



# **Estácio**

**UNIVERSIDADE ESTÁCIO DE SÁ**

**CAMPUS PARANGABA**

**Uma abordagem de Aprendizado de Máquina  
para previsão de tendência climáticas**

**Antonio Wellington - 202308719025**

**Davi Guerra - 202308292201**

**Fernando Chaves - 202308423991**

**Gustavo Ferreira - 202308424459**

**Isaac Ferreira - 202308424513**

**Lucas Vieira - 20230824442**

**Orientador: Cynthia Moreira Maia**

**2024**

**FORTALEZA/CE**

## Sumário

1. DIAGNÓSTICO E TEORIZAÇÃO .....	3
1.1. Identificação das partes interessadas e parceiros .....	3
1.2. Problemática e/ou problemas identificados .....	3
1.3. Justificativa.....	4
1.4. Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos) .....	5
1.5. Referencial teórico (subsídio teórico para propositura de ações da extensão) .....	5
2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO .....	8
2.1. Plano de trabalho (usando ferramenta acordada com o docente) .....	8
2.2. Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as 1010 .....	
2.3. Grupo de trabalho (descrição da responsabilidade de cada membro) .....	11
2.4. Metas, critérios ou indicadores de avaliação do projeto .....	12
2.5. Recursos previstos .....	13
2.6. Detalhamento técnico do projeto .....	14
2.7. Visão geral do projeto.....	17
3. ENCERRAMENTO DO PROJETO .....	21
3.1. Relatório Coletivo (podendo ser oral e escrita ou apenas escrita) .....	21
3.2. Relato de Experiência Individual .....	22
3.2.1. CONTEXTUALIZAÇÃO.....	22
3.2.2. METODOLOGIA .....	23
3.2.3. RESULTADOS E DISCUSSÃO: .....	25
3.2.4. REFLEXÃO APROFUNDADA.....	26
3.2.5. CONSIDERAÇÕES FINAIS.....	26
4. BIBLIOGRAFIA.....	27
5. CÓDIGO.....	29

## 1. DIAGNÓSTICO E TEORIZAÇÃO

### 1.1. Identificação das partes interessadas e parceiros

O projeto não envolveu interação direta com stakeholder. Para compreender as exigências climáticas das plantações, foram realizadas pesquisas com base em estudos de pesquisadores da Embrapa (Empresa Brasileira de Pesquisa Agropecuária). Para a coleta dos dados meteorológicos utilizou informações fornecidas pelo INMET (Instituto Nacional de Meteorologia).

### 1.2. Problemática e/ou problemas identificados

Em vista das mudanças climáticas dos últimos anos, pretendemos avaliar o impacto delas no Brasil. Escolhemos a região Centro-Oeste por ser o principal produtor agrícola do Brasil, e as alterações climáticas podem atingir diretamente a produção dessa Região, as oscilações nas condições meteorológicas impactam não apenas a produtividade, mas também a segurança alimentar e a sustentabilidade das práticas agrícolas.

A variabilidade da chuva tem gerado problemas para os agricultores, que frequentemente enfrentam períodos de extrema seca ou chuvas excessivas, ambos prejudiciais à colheita.

O aumento das temperaturas tem efeito direto sobre a saúde das culturas. Muitas vezes, temperaturas mais elevadas comprometem a produtividade e podem afetar a qualidade dos produtos agrícolas.

A umidade é um fator crítico para o desenvolvimento das plantas. Tanto a falta quanto o excesso de umidade podem levar a doenças, pragas e outras complicações, impactando diretamente a viabilidade econômica dos agricultores.

Esses fenômenos climáticos não afetam apenas os produtores rurais, mas têm um impacto profundo em toda a população local. A seca extrema, por exemplo, compromete o abastecimento de água, afetando tanto a agricultura quanto o consumo diário das famílias, as chuvas torrenciais, por outro lado, causam não apenas danos às lavouras, mas também à infraestrutura urbana e as altas temperaturas e a baixa umidade agravam ainda mais a

situação, contribuindo para o surgimento de problemas de saúde pública, como desidratação e doenças relacionadas ao calor.

Assim, a elaboração deste projeto de extensão visa abordar essas questões, utilizando Aprendizado de Máquina para analisar padrões climáticos e oferecer informações úteis para a comunidade agrícola. O objetivo é não apenas reduzir os impactos negativos das condições climáticas, mas também informar as autoridades públicas para tomar decisões mais informadas.

### 1.3. Justificativa

A influência do clima nas lavouras é pertinente academicamente, vários cientistas ao longo dos anos vêm dirigindo estudos sobre os seus efeitos, a aprendizagem de máquina e estatística seria uma ferramenta valiosa que mostraria na prática os danos causados, é mais um recurso que os cientistas poderiam usar. Essa abordagem educacional não apenas enriquece a formação teórica, mas também os envolve em práticas que têm um impacto direto na comunidade.

O estudo das variáveis climáticas: precipitação, temperatura e umidade, proporciona uma oportunidade para que o grupo desenvolva competências em aprendizagem de máquina e análise estatística, essenciais para o mercado de trabalho atual. Além disso, ao abordar uma questão tão relevante e urgente, os alunos se tornam mais conscientes dos problemas, levando a uma análise crítica da situação.

As motivações do grupo de trabalho são múltiplas. Em primeiro lugar, há um forte desejo de contribuir para a comunidade, oferecendo soluções que possam esclarecer como os impactos das mudanças climáticas atuam no Brasil, assim, trazer resoluções para a problemática.

No geral, a abordagem da questão climática se alinha perfeitamente com os objetivos acadêmicos do curso, promovendo uma experiência de aprendizagem rica e significativa. Essa experiência não só capacita os alunos em suas áreas de estudo, mas também os prepara para enfrentar desafios reais e contribuir para soluções inovadoras, reafirmando o papel da academia como um agente ativo na produção de conhecimento.

#### 1.4. Objetivos, resultados e efeitos a serem alcançados

##### Objetivos do Projeto de Extensão

- 1. Analisar:** Utilizar técnicas de estatística e *Big Data* para coletar e analisar dados sobre precipitação, temperatura e umidade, mapear padrões climáticos identificando tendências e padrões que impactam a região local.
- 2. Treinar:** Treinar os algoritmos de *machine learning* com os dados fornecidos, gerando resultados de previsões para o clima, com esse material podemos calcular as chuvas dos próximos meses através de uma ferramenta de monitoramento que fornece previsões climáticas e alertas para a comunidade.
- 3. Validar:** Validar se os resultados obtidos pelo treinamento estão prevendo corretamente, observar como está comportando o modelo e se está realmente trazendo um impacto para sociedade.

##### Instrumentos de Avaliação

Para garantir a efetividade dos objetivos, o grupo de trabalho utilizará os seguintes instrumentos de avaliação:

Questionários de Feedback: Aplicar questionários antes e após a implementação das ferramentas de aprendizado de máquinas para ver se os resultados são condizentes.

Relatórios de Acompanhamento: Elaborar relatórios periódicos que documentem as análises climáticas e a eficácia do modelo, permitindo ajustes e melhorias contínuas.

#### 1.5. Referencial teórico

Ivan Rodrigues de Almeida da Universidade Estadual Paulista, mostrou através de um estudo científico o papel das condições climáticas na organização das atividades humanas e na formação do ambiente terrestre. A agricultura, como uma das principais atividades humanas, é diretamente influenciada pelas condições climáticas, especialmente quando se trata da produção de alimentos. No contexto brasileiro, a soja ocupa posição de destaque, sendo um dos principais produtos agrícolas do país, com significativa relevância tanto no mercado

interno quanto no comércio internacional. O Brasil se destaca como o segundo maior produtor mundial de soja, a produção brasileira durante anos concentrou-se na região Sul, mas com o passar dos anos migrou-se para região Centro-Oeste, principalmente no estado de Mato Grosso. A hipótese deste estudo sugere que a diversificação das condições climáticas no território brasileiro tem influenciado essa migração da soja, com implicações não apenas na produção agrícola, mas também no meio ambiente e nas condições socioeconômicas das regiões envolvidas. A transição dos polos de produção de soja da Região Sul para a Região Centro-Oeste tem gerado impactos ambientais significativos, como o desmatamento e a degradação do solo, além de acentuar desigualdades sociais, com a concentração de grandes propriedades em algumas áreas e a marginalização de pequenos produtores em outras.

A análise das condições climáticas das regiões produtoras de soja, com ênfase no estado de Mato Grosso, revelou uma menor variabilidade da pluviosidade comparada a outras regiões do país, favorecendo a produção contínua de soja durante várias épocas do ano. A pluviometria no Mato Grosso, caracterizada por um regime de chuvas mais concentrado, torna a região mais atrativa para os produtores, ao mesmo tempo em que modifica a dinâmica de ocupação do solo e a distribuição fundiária. A utilização de ferramentas de geoprocessamento permitiu a criação de um sistema de análise espacial que possibilitou a visualização das mudanças na distribuição da produção e os efeitos climáticos sobre a organização das atividades agrícolas.

Este estudo confirma que as condições climáticas, especialmente a pluviosidade, têm grande influência na expansão da soja para o Centro-Oeste brasileiro, provocando alterações significativas no ambiente local e regional.

Com uma estação chuvosa bem definida e com baixa variabilidade pluviométrica, o estado do Mato Grosso apresenta um cenário capaz de modificar substancialmente a participação das unidades da federação na produção agrícola do país.

Outro referencial teórico utilizado foi a tese de Thiago Bigolin da Universidade Federal do Rio Grande do Sul, sua tese mostra que o sistema de duplo cultivo, que integra as culturas de milho e soja em uma mesma área ao longo do ano, é uma prática agrícola que colocou o Brasil em uma posição de destaque como um dos maiores produtores agrícolas. No entanto, as mudanças climáticas representam uma ameaça significativa ao futuro desse sistema, afetando

tanto o crescimento e o rendimento das plantações individualmente, quanto a dinâmica geral dos sistemas agrícolas no Brasil.

O trabalho de Thiago Bigolin foi centrado em analisar os impactos das mudanças climáticas passadas e futuras sobre o desenvolvimento e a produtividade do milho e da soja, especialmente no contexto do sistema de duplo cultivo, em regiões de clima subtropical. Sua análise se concentra nas possíveis implicações das mudanças climáticas para a adoção e a viabilidade deste sistema agrícola no futuro próximo, considerando os efeitos do clima sobre o desempenho das culturas e as mudanças nos padrões de produção.

Utilizando de metodologia científica sua tese conclui que o sistema de duplo cultivo poderá enfrentar dificuldades no futuro, haja vista que as previsões climáticas apontam para diminuição das chuvas e encurtamento da estação chuvosa associado ao aumento de temperatura que pode ser potencialmente danoso ao plantio de soja e milho.

Ademais, segundo a pesquisadora da Embrapa, Samira França Oliveira, os impactos ambientais relacionados às mudanças climáticas afetaram a maioria dos setores da sociedade. Entre os impactos relacionados às mudanças climáticas, destaca-se o aumento da frequência e intensidade dos eventos climáticos extremos (ECE), em relação aos extremos de calor e precipitação, com ocorrência de inundações, alagamentos, deslizamentos, movimentos de massa e seca para algumas regiões. No setor da agricultura familiar, em função do menor acesso aos recursos financeiros e de dependerem diretamente da produção agrícola para sua subsistência, em geral, apresentaram maior vulnerabilidade e menor condição de adaptação diante da variabilidade climática atual e projetada em função do aquecimento global.

Ainda que pareça óbvio, o rendimento e desenvolvimento das culturas agrícolas está intimamente ligado às condições climáticas proporcionadas pelo ambiente. É amplamente discutido e aceito pela comunidade acadêmica a relação entre clima e agricultura, esses três trabalhos acadêmicos citados mostram justamente essa estreita relação e como eles foram a base do nosso projeto, prever e verificar as variáveis climáticas que impactam a produção agrícola brasileira, para grandes e pequenos produtores.

## 2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

### 2.1. Plano de trabalho

O plano de trabalho foi dividido em cinco etapas.

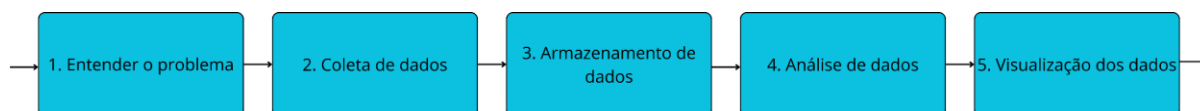


Figura 1 - Etapas do plano de trabalho.

O projeto tem como objetivo prever a condição climática para os próximos meses, assim os agricultores poderão tomar decisões para garantir o sucesso da próxima safra.

#### 1. Entendimento do problema.

Em um primeiro momento, a equipe se reuniu para responder uma pergunta: “Qual é o objetivo do projeto?”. O Brasil é um dos principais produtores e exportadores agrícola do mundo, segundo os pesquisadores da Embrapa, José Renato Bouças Farias, Norman Neumaier e Alexandre Lima Nepomuceno, a soja possui algumas exigências climáticas devido a características da espécie, o plantio da soja é adequado às regiões cuja temperatura oscila entre 20°C e 30°C sendo a temperatura ideal para seu desenvolvimento. Com isso em mente, pensamos em desenvolver modelos que possam prever as variáveis climáticas que possam impactar no plantio das regiões do centro-oeste, local onde há as maiores áreas agrícolas.

#### 2. Coleta de dados.

Em segundo lugar, precisamos identificar a fonte dos dados, escolhemos o Instituto Nacional de Meteorologia (INMET), delimitamos o escopo da coleta dos dados para ficar entre o dia primeiro de janeiro de 2016 até o dia trinta de setembro de 2024 do município de Campo Grande, por ser um polo agrícola. A coleta foi realizada pelo Lucas Vieira, os dados coletados foram estruturados no formato CSV com informações diárias de diversas variáveis que influenciam na agricultura, como por exemplo, precipitação total, temperatura do ar, umidade relativa, radiação global, pressão atmosférica, e algumas informações sobre o vento, como direção, velocidade e rajada máxima.



### **3. Armazenamento de dados.**

O armazenamento e compartilhamento dos dados foi feito utilizando o Drive, um serviço de armazenamento de arquivos, criado pelo Google. O compartilhamento foi realizado para que todos os membros do grupo tivessem acesso e pudessem visualizar os dados que estavam sendo utilizados, bem como os modelos de regressão linear e de séries temporais.

### **4. Análise de dados.**

Antes de iniciar a análise dos dados, foi preciso fazer um pré-processamento dos dados, ao fazer uma exploração inicial, identificamos e visualizamos valores ausentes e inconsistentes, usamos as Bibliotecas do Python, Pandas e Numpy, para solucionar esse problema. Após a etapa de exploração foi preciso fazer o preenchimento dos valores nulos, para isso foi utilizado os métodos *'ffill'* e *'bfill'* para o preenchimento das colunas, pegando o valores anteriores e posteriores da coluna para tirar uma média e preencher os espaços vazios, após o preenchimento dos valores faltosos, foi realizado uma média para normalizar os valores de todas as colunas que seriam utilizadas para o treinamento do modelo, isso se fez necessário pois havia dados para cada hora do dia, fornecendo vinte e quatro valores para um único dia, causando uma confusão no treinamento. Depois da normalização dos dados foi feita a conversão de dados para formatos adequados, ajustando a data para o formato *'datetime'* e as outras colunas numéricas para *float*.

Após o pré-processamento, os dados foram divididos entre dados de treino (70% dos dados) e teste (30% dos dados), para a criação dos modelos de predição foram utilizadas técnicas de aprendizado de máquina supervisionado com os algoritmos de regressão linear, os algoritmos utilizados foram o de árvore de decisão (*Decision Tree*), floresta aleatória (*Random Forest*) e KNN (*k-Nearest Neighbors*). Porém para a previsão da precipitação os modelos regressão linear não foram eficazes e foi preciso utilizar séries temporais com os modelos ARMA (*Autoregressive Moving Average*), ARIMA (*Autoregressive Integrated Moving Average*) e SARIMA (*Autoregressive Integrated Moving Average with Seasonality*) para analisar padrões, prever tendência e entender o comportamento de fenômenos ao longo do tempo.

A etapa de pré-processamento e criação dos modelos de regressão linear e séries temporais foi realizada por Fernando Chaves, enquanto treinamento dos modelos para a previsão das temperaturas e umidade foi realizado por Gustavo Ferreira e Antonio Wellington.

## **5. Visualização dos dados.**

Após a realização do treinamento dos modelos, foram gerados gráficos de dispersão e *forecast plot* para a visualização e análise dos resultados. Para validar a precisão e a confiabilidade dos resultados foram utilizadas nos modelos de regressão linear as métricas de avaliação de Coeficiente de Determinação ( $R^2$ ) e Erro Absoluto Médio (MAE), para os modelos de séries temporais foi utilizado o Erro Quadrático Médio (RMSE) e Erro Absoluto Médio (MAE).

A geração dos gráficos das temperaturas, umidade e precipitação foi realizada por Gustavo Ferreira, Antonio Wellington e Fernando Chaves, respectivamente.

O cronograma do trabalho e realização das etapas não foi previamente estabelecido, a construção dos modelos de regressão linear foi feita ao longo do período letivo, quanto aos modelos de séries temporais foi feita com apoio de aulas extracurriculares. O acompanhamento dos resultados foi feito com o compartilhamento dos gráficos via grupo do trabalho no whatsapp e o armazenamento compartilhado pelo Drive.

### **2.2. Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias utilizadas pelo grupo para mobilizá-los.**

Não tivemos um contato com a parte sociocomunitária por conta da natureza do trabalho, a avaliação do projeto foi feita pela nossa orientadora, Cynthia Moreira Maia. Nossos encontros foram feitos durante a aula, onde ela passava orientação sobre o trabalho, essa interação entre aluno e professor foi fundamental para o desenvolvimento do projeto. Nossa orientadora trouxe dicas que foram usadas com êxito por nossa equipe, focar só em campo grande, tirar a média dos valores e criar uma tabela são exemplos das orientações que trouxeram êxito ao projeto.

### **2.3. Grupo de trabalho**

**Antonio Wellington:**

- Participou da coleta de dados.
- Participou do pré-processamento de dados.
- Participou da criação dos modelos de regressão linear.
- Participou da criação dos modelos de séries temporais.
- Fez o treinamento do aprendizado de máquina para umidade relativa, máxima e mínima.

**Davi Guerra:**

- Participou da coleta de dados.
- Participou na análise dos gráficos.
- Realizou pesquisas sobre o setor da agricultura.

**Fernando Chaves:**

- Participou da coleta de dados.
- Fez o pré-processamento dos dados.
- Criou os modelos de regressão linear.
- Criou os modelos de séries temporais.
- Fez o treinamento dos modelos para precipitação total.

**Gustavo Ferreira:**

- Participou da coleta de dados.
- Participou do pré-processamento de dados.
- Participou da criação dos modelos de regressão linear.
- Participou da criação dos modelos de séries temporais.
- Fez o treinamento do aprendizado de máquina para temperatura do ponto de orvalho e temperatura do ar - bulbo seco.

**Isaac Ferreira:**

- Participou da coleta de dados.

- Participou da criação dos modelos de regressão linear
- Participou na análise dos gráficos.
- Ajudou nas pesquisas sobre o setor da agricultura.

#### **Lucas Vieira:**

- Fez a coleta de dados.
- Fez o pré-processamento de dados.
- Fez uma análise exploratória dos dados.
- Participou da criação dos modelos de regressão linear.
- Fez os modelos de séries temporais.
- Fez a análise dos gráficos gerados pelo treinamento dos modelos de regressão linear e séries temporais.

### **2.4. Metas, critérios ou indicadores de avaliação do projeto**

#### **1. Análise dos dados:**

Para a análise dos dados, reunimos informações históricas sobre precipitação, temperatura e umidade de fontes confiáveis, como estações meteorológicas e bases de dados públicas do INMET, depois realizamos uma análise exploratória para identificar e tratar os valores nulos ou inconsistentes, seguimos com o preenchimento dos dados faltosos efetuamos e normalização dos dados, convertendo as datas para o tipo '*datetime*' e os valores numéricos para '*float*'. Em seguida aplicamos técnicas de estatísticas para identificar padrões e tendências climáticas relevantes na região local, buscamos a relação entre as variáveis meteorológicas para buscar fatores que influenciam as mudanças climáticas.

As métricas de avaliação da análise de dados seria a qualidade dos dados, tendo os valores nulos preenchidos e em segundo ponto seria a consistência dos dados, mantendo todos os valores convertidos para o tipo de dado adequado de forma padronizada.

#### **2. Treinamento dos modelos de machine learning:**

Para o treinamento dos modelos preditivos para estimar as estatísticas climáticas escolhemos algoritmos apropriados para a predição do clima, como regressão linear, e série temporal,

após a criação dos modelos separamos os dados em conjuntos de treinamento e teste para garantir a robustez do modelo. Foi dividido em 70% para treinamento e 30% para teste, após a divisão, veio a aplicação dos modelos para realizar as previsões de precipitação, temperatura e umidade.

### 3. Validação do Modelo e Avaliação de Impacto

Para validar a precisão dos modelos preditivos foi utilizado o coeficiente de determinação ( $R^2$ ), Erro Quadrático Médio ( $RMSE$ ) e Erro Médio Absoluto ( $MAE$ ).

Quanto às métricas de avaliação, é necessário realizar ajustes nos modelos para evitar o overfitting com o objetivo de garantir a veracidade dos resultados.

#### 2.5. Recursos previstos

**Recursos materiais:** Recursos utilizados foram computadores e notebooks pessoais.

**Recursos institucionais:** Computadores para aprendizado, Citrix.

**Recursos humanos:** A equipe do projeto é formada por seis integrantes, cada membro trouxe conhecimentos valiosos para cada fase do trabalho, os participantes realizaram pesquisas para a construção dos modelos, as pesquisas incluíam informações das variáveis e como elas impactam no desenvolvimento das safras, o projeto foi supervisionado pela orientadora Cynthia Moreira Maia. Sua orientação tem sido crucial para garantir que a equipe esteja alinhada com as melhores práticas. A professora tem contribuído com feedback valioso, ajudando a equipe a superar desafios e a desenvolver soluções inovadoras. Sua vasta experiência acadêmica e prática proporciona um suporte, permitindo que a equipe maximize seu potencial.

**Recursos financeiros:** Não houve gastos financeiros diretos no projeto, todos os recursos necessários foram fornecidos pelos alunos. O uso de software open-source e dados fornecidos são de domínio público, garantindo a viabilidade do projeto sem a necessidade de investimentos adicionais.

## 2.6. Detalhamento técnico do projeto

### 2.6.1. Pandas

Para realizar a análise exploratória do conjunto de dados do projeto foi utilizado a biblioteca *open source* Pandas que forneceu à equipe do trabalho ferramentas necessárias para realizar a limpeza e o tratamento dos dados deu suporte nas atividades de *Machine Learning* permitindo consultas e a visualização dos dados.

### 2.6.2. Ploty Express

Para a criação dos gráficos de dispersão, entre outros, foi utilizado um módulo da biblioteca Ploty, o Express, ou px, foi utilizamos para analisar visualmente o aprendizado do modelo com conjunto de dados fornecidos pelo INMET, ele possui uma sintaxe simples, mas intuitiva o que tornou seu uso simples e eficaz.

### 2.6.3. Sklearn

A equipe do projeto utilizou vários módulos da biblioteca Sklearn para a construção e avaliação dos modelos de aprendizagem de máquina e eles são:

#### 2.6.3.1. model\_selection

Utilizamos a função `train_test_split()` para dividir o conjunto de dados em partes de treinamento (70% dos dados) e teste (30% dos dados), permitindo uma avaliação confiável do desempenho do modelo. Outra funcionalidade do módulo utilizada, foi a função `max_depth` importada do `GridSearchCV`, essa ferramenta é utilizada para encontrar a melhor combinação de hiperparâmetros, um desses é o `max_depth`, que foi utilizado para controlar a profundidade máxima da árvore aleatória, limitando o número de divisões sucessivas nós tentamos reduzir o overfitting, já que o modelo se adaptou excessivamente, porém mesmo com a redução das divisões, o coeficiente de determinação e o erro médio absoluto além do gráfico apresentaram valores excepcionais.

#### 2.6.3.2. Neighbors

Para a criação do modelo k-NN foi preciso importar o modelo `KNeighborsRegressor`, já que estamos tratando de um problema de regressão. O *k-Nearest Neighbors* é um modelo simples, ele é usado em casos em que a explicação do modelo é tão importante

quanto a precisão, ele tenta classificar cada amostra do conjunto de dados avaliando sua distância em relação aos vizinhos mais próximos. Embora, no projeto tenha sido o modelo que apresentou os piores resultados, sua implementação foi fácil e seus resultados foram válidos para comparação com outros modelos.

#### 2.6.3.3. Tree

O modelo de árvore de decisão foi implementado usando o *DecisionTreeRegressor* do módulo *tree*. Dentre todos os algoritmos de *machine learning* o que se destaca pela sua estrutura visual é a árvore de decisão (*Decision Tree*). Como o próprio nome sugere, o modelo se assemelha a uma árvore, estabelecendo-nos, relacionando-se por uma hierarquia. Existe um nó raiz, que é o mais importante e os nós-folhas, que são os resultados, sim ou não, gerado a partir de uma pergunta. Sua implementação foi simples e seus resultados foram bons em relação ao modelo K-NN, mas inferiores quando comparado ao modelo de árvore aleatória

#### 2.6.3.4. Ensemble

Para a implementação do modelo de árvore aleatória, foi importado do módulo *ensemble* o *RandomForestRegressor*, esse algoritmo cria de forma aleatória várias árvores de decisão e combina os resultados de todas elas para chegar em um resultado. Esse modelo foi o que apresentou os melhores resultados, com o coeficiente de determinação ( $R^2$ ) bem próximo de 1, e um erro médio absoluto (MAE) quase 0.

#### 2.6.3.5. metrics

Para a avaliação dos modelos de regressão linear, foram abordadas do módulo *metrics* as métricas, coeficiente de determinação ( $R^2$ ) e Erro Absoluto Médio. O  $R^2$  representa o percentual da variância dos dados, quanto maior o valor, mais explicativo é o modelo, sendo o máximo 1, e quanto menor o valor, menos explicativo é o modelo, sendo o mínimo 0. O Erro Absoluto Médio ou *MAE* do inglês, retorna a média da diferença entre o valor real e o valor predito, essa métrica não é afetada por valores discrepantes - os *outliers*, o valor de saída diz quanto para mais ou para menos em relação ao valor real, o modelo pode estar errado.

#### 2.6.3.6. linear\_model

Utilizamos o módulo `linear_model` para fazer a importação do modelo de regressão linear padrão, para realizar o modelo autorregressivo, que possui valores aparentemente bons, mas os gráficos não apresentam resultados positivos.

#### 2.6.4. `numpy`

Numpy ou np, é uma biblioteca que utilizamos para trabalhar com arrays na criação do modelo autorregressivo (AR), seus resultados servirão para comparação com os modelos de séries temporais.

#### 2.6.5. `statsmodels.graphics.tsaplots`

Para realizar uma análise de séries temporais é necessário avaliar a correlação de um valor com outro valor, no nosso caso precisávamos encontrar alguma variável climática que se correlaciona-se com a precipitação total, e para isso foi preciso usar duas funções do `statsmodels.graphics.tsaplots`, o `plot_acf` (autocorrelação ou *ACF*, quando comparamos o valor do presente com valores do passado) e o `plot_pacf` (autocorrelação parcial ou *PACF*), a diferença entre essas autocorrelações é apenas um detalhe, o *ACF* temos a autocorrelação direta e indireta e em uma *PACF* apenas direta. Tentamos com umidade relativa e com as temperaturas de bulbo - seco e ponto de orvalho, mas a principal relação foi com a data.

#### 2.6.6. `statsmodels.tsa.statespace.sarimax`

Para a criação do modelo de séries temporais *ARMA* (*Autoregressive Moving Average*) foi preciso importar o *SARIMAX*, um modelo estatístico usado para prever séries temporais com sazonalidade e variáveis extremas. O modelo *ARMA* está incluso no *SARIMAX* e é uma versão mais simples pois não trabalha com sazonalidade e sem variáveis exógenas. O modelo *ARMA* combina o modelo AutoRegressivo (*AR*), onde os cada valor vai depender dos valores passados, e a Média Móvel (*MA*), onde o valor atual também dependerá dos erros passados. Como a precipitação apresenta uma sazonalidade, o modelo não apresentou resultados consistentes uma vez que o Erro Médio Absoluto (*MAE*) e o Erro Quadrático Médio (*RMSE*) - Desvio padrão dos valores residuais (erros de previsão) - apresentaram bons valores, ainda sim seus resultados foram utilizados para comparação com outros modelos. Além do modelo *ARMA*, o modelo *SARIMA* também está incluso da biblioteca, o *SARIMA* (*Seasonal Autoregressive Integrated Moving Average*) é o modelo mais completo e complexo, usado



também para prever séries temporais, diferentemente do modelo *ARMA* ele trabalha com sazonalidade, sendo uma extensão dos modelos *ARMA* e *ARIMA*, ele consegue identificar dependências entre valores passados e repetições sazonais, no nosso exemplo existe tanto uma variação constante anual, onde no início do ano chove muito e vai decaindo com o passar dos meses, e isso se repete para todos os anos. Porém os resultados não foram positivos, o forecast (previsões futuras) acabaram se repetindo, com vários dias possuindo a mesma média de chuva.

#### 2.6.7. statsmodels.tsa.arima.model

Para a utilização do modelo *ARIMA* tivemos que importar a biblioteca `statsmodels.tsa.arima.model`, o *ARIMA* é um modelo de previsão que para prever valores futuros ele se baseia em dados passados, então se a média de chuva de hoje é parecida com a de ontem, ele segue esse padrão, esse é o componente AutoRegressivo (*AR*) presente no próprio nome do modelo, o 'I' é de integrado, ajudando a tornar a série temporal estacionária, ou seja 'estável', removendo tendências para facilitar a previsão, essa remoção é feita a partir da subtração de valores sucessivos da série. E o pôr fim a Média Móvel (*MA*) que usa dos erros passados para ajustar as previsões futuras. Entre os três modelos utilizados o *ARIMA* foi o que apresentou os melhores resultados, tanto em relação às métricas de avaliação MAE e RMSE, quanto o gráfico que embora tenha subestimado os valores da chuva, conseguiu captar bem as tendências da chuva durante o ano, e o forecast apresentou resultados positivos, sem valores repetidos que aconteceu com o *SARIMA* ou inconsistentes como aconteceu com o modelo *ARMA*.

### 2.7. Visão geral do projeto

Para a criação dos modelos para realizar futuras previsões das variáveis, temperatura de orvalho, temperatura do ar, umidade relativa e precipitação total foi utilizado o google colab, um serviço hospedado do Jupyter Notebook que fornece um ambiente adequado para aprendizado de máquina e ciência de dados.

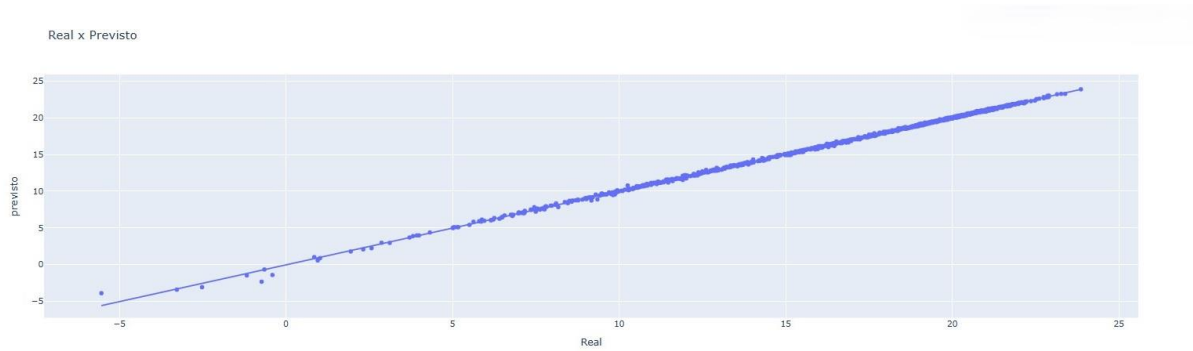


Figura 2 - Gráfico de dispersão do modelo *random forest* para previsão da temperatura do ponto de orvalho.

As métricas de avaliação  $R^2$  e  $MAE$  do modelo de floresta randômica obteve os melhores resultados obtendo 0.9993 para o coeficiente de determinação e 0.0720 para o erro médio absoluto.

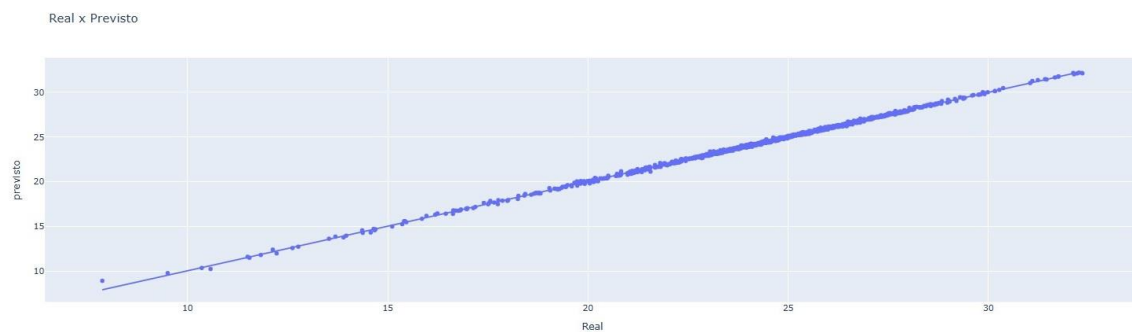


Figura 3 - Gráfico de dispersão do modelo *random forest* para previsão da temperatura do ar.

As métricas de avaliação  $R^2$  e  $MAE$  do modelo de floresta randômica obteve os melhores resultados obtendo 0.9991 para o coeficiente de determinação e 0.0713 para o erro médio absoluto.

Os gráficos gerados pelo modelo *ARIMA* são chamados de gráficos *forecast*, uma representação visual de uma previsão.

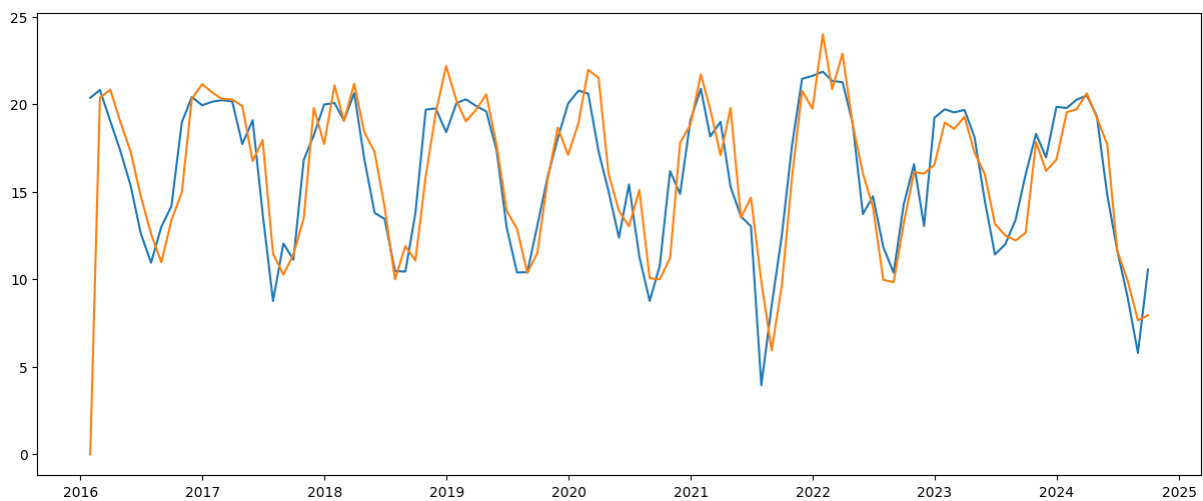


Figura 4 - gráfico *forecast* do modelo *ARIMA* para previsão da média mensal para temperatura do ponto de orvalho.

As métricas de avaliação *MSE* e *MAE* do modelo de *ARIMA* obteve como resultado para a média mensal, 8.0153 para o erro quadrático médio — desvio padrão dos valores residuais, ou seja, erros de previsão — e 1.7625 para o erro médio absoluto.

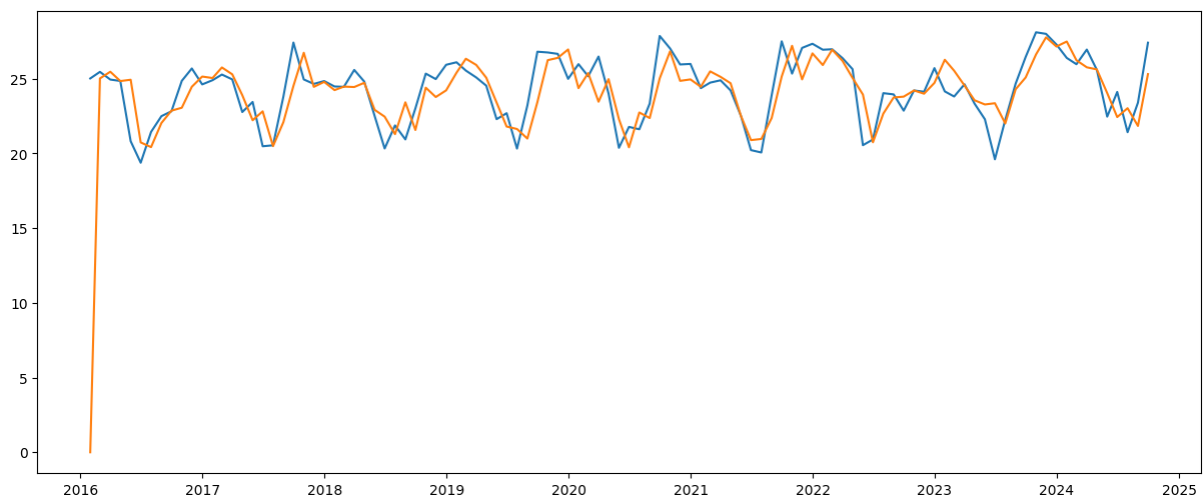


Figura 5 - Gráfico *forecast* do modelo *ARIMA* para previsão da média mensal para temperatura do ar.

As métricas de avaliação *MSE* e *MAE* do modelo de *ARIMA* obteve como resultado para previsão da média mensal para a temperatura do ar, 8.0153 para o erro quadrático médio e 1.7625 para o erro médio absoluto.

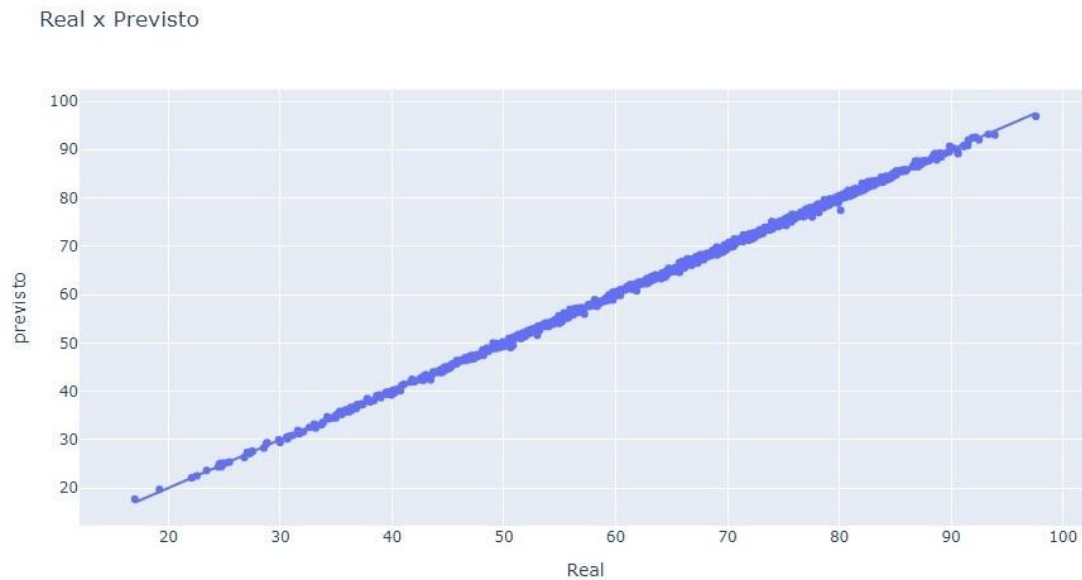


Figura 6 - Gráfico de dispersão do modelo *random forest* para previsão da umidade relativa.

As métricas de avaliação  $R^2$  e  $MAE$  do modelo de floresta randômica obteve os melhores resultados obtendo 0.9993 para o coeficiente de determinação e 0.2930 para o erro médio absoluto.

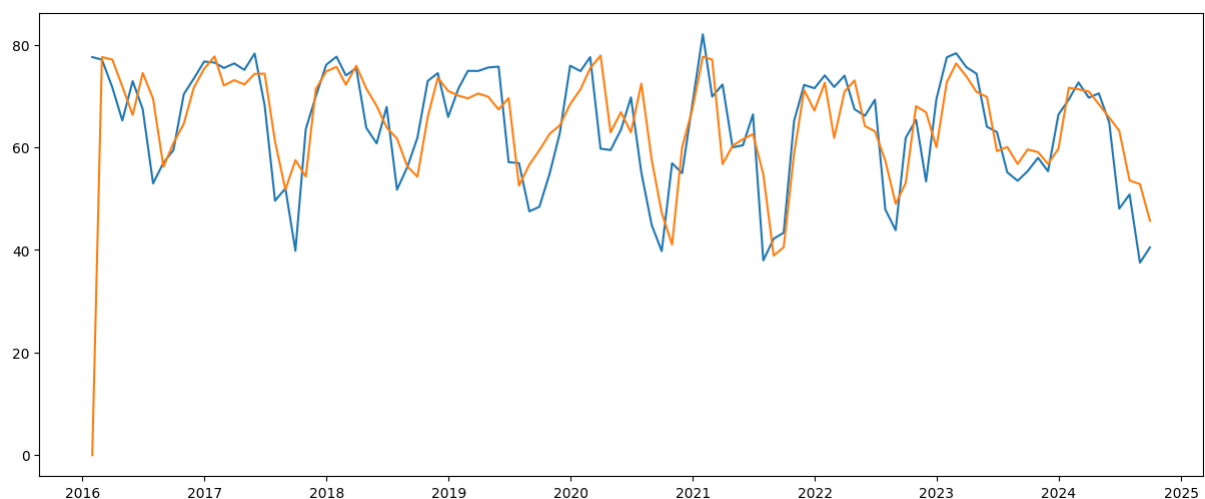


Figura 7 - Gráfico *forecast* do modelo *ARIMA* para previsão da média mensal para umidade relativa.

As métricas de avaliação  $MSE$  e  $MAE$  do modelo de *ARIMA* obteve como resultado para previsão da média mensal para a umidade relativa, 108.5 para o erro quadrático médio e 6.2583 para o erro médio absoluto.

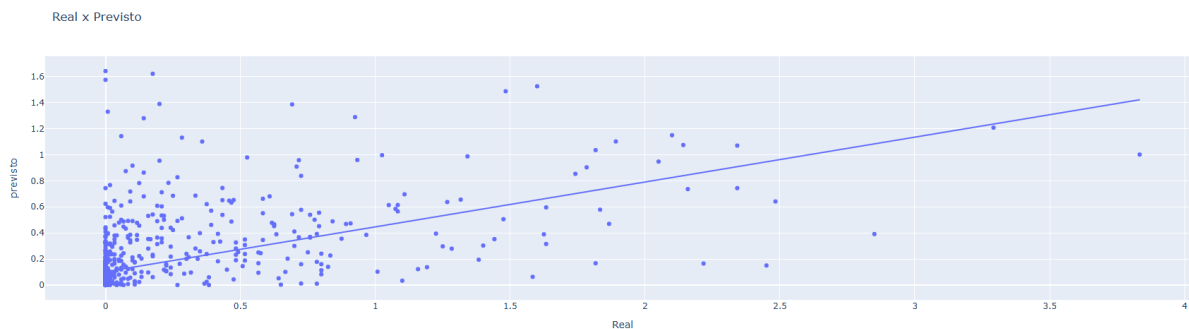


Figura 8 - Gráfico de dispersão do modelo random forest para previsão da precipitação total.

As métricas de avaliação  $R^2$  e  $MAE$  do modelo de floresta randômica obteve 0.2948 para o coeficiente de determinação e 0.1578 para o erro médio absoluto. Os valores das métricas de avaliação apresentavam valores positivos, mas o gráfico revela que o modelo tem um desempenho limitado, por isso foi necessário utilizar séries temporais.

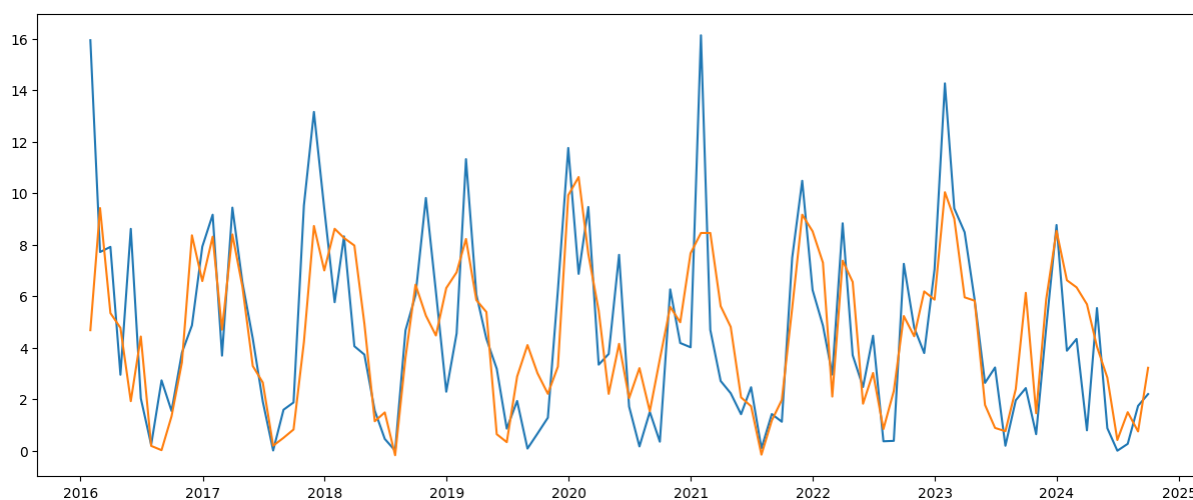


Figura 7 - Gráfico *forecast* do modelo *ARIMA* para previsão da média da precipitação mensal.

As métricas de avaliação  $MSE$  e  $MAE$  do modelo de *ARIMA* obteve os melhores resultados obtendo 6.6303 para o erro quadrático médio e 1.9036 para o erro médio absoluto. Como o gráfico mostra, o modelo *ARIMA* conseguiu capturar bem a tendência da chuva, demonstrando ser o melhor modelo para a previsão da média mensal da chuva para meses futuros.

### 3. ENCERRAMENTO DO PROJETO

#### 3.1. Relato Coletivo:

O presente trabalho teve por intuito a construção de modelos de aprendizado de máquina com regressão linear e modelos de séries temporais para prever as condições climáticas que impactam diretamente na agricultura, para a realização dessa atividade foi utilizado os modelos de regressão linear, *K-NN*, *Decision Tree* e *Random Forest*, e os modelos de séries de séries temporais *ARMA*, *ARIMA* e *SARIMA*. As métricas de avaliação apresentaram resultados excelentes e os gráficos demonstraram como os modelos conseguiram se adequar aos dados, dessa forma acredita-se que o objetivo desse trabalho de extensão foi alcançado com êxito.

## 3.2. Relato de Experiência Individual

### 3.2.1. CONTEXTUALIZAÇÃO

Neste trabalho, participei ativamente durante toda a etapa da coleta e formação do conjunto de dados, eu assumi a responsabilidade de criar os modelos de aprendizado de máquina com regressão linear, *k-NN*, árvore de decisão e floresta aleatória bem como as métricas de avaliação, com o apoio da professora, a criação foi rápida e simples, mas sua utilização se mostrou uma atividade desafiadora, havia vinte e quatro valores para cada variável climática, para um único dia. Junto com o Lucas, realizei a etapa de pré-processamento dos dados, preenchendo os valores faltosos e tirando a média diária para cada dia, após a etapa de pré-processamento foi gerado um novo *dataframe* e com os novos valores, o resultados começaram a ser positivos, porém um novo desafio apareceu, *Decision Tree* e *Random Forest* apresentaram um super ajuste ao dados, Lucas e eu tentamos resolver o overfitting reduzindo o número máximo de ramificação, mas os modelos ainda apresentaram um coeficiente de determinação muito próximo a um, e um erro absoluto médio muito próximo de zero, não sabendo mais como progredir passei para a próxima etapa. Colaborei com o Gustavo e Wellington para realizar os testes dos modelos, aproveitando para explicar o funcionamento de cada modelos e como eles poderiam utilizá-los para realizar seus próprios testes, todos conseguiram executar sem problemas. Na hora de realizar os testes de precipitação, ao qual fiquei responsável, percebi que os modelos não conseguiam capturar as tendências da chuva devido ao seu comportamento imprevisível, por recomendação da orientadora, fui em busca de aulas extracurriculares que ensinavam como utilizar os modelos de séries temporais, me voluntariei para estudar e criar o modelo *ARIMA* e com a ajuda do Lucas criamos também os

modelos *ARMA* e *SARIMA*, os modelos apresentaram resultados muito positivos, e achei interessante repassar eles para os outros membros do grupo, realizando o mesmo acompanhamento e explicações de como executar cada modelo, após o treinamento, enviei os melhores resultados para o Lucas realizar a análise dos gráficos. Além de minhas responsabilidades técnicas, também participei ativamente na elaboração e correção do relatório e banner do projeto, garantindo que todas as informações relevantes fossem incluídas e apresentadas de forma clara e concisa.

### 3.2.2. METODOLOGIA

#### 1. Coleta de dados.

A coleta de dados foi feita primeiramente escolhendo uma fonte de dados confiável sobre as informações do clima, e uma fonte que nos informasse sobre como as plantações são diretamente afetadas pelas condições climáticas, como temperatura do ar ou do solo, umidade, radiação, precipitação entre outros. As fontes encontradas foram o INMET para o clima e a EMBRAPA sobre a pré-produção dos cultivos. A coleta foi feita com a participação de todos os membros da equipe que se reuniram no Discord para a realização dessa etapa. Essa tarefa levou algumas horas, pois o site do INMET fornece duas formas de coletar os dados, a primeira maneira é bem limitada, pois podemos selecionar uma região específica, mas com um intervalo de tempo de apenas seis meses, e a outra maneira é capturando os doze meses para todas os estados e todos os municípios de cada estado, o que gerou um grande volume de dados, isso levou a equipe a realizar uma busca pelo melhor local a ser analisado, levando em consideração tanto no quesito de qualidade dos dados, quanto no quesito de ser um polo agrícola, Campo Grande foi o município escolhido, pois no ano de 2023 o município foi declarado como “Capital do Agro” segundo a Câmara Municipal de Campo Grande.

#### 2. Etapa de Pré-processamento.

A etapa de pré-processamento foi uma tarefa também realizada em grupo, em primeiro lugar fizemos uma análise para saber se havia linhas com valores nulos, após a confirmação nós convertimos todas as colunas para o tipo *float64*, com exceção das colunas, data que foi

convertido para *datetime*, depois disso utilizamos as funções '*ffill*' (preenchendo para frente usando o valor anterior) e '*bfill*' (preenchendo para trás usando o valor posterior). Após o preenchimento dos valores faltosos e a conversão dos tipos de dados, tiramos a mediana da coluna da radiação, porque os valores discrepantes, alteravam o valor da média, e para as outras colunas foi tirado uma média, após o tratamento dos dados um novo *dataframe* foi gerado para treinar os modelos.

### **3. Análise exploratória.**

Após a etapa de tratamento de dados, realizamos uma pesquisa utilizando o site da EMBRAPA para buscar as principais condições climáticas que influenciam na atividade agrícola, e como elas se correlacionam, as principais variáveis para os principais produtos exportados no Brasil, soja, frutas em geral e milho foram a temperatura, umidade e precipitação. A temperatura diz se o plantio da soja é adequado ou não, alta umidade pode acarretar o aparecimento de fungos nas frutas, e a baixa precipitação pode afetar as etapas de germinação e desenvolvimento do milho. Além disso, essas variáveis também estão diretamente relacionadas, pois com uma alta umidade provoca uma queda da temperatura, que ao se aproximar da temperatura do ponto de orvalho, ocasiona a condensação do vapor causando uma possível chuva.

### **4. Criação e teste dos modelos.**

A criação dos modelos *k-NN*, *Decision Tree* e *Random Forest* foi feita com base nos modelos construídos em sala com a orientadora Cynthia Maia, já a criação dos modelos de séries temporais foi feita utilizando conhecimentos adquiridos em sala junto de aulas, oficinas e pesquisas extracurriculares. Para o treinamento dos modelos, a divisão foi feita de forma simples, cada um pegou o que gostaria de analisar, Wellington se voluntariou para fazer os testes em relação a umidade máxima, mínima e relativa, Gustavo se prontificou para ficar com a temperatura do ponto de orvalho e temperatura do ar, e como eu me dispus a construir os modelos de séries temporais, eu fiquei responsável por treinar os modelos *ARMA*, *ARIMA* e *SARIMA* para buscar uma média de chuva dos meses seguintes. O treinamento foi feito em três



dias, no primeiro dia eu ajudei Wellington a utilizar todos os modelos para umidade, os comandos de regressão linear foram simples de aplicar, por termos aprendido em sala, mas a utilização dos modelos de séries temporais levou algum tempo, pois tive que explicar como era feito a utilização e como eles funcionavam. No dia seguinte o mesmo processo se repetiu, mas dessa vez com o Gustavo, treinando os modelos para análise da temperatura, e no dia final, eu realizei junto aos dois o treinamento de todos os modelos para análise da precipitação. No final de cada treinamento eram gerados e armazenados os gráficos para explicação visual do aprendizado dos modelos e o resultado numérico das métricas de avaliação  $R^2$ ,  $MAE$  e  $MSE$ .

## **5. Métricas de avaliação dos modelos.**

Nós utilizamos as métricas de avaliação aprendidos em sala para regressão linear.  $R^2$  (Coeficiente de Determinação) para ter um valor que nos ajudasse a entender o quão bem o modelo se ajusta aos dados, e o  $MAE$  (Erro Médio Absoluto) para avaliar o desempenho do modelo, só que para os modelos de séries temporais o coeficiente de determinação apresentou algumas limitações, pois ele não reflete bem o desempenho do modelo, já que a precipitação possui uma forte tendência, então no seu lugar usamos o  $MSE$  (Erro Quadrático Médio), que permite uma interpretação mais intuitiva dos erros.

Além das métricas quantitativas, nós geramos gráficos para podermos analisar de forma visual o aprendizado dos modelos, já que os valores apresentando resultados bons não significava necessariamente que o modelo se adequou aos dados, isso aconteceu bastante quando tentei usar os modelos de regressão linear para analisar a precipitação, o  $R^2$  e o  $MAE$  apresentavam valores positivos, mas o gráfico apresentava uma má adequação do modelo aos dados.

### **3.2.3. RESULTADOS E DISCUSSÃO**

As etapas iniciais do projeto foram as mais difíceis, primeiramente, queríamos analisar o clima de Fortaleza, porém, o INMET não capturou a maioria das variáveis necessárias para o estudo, o conjunto de dados estava quase vazio então tivemos que mudar o local da análise, escolhemos Campo Grande, município do Mato Grosso do Sul, por ser um polo agrícola que

possuía um conjunto de dados mais completo em comparação ao de Fortaleza, entretanto o novo conjunto de dados ainda apresentava uma formatação inadequada para o uso, havia variáveis com tipos inconsistentes, dados faltosos, e múltiplos valores para um mesmo dia. A utilização desses dados somados com a utilização dos modelos de classificação, que não se adequava ao nosso caso, foi muito frustrante, tudo o que podíamos fazer era uma análise exploratória e realizar o tratamento dos dados, tarefa realizada sem muitas dificuldades. A orientação da professora foi fundamental para a continuidade do projeto, os conselhos de criar um *dataframe* com a média dos valores diários e a utilização dos modelos de regressão linear foi o que garantiu o sucesso do projeto. Toda a teoria apresentada em sala de aula trouxe uma nova visão para como progredir com o projeto. Esperamos que nossos modelos sejam capazes de prever com sucesso as condições climáticas para os próximos meses, ajudando não somente os agricultores, grandes ou pequenos, mas também a população da região analisada.

#### 3.2.4. REFLEXÃO APROFUNDADA

O projeto se demonstrou uma experiência desafiadora, levou horas para conseguir gerar o conjunto de dados adequado para uso, também tivemos que realizar diversas pesquisas no INMET na EMBRAPA em artigos e sites web, toda essa etapa foi muito complicada, contudo, graças a orientadora, os modelos foram de fácil construção, embora os de séries temporais tenha dado certa dor de cabeça e gerado desconfiança nos membros da equipe pelo seu super ajuste aos dados. Complicações à parte, eu acredito que o projeto cumpriu o que prometeu, conseguimos criar modelos para prever as condições do clima para os próximos meses. Em geral esse projeto foi uma experiência enriquecedora e incentivadora para os membros do grupo, utilizaremos todo o conhecimento aprendido para realizar futuras análises dos dados que capturaremos no trabalho de IOT.

#### 3.2.5. CONSIDERAÇÕES FINAIS

Durante as pesquisas sobre os impactos do clima e como a previsão das variáveis climáticas podem ajudar os agricultores, nos deparamos com uma tecnologia alternativa para essas previsões, utilizando

de redes neurais para a projeção climática no Brasil. A utilização seria feita pelo treinamento de uma rede neural artificial usando o algoritmo de *backpropagation* relacionando as variações de temperatura com alterações locais (latitude, longitude e altitude) e alterações temporais (ano e mês), os dados poderiam ser séries históricas de estações meteorológicas convencionais do INMET. O uso de redes neurais permitiria a modelagem das variações espaço-temporais da temperatura e a projeção de tendências futuras.

## 4. BIBLIOGRAFIA

EMBRAPA (Brasília). Ministério da Agricultura e Pecuária. **Cultivos**. 1973.

Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos>. Acesso em: 14 ago. 2024.

BRASÍLIA. Instituto Nacional de Meteorologia. Ministério da Agricultura e Pecuária. **DADOS HISTÓRICOS ANUAIS**. 1909. Disponível em:

<https://portal.inmet.gov.br/dadoshistoricos>. Acesso em: 14 ago. 2024.

MCKINNEY, Wes. **Pandas - Python Data Analysis Library**. 2008. Disponível em: <https://pandas.pydata.org/docs/>. Acesso em: 15 ago. 2024.

PLOTLY INC.. **Plotly Express**. 2019. Disponível em:

<https://plotly.com/python/plotly-express/>. Acesso em: 14 ago. 2024.

COURNAPEAU, David. **Scikit-learn: machine learning in Python**. 2007.

Disponível em: <https://scikit-learn.org/stable/>. Acesso em: 14 ago. 2024.

SCIKIT-LEARN DEVELOPERS (França). **Model selection and evaluation**.

2007. Disponível em: [https://scikit-learn.org/stable/model\\_selection.html](https://scikit-learn.org/stable/model_selection.html).

Acesso em: 14 ago. 2024.

SCIKIT-LEARN DEVELOPERS (França). **NearestNeighbors**. 2007. Disponível

em: [https://scikit-](https://scikit-learn.org/dev/modules/generated/sklearn.neighbors.NearestNeighbors.html)

[learn.org/dev/modules/generated/sklearn.neighbors.NearestNeighbors.html](https://scikit-learn.org/dev/modules/generated/sklearn.neighbors.NearestNeighbors.html).

Acesso em: 14 ago. 2024.

SCIKIT-LEARN DEVELOPERS (França). **Decision Trees**. 2007. Disponível em: <https://scikit-learn.org/1.5/modules/tree.html>. Acesso em: 14 ago. 2024.

SCIKIT-LEARN DEVELOPERS (França). **Ensemble methods**. 2007. Disponível em: <https://scikit-learn.org/1.5/modules/ensemble.html>. Acesso em: 14 ago. 2024.

SCIKIT-LEARN DEVELOPERS (França). **Metrics and scoring**: quantifying the quality of predictions. quantifying the quality of predictions. 2007. Disponível em: [https://scikit-learn.org/1.5/modules/model\\_evaluation.html](https://scikit-learn.org/1.5/modules/model_evaluation.html). Acesso em: 14 ago. 2024.

SCIKIT-LEARN DEVELOPERS (França). **Sklearn.linear\_model**. 2007. Disponível em: [https://scikit-learn.org/1.5/api/sklearn.linear\\_model.html](https://scikit-learn.org/1.5/api/sklearn.linear_model.html). Acesso em: 14 ago. 2024.

OLIPHANT, Travis. **Numpy**. 2005. Disponível em: <https://numpy.org/>. Acesso em: 14 ago. 2024.

PERKTOLD, Josef *et al.* **Statsmodels.graphics.tsaplots.plot\_acf**. 2009. Disponível em: [https://www.statsmodels.org/dev/generated/statsmodels.graphics.tsaplots.plot\\_acf.html](https://www.statsmodels.org/dev/generated/statsmodels.graphics.tsaplots.plot_acf.html). Acesso em: 14 ago. 2024.

PERKTOLD, Josef; SEABOLD, Skipper; TAYLOR, Jonathan. **Statsmodels.graphics.tsaplots.plot\_pacf**. 2009. Disponível em: [https://www.statsmodels.org/dev/generated/statsmodels.graphics.tsaplots.plot\\_pacf.html](https://www.statsmodels.org/dev/generated/statsmodels.graphics.tsaplots.plot_pacf.html). Acesso em: 14 ago. 2024.

PERKTOLD, Josef; SEABOLD, Skipper; TAYLOR, Jonathan. **Statsmodels.tsa.statespace.sarimax.SARIMAX**. 2009. Disponível em: <https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html>. Acesso em: 14 ago. 2024.

PERKTOLD, Josef; SEABOLD, Skipper; TAYLOR, Jonathan. **Statsmodels.tsa.arima.model.ARIMA**. 2009. Disponível em:

<https://www.statsmodels.org/dev/generated/statsmodels.tsa.arima.model.ARIMA.html>. Acesso em: 14 ago. 2024.

ALMEIDA, Ivan Rodrigues de. *O clima como um dos fatores de expansão da cultura da soja no Rio Grande do Sul, Paraná e Mato Grosso*. 2005. 119 f. : il. Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia, 2005.

BIGOLIN, Tiago. *Mudanças climáticas e o sistema de duplo cultivo milho-soja em clima subtropical*. 2023. 54 f. Tese (Doutorado) - Curso de Programa de Pós-Graduação em Agronegócios, Centro de Estudos e Pesquisas em Agronegócios, Universidade Federal do Rio Grande do Sul, Porto Alegre, BR-RS, 2023. Cap. 1.

OLIVEIRA, Samira França; PRADO, Rachel Bardy; MONTEIRO, Joyce Maria Guimarães. **Impactos das mudanças climáticas na produção agrícola e medidas de adaptação sob a percepção de atores e produtores rurais de Nova Friburgo, RJ**. Campo Grande: Creative Commons Attribution, 2022. 24 p. Disponível em:  
<https://ainfo.cnptia.embrapa.br/digital/bitstream/doc/1150320/1/Impactos-das-mudancas-climaticas-na-producao-agricola-2022.pdf>. Acesso em: 14 ago. 2024.

EMBRAPA, 2009, Brasília. **Uso de redes neurais para a projeção climática no Brasil: temperatura mínima**. Disponível em:  
<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/58117/1/Usos-redes.pdf>. Acesso em: 14 ago. 2024.

## 5. CÓDIGO

### 1. Pré-processamento

<https://colab.research.google.com/drive/1BrpPMoxB-ZfRdRh4JDqR6NhuqvkeROBh?usp=sharing>

### 2. Regressão Linear

[https://colab.research.google.com/drive/1C5xKPfkZijWu8bdst0hKp\\_Xo2usA02lO?usp=sharing](https://colab.research.google.com/drive/1C5xKPfkZijWu8bdst0hKp_Xo2usA02lO?usp=sharing)

### **3. Séries Temporais**

<https://colab.research.google.com/drive/1Nb1grvBfg2R6-BtrEFMNKJysw5bB7D88?usp=sharing>

### **4. Dataframe**

<https://docs.google.com/spreadsheets/d/1NMuXEJdqdHtl-JhcUAku520CKwYBxRe/edit?usp=sharing&ouid=102395861621363319264&rtpof=true&sd=true>

### **5. Análise dos dados.**

#### **5.1. Análise da temperatura do ponto de orvalho**

[https://drive.google.com/file/d/1lhba2r07L8yZU4porJhcn3KuRY8\\_vFhZ/view?usp=sharing](https://drive.google.com/file/d/1lhba2r07L8yZU4porJhcn3KuRY8_vFhZ/view?usp=sharing)

#### **5.2. Análise da temperatura do ar - bulbo seco**

<https://drive.google.com/file/d/1WXKPdwSvRoSpgqpe84Is73VVwl81rmID/view?usp=sharing>

#### **5.3 Análise da umidade do ar**

<https://drive.google.com/file/d/1KMWZsJkDExcFXizwlkOycCAj4e4XJs9j/view?usp=sharing>

#### **5.4 Análise da chuva**

<https://drive.google.com/file/d/1t7bm-hF1FX9wKwwpVI6Mv92K0NjvHg1/view?usp=sharing>