

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
PROJETOS EMBARCADOS COM CIÊNCIA DE DADOS

HUGO SILVA RODRIGUES

**PROJETO**

CORNÉLIO PROCÓPIO  
2023

## SUMÁRIO

<b>1 – Introdução</b>	<b>1</b>
1.1 Objetivo	1
<b>2 – Referencial Teórico</b>	<b>2</b>
<b>3 – Metodologia</b>	<b>3</b>
3.1 Materiais	3
3.2 Coleta de Dados	3
3.3 Repositório dos dados	4
<b>4 – Análise dos dados</b>	<b>5</b>
4.1 Impacto do Momento da Coleta	7
<b>5 – Criação do Modelo</b>	<b>8</b>
<b>6 – Conclusão</b>	<b>9</b>
<b>Referências</b>	<b>10</b>

## 1 Introdução

Tomar decisões faz parte do cotidiano, sejam elas simples ou complexas. Frequentemente, é utilizado suposições ou conhecimentos adquiridos anteriormente para escolher a melhor alternativa. Contudo, seria essa realmente a melhor estratégia? Com os avanços atuais, a tecnologia se tornou uma ferramenta para auxiliar na tomada de decisões mais eficazes (CHAVES et al., 2022).

O campo de Tecnologia da Informação (TI) desempenha um papel crucial nas empresas modernas, gerenciando dispositivos que processam e compartilham dados. Resumidamente, ele organiza e interpreta informações para extrair insights valiosos, que, por sua vez, apoiam a tomada de decisões estratégicas (CHAVES et al., 2022, p. 7).

Segundo Chaves et al. (2022, p. 9), a TI envolve três etapas principais: a geração de dados, o processamento desses dados e, finalmente, a produção de informações que são utilizadas para tomar decisões.

Vários setores já estão utilizando a tecnologia em seus processos decisórios, e a agricultura é um exemplo onde essa integração se mostra cada vez útil, (RURAL, 2004).

Com a evolução tecnológica, surgiram novos termos e conceitos, sendo um dos mais notáveis o IoT (Internet das Coisas). Essa inovação representa uma evolução da internet convencional, onde dispositivos conectados entre si pela internet coletam e compartilham dados, ampliando o alcance e o impacto da rede mundial (SANTOS et al., 2019).

A aquisição de dados vem se tornando cada vez mais fácil e prática, graças evolução gerada pela IoT. Atualmente, existem dispositivos que podem ser aplicados de diferentes formas para a obtenção de dados e, posteriormente, tratá-los para ajudar na tomada de decisão. Um dos mais conhecidos é o ESP32 (MORAIS; SHAEFFER; MARCELINO, 2023, pp.7–9).

Com diversos tipos de conectividade, como Wi-Fi e Bluetooth, o ESP32 oferece uma ampla gama de aplicações, sendo adequado tanto para projetos estudantis quanto para aplicações profissionais (HERCOG et al., 2023).

### 1.1 Objetivo

Diante desse cenário de evolução tecnológica e o uso de dados para tomada de decisões, o presente estudo tem como foco apresentar a utilização do ESP32 em um sistema de aquisição de dados meteorológicos, juntamente com a aplicação de técnicas de aprendizado de máquina para prever a ocorrência de chuva.

## 2 Referencial Teórico

Conforme apresentado em [Chaves et al. \(2022\)](#) , as empresas precisam de dados tratados que geram informações essenciais para se manterem competitivas, esse processo está se expandindo para além do mundo empresarial.

A tomada de decisão, orientada por insights provenientes dessas informações, resulta principalmente em melhorias na produtividade e no valor agregado ao produto ([CHAVES et al., 2022](#)).

Com a necessidade crescente de analisar informações e tomar decisões com base nelas, os softwares de apoio à decisão se tornaram mais relevantes e úteis, especialmente com o aumento da quantidade de dados disponíveis ([CHAVES et al., 2022](#)). Um dos softwares que mais se destaca é o Power BI, da Microsoft, que facilita a criação de dashboards práticos e fornece insights valiosos para a tomada de decisão ([GARTNER, 2022](#)) .

Na busca por maximizar o potencial dos dados, as decisões automatizadas ganharam espaço, utilizando inteligência artificial para dar às máquinas a capacidade de aprender e tomar decisões ([SANTOS, 2023](#), pp.26–27).

Com o uso de inteligência artificial, é possível acelerar ainda mais o processamento e a tomada de decisão, permitindo que a IA tome decisões de forma instantânea com base nos padrões aprendidos durante seu treinamento ([SANTOS, 2023](#), pp.14).

### 3 Metodologia

Conforme apresentado na [Seção 1.1](#), o foco deste estudo é a classificação de chuva com base nas condições atmosféricas específicas de um dado momento. Para isso, será implementado um sistema de aquisição de dados utilizando um microcontrolador ESP32 e uma série de sensores. Os dados coletados serão enviados e armazenados na plataforma ThingSpeak, que servirá como repositório para análises futuras por meio da linguagem de programação Python.

#### 3.1 Materiais

A montagem do sistema de aquisição de dados exigiu os seguintes componentes:

- 1 microcontrolador ESP32;
- 1 Sensor BME280, responsável pela coleta de dados de temperatura, umidade, altitude e pressão atmosférica;
- 1 Sensor de detecção de chuva;
- 2 Protoboards de 830 pontos cada;
- 1 Fonte de alimentação compatível com as protoboards.

Inicialmente, planejou-se incorporar o sensor GUV-A-S12SD para medir a intensidade dos raios UV. No entanto, devido a um defeito no componente, ele foi excluído do projeto.

A programação do ESP32 e a integração com os sensores foram realizadas através da plataforma Arduino. Embora o armazenamento de dados no Google Sheets fosse uma opção inicial, optou-se pelo ThingSpeak devido à sua maior eficiência e praticidade.

#### 3.2 Coleta de Dados

Para assegurar a robustez dos dados, realizaram-se medições em diferentes condições climáticas: dias ensolarados, chuvosos, nublados e também durante a noite. As medições em dias quentes ocorreram sem contratempos. Contudo, a coleta de dados em dias chuvosos foi menos frequente e exigiu cuidados adicionais para evitar danos aos equipamentos.

De forma a preservar a integridade dos componentes eletrônicos, as medições de chuva foram realizadas sob uma cobertura. Com cuidado, posicionou o sensor de chuva de maneira que pudesse registrar as precipitações de forma adequada. No entanto, em momentos em que havia risco de danos aos componentes, a coleta de dados continuou a ser feita pelo sensor BME280, enquanto as leituras do sensor de chuva foram ajustadas manualmente para refletir as condições esperadas.

### 3.3 Repositório dos dados

Os dados coletados foram armazenados no ThingSpeak e possuem o formato apresentado na [Tabela 1](#).

Nome da Coluna	Tipo de Dado
created_at	object
temperatura	float64
pressao	float64
altitude	float64
humidade	float64
chuva (Variável a prever)	int64

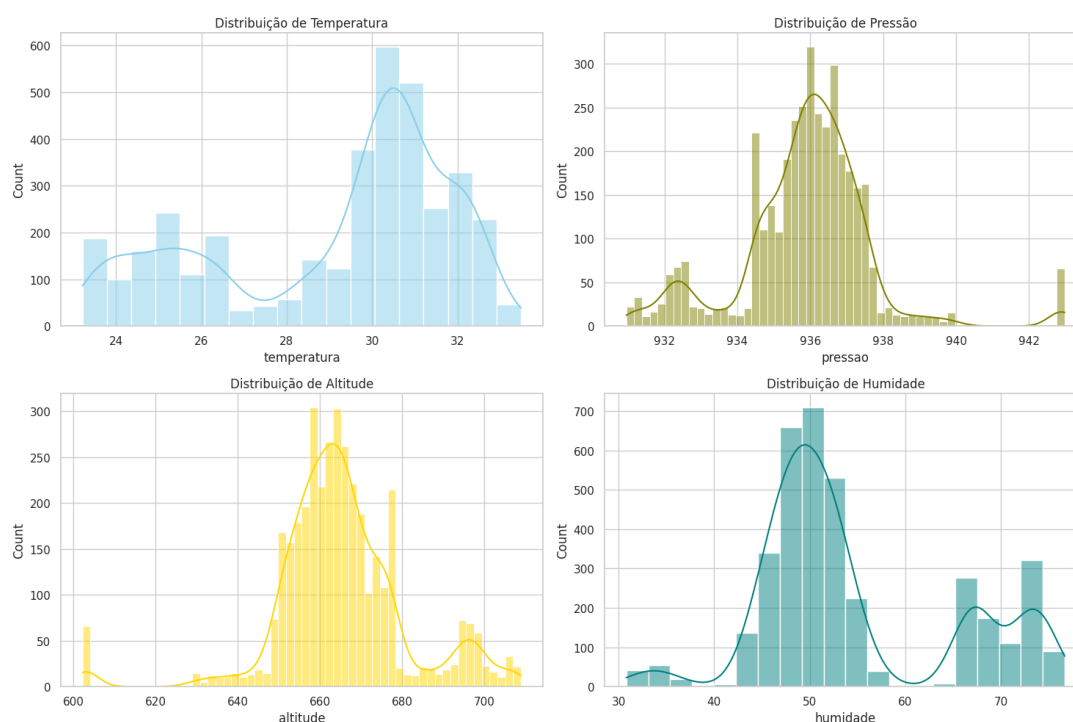
Tabela 1 – Dados

Sendo a coluna "created" responsável por armazenar os dados de dia, hora e minuto de cada coleta.

## 4 Análise dos dados

Ao iniciar a análise, foi realizado o tratamento de dados ausentes, e posteriormente analisado a distribuição das variáveis temperatura, humidade, pressão e altitude, [Figura 1](#).

Figura 1 – Plano de fase resultante da aplicação do SMC em um sistema não linear



Fonte: Autoria Própria

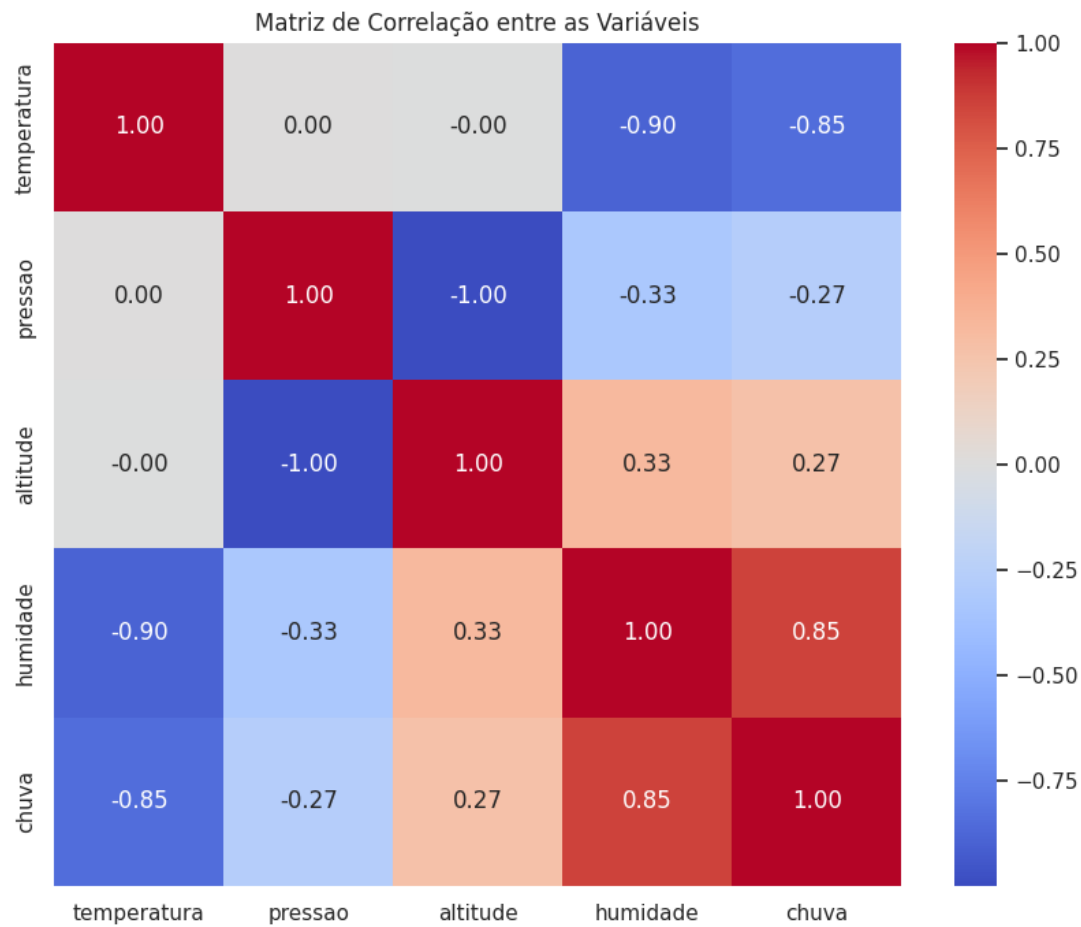
Conforme ilustrado na [Figura 1](#), as variáveis temperatura e humidade demonstraram as maiores variações, sugerindo que elas podem ter um impacto significativo no aprendizado do modelo de previsão de chuva.

Para explorar a correlação entre as variáveis, foi plotado uma matriz de correlação [Figura 2](#).

Como esperado, as variáveis que mostraram a maior correlação com a chuva foram a humidade, com uma relação positiva de 85%, e a temperatura, com uma relação negativa de -85%. Para um entendimento mais aprofundado dessas duas variáveis, foram criados gráficos de violino [Figura 3](#).

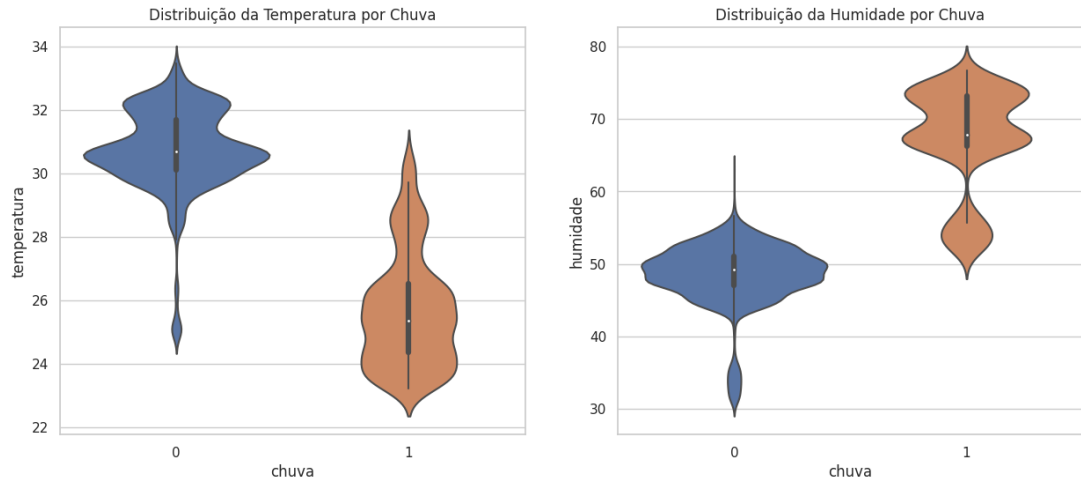
A [Figura 3](#) reforça as observações anteriores. No caso da chuva, é observado variações nos valores de temperatura, com períodos chuvosos geralmente apresentando temperaturas mais baixas, o que indica uma correlação negativa. Por outro lado, para a humidade, valores mais elevados são esperados em períodos chuvosos, confirmando a relação positiva entre chuva e humidade.

Figura 2 – Matriz de Correlação



Fonte: Autoria Própria

Figura 3 – Distribuição Temperatura e Chuva



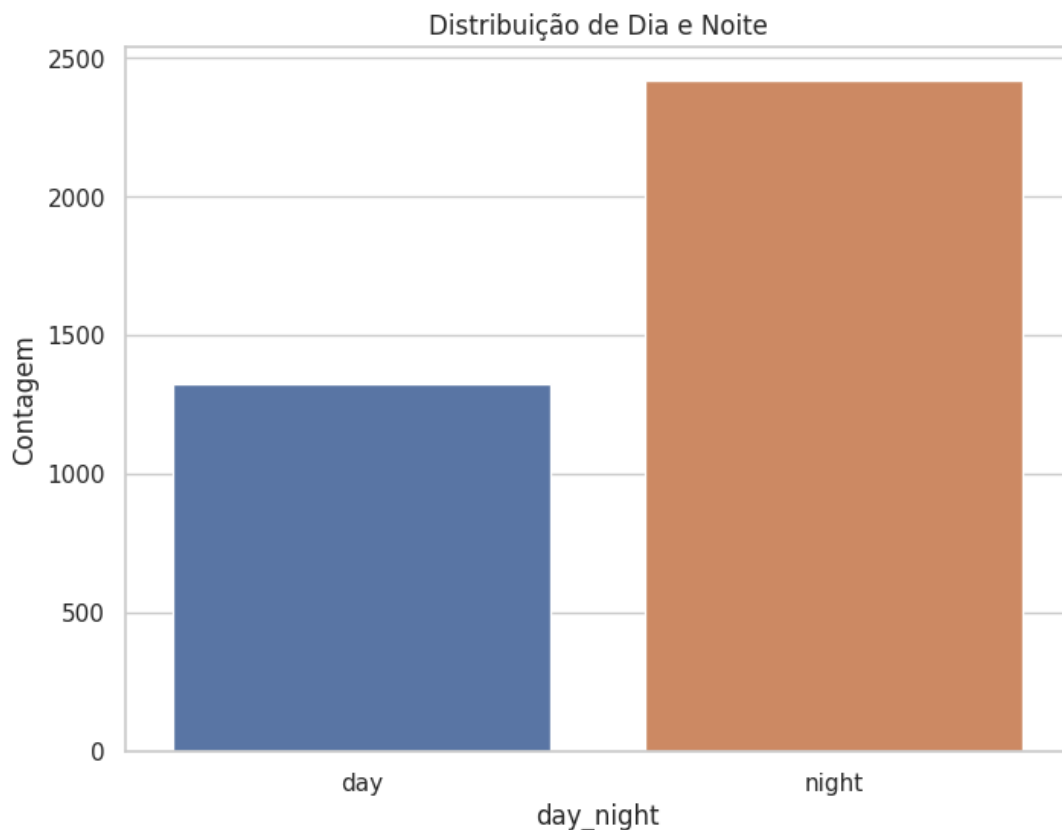
Fonte: Autoria Própria



#### 4.1 Impacto do Momento da Coleta

De acordo com UFPR (), normalmente as temperaturas mais altas são registradas durante o dia, enquanto as mais baixas ocorrem durante a noite. Portanto, foi reestruturado a coluna 'created' para gerar uma variável categórica, representando dia ou noite. Isso permite que o modelo diferencie entre temperaturas baixas que são típicas do período noturno e variações térmicas durante o dia. A Figura 4 apresenta a distribuição dessa nova coluna.

Figura 4 – Distribuição Temperatura e Chuva



Fonte: Autoria Própria

Conforme apresentado na Figura 4, observa-se que existem mais dados coletados durante o período noturno.

## 5 Criação do Modelo

Após a conclusão da análise dos dados, foi dado início a criação do primeiro modelo, um RandomForestClassifier padrão, sem qualquer modificação em seus parâmetros, foi aplicado uma validação cruzada de 5 etapas. Ao expor o modelo aos dados de teste, foram obtidos os resultados apresentados na [Tabela 2](#).

Métrica	Valor
Precision	0.9882
Recall	0.9654
F1 Score	0.9768

Tabela 2 – Resultado da Avaliação do Modelo RandomForestClassifier

Como é possível constatar, mesmo sendo um modelo simples, ele já demonstrou a capacidade de classificar os dados com excelência.

Foram criados mais três modelos de classificação para fins de teste: Logistic Regression, Naive Bayes, e Support Vector Machine. Todos os modelos foram submetidos a uma validação cruzada de 5 etapas e configurados sem nenhuma modificação em seus parâmetros. Ao serem expostos aos dados de teste, os resultados obtidos para cada modelo estão apresentados na [Tabela 3](#).

Modelo	Precision	Recall	F1 Score
Logistic Regression	0.9706	0.9706	0.9705
Naive Bayes	0.9323	0.9324	0.9314
Support Vector Machine	0.9742	0.9742	0.9741

Tabela 3 – Resultado da Avaliação

## 6 Conclusão

A utilização de dados para tomada de decisões é uma realidade e, hoje mais do que nunca, está em constante evolução, devido ao aumento de dados e às maneiras de obtê-los. Durante este projeto, utilizou-se o ESP32 para obter dados meteorológicos, visando a posterior classificação. Demonstrando que uma abordagem simples com um microcontrolador e alguns sensores pode gerar um grande estudo.

O modelo RandomForestClassifier apresentou um ótimo resultado na classificação dos dados de teste, sendo escolhido como o modelo final para o projeto. Porém, vale ressaltar que, na realidade, todos os modelos testados tiveram resultados satisfatórios.

A base de dados se apresentou simples, visto que todos os modelos tiveram bons scores. Portanto, expandir os dados para outras localizações geraria mais informações de aprendizado para o modelo, permitindo sua generalização. Contudo, por se tratar de um estudo, a forma como os dados são obtidos gera complicações, especialmente para captar dados de dias chuvosos, devido aos riscos para os componentes.

Com a utilização de mais dados, seria possível realizar o deploy do modelo para uso em uma API de previsão meteorológica.

## Referências

CHAVES, L. et al. A importância da tecnologia da informação nas tomadas de decisões em empresas. **Tecnologia da Informação e Comunicação: pesquisas em inovações tecnológicas**, v. 3, 2022. Citado 2 vezes nas páginas 1 e 2.

GARTNER. **Gartner Magic Quadrant for Analytics and Business Intelligence Platforms**. 2022. Acesso em: 3 de novembro de 2023. Citado na página 2.

HERCOG, D. et al. Design and implementation of esp32-based iot devices. **Sensors**, v. 6739, 2023. Citado na página 1.

MORAIS, J. B. M. de; SHAEFFER, L.; MARCELINO, R. Utilização de iot para aquisição de dados e monitoramento no processo de estampagem industrial. **Revista Foco**, v. 16, 2023. Citado na página 1.

RURAL, I. de Planejamento e Economia Agrícola de Santa Catarina Secretaria de Estado da Agricultura e P. Impacto da tecnologia da informação (ti) sobre o processo decisório do agricultor familiar. 2004. Citado na página 1.

SANTOS, A. S. D. **Machine Learning: Uma Proposta De Guia Orientativo Para O Exercício Do Direito De Revisão Das Decisões Automatizadas**. 77 p. Dissertação (Mestrado) — Fundação Getulio Vargas, São Paulo, 2023. Citado na página 2.

SANTOS, B. P. et al. Internet das coisas: da teoria à prática. **Departamento de Ciência da Computação Universidade Federal de Minas Gerais (UFMG)**, 2019. Citado na página 1.

UFPR. **Controles Da Temperatura**. Citado na página 7.