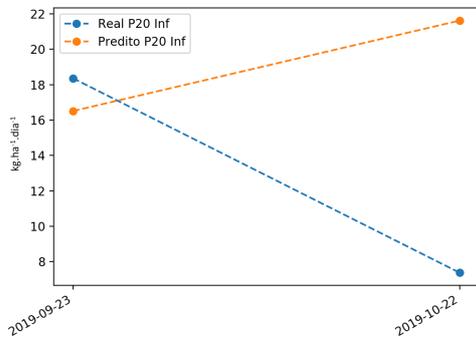
igual na metade dos casos.

A figura 69 apresenta os resultados gerados pelo modelo de predição, e é possível verificar o comportamento da curva com os valores reais de taxa de acúmulo *versus* o valor predito para o intervalo de 15 dias. As figuras 69(a) e 69(b) apresentam os resultados obtidos para o Potreiro 20, no tratamento Infestado e MIRAPASTO, respectivamente, é possível notar que o valor de taxa de acúmulo predito está fora do intervalo dos valores reais presentes nos pontos de anteriores. As figuras 69(c) e 69(d) apresentam os resultados obtidos para o Potreiro 21, no tratamento Infestado e MIRAPASTO, respectivamente, e também é possível verificar que os valores preditos extrapolam o intervalo de valores reais.

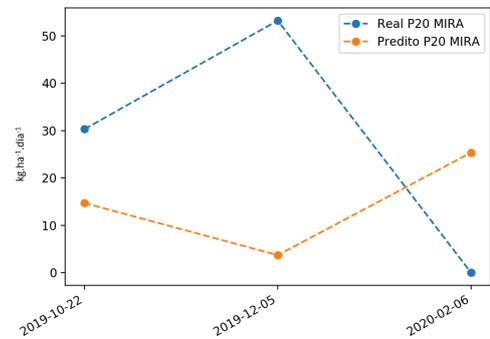
Complementarmente, em função da disponibilidade das imagens da área de interesse, foram realizados experimentos preliminares buscando identificar uma melhor divisão das subáreas. Uma das limitações do método de amostragem direta se deve ao fato das gaiolas, utilizadas para isolar a área de corte, serem posicionadas onde a pastagem representa melhor a subárea em questão, a partir de uma avaliação visual *in loco* por um profissional. Com o objetivo de aperfeiçoar as divisões das subáreas, foi criado um *script* para a classificação das imagens e com isso verificar se existia a possibilidade de utilização de um método automático para dividir as áreas. Através do pacote Python chamado

Figura 68 – Valores de taxa de acúmulo preditos pela RNN LSTM, utilizando NDVI, para o presente para cada uma das subáreas de estudo

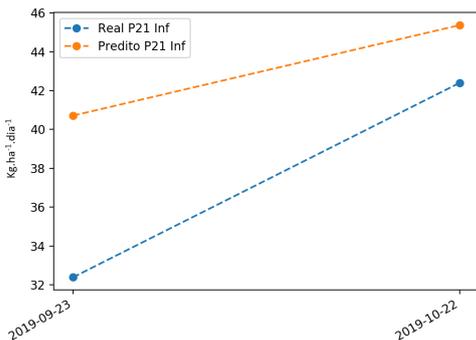
(a) Valores de NDVI reais *versus* preditos pela RNN LSTM para o presente para o Potreiro 20 Infestado



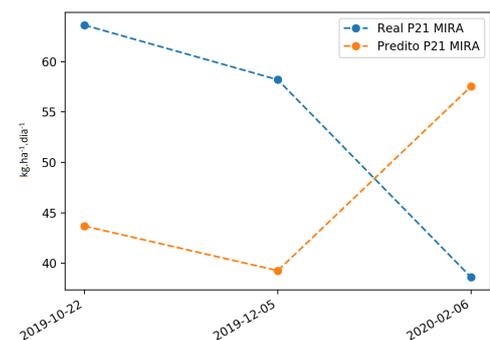
(b) Valores de taxa de acúmulo reais *versus* preditos pela RNN LSTM para o presente para o Potreiro 20 MIRAPASTO



(c) Valores de NDVI reais *versus* preditos pela RNN LSTM para o presente para o Potreiro 21 Infestado



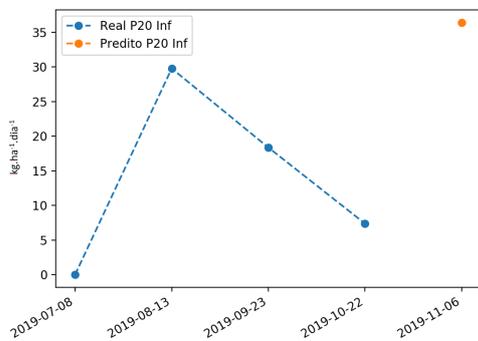
(d) Valores de NDVI reais *versus* preditos pela RNN LSTM para o presente para o Potreiro 21 MIRPASTO



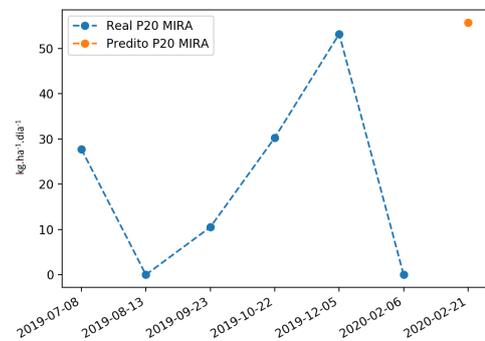
Fonte: Autor (2021)

Figura 69 – Valores de taxa de acúmulo preditos pela RNN LSTM para o futuro, utilizando valores de NDVI, para cada uma das subáreas de estudo

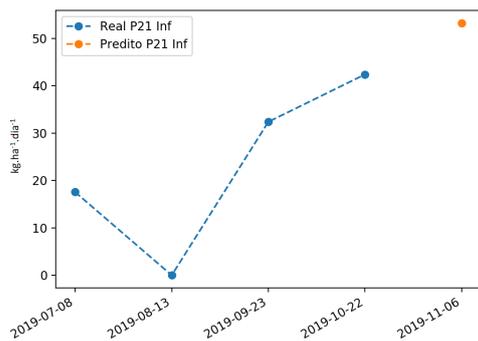
(a) Gráfico de linha apresentando valores reais e predito do NDVI para a subárea, chamada de Potreiro 20 Infestado



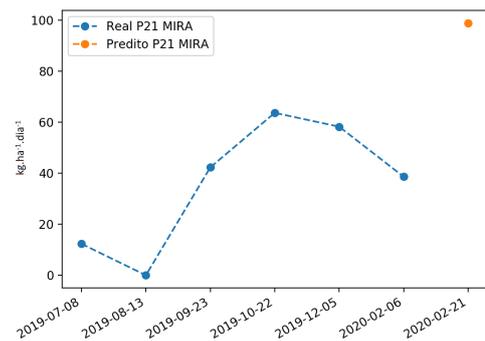
(b) Gráfico de linha apresentando valores reais e predito do NDVI para a subárea, chamada de Potreiro 20 MIRAPASTO



(c) Gráfico de linha apresentando valores reais e predito do NDVI para a subárea, chamada de Potreiro 21 Infestado



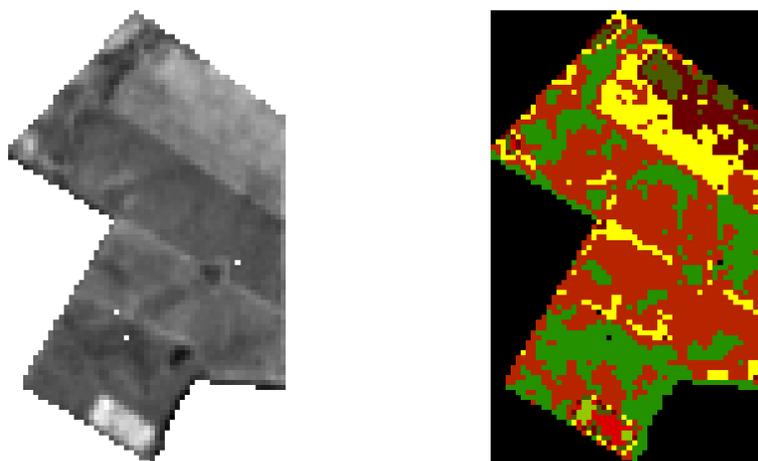
(d) Gráfico de linha apresentando valores reais e predito do NDVI para a subárea, chamada de Potreiro 21 MIRAPASTO



Fonte: Autor (2021)

sklearn é possível utilizar um algoritmo de *clustering*, chamado K-means, onde através de operação estatística, especificamente a média, agrupa os pixels mais similares, em k grupos, onde o número k é definido pelo usuário, foi testado diferentes valores de k, sendo o valor 8 apresentando um resultado mais satisfatório. A figura 70 mostra na sua esquerda a imagem original e na sua direita o resultado da classificação, cada uma das cores representa um grupo diferente e a cor preta representa a área em que não faz parte do estudo.

Figura 70 – Imagem original e o resultado do algoritmo de classificação K-means



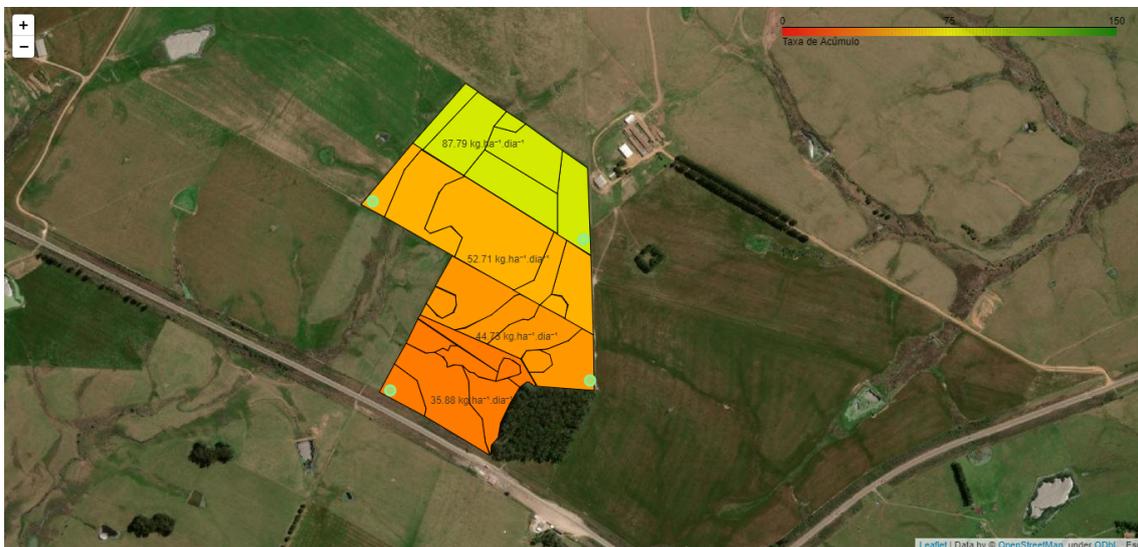
Fonte: Autor (2021)

5.6 Visualização com Mapa Temático e GIF animada

A figura 71 apresenta o mapa temático, criado com o objetivo de tornar mais intuitivo para o usuário a visualização dos resultados, através das cores, além de gerar alertas sobre o ajuste da carga animal presente em cada um dos potreiros. O alerta é representado pelo círculo no limite inferior direito (para os potreiros P20 Infestado e P21 Infestado) e esquerdo (para os potreiros P20 MIRAPASTO e P21 MIRAPASTO). Caso o usuário clique no círculo, é apresentada uma mensagem, variando de acordo com a cor do círculo, para a cor seja verde, a mensagem será de inserir animais na área, para a cor amarela, será de manter a quantidade de animais e vermelha será de retirar animais. A coloração do mapa é feita através do intervalo de $[0, +150]$, recebendo a cor vermelha no

caso do valor ser mínimo extremo (0) e verde caso o valor seja máximo extremo (150). A coloração do círculo de alerta será definida a partir da soma entre a MF do mês anterior e a taxa de acúmulo predita, caso esse valor seja maior do que 12% da carga animal, a cor será verde, caso o resultado seja menor do que 12% da carga animal, a cor será vermelha e se o resultado for igual, a cor será amarela. No mapa apresentado (figura 71) mostra que para todos os poteiros a soma entre MF e taxa de acúmulo se manteve superior aos 12% da carga animal.

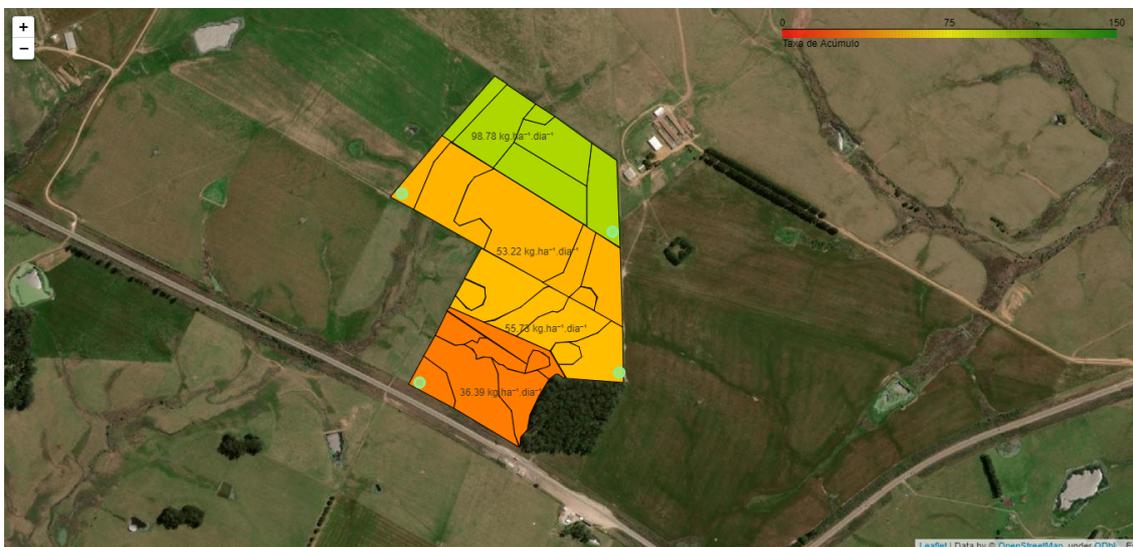
Figura 71 – Mapa temático com valores de taxa de acúmulo preditos pela RNN LSTM para o futuro, utilizando dados históricos de pastagem, para cada uma das subáreas de estudo apresentadas no mapa de vegetação



Fonte: Autor (2021)

A figura 72 apresenta o resultado do modelo de predição para o futuro, utilizando valores de NDVI, na forma de mapa de vegetação. Vale ressaltar que os parâmetros e o método de criação do mapa com valores de taxa de acúmulo preditos utilizando valores de MF (figura 71) e valores de taxa de acúmulo preditos utilizando valores de NDVI (figura 72) foi o mesmo, ou seja, o intervalo utilizado para coloração do mapa e dos círculos de alerta, foram iguais.

Figura 72 – Mapa temático com valores de taxa de acúmulo previstos pela RNN LSTM para o futuro, utilizando NDVI, para cada uma das subáreas de estudo apresentadas no mapa de vegetação



Fonte: Autor (2021)

Com o objetivo de visualizar o comportamento temporal do NDVI calculado para a área total, foi criada uma GIF animada. Através da figura 73 é possível analisar as seis primeiras sequências de imagens obtidas na animação, podendo assim, analisar o comportamento ao longo de seis amostras.

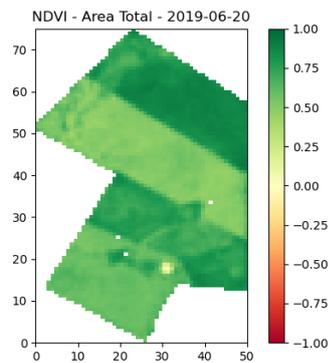
5.7 Considerações Finais

A partir dos resultados é possível verificar que a acurácia do módulo de predição foi superior à obtida por Schulte (2019) e se manteve com percentuais satisfatórios na maioria dos casos. Também foi possível validar a biblioteca PyETo que realiza o cálculo de evapotranspiração através do método de Penman-Monteith (1998). Os valores de evapotranspiração estão significativamente próximos dos valores calculados manualmente na Embrapa - Pecuária Sul, o que confirmou a viabilidade e relevância de automatizar essa etapa do processo, tendo em vista a sua complexidade e necessidade de interpretação de um número amplo de variáveis.

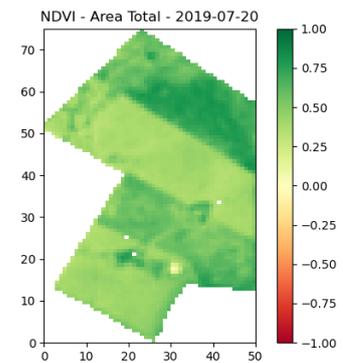
Também foi possível mensurar o erro dos modelos de previsão meteorológica. Os valores de previsão e os valores reais obtidos através da estação automática podem ser vistos na tabela 4. Além disso, foi possível validar o módulo de coleta automatizada de dados meteorológicos, de previsão meteorológica e do satélite Sentinel-2.

Figura 73 – Seis primeiras imagens presentes na GIF animada criada para visualização do comportamento do NDVI ao longo do tempo

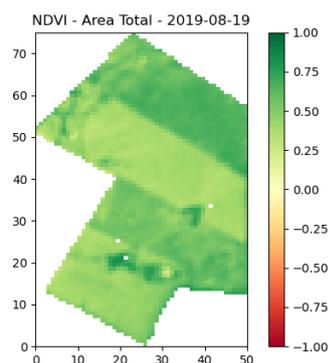
(a) Sequência 1



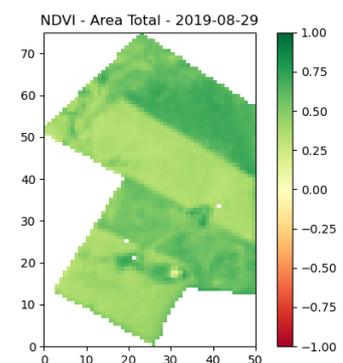
(b) Sequência 2



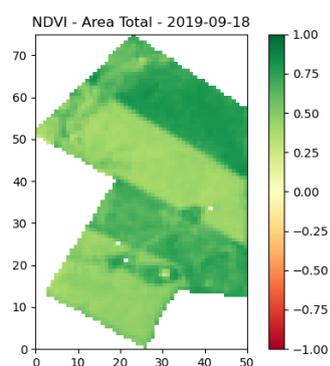
(c) Sequência 3



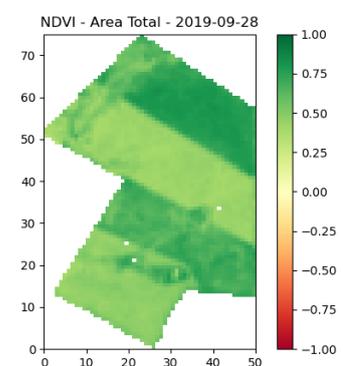
(d) Sequência 4



(e) Sequência 5



(f) Sequência 6



Fonte: Autor (2021)

Tabela 4 – Valores de previsão meteorológica e dados reais da estação automática do INMET para o ano de 2019 (°C)

Dia	Estação automática		NOAA		CPTEC	
	Tmax	Tmin	Tmax	Tmin	Tmax	Tmin
15/11	19	17,6	23,88	18,27	24	10
16/11	20,3	18,5	19,09	13,23	27	11
17/11	23,4	21,6	21,5	15,45	28	13
18/11	25	23,5	23,09	16,85	29	13
19/11	24,8	23,1	20,07	13,95	30	14
20/11	25,2	23,7	19,75	15,62	33	16

Fonte: Autor (2021)

Através dos valores apresentados (tabela 4) é possível notar que o modelo global do NOAA apresenta valores mais próximos dos reais coletados pela estação automática. Também é possível notar que a diferença entre os valores de previsão x real são maiores em relação a temperatura mínima comparado a diferença em relação a temperatura máxima.

Em virtude dos experimentos realizados no módulo de predição, constatou-se que havia a oportunidade de implementar melhorias para tornar o modelo apresentasse resultados mais satisfatórios. A acurácia da predição do modelo original, com os refinamentos feitos nas entradas do modelo, apresentou melhoras significativas, porém, observou-se a necessidade de alterar a variável de saída do modelo para taxa de acúmulo com o intuito de aprimorar a generalização do modelo, ficando assim mais fácil para o usuário final aplicar diretamente a taxa de acúmulo predito e a MF atual de fora da gaiola para assim poder realizar o ajuste na taxa de lotação da área.

O uso do NDVI representa, a médio prazo, a oportunidade de adquirir uma maior quantidade de amostras sobre uma área de interesse (unidade de manejo de pastejo) para utilizar no modelo de predição, caso seja realizado o cálculo do NDVI para o ponto exato onde está presente a gaiola e não só o NDVI médio da subárea, com isso, substituindo tanto o valor de MF de fora da gaiola quando o MF de dentro da gaiola.

A visualização dos resultados através de mapas se torna muito mais agradável e intuitiva, além de dar informações que auxiliem o produtor na tomada de decisão, ou seja, informando-o se há necessidade de retirar, manter ou inserir animais na área naquele período em estudo. A GIF animada auxilia o produtor na visualização do comportamento

temporal da pastagem, visto que o NDVI vai ser diretamente influenciado se a pastagem estiver seca, estiver verde, se houver pouca cobertura do solo, etc.

A realização da divisão da área de forma automatizada se mostrou insatisfatória, visto que classifica áreas similares muito espaçadas geograficamente e de maneira muito heterogênea, com isso resultaria em muitas subdivisões na área.

6 DISCUSSÃO

A automatização da coleta de dados se mostra uma funcionalidade muito importante do sistema, visto que não demanda muito esforço do usuário e diminui a interação com o sistema, além da redução no tempo de aquisição.

A adaptação da inserção de dados de pastagens referentes a outras áreas, ou seja, a utilização de diferentes conjuntos de dados de pastagem para treinamento e teste, mostrou que o modelo de predição mantém o comportamento esperado, já que obteve melhores resultados (treinamento estratificado por tratamento (T)) da mesma maneira que nos testes utilizando somente o conjunto de dados A. Além disso, a união entre os conjuntos de dados A e B, mostrou que o modelo manteve acurácia satisfatória, apesar das diferentes fontes de dados. Havendo uma relativa melhora devido ao aumento no número de amostras de treinamento. A comprovação de que o modelo apresenta melhores resultados quando treinado com tratamentos similares, se dá pelo fato de que quando treinado com conjunto de dados Infestado ou MIRAPASTO, e testado com respectivos poteiros que recebem tratamento similar, a acurácia do modelo se mantém satisfatória.

Através dos experimentos realizados no módulo de predição, foram observados alguns fatores importantes que ainda devem ser tratados no sistema para que atinja a melhor acurácia possível, como por exemplo, como lidar com valores negativos preditos pelo modelo, além, da modelagem do conjunto de dados de treinamento, verificando quais variáveis causam mais impacto no modelo de predição. Para a realização da predição futura, atualmente feita utilizando o valor de taxa de acúmulo do mês correspondente, porém, no ano anterior, pode acarretar numa redução na acurácia do modelo, visto que mesmo que sejam meses iguais, pode ocorrer uma diferença climática de um ano para o outro, impactando significativamente no valor de taxa de acúmulo real.

Atualmente o modelo de predição é dito como *uni-variate*, ou seja, realiza a predição somente de uma variável, chamada de variável de saída. Como existem relação de dependência entre MF e taxa de acúmulo, caso não se tenha o valor de de MF de fora da gaiola, por exemplo, seria importante adaptar ao modelo a capacidade de predição, chamada *multivariate*, ou seja, realizar a predição, ou aceitar mais de uma variável como saída do modelo.

Para realizar o refinamento do modelo de predição foi feita uma análise do conjunto de dados de treinamento e teste, com a utilização da data de maneira integral no modelo original, foi buscada uma alternativa em como representar de maneira mais clara

para a RNN LSTM a informação das diferentes estações do ano, já que a data causava muita variabilidade no conjunto de dados, já que dificilmente havia coleta de amostras de pastagem em dias iguais de anos diferentes. Após isso, observou-se que a necessidade de incorporar ao modelo de maneira explícita, a informação de taxa de acúmulo, mesmo que anteriormente existissem todas as informações necessárias para o cálculo de taxa de acúmulo, ou seja, MF de fora e dentro da gaiola e também a quantidade de dias entre uma coleta e outra, para que a RNN LSTM obtivesse a informação do quanto a pastagem cresce em condições meteorológicas específicas. Com a mesma linha de pensamento observou-se a necessidade de explicitar a informação de carga animal no conjunto de dados, para que a RNN LSTM coletasse a informação do quanto a carga animal impactava no valor de MF de fora da gaiola, não atribuindo somente as condições climáticas.

Para estimar a confiabilidade do método indireto foi feita a regressão linear entre os valores de NDVI e reais de corte da pastagem. Para expressar a correlação entre as variáveis, foi utilizado o teste que mede a relação estatística, chamado coeficiente de correlação de Pearson, que varia de -1 a 1. Foram feitos os testes com os valores de cada ponto onde estão localizadas as gaiolas e não os valores médios de cada subárea. Importante ressaltar que como as datas de corte e das imagens não são iguais, foram utilizadas as imagens obtidas na data mais próxima do corte. Caso o valor do coeficiente seja maior que 0, representa que as variáveis estão diretamente relacionadas, ou seja, quando uma aumenta a outra também aumenta e quanto mais próximo de 1 maior é a relação entre elas. Se o coeficiente for menor que 0, significa que as variáveis são inversamente relacionadas, ou seja, quando uma diminui a outra aumenta ou vice-versa e quanto mais próximo de -1 maior é a relação entre elas. Já no caso do coeficiente ser 0, significa que as variáveis não estão correlacionadas.

Através da tabela 5 é possível verificar os valores do coeficiente para cada uma das gaiolas posicionadas nos poteiros. A tabela 5(a) apresenta os valores para o poteiro 20 Infestado, nesse cenário os valores de NDVI e de corte estão diretamente relacionados na Gaiola 1, inversamente relacionados na Gaiola 4 e não estão relacionados na Gaiola 5. Para o tratamento MIRAPASTO é possível verificar, através da tabela 5(b), que na Gaiola 1 os valores não estão relacionados e nas Gaiolas 3 e 5 estão diretamente relacionados. Os valores apresentados na tabela 5(c) são referentes ao poteiro 21 Infestado e é possível notar que na Gaiola 1 a relação é direta, na Gaiola 3 não há relação e na Gaiola 5 a relação é inversa. Para o tratamento MIRAPASTO a relação é inversa em todas as gaiolas, como mostra a tabela 5(d). A partir disso pode-se constatar que na maioria dos casos o NDVI

apresentou correlação com o valor real de pastagem obtido através do corte, seja uma relação direta ou inversa.

Tabela 5 – Correlação entre valores de NDVI e reais para cada uma das gaiolas utilizadas nos poteiros 20 e 21

(a) Valores referentes ao poteiro 20 Infestado

Gaiola 1	Gaiola 4	Gaiola 5
0.88	-0.28	0

(b) Valores referentes ao poteiro 20 MIRAPASTO

Gaiola 1	Gaiola 3	Gaiola 5
-0.06	0.28	0.64

(c) Valores referentes ao poteiro 21 Infestado

Gaiola 1	Gaiola 3	Gaiola 5
0.35	0.05	-0.14

(d) Valores referentes ao poteiro 21 MIRAPASTO

Gaiola 1	Gaiola 2	Gaiola 4
-0.31	-0.47	-0.15

Fonte: Autor (2021)

A apresentação dos resultados de maneira gráfica, através da utilização de mapas de vegetação e a GIF, torna o sistema mais amigável ao usuário e a compreensão dos resultados fica mais intuitiva, já que através da coloração do mapa o usuário já pode visualizar o estado da pastagem, mesmo que não faça uma análise mais profunda. A visualização gráfica também possibilitou notificar o usuário sobre a taxa de lotação para cada uma das áreas, com isso, auxilia efetivamente na tomada de decisão. A utilização de uma GIF animada para o usuário poder ver o comportamento temporal da pastagem além de auxiliar a análise da pastagem em diferentes épocas e estações do ano, expõe os resultados do método indireto desenvolvido.

Existem dois sistemas comerciais chamados de Farmote System¹ e Pasture.io². Farmote System faz uma combinação entre imagens multiespectrais obtidas por satélite e sensores posicionados na área da pastagem, com isso, é possível obter informações sobre a pastagem, como crescimento. Os resultados obtidos são disponibilizados em uma plataforma web compatível com dispositivos móveis. Já o sistema Pasture.io não é focado somente no crescimento da pastagem, também oferece informações sobre a oferta de forragem em um determinado local, unindo dados climáticos e imagens de satélite. A tabela 6 apresenta uma comparação dos sistemas, exibindo funcionalidades e aspectos presentes em cada um dos sistemas, devido a não ter acesso a algumas informações, nem todos os aspectos puderam ser comparados e estão representados por um traço na tabela.

¹<https://farmote.com/>

²<https://pasture.io/>

Tabela 6 – Funcionalidades e aspectos dos sistemas

	TouceiraTech Pasture.io		Farmote System
Pastagem	Nativa	-	-
Método indireto	Sim	Sim	Sim
Sensores	Não	Não	Sim
Mapa de vegetação	Sim	Sim	Sim
GIF animada	Sim	Não	Não
Dados meteorológicos	Sim	Sim	Sim
Web Service/App	Não	Sim	Sim
Dados históricos	Sim	Sim	-
Imagens de satélite	Sim	Sim	Sim

Fonte: Autor (2021)

A existência desses sistemas comerciais enfatiza a importância do TouceiraTech para o contexto de pecuária e também expõe, a partir das semelhanças entre as funcionalidades dos sistemas, que o sistema atende a maioria dos requisitos necessários para que possa auxiliar o pecuarista na tomada de decisão.

6.1 Considerações Finais

Considerando os resultados apresentados e as discussões, é possível fazer um contraste com as questões de pesquisa e os objetivos do trabalho.

Em relação à primeira questão que se buscava responder, é notável que ficou comprovado que é possível propor, implementar e avaliar um protótipo de prova de conceito de FMIS partindo de princípios de pecuária de precisão, capaz de coletar, armazenar, processar e visualizar dados das pastagem - massa de forragem (MF) instantânea - necessários para o ajuste de taxa de lotação. Sobre a segunda questão, encontrou-se limitações em acessar bases de dados abertas com previsão superior a 15 dias, embora mais recentemente tenha sido identificada uma possível solução por meio do serviço OpenWeather³ que, em tese, fornece previsões meteorológicas para um período de 30 dias subsequente. A terceira questão também pôde ser respondida afirmativamente, de forma preliminar, tendo em vista os resultados promissores obtidos com as funcionalidade relacionadas à coleta automática de imagens do Sentinel-2 com 10 metros de resolução espacial e a au-

³<https://openweathermap.org/api/forecast30>

tomatização do cálculo do NDVI para uma polígono correspondente à área de manejo.

Sobre o objetivo geral, os resultados comprovam que houve evolução no processo de desenvolvimento de um protótipo de prova de conceito, no qual foi aprimorado o método de predição originalmente desenvolvido por Schulte (2019), não apenas com a inclusão da possibilidade de uso de amostras indiretas a partir de imagens de satélite, mas também pelo ajuste das variáveis de entrada e do treinamento estratificado. Assim como foi implementada e validada a funcionalidade que permitiu incorporar ao modelo de banco de dados espacial proposto os dados de previsão meteorológica referentes às regiões de interesse, acompanhada pelas funcionalidades de apresentação visual dos resultados da predição, inclusive por meio de mapas da vegetação usando animações no formato GIF. Neste contexto, não apenas os objetivos específicos essenciais foram integralmente atingidos, como também um dos objetivos desejáveis foi atendido, relacionado ao refinamento do modelo RNN LSTM. Por outro lado, ainda não foi possível avançar significativamente na estimativa da confiabilidade de métodos indiretos já existentes, tampouco na coleta de dados de sensores ultrassom para estimar altura de pastagens.

7 CONCLUSÃO

De forma geral, os estudos mostraram a importância de um método de predição de disponibilidade de forragem e, com o desenvolvimento do trabalho, é possível indicar um grande potencial de uso de um sistema baseado em FMIS no auxílio na tomada de decisão de uma propriedade cuja atividade principal seja a pecuária.

Com a implementação da coleta automatizada de dados meteorológicos, um dos objetivos essenciais do trabalho, é possível notar a significativa relevância para o desenvolvimento do sistema baseado em FMIS, visto que, a inserção manual de dados é bastante custosa em relação ao tempo demandado para tal tarefa.

Os avanços nos ajustes no modelo, com destaque para o treinamento estratificado por tratamento e para a inclusão de outras variáveis relevantes para a predição, foi possível melhorar de forma expressiva a acurácia em comparação com os resultados do modelo original (SCHULTE, 2019, p. 70). Mais especificamente, enquanto que no modelo original, por exemplo, a raiz do erro quadrático médio era próxima a $2000 \text{ kg}\cdot\text{ha}^{-1}$ em uma das áreas de manejo com tratamento equivalente ao natural (P20 Infestado), no melhor caso do modelo ajustado com treinamento estratificado por tratamento (T) este indicador de acurácia reduziu para cerca de $600 \text{ kg}\cdot\text{ha}^{-1}$, ou seja, em cerca de 70%. Além disso, o modelo ajustado foi posteriormente alterado para a predição da taxa de acúmulo, o que permite uma melhor comparação de resultados entre áreas de estudo.

Com o desenvolvimento do método indireto foi possível constatar a capacidade de gerar mais amostras de dados referentes à pastagem, assim, aumentar o número de entradas para a RNN LSTM. Como a quantidade de entradas é diretamente proporcional a acurácia do modelo, por esse fato justifica-se a criação do método indireto, além da redução aos danos na pastagem e diminuição no tempo de aquisição das informações.

Também constatou-se, a partir dos resultados gerados, que a predição no cenário onde a área está infestada, na maioria dos casos, apresenta uma subestimação da disponibilidade de forragem. Já para o cenário onde há o controle do capim-annoni, na maioria dos casos, há uma superestimação da disponibilidade de massa de forragem.

Embora os resultados sejam preliminares, foi possível identificar que o comportamento do NDVI calculado está de acordo com as épocas em que a pastagem fica mais seca, ou seja, nessas épocas o valor se aproxima de 0 e em épocas em que a pastagem fica mais verde se aproxima de 1.

7.1 Trabalhos Futuros

Através do desenvolvimento do sistema, foi possível elencar funcionalidades com o objetivo de aprimorar e tornar o sistema cada vez mais robusto para que seja possível auxiliar o produtor da maneira mais eficaz e eficiente possível.

As funcionalidades elencadas e possíveis abordagens de implementação para o módulo de fonte de dados, especificamente o método indireto de coleta de dados da pastagem e também os dados de previsão meteorológica, são apresentadas a seguir:

- Criar um método mais eficiente de classificação de imagens, para a divisão automatizada de áreas. Uma das potenciais alternativas para essa tarefa é a utilização de RNN, realizando o treinamento de acordo com os pixels das imagens;
- A automação na criação dos arquivos shapefile para cada uma das áreas de estudo. Existem alguns pacotes em Python que possibilitam a criação, edição e manipulação de arquivos shapefile, como por exemplo, shapely, fiona, shapelib, osgeo, dentre outros;
- Armazenar o valor do NDVI para um ponto em específico e não o NDVI médio da subárea. Isso é possível através da coordenada geográfica onde está posicionada a gaiola de corte, assim é possível obter o valor de NDVI dos pixels que representam apenas aquele ponto;
- A criação de um *web-service* onde o usuário vai informar um arquivo de pontos e coordenadas e com isso será possível delimitar a área de estudo, obter as imagens de satélite, realizar o recorte, calcular o ndvi, passar as entradas para o modelo de predição e exibir os resultados;
- Obter dados de previsão meteorológica a partir de uma fonte que forneça em um intervalo maior que 15 dias, também comparar os modelos estudados até então com outros, como o European Center for Medium-Range Weather Forecast (ECMWF), que realiza previsões para todo o globo terrestre em uma amostragem aproximadamente 25x25km ou 0,25°.

As funcionalidades elencadas para o módulo de predição e possíveis abordagens de implementação são apresentadas a seguir:

- Adaptar o modelo de predição para ser *multivariate*, gerando como resultado mais de uma variável predita;
- Realizar análise de quais variáveis são mais impactantes ao modelo, e também inse-

rir informações sobre a fertilidade e o tipo de solo, tipo de pasto, resíduo de pastejo, dentre outros;

- Realizar a suavização do valor predito, ou seja, caso seja negativo, ser igualado a zero nas estações em que não deve ocorrer taxa de acúmulo negativa;
- Adaptar o modelo de predição futura para que, na última entrada, leve em conta o valor médio do histórico de taxas de acúmulo do mês de interesse;
- Explorar alternativas que visem o melhor uso dos conjuntos de dados, ou seja, a melhor divisão do conjunto em treinamento e teste.

REFERÊNCIAS

- ACADEMY, D. S. **Deep Learning Book**. 2021. Disponível em: <https://www.deeplearningbook.com.br/>. Acesso em: 04 mai. 2021.
- 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**, IEEE, v. 10, n. 7, p. 3254–3264, 2016.
- ALLEN, R. G. et al. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. **Fao, Rome**, v. 300, n. 9, p. D05109, 1998.
- ALVES, E.; CAVALCANTI, W. Interveniências sobre a temática físico-cibernética em tempos de indústria 4.0. In: **Anais do VI Simpósio de Engenharia de Produção - SIMEP 2018**. [S.l.: s.n.], 2018.
- AMO, S. D. Técnicas de mineração de dados. **Jornada de Atualização em Informatica**, 2004.
- ANDRADE, C. M. S. d. et al. Padrões de desempenho e produtividade animal para recria-engorda de bovinos de corte no acre. **Embrapa Acre-Documentos (INFOTECA-E)**, Rio Branco, AC: Embrapa Acre, 2005., 2005.
- AREND, M. R. Uso de rotinas de apoio à decisão à implantação de sistemas orgânicos de produção agropecuária no município de arroio do meio-rs. **Salão de Iniciação Científica (22.: 2010 out. 18-22: Porto Alegre, RS). Livro de resumos. Porto Alegre: UFRGS, 2010.**, 2010.
- ARRUDA, D. S. R. et al. Métodos de avaliação de massa de forragem em pastagens de capim-estrela submetidas a intensidades de pastejo. **Ciência Rural**, SciELO Brasil, v. 41, n. 11, p. 2004–2009, 2011.
- BAMBINI, M. D.; ROMANI, L. A. S. Mercado de agrotic e transferência de tecnologia. **Embrapa Informática Agropecuária-Capítulo em livro científico (ALICE)**, In: MASSRUHÁ, SMFS; LEITE, MA de A.; LUCHIARI JUNIOR, A.; ROMANI, LAS (Ed . . . , 2014.
- BATISTA, G. E. d. A. P. **Pré-processamento de dados em aprendizado de máquina supervisionado**. Tese (Doutorado) — Universidade de São Paulo, 2003.
- BCRC. **Grazing Management**. 2019. Disponível em: <https://www.beefresearch.ca/research-topic.cfm/grazing-management-48>. Acesso em: 21 nov. 2019.
- BELWARD, A. S.; SKØIEN, J. O. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 103, p. 115–128, 2015.
- BOEHLJE, M.; EIDMAN, V.; ERRINGTON, A. Farm management. **Agricultural Systems**, 1985.

BOSCARIOLI, C. et al. Uma reflexão sobre banco de dados orientados a objetos. In: **Congresso de Tecnologias para Gestão de Dados e Metadados do Cone Sul, Paraná, Brasil**. [S.l.: s.n.], 2006.

BRAGA, L. P. V. B. **Introdução à Mineração de Dados-2a edição: Edição ampliada e revisada**. [S.l.]: Editora E-papers, 2005.

CAMILO, C. O.; SILVA, J. C. d. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás (UFG)**, p. 1–29, 2009.

CARVALHO, R. Método de determinação da disponibilidade de forragem. **Ciência et Praxis**, v. 1, n. 02, p. 7–10, 2008.

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

CAUDURO, G. F. et al. Comparação de métodos de medida indireta de massa de forragem em pasto de azevém anual (*lolium multiflorum lam.*). **Ciência Rural**, v. 36, p. 1617 – 1623, 2006.

CAVALHEIRO, D. et al. A tecnologia da informação no agronegócio: uma revisão bibliográfica. In: **XVIII Mostra de Iniciação Científica, Pós-Graduação, Pesquisa e Extensão**. [S.l.: s.n.], 2018. p. 1–16.

CEOLIN, A. C. et al. Sistemas de informação sob a perspectiva de custos na gestão da pecuária de corte gaúcha. **Custos e @gronegócio on-line**, v. 4, 2008.

CHAN, S. H. et al. Decision support system (dss) use and decision performance: Dss motivation and its antecedents. **Information & Management**, Elsevier, v. 54, n. 7, p. 934–947, 2017.

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.

COELHO, A. P.; ROSALEN, D. L.; FARIA, R. T. d. Vegetation indices in the prediction of biomass and grain yield of white oat under irrigation levels. **Pesquisa Agropecuária Tropical**, SciELO Brasil, v. 48, n. 2, p. 109–117, 2018.

CÓRDOVA, U. d. A. et al. Validação da tecnologia de melhoramento de pastagens naturais no planalto sul de santa catarina. **Revista de Ciências Agroveterinárias**, v. 11, n. 1, p. 54–62, 2014.

CÓSER, A. C.; MARTINS, C. E.; DERESZ, F. Metodologias para estimativa da produção de forragem em pastagem de capim-elefante. **Embrapa Gado de Leite-Circular Técnica (INFOTECA-E)**, Juiz de Fora: Embrapa Gado de Leite, 2002., 2002.

COSTA, J. L. B. **Níveis de intensificação em pastagem natural : desempenho animal e produção vegetal**. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul, 2019.

- COSTA, N. d. L. et al. Estimativa do rendimento potencial de azevém anual (*Iolium multiflorum lam.*) através de modelos matemáticos. **Embrapa Roraima-Artigo em periódico indexado (ALICE)**, PUBVET, Londrina, V. 5, N. 9, Ed. 156, Art. 1051, 2011., 2011.
- CRUZ, P. G. d. et al. Modelos empíricos para estimar o acúmulo de matéria seca de capim-marandu com variáveis agrometeorológicas. **Pesquisa Agropecuária Brasileira**, v. 46, n. 7, p. 675–681, 2011.
- CUNHA, W. F. d. **Métodos indiretos para estimativa de massa de forragem em pastagens de *Cynodon spp.*** Dissertação (Mestrado) — Universidade de São Paulo, 2002.
- DAMASCENO, V. G. et al. Índice de vegetação por diferença normalizada (ndvi) para estimativa do potencial produtivo de forrageiras exóticas. In: **V Congresso Internacional das Ciências Agrárias – V COINTER PDVAgro**. [S.l.: s.n.], 2020. ISSN 2526-7701.
- DASORIYA, R. A review of big data analytics over cloud. In: IEEE. **2017 IEEE international conference on consumer electronics-Asia (ICCE-Asia)**. [S.l.], 2017. p. 1–6.
- EDVAN, R. L. et al. Methods for estimating forage mass in pastures in a tropical climate. **Revista de Ciências Agrárias**, scielopt, v. 39, p. 36 – 45, 03 2016. ISSN 0871-018X.
- ELMASRI, R. et al. Sistemas de banco de dados. Pearson Addison Wesley São Paulo, 2005.
- ESTRADA, L. H. C.; JÚNIOR, D. N.; REGAZZI, A. J. 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.
- FERNANDES, E. N.; PESSOA, M. C. Y. Sistemas de apoio à decisão na gestão ambiental de propriedades leiteiras. **Embrapa Meio Ambiente-Capítulo em livro científico (ALICE)**, In: CARVALHO, L. de A.; ZOCCAL, R.; MARTINS, P. do C.; ARCURI, PB; MOREIRA . . . , 2005.
- FERNANDES, H. J. et al. Ganho de peso, conversão alimentar, ingestão diária de nutrientes e digestibilidade de garrotes não-castrados de três grupos genéticos em recria e terminação. **Revista Brasileira de Zootecnia**, v. 33, p. 2403 – 2411, 2004.
- FJERMESTAD, J. Handbook on decision support systems 1: Basic themes. **Interfaces**, Institute for Operations Research and the Management Sciences, v. 39, n. 3, p. 291, 2009.
- FLORENZANO, T. G. **Iniciação em sensoriamento remoto**. [S.l.]: Oficina de textos, 2007.
- FONSECA, E. L. d.; FORMAGGIO, A. R.; PONZONI, F. J. Estimativa da disponibilidade de forragem do bioma campos sulinos a partir de dados radiométricos orbitais: parametrização do submodelo espectral. **Ciência Rural**, Universidade Federal de Santa Maria, v. 37, n. 6, p. 1668–1674, 2007.

- FREITAS, V. A. L. d.; BASSO, L. d. O.; FERRI, R. Sensoriamento remoto e geoprocessamento. **Universidade Tecnológica Federal do Paraná**, 2015.
- GALVÃO, N. D.; MARIN, H. d. F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, Universidade Federal de São Paulo, v. 22, n. 5, p. 686–690, 2009.
- GARCIA, S. **Geração agPastto: Por que e Como calcular a massa de forragem?** 2017. Disponível em: <https://agrocereasmultimix.com.br/blog/como-calculer-massa-de-forragem/>. Acesso em: 25 de outubro de 2019.
- GARDNER, A. **Técnicas de pesquisa em pastagens e aplicabilidade de resultados em sistemas de produção**. [S.l.]: IICA Biblioteca Venezuela, 1986.
- HEAIVILIN, H. **Big Data in Science & Problem Solving**. 2018. Disponível em: <https://international-soil-radiocarbon-database.github.io/SOC-Hub/2018/06/27/Bigdata-Ag/>. Acesso em: 28 de outubro de 2019.
- HODGSON, J. Nomenclature and definitions in grazing studies. **Grass and Forage Science**, Blackwell Scientific Publications, v. 34, n. 1, p. 11 – 18, 3 1979.
- HODGSON, J. et al. **Sward measurement handbook**. [S.l.], 1981.
- IBGE, I. B. d. G. e. E. Produção da pecuária municipal 2019. 2020.
- ISMAIL, A. A.; WOOD, T.; BRAVO, H. C. Improving long-horizon forecasts with expectation-biased lstm networks. **arXiv preprint arXiv:1804.06776**, 2018.
- JAYARAMAN, 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.
- JHURIA, M.; KUMAR, A.; BORSE, R. Image processing for smart farming: Detection of disease and fruit grading. In: IEEE. **2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)**. [S.l.], 2013. p. 521–526.
- JONES, R. **Field and laboratory methods for grassland and animal production research**. [S.l.]: CABI, 2000.
- JUNIOR, W. T. L. Big data, jornalismo computacional e data journalism: estrutura, pensamento e prática profissional na web de dados. **Estudos em Comunicação**, v. 12, p. 207–222, 2012.
- KAMILARIS, A.; KARTAKOULLIS, A.; PRENAFETA-BOLDÚ, F. A review on the practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, Elsevier, v. 143, p. 23–37, 2017.
- KICHEL, A. N.; MIRANDA, C. H. B.; ZIMMER, A. H. Degradação de pastagens e produção de bovinos de corte com a integração agricultura x pecuária. **Simpósio de produção de gado de Corte**, 1999.
- LEE, H. D. **Seleção e construção de features relevantes para o aprendizado de máquina**. Tese (Doutorado) — Universidade de São Paulo, 2000.

- LOPES, M. A. **Informática aplicada á bovinocultura**. Jaboticabal: FUNEP, 1997.
- MACHADO, J. G. d. C. F.; NANTES, J. F. D. Tecnologia de informação em organizações rurais: um estudo na pecuária de corte. **Informações Econômicas, São Paulo**, v. 38, n. 10, p. 45–56, 2008.
- MACHADO, J. G. de C. F. **A adoção da identificação eletrônica de animais na gestão do empreendimento rural: um estudo multicaso na pecuária de corte**. 129 p. Dissertação (Mestrado) — Universidade Federal de São Carlos, São Carlos, 2002.
- MACHADO, J. G. de C. F.; NANTES, J. F. D. Adoção da tecnologia da informação em organizações rurais: o caso da pecuária de corte. **Gest. Prod.**, v. 18, n. 3, p. 555–570, 2011.
- MACHADO, L. A. Z.; KICHEL, A. N. **Ajuste de lotação no manejo de pastagens**. [S.l.]: Embrapa Agropecuária Oeste Dourados, 2004.
- MASSRUHÁ, S. M. F. S. et al. Tecnologias da informação e comunicação e suas relações com a agricultura. **Embrapa Informática Agropecuária-Livro científico (ALICE)**, Brasília, DF: Embrapa, 2014., 2014.
- MASSRUHÁ, S. M. F. S.; LEITE, M. A. 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 ... , 2014.
- MEDEIROS, S. R. de et al. Ferramentas de pecuária de precisão voltadas à nutrição de bovinos de corte. In: **Simpósio Brasileiro de Pecuária de Precisão Aplicada à Bovinocultura de Corte, Campo Grande, Brasil**. [S.l.: s.n.], 2014.
- MELADO, J. Pastagem ecológica e serviços ambientais da pecuária sustentável. **Revista de Política Agrícola**, v. 16, n. 3, p. 113–118, 2007.
- MELL, P.; GRANCE, T. The nist definition of cloud computing. Computer Security Division, Information Technology Laboratory, National ... , 2011.
- MENESES, P. R.; ALMEIDA, T. d. Introdução ao processamento de imagens de sensoriamento remoto. **Universidade de Brasília, Brasília**, 2012.
- MIRKIN, B. Data analysis, mathematical statistics, machine learning, data mining: Similarities and differences. In: IEEE. **2011 International Conference on Advanced Computer Science and Information Systems**. [S.l.], 2011. p. 1–8.
- MOOJEN, E. L.; MARASCHIN, G. E. Potencial produtivo de uma pastagem nativa do rio grande do sul submetida a níveis de oferta de forragem. **Ciência rural. Santa Maria. Vol. 32, n. 1 (jan./fev. 2002), p. 127-132**, 2002.
- MOURA, G. B. d. et al. **Redes neurais recorrentes para a classificação de estruturas retóricas**. Dissertação (Mestrado) — Universidade Estadual de Maringá., 2018.
- MURAKAMI, E. et al. An infrastructure for the development of distributed service-oriented information systems for precision agriculture. **Computers and Electronics in agriculture**, Elsevier, v. 58, n. 1, p. 37–48, 2007.

NABINGER, C. Eficiência do uso de pastagens: disponibilidade e perdas de forragem. **Simpósio sobre manejo da pastagem**, v. 14, p. 213–251, 1997.

NABINGER, C. Manejo e produtividade das pastagens nativas do subtropica brasileiro. In: **I Simpósio de Forrageiras e Produção Animal**. [S.l.: s.n.], 2006. p. 25–75.

NEIVA, F. W. Revisão sistemática da literatura em ciência da computação um guia prático. Universidade Federal de Juiz de Fora, 2016.

NEVES, M. et al. Avaliação de sensor multiespectral em mapeamentos da infestação de pastagens naturais pelo capim-annoni. In: **Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto**. [S.l.]: INPE, 2020. p. 1548–1551. ISBN 978-85-17-00097-3.

PEDREIRA, C. G. S. Avanços metodológicos na avaliação de pastagens. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 39., 2002, Recife. **Anais...** [S.l.]: Sociedade Brasileira de Zootecnia, 2002. p. 100–150.

PIRES, P. et al. Plataformas para a internet das coisas. **Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, 2015.

PRO, G. **User Guide**. 2016. Disponível em: https://www.novel.co.nz/Downloads/GMPro/GM_Pro_Manual_v3-09.pdf. Acesso em: 25 de outubro de 2019.

PRODANOV, C. C.; FREITAS, E. C. de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição**. [S.l.]: Editora Feevale, 2013.

PUNALEKAR, S. M. et al. Application of sentinel-2a data for pasture biomass monitoring using a physically based radiative transfer model. **Remote Sensing of Environment**, Elsevier, v. 218, p. 207–220, 2018.

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.

RAJASEKARAN, T.; ANANDAMURUGAN, S. Challenges and applications of wireless sensor networks in smart farming—a survey. In: **Advances in Big Data and Cloud Computing**. [S.l.]: Springer, 2019. p. 353–361.

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

SAGHEER, A.; KOTB, M. Time series forecasting of petroleum production using deep lstm recurrent networks. **Neurocomputing**, Elsevier, v. 323, p. 203–213, 2019.

SALAMI, P.; AHMADI, H. Review of farm management information systems (fmis). **New York Science Journal**, v. 3, n. 5, p. 87–95, 2010.

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

- SANTAELLA, L. et al. Desvelando a internet das coisas. **Revista GEMInIS**, v. 4, n. 2, p. 19–32, 2013.
- SANTOS, B. et al. Internet das coisas: da teoria à prática. **Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, p. 31, 2016.
- SANTOS, G. **Uma aplicação de redes neurais recorrentes do tipo LSTM à previsão dos preços de curto prazo do mercado de energia elétrica brasileiro**. 76 p. Dissertação (Mestrado) — Escola de Economia de São Paulo, 2019.
- SANTOS, S. A. et al. Metodologia de amostragem para avaliação da qualidade das pastagens nativas consumidas por bovinos no pantanal. **Embrapa Pantanal-Documents (INFOTECA-E)**, Corumbá: Embrapa Pantanal, 2002., 2002.
- SCHULTE, L. G. **Suporte a Decisão em Pastagens: Análise Espaço-temporal e Aprendizado de Máquina para Predição da Disponibilidade de Forragem no Contexto de Smart Farming**. Dissertação (Mestrado) — Universidade Federal do Pampa, 2019.
- SILVEIRA, V. C. P. et al. Qualidade da pastagem nativa obtida por diferentes métodos de amostragem e em diferentes solos na APA do Ibirapuitã, Brasil. **Ciência Rural**, SciELO Brasil, v. 35, n. 3, p. 582–588, 2005.
- SLATER, P. N. Remote sensing: optics and optical systems. **Reading, Mass., Addison-Wesley Publishing Co. (Remote Sensing. No. 1), 1980. 593 p.**, v. 1, 1980.
- SØRENSEN, C. G. et al. Conceptual model of a future farm management information system. **Computers and electronics in agriculture**, Elsevier, v. 72, n. 1, p. 37–47, 2010.
- SOUTO, M. D. et al. Técnicas de aprendizado de máquina para problemas de biologia molecular. **Sociedade Brasileira de Computação**, 2003.
- STERGIOU, C. et al. Secure integration of IoT and cloud computing. **Future Generation Computer Systems**, Elsevier, v. 78, p. 964–975, 2018.
- TANAKA, L. C.; CAMARGO, F. M.; GOTARDO, R. A. Sistema gerenciador de banco de dados: Sgbd exist xml. **Revista Eletrônica de Sistemas de Informação e de Gestão Tecnológica**, v. 2, n. 1, 2012.
- TECNOLOGIA, P. P. e. **Welcome to Python**. 2019. Disponível em: <https://www.polite.com.br/regua-para-manejo-de-pastagem-embrapa-inteirica-pr-1195-168429.htm>. Acesso em: 19 de novembro de 2019.
- THORNTHWAITE, C. W. The water balance. **Publ Clim**, v. 8, p. 1–104, 1957.
- TONATO, F. et al. Desenvolvimento de modelos preditores de acúmulo de forragem em pastagens tropicais. **Pesquisa Agropecuária Brasileira**, v. 45, n. 5, p. 522–529, 2011.
- TONATO, F. et al. Desenvolvimento de modelos preditores de acúmulo de forragem e sustentabilidade de sistemas de produção em pastejo. In: IN: CONGRESSO BRASILEIRO DE ZOOTECNIA, 22., 2012, CUIABÁ. ZOOTEC 2012: ANAIS **Embrapa Amazônia Ocidental-Artigo em anais de congresso (ALICE)**. [S.l.], 2012.

VASSALLI, L. C. Aplicação de redes neurais lstm para a previsão de curto prazo de vazão do rio paraíba do sul. 2018.

VERAS, M. **Cloud Computing: nova arquitetura da TI**. [S.l.]: Brasport, 2012.

VERDI, F. L. et al. Novas arquiteturas de data center para cloud computing. **Minicursos do XXVIII SBRC**, p. 103–152, 2010.

WANG, Y. et al. Probabilistic individual load forecasting using pinball loss guided lstm. **Applied Energy**, Elsevier, v. 235, p. 10–20, 2019.

WHITE, J.; HODGSON, J. **New Zealand pasture and crop science**. [S.l.]: Auckland, N.Z : Oxford University Press, 1999.

WOLFERT, S. et al. Big data in smart farming - a review. **Agricultural Systems**, Elsevier, v. 153, p. 69–80, 2017.

WOLFERT, S.; GOENSE, D.; SØRENSEN, C. A. G. A future internet collaboration platform for safe and healthy food from farm to fork. In: IEEE. **2014 Annual SRII Global Conference**. [S.l.], 2014. p. 266–273.

APÊNDICE A – SCRIPT DE SUMARIZAÇÃO DOS DADOS DE PASTAGEM

```

1 import psycopg2
import numpy as np
3 from datetime import datetime
import matplotlib.pyplot as plt
5 from scipy.stats.mstats import mquantiles

7 now = datetime.now()

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

11
try:
13 #Conecta com o BD
con = psycopg2.connect("host='localhost' port='5432' dbname='db_2019'
user='postgres' password='12345'")
15 #con = psycopg2.connect("host='localhost' port='5432' dbname='
pastagemOutlier' user='postgres' password='123456'")
cur = con.cursor()

17
except:
19 log.write("\nFalha na conexÃ£o com o BD!")

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

23
cur.execute(sql)
25 con.commit()

27 try:
tdo= cur.fetchall()

29
ent = []
31 ant=None

33 except:
print("Erro - Nao foi possivel acessar os dados no BD")

35
file1 = open("selectmstotalD.sql", 'r')

```

```
37 sql1 = " ".join(file1.readlines())
39 cur.execute(sql1)
   con.commit()
41
   try:
43     dados1= cur.fetchall()
   except:
45     print("Erro - Nao foi possivel acessar os dados no BD")
47 file2 = open("selectmstotalF.sql", 'r')
   sql2 = " ".join(file2.readlines())
49
   cur.execute(sql2)
51 con.commit()
53
   try:
       dados2= cur.fetchall()
55 except:
       print("Erro - Nao foi possivel acessar os dados no BD")
57
59 mstotalD = [float(d[0]) for d in dados1]
   quartilD = mquantiles(mstotalD , prob=[0.75])
61 print(quartilD)
63 mstotalF = [float(d[0]) for d in dados2]
   quartilF = mquantiles(mstotalF , prob=[0.75])
65 print(quartilF)
67 for x in tdo:
69     if (ant == None):
       ant=x
71     continue
       #print(x)
73     if (ant[1] != x[1]) and x[0]== 'D':
       c=x[1]-ant[1]
75
       file = open("selectclima.sql", 'r')
77 sqlclima = " ".join(file.readlines())
```

```
79     data = (ant[1].isoformat(), x[1].isoformat())
81
82     cur.execute(sqlclima, data)
83     con.commit()
84
85     clima= cur.fetchall()
86
87     DG = x[3]
88     FG = ant[3]
89
90     DG = float(DG)
91     FG = float(FG)
92
93     if DG < quartilD[0] and FG < quartilF[0]:
94         TA = (DG-FG)/(x[1]-ant[1]).days
95     if DG < quartilD[0] and FG > quartilF[0]:
96         TA = (DG-quartilF[0])/(x[1]-ant[1]).days
97     if DG > quartilD[0] and FG < quartilF[0]:
98         TA = (quartilD[0]-FG)/(x[1]-ant[1]).days
99     if DG > quartilD[0] and FG > quartilF[0]:
100         TA = (quartilD[0]-quartilF[0])/(x[1]-ant[1]).days
101
102     if x[7] == 1:
103         trat1 = 1 #infestado
104         trat2 = 0 #mirapasto
105     elif x[7] == 2:
106         trat1 = 0
107         trat2 = 1
108     elif x[7] == 3:
109         trat1 = 1
110         trat2 = 0
111     else:
112         trat1 = 0
113         trat2 = 1
114
115     if TA<0:
116         TA=0
117
118     if x[3] > quartilD:
119         if ant[3] > quartilF:
```

```

119 #presente
#ent.append([x[1], (x[1]-ant[1]).days, x[2], ant[2], x[4], ant
    [3], clima[0][0], clima[0][7],
121 #     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],
123 #     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],
125 #     clima[0][26], clima[0][27], clima[0][28], clima[0][29], x
    [3]])
#futuro
127 ent.append([x[5], x[6], (x[1]-ant[1]).days, ant[2], '%.2f' %
    quartilF[0], '%.2f' % quartilD[0], '', 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],
129 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],
131 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], '%.2f' % TA])
133 else:
#presente
135 #ent.append([x[1], (x[1]-ant[1]).days, x[2], ant[2], x[4], 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],
137 #     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],
139 #     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]])
141 #futuro

```

```

ent.append([x[5], x[6], (x[1]-ant[1]).days, ant[2], ant[3],
           '%.2f' % quartilD[0], '', clima[0][0], clima[0][7], clima
143         [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],
145         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],
147         clima[0][27], clima[0][28], clima[0][29], '%.2f' % TA])

else:
149     if ant[3] > quartilF:
           #presente
151         #ent.append([x[1], (x[1]-ant[1]).days, x[2], ant[2], x[4], 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],
153         #         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],
155         #         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
157         ent.append([x[5], x[6], (x[1]-ant[1]).days, ant[2], '%.2f' %
           quartilF[0], x[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],
161         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],
163         clima[0][27], clima[0][28], clima[0][29], '%.2f' % TA])

```

```

else:
165     #presente
    #ent.append([x[1], (x[1]-ant[1]).days, x[2], ant[2], x[4], ant
        3], clima[0][0], clima[0][7],
167     #     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],
169     #     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],
171     #     clima[0][26], clima[0][27], clima[0][28], clima[0][29], x
        [3]))
    #futuro
173     ent.append([x[5], x[6], (x[1]-ant[1]).days, ant[2], ant[3], x
        [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],
175     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],
177     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], '%.2f' % TA])
179
    ant=x
181
183 import csv
185 #presente
    #header = ['data', 'numerodias', 'alturamedia', 'alturamediaanterior',
        'pcaonni', 'mstotalanterior',
187     #     'tmin', 'dptmin', 'vartmin', 'tmed', 'dptmed', 'vartmed', 'tmax',
        'dptmax', 'vartmax',
    #     'umidade', 'dpumidade', 'varumidade', 'velocidadevento', '
        dpvelocidadevento', 'varvelocidadevento',
189     #     'radiacaosolar', 'dpradiacaosolar', 'varradiacaosolar', 'chuva',
        'dpchuva', 'varchuva', 'somatermica',

```

```

#      'dpsomatermica', 'varsomatermica', 'def', 'dpdef', 'vardef', 'exc
      ', 'dpexc', 'varexc', 'taxaacumulo']
191
#futuro
193 header = ['mes', 'ano', 'numerodias', 'alturamediaanterior', '
      mstotalanterior', 'mstotal', 'CA', 'tmin', 'dptmin', 'vartmin', '
      tmed',
      'dptmed', 'vartmed', 'tmax', 'dptmax', 'vartmax', 'umidade', '
      dpumidade', 'varumidade', 'velocidadevento',
195 'dpvelocidadevento', 'varveloividadevento', 'radiacaosolar', '
      dpradiacaosolar', 'varradiacaosolar', 'chuva',
      'dpchuva', 'varchuva', 'somatermica', 'dpsomatermica', '
      varsomatermica', 'def', 'dpdef', 'vardef', 'exc',
197 'dpexc', 'varexc', 'TA']

199
#REVERSE
201 #ent.reverse()

203 with open('e1.csv', 'w', newline='') as myfile:
      wr = csv.writer(myfile, quoting=csv.QUOTE_ALL, delimiter=';')
205 wr.writerow(header)
      wr.writerows(ent)

207
now = datetime.now()

209
log = open('log.txt', 'a')
211 log.write("-----Consolida dados ModeloAjustado ExecuÃ§Ã£o encerrada
      em: " + str(now) + "\n\n")

```

Select para recuperar os dados de pastagem presentes no banco de dados

```

1 select
p.dentrofora ,
3 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 ,
5 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 ,
7 EXTRACT(MONTH FROM m.data) as "Mes" ,
EXTRACT(YEAR FROM m.data) as "Ano" ,
9 p.idpotreiro
from subarea s, potreiro t, pastagem p, medicao m
11 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=1
group by p.dentrofora , m.data , p.idpotreiro
13 order by m.data ;

```

Select para recuperar os dados meteorológicos presentes no banco de dados

```

1 select
cast((avg(tmin)) as NUMERIC(17,2)) as tmin ,
3 cast((avg(tmed)) as NUMERIC(17,2)) as tmed ,
cast((avg(tmax)) as NUMERIC(17,2)) as tmax ,
5 cast((avg(umidade)) as NUMERIC(17,2)) as umidade ,
cast((avg(velocidadevento)) as NUMERIC(17,2)) as velocidadevento ,
7 cast((sum(radiacaosolar)) as NUMERIC(17,2)) as radiacaosolar ,
cast((sum(chuva)) as NUMERIC(17,2)) as chuva ,
9 cast((stddev(tmin)) as NUMERIC(17,2)) as dptmin ,
cast((stddev(tmed)) as NUMERIC(17,2)) as dptmed ,
11 cast((stddev(tmax)) as NUMERIC(17,2)) as dptmax ,
cast((stddev(umidade)) as NUMERIC(17,2)) as dpumidade ,
13 cast((stddev(velocidadevento)) as NUMERIC(17,2)) as dpvelocidadevento ,
cast((stddev(radiacaosolar)) as NUMERIC(17,2)) as dpradiacaosolar ,
15 cast((stddev(chuva)) as NUMERIC(17,2)) as dpchuva ,
cast((variance(tmin)) as NUMERIC(17,2)) as vartmin ,
17 cast((variance(tmed)) as NUMERIC(17,2)) as vartmed ,
cast((variance(tmax)) as NUMERIC(17,2)) as vartmax ,
19 cast((variance(umidade)) as NUMERIC(17,2)) as varumidade ,
cast((variance(velocidadevento)) as NUMERIC(17,2)) as
  varvelocidadevento ,
21 cast((variance(radiacaosolar)) as NUMERIC(17,2)) as varradiacaosolar ,
cast((variance(chuva)) as NUMERIC(17,2)) as varchuva ,
23 cast((sum(somatermica)) as NUMERIC(17,2)) as somatermica ,
cast((stddev(somatermica)) as NUMERIC(17,2)) as dpsomatermica ,
25 cast((variance(somatermica)) as NUMERIC(17,2)) as varsomatermica ,
cast((sum(def)) as NUMERIC(17,2)) as def ,

```

```
27 cast((stddev(def)) as NUMERIC(17,2)) as dpdef ,  
    cast((variance(def)) as NUMERIC(17,2)) as vardef ,  
29 cast((sum(exc)) as NUMERIC(17,2)) as exc ,  
    cast((stddev(exc)) as NUMERIC(17,2)) as dpexc ,  
31 cast((variance(exc)) as NUMERIC(17,2)) as varexc  
    from clima_evapo  
33 WHERE data between %s and %s
```

**APÊNDICE B – SCRIPT DO MÓDULO DE PREDIÇÃO DO MODELO
AJUSTADO UTILIZANDO O CONJUNTO DE DADOS A, PARA O POTREIRO
P20 INFESTADO, UTILIZANDO TREINAMENTO ESTRATIFICADO POR
TRATAMENTO**

```
1 import os
  from math import sqrt
3 from numpy import concatenate
  from matplotlib import pyplot
5 from pandas import read_csv
  from pandas import DataFrame
7 from pandas import concat
  import pandas as pd
9 from sklearn.preprocessing import MinMaxScaler
  from sklearn.preprocessing import LabelEncoder
11 from sklearn.metrics import mean_squared_error
  from keras.models import load_model
13 from keras.models import Sequential
  from keras.layers import Dense
15 from keras.layers import LSTM
  from keras.layers import Dropout
17 import numpy as np
  from datetime import date
19 import csv
  from folium.plugins import HeatMap
21 import os
  import folium
23 import time
  import webbrowser
25 import branca.colormap as cmp
  from folium.features import DivIcon
27 from folium.plugins import FloatImage
  from scipy import stats
29 from sklearn.metrics import r2_score
  from datetime import datetime
31
  now = datetime.now()
33
  log = open('log.txt', 'a')
```

```

35 log.write("----- Execucao iniciada em: " + str(now) + "\n\n")
37 caminho = r'C:\Users\Anderson Fiscoeder\Desktop\Tese\Modelos\
    ModeloAjustado'
cmh = os.path.join(caminho, 'Total')
39 #####Potreiro 20 Infestado
    #####
# load dataset
41 dataset1 = read_csv('entrada_Inf_CN.csv', header=0, index_col=0,
    delimiter=';')
values1 = dataset1.values
43 # integer encode direction
encoder = LabelEncoder()
45
dataset2 = read_csv('e1_Teste.csv', header=0, index_col=0, delimiter=';
    ')
47 values2 = dataset2.values
# integer encode direction
49 encoder = LabelEncoder()
51 #Drop NA Values
values1=values1[~np.isnan(values1).any(axis=1)]
53 values2=values2[~np.isnan(values2).any(axis=1)]
55 # ensure all data is float
values1 = values1.astype('float32')
57 values2 = values2.astype('float32')
59 real1 = values1[:, -1]
61 # normalize features
scaler = MinMaxScaler()
63 scaled1 = scaler.fit_transform(values1)
scaled2 = scaler.fit_transform(values2)
65
scaled1 = DataFrame(scaled1)
67 scaled2 = DataFrame(scaled2)
#print(scaled)
69 #print(scaled2)
# split into train and test sets
71 values1 = scaled1.values

```

```

values2 = scaled2.values
73 #n_train = 24
#n_train = 48
75 train1 = values1
test1 = values2
77 #print(train)
#print(test)
79 # split into input and outputs
train_X1, train_y1 = train1[:, :-1], train1[:, -1]
81 test_X1, test_y1 = test1[:, :-1], test1[:, -1]

83 #print(train_X)
#print(train_y)
85 #print(test_X1)
#print(test_y1)
87
# reshape input to be 3D [samples, timesteps, features]
89 train_X1 = train_X1.reshape((train_X1.shape[0], 1, train_X1.shape[1]))
test_X1 = test_X1.reshape((test_X1.shape[0], 1, test_X1.shape[1]))
91
#print(train_X)
93 #print(train_y)
#print(test_X)
95 #print(test_y)

97 print(train_X1.shape, train_y1.shape, test_X1.shape, test_y1.shape)
print(train_X1.shape[1])
99
if not os.path.exists('M_INF_CN.h5'):
101 print('modelo inexistente... criando modelo...')
# design network
103 model = Sequential()
#model.add(LSTM(30, input_shape=(train_X.shape[1], train_X.shape[2]),
#kernel_initializer='normal', activation='tanh',
#return_sequences = True))#bom para 1 e 2
105 #model.add(LSTM(15, input_shape=(train_X.shape[1], train_X.shape[2]),
#kernel_initializer='normal', activation='tanh',
#return_sequences = True))#bom para 1 e 2
#model.add(LSTM(7, input_shape=(train_X.shape[1], train_X.shape[2]),
#kernel_initializer='normal', activation='tanh'))
107

```

```

109 model.add(LSTM(30, input_shape=(train_X1.shape[1], train_X1.shape[2])
    , kernel_initializer='normal', return_sequences = True))#bom
    para 1 e 2
model.add(LSTM(15, input_shape=(train_X1.shape[1], train_X1.shape[2])
    , kernel_initializer='normal'))
111
#model.add(LSTM(50, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', activation='sigmoid'))
113 #model.add(LSTM(10, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', activation='softmax'))
#model.add(LSTM(50, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', activation='softplus'))
115 #model.add(LSTM(50, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', activation='softsign'))
#model.add(LSTM(50, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', activation='hard_sigmoid'))
117 #model.add(LSTM(50, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', activation='linear'))
#model.add(LSTM(10, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', activation='selu'))#bom para 2
119 #model.add(LSTM(15, input_shape=(train_X.shape[1], train_X.shape[2]),
    kernel_initializer='normal', activation='relu'))#bom para 2

121 #model.add(Dense(500, kernel_initializer='normal', activation='tanh')
    )#para 2 melhorou

123 # Hidden - Layers

125 #model.add(Dropout(0.3, noise_shape=None, seed=None))
#model.add(Dense(50, activation = "tanh"))
127 #model.add(Dropout(0.2, noise_shape=None, seed=None))
#model.add(Dense(10, activation = "tanh"))

129
model.add(Dense(1, kernel_initializer='normal'))
131
#model.compile(loss='mse', optimizer='SGD')
133 model.compile(loss='mean_squared_error', optimizer='rmsprop')#mto bom
    para 1 e 2
#model.compile(loss='mse', optimizer='Adagrad')#mto bom para 1
135 #model.compile(loss='mse', optimizer='Adadelta')#mto bom para 1 e 2

```

```

#model.compile(loss='mse', optimizer='Adam')#mto bom para 1
137 #model.compile(loss='mse', optimizer='Adamax')#mto bom para 1
#model.compile(loss='mse', optimizer='Nadam')

139
# fit network
141 history = model.fit(train_X1, train_y1, epochs=5000, batch_size=72,
    validation_data=(test_X1, test_y1), verbose=0, shuffle=False)
#-----
143 # save the network's architecture
print('Salvando modelo...')
145 model.save('M_INF_CN.h5')
#-----
147 else:
#-----
149 # load the network's architecture
model = load_model('M_INF_CN.h5')
151 history = model
#-----
153 print('Modelo encontrado e carregado')

155 # make a prediction
yhat1 = model.predict(test_X1)
157 test_X1 = test_X1.reshape((test_X1.shape[0], test_X1.shape[2]))
# invert scaling for forecast
159 inv_yhat1 = concatenate((test_X1, yhat1), axis=1)
inv_yhat1 = scaler.inverse_transform(inv_yhat1)
161 inv_yhat1 = inv_yhat1[:, -1]
# invert scaling for actual
163 test_y1 = test_y1.reshape((len(test_y1), 1))
inv_y1 = concatenate((test_X1, test_y1), axis=1)
165 inv_y1 = scaler.inverse_transform(inv_y1)
inv_y1 = inv_y1[:, -1]

167
print(model.summary())
169
from keras.utils import plot_model
171 plot_model(model, to_file='model_plot.png', show_shapes=True,
    show_layer_names=True)

173 #from ann_visualizer.visualize import ann_viz

```

```

175 #ann_viz(model, title="My first neural network")

177 # calculate RMSE
rmse = sqrt(mean_squared_error(inv_y1, inv_yhat1))
179 desvioAmostraldpred1 = np.std(inv_yhat1) #desvio padrÃ£o populacional
varianciaAmostraldpred1 = inv_yhat1.var() #variancia populacional
181
desvioAmostraldreal1 = np.std(inv_y1) #desvio padrÃ£o populacional
183 varianciaAmostraldreal1 = inv_y1.var() #variancia populacional

185 slope, intercept, r_value, p_value, std_err = stats.linregress(inv_y1,
inv_yhat1)

187
coeffs1 = np.polyfit(inv_y1, inv_yhat1, 5)
189 p1 = np.poly1d(coeffs1)
# fit values, and mean
191 yhat1 = p1(inv_y1) # or [p(z) for z in x]
ybar1 = np.sum(inv_yhat1)/len(inv_yhat1) # or sum(y)/len(y)
193 ssreg1 = np.sum((yhat1-ybar1)**2) # or sum([(yihat - ybar)**2
for yihat in yhat])
sstot1 = np.sum((inv_yhat1 - ybar1)**2) # or sum([(yi - ybar)**2
for yi in y])
195 r1 = ssreg1 / sstot1

197 #####CriÃ§Ã£o do grÃ¡fico de linha
#####

#####GrÃ¡fico P20 Infestado
#####
199 arq = read_csv(r'C:\Users\Anderson Fiscoeder\Desktop\Tese\Modelos\
ModeloOriginal_Variaveis\lotacao1.csv', header=0, delimiter=';')
dados = arq.values
201 tam = len(arq)
tam2 = len(dataset2)
203 peso = dados[:, -1]
data = dados[:, 0]#E necessario o arquivo csv estar com o formato da
data em: Ano-Mes-Dia
205 dia = data[(tam-12):,]
fig, ax = pyplot.subplots()
207 pyplot.ylabel('kg/ha/dia')

```

```

fig.autofmt_xdate()
209 ax.plot(dia, inv_y1, label='Real P20 Inf', linestyle='--', marker='o')
ax.plot(dia, inv_yhat1, label='Predito P20 Inf', linestyle='--', marker
        = 'o')
211 #pyplot.plot_date(dia, peso[(tam-12):], label='Taxa de lotaÃ§Ã£o',
        linestyle='--', marker='o')#[ (tam-12):] para pegar o 12 Ãºltimos
        dados da coluna dias
pyplot.legend()
213 dpi = fig.get_dpi()
pyplot.savefig(r'C:\Users\Anderson_Fischoeder\Desktop\Tese\Modelos\
        ModeloAjustado\Inf_CN\P20Inf_InfCN_ModeloAjus.png', dpi=dpi*2)
215 pyplot.close()

217 erro = []
for x in inv_yhat1:
219     erro = inv_yhat1 - inv_y1
#Boxplot
221 fig1, ax1 = pyplot.subplots()
pyplot.boxplot([inv_y1, inv_yhat1, erro], labels=['Real', 'Predito', '
        Erro'])
223 pyplot.title('P20 - Infestado')
dpi = fig1.get_dpi()
225 pyplot.savefig(r'C:\Users\Anderson_Fischoeder\Desktop\Tese\Modelos\
        ModeloAjustado\Inf_CN\P20Inf_Boxplot_InfCN_ModeloAjus.png', dpi=dpi
        *2)
pyplot.close()
227 #GrÃ¡fico de DispersÃ£o
pyplot.scatter(inv_y1, inv_yhat1)
229 range = [inv_y1.min(), inv_yhat1.max()]
pyplot.plot(range, range, 'red')
231 pyplot.title('P20 - Infestado - Real x Predito')
pyplot.ylabel('Predito')
233 pyplot.xlabel('Real')
pyplot.savefig(r'C:\Users\Anderson_Fischoeder\Desktop\Tese\Modelos\
        ModeloAjustado\Inf_CN\P20Inf_Dispersao_InfCN_ModeloAjus.png', dpi=
        dpi*2)
235 pyplot.close()

237 print('P20 - Infestado')
print("R2 linear", r_value ** 2)
239 print("R2 Polinomial:", r1)

```

```

print('Test RMSE: %.3f' % rmse)
241 print("Desvio Real", desvioAmostrareal1)
print("Variância Real", varianciaAmostrareal1)
243 print("Desvio Predito", desvioAmostrapred1)
print("Variância Predito", varianciaAmostrapred1)
245 print('Real')
print(inv_y1.round(1))
247 print('Predito')
print(inv_yhat1.round(0))
249
file = open(r"C:\Users\Anderson Fiscoeder\Desktop\Tese\Modelos\
        ModeloAjustado\Inf_CN\dados.txt", "a")
251 file.write('\n\r' + 'P20 - Infestado' + '\n')
file.write('Test RMSE: ' + '%.3f' % rmse + '\n')
253 file.write('R2 linear: ' + str(r_value ** 2) + '\n')
file.write("R2 Polinomial:" + str(r1) + '\r')
255 file.write("Desvio Real:" + str(desvioAmostrareal1) + '\n')
file.write("Variância Real:" + str(varianciaAmostrareal1) + '\n')
257 file.write("Desvio Predito:" + str(desvioAmostrapred1) + '\n')
file.write("Variância Predito:" + str(varianciaAmostrapred1) + '\n')
259 file.write('Real' + '\n')
for a in inv_y1:
261     file.write(str(a) + ',' + ' ')
file.write('\n' + 'Predito' + '\n')
263 for b in inv_yhat1:
    file.write(str(b) + ',' + ' ')
265 file.close()

267 now = datetime.now()

269 log = open('log.txt', 'a')
log.write("----- ExecuçÃ£o Total encerrada em: " + str(now) + "\n\n"
        )

```