

# **ROTEIRO DO RELATO INDIVIDUAL**



# **Estácio**

**UNIVERSIDADE ESTÁCIO DE SÁ**

**CAMPUS PARANGABA**

**ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**

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

**Fernando Chaves - 202308423991**

**Orientador: Cynthia Moreira Maia**

**2024**

**FORTALEZA/CE**

## 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, os 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.

## **2. OBJETIVOS**

O objetivo do projeto era poder analisar as condições climáticas para ajudar algum setor da sociedade, principalmente com os efeitos do aquecimento global, que dificulta a análise do clima e acaba afetando a população em geral. A equipe escolheu levar a análise para o setor da agricultura, os resultados gerados pelos modelos podem ajudar não somente pequenos e grandes agricultores a tomarem medidas preventivas para evitar a perda de parte de sua safra, mas também a população, com a informação da precipitação, regiões em áreas de risco podem tomar medidas preventivas para tentar reduzir os danos gerados por enchentes e alagamentos, sabendo que a umidade e temperatura altas podem causar sérias consequências à saúde das pessoas, o governo pode alertar a população sobre essas condições, além de orientar sobre como se proteger, como beber mais água e evitar exposição direta ao sol.

## **3. METODOLOGIA**

### **3.1. 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 todos 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.

### **3.2. 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 convertemos 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.

### **3.3. 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.

### **3.4. 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*.

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

Nós utilizamos as métricas de avaliação aprendidas 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 significa 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.

#### **4. 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.

#### **5. 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.

## 6. 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.