



**FACULDADE DE CIÊNCIAS**  
**Departamento de Matemática e Informática**

Trabalho de Licenciatura em Informática

**Sistema de Previsão do Preço do Milho em  
Moçambique baseado em Machine  
Learning**

**Autor: Hérco Fernando Taimo Zau Zau**

Maputo, 05 de Dezembro de 2025



**FACULDADE DE CIÊNCIAS**  
**Departamento de Matemática e Informática**

Trabalho de Licenciatura em Informática

**Sistema de Previsão do Preço do Milho em  
Moçambique baseado em Machine Learning**

**Autor:** Hérco Fernando Taimo Zau Zau

**Supervisor:** Prof. Doutor Cachimo Combo Assane

Maputo, 05 de Dezembro de 2025

## **Dedicatória**

*A todas as pessoas que usam a ciência a serviço  
da humanidade.*

## Declaração de Honra

Declaro por minha honra que o presente Trabalho de Licenciatura é resultado da minha investigação e que o processo foi concebido para ser submetido apenas para a obtenção do grau de Licenciado em Informática na faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, 05 de dezembro de 2025

---

Hérco Fernando Taimo Zau Zau

## Agradecimentos

A Deus, que nos permitiu alcançar os mistérios da criação por meio da ciência e nos abençoou com mentes brilhantes ao longo dos séculos, e que, por intermédio de Cristo Jesus, tem sido a base da minha carreira científica desde a infância.

Aos meus pais, Jerónimo Zau Zau e Arminda Zau Zau, que me deram o sonho de ser cientista e todas as bases para conseguir.

À minha namorada, Cecília Langa, pelo suporte incondicional.

Aos meus colegas, sem os quais não teria conseguido terminar o curso: Keil Bambo, Milton Maleiane, Stanley Nhampossa, Tatiana Langa, Francisco Domingos, Edilson Ricardo, Augusto Chissano, Jaime Beatriz, Klésia Chiwissa, António Sambo, Alberto Dgedge, Grácio Macuácuca, Hélio Simango, Hermon da Cruz, Luís Milice, Malagy Casimo, Marcelo Boca, Milagre Chavanguane, Milton Pereira, Paulo Massingue, Igor Idércio, Ziad Ibrahim, Barnabé Bila, Cecília Moiane, Berto Timana e outros.

Aos docentes que fizeram parte desta jornada com muito apoio, em especial ao Prof. Doutor Cachimo Assane e à Dra. Judite Mandlate.

Às pessoas que apostaram no meu potencial como cientista: Ana Mangana, Eng. Rogério Lam, Fernando Gomes e Joaquim Gomes.

E em especial, ao amigo que me direcionou para o curso de informática e deu todo o suporte,  
Pereira F. Lupaka.

## Resumo

O presente trabalho visa desenvolver e implementar um sistema web para previsão dos preços do milho no mercado moçambicano, recorrendo a técnicas de Machine Learning (aprendizagem de máquina) e uso de inteligência artificial, apresentando uma solução moderna como diferencial às abordagens tradicionais. Com o avanço corrente das plataformas web que disponibilizam dados públicos, instituições como ONG's têm disponibilizado dados de Moçambique, entre estes, dados históricos do mercado agrícola. Estes dados, cada vez mais crescentes e acessíveis, incluem variáveis-chave para responder ao problema da crescente necessidade de previsões precisas das flutuações do preço do milho em Moçambique. Deste modo, torna-se necessário o estudo aprofundado de estratégias de Machine Learning baseadas em modelos estatísticos. A identificação do modelo mais apropriado, com parâmetros otimizados e resultados acurados, é um desafio significativo para garantir uma solução global para o problema da pesquisa. Dentro de conjuntos de experimentos conduzidos, foram comparados três algoritmos de aprendizagem de máquina, nomeadamente Regressão Multilinear, LSTM e Random Forest, aplicados a volumes de dados históricos do mercado do milho em Moçambique. O modelo LSTM foi selecionado como o mais eficaz para ser o mecanismo central de previsão do sistema proposto. O sistema final integra um backend desenvolvido em Django, responsável por disponibilizar uma plataforma web funcional para visualização de previsões. Adicionalmente, o sistema final incorpora um modelo de IA para chatbot, concebido para interagir com os utilizadores, permitindo esclarecer dúvidas, explicar resultados das previsões e facilitar o acesso à informação de forma intuitiva e conversacional. A combinação de modelos robustos e dados de alta qualidade oferece um caminho para prever com mais precisão as dinâmicas do mercado, promovendo a estabilidade económica e uma melhor gestão dos recursos agrícolas.

**Palavras-chave:** previsão de preços, mercado agrícola, Machine Learning, regressão multilinear, random forest, LSTM.

## Abstract

The present work aims to develop and implement a web-based system for forecasting maize prices in the Mozambican market, using Machine Learning techniques and artificial intelligence, introducing a modern solution as a differentiating approach compared to traditional methods. With the ongoing growth of web platforms that provide public data, institutions such as NGOs have been making data on Mozambique available, including historical records of the agricultural market. These increasingly abundant and accessible datasets contain key variables necessary to address the growing need for accurate forecasts of maize price fluctuations in the country. Thus, an in-depth study of Machine Learning strategies based on statistical models becomes essential. Identifying the most appropriate model with optimised parameters and reliable results is a significant challenge to ensure a comprehensive solution to the research problem. In a set of conducted experiments, three machine learning algorithms were compared, namely Multiple Linear Regression, LSTM and Random Forest, applied to historical data on Mozambique's maize market. The LSTM model was selected as the most effective and was adopted as the core forecasting engine of the proposed system. The final system integrates a backend developed in Django, which provides a functional web platform for visualising predictions. Additionally, it incorporates an AI-based chatbot designed to interact with users, clarify questions, explain forecast results and facilitate intuitive and conversational access to information. The combination of robust models and high-quality data provides a pathway for improving the accuracy of market dynamics forecasting, promoting economic stability and supporting better management of agricultural resources.

**Keywords:** Price forecasting, agricultural market, machine learning, multiple linear regression, random forest, LSTM.

## Lista de Abreviaturas

<b>Abreviatura</b>	<b>Descrição Original</b>	<b>Descrição Traduzida</b>
<b>API</b>	Application Programming Interface	Interface de Programação de Aplicações
<b>ARIMA</b>	AutoRegressive Integrated Moving Average	Média Móvel Integrada Auto-Regressiva
<b>FAO</b>	Food and Agriculture Organization	Organização das Nações Unidas para a Alimentação e Agricultura
<b>IA</b>	Inteligência Artificial	Inteligência Artificial
<b>INE</b>	Instituto Nacional de Estatística	Instituto Nacional de Estatística
<b>LSTM</b>	Long Short-Term Memory	Memória de Longo e Curto Prazo
<b>MADER</b>	Ministério da Agricultura e Desenvolvimento Rural	Ministério da Agricultura e Desenvolvimento Rural
<b>MAPE</b>	Mean Absolute Percentage Error	Erro Percentual Absoluto Médio
<b>ML</b>	Machine Learning	Aprendizagem de Máquina
<b>PIB</b>	Produto Interno Bruto	Produto Interno Bruto
<b>RF</b>	Random Forest	Floresta Aleatória
<b>RNN</b>	Redes Neurais Recorrentes	Redes Neurais Recorrentes
<b>SVR</b>	Support Vector Regression	Regressão por Vetores de Suporte

## Glossário

**Algoritmo:** Sequência lógica e estruturada de instruções computacionais que resolvem um problema ou realizam uma tarefa específica.

**Aprendizagem de Máquina (Machine Learning):** Área da inteligência artificial que permite aos sistemas computacionais aprenderem com dados e experiências anteriores, identificando padrões para realizar previsões ou decisões.

**Aprendizagem Supervisionada:** Técnica de aprendizagem automática em que os modelos são treinados com dados rotulados, contendo tanto as variáveis de entrada como as de saída desejadas.

**Aprendizagem Não Supervisionada:** Técnica em que os modelos são treinados apenas com variáveis de entrada não rotuladas, com o objectivo de identificar padrões ou agrupamentos ocultos.

**Aprendizagem por Reforço:** Técnica na qual o modelo aprende através de recompensas e penalizações recebidas pelas suas acções, com o intuito de otimizar decisões futuras.

**Base de Dados:** Conjunto estruturado de dados organizados em tabelas ou formatos específicos, permitindo o armazenamento, a manipulação e a consulta eficiente da informação.

**Classificação:** Processo pelo qual modelos predizem categorias discretas, atribuindo entradas a classes específicas.

**Dados Históricos:** Informação armazenada ao longo do tempo, utilizada para identificar tendências e padrões que auxiliam no treino de modelos preditivos.

**Dataframe:** Estrutura bidimensional utilizada em linguagens de programação, especialmente em Python, para manipular e analisar dados em formato de tabela.

**Feature (Variável ou Atributo):** Características ou propriedades utilizadas como entradas dos modelos de aprendizagem automática para realizar previsões ou classificações.

**Hiperparâmetros:** Parâmetros externos ao modelo que são definidos antes do treino e influenciam directamente o processo de aprendizagem.

**Inteligência Artificial (IA):** Área da ciência da computação que visa desenvolver sistemas capazes de realizar tarefas que exigem inteligência humana, como reconhecimento de padrões, raciocínio lógico e tomada de decisão.

**Modelo Preditivo:** Modelo matemático ou computacional que utiliza dados históricos para prever resultados futuros ou desconhecidos.

**Normalização dos Dados:** Processo de transformação das variáveis para uma escala comum, evitando que variáveis com magnitudes diferentes dominem o processo de aprendizagem.

**Previsão:** Processo de utilização de modelos estatísticos ou de aprendizagem automática para estimar valores futuros com base em dados históricos.

**Preço do Milho:** Valor monetário pelo qual o milho é comercializado no mercado, sendo o foco da previsão neste estudo.

**Processamento de Dados:** Conjunto de técnicas utilizadas para preparar, limpar e transformar dados brutos em formatos adequados para análise e modelação.

**Redes Neurais Artificiais (RNA):** Modelos computacionais inspirados no funcionamento das redes neurais biológicas, capazes de aprender relações complexas entre variáveis.

**Regressão:** Técnica estatística ou de aprendizagem automática usada para prever valores numéricos contínuos, como o preço futuro do milho.

**Série Temporal:** Sequência ordenada de observações ou valores recolhidos ao longo do tempo em intervalos regulares, usada para análises e previsões ao longo do tempo.

**Treino do Modelo:** Processo de fornecimento de dados históricos ao modelo para que este aprenda padrões e ajuste os seus parâmetros internos com o objectivo de realizar previsões precisas.

**Validação Cruzada:** Técnica usada para avaliar a capacidade de generalização do modelo, dividindo os dados em subconjuntos para treino e validação.

**Variáveis Independentes (ou Preditoras):** Variáveis que explicam ou influenciam o comportamento da variável-alvo (preço do milho).

**Variável Dependente (ou Alvo):** Variável que se pretende prever (preço do milho) com base nas variáveis independentes.

# Índice

Dedicatória .....	iii
Declaração de Honra .....	iv
Agradecimentos .....	v
Resumo .....	vi
Abstract.....	vii
Lista de Abreviaturas .....	viii
Glossário.....	ix
Lista de Figuras.....	xv
1. Introdução.....	1
1.1. Contextualização.....	1
1.1.1. Delimitação do Tema .....	3
1.2. Definição do Problema.....	3
1.3. Importância do Estudo e suas Soluções.....	4
1.3.1. Impacto Social e Econômico.....	4
1.3.2. Contribuições Tecnológicas e Acadêmicas.....	5
1.4. Objectivos.....	6
1.4.1. Objectivo Geral .....	6
1.4.2. Objectivos Específicos .....	6
1.5. Estrutura do Trabalho.....	6
2. Revisão de Literatura .....	8
2.1. Segurança Alimentar.....	8
2.1.1. Disponibilidade de Alimentos .....	8
2.1.2. Acesso a Alimentos.....	9
2.1.3. Utilização de Alimentos.....	9
2.1.4. Estabilidade .....	9

2.1.5.	Desafios da Segurança Alimentar .....	9
2.1.6.	Importância da Segurança Alimentar .....	10
2.2.	Aprendizagem de Máquina.....	10
2.2.1.	Modelos de Machine Learning .....	12
2.3.	Previsão do Preço do Milho.....	17
2.3.1.	Estudos de Previsão em Machine Learning .....	18
3.	Material e Métodos.....	20
3.1.	Metodologia de Pesquisa .....	20
3.2.	Classificação da Pesquisa .....	21
3.2.1.	Quanto à abordagem .....	21
3.2.2.	Quanto aos objectivos .....	21
3.2.3.	Quanto à natureza .....	21
3.2.4.	Quanto ao método .....	22
3.2.5.	Material e Ficha Técnica.....	22
3.3.	Procedimentos Técnicos.....	23
3.3.1.	Recolha e Análise Exploratória de Dados .....	23
3.3.2.	Escolha dos Modelos de Machine Learning.....	24
3.3.3.	Divisão da Amostra de Treino e Teste .....	29
3.3.4.	Processo de Ajuste de Parâmetros e Métricas.....	30
3.3.5.	Medidas de Avaliação de Desempenho do Modelo .....	30
3.3.6.	Processo de Melhoramento de Modelo .....	32
3.3.7.	Processo de Tomada de Decisão .....	33
3.4.	Métodos de Implementação.....	33
3.4.1.	Linguagem de Programação e Frameworks .....	33
3.4.2.	Integração de Modelos e Dados.....	34
3.4.3.	Implementação da Interface Gráfica .....	35
3.4.4.	Testes e Validação .....	35
3.4.5.	Diagramas .....	36

4.	Experimentos e Resultados .....	37
4.1.	Estatística Analítica Descritiva (EAD) .....	37
4.1.1.	Tratamento Preliminar dos Dados .....	37
4.1.2.	Compreensão Inicial dos Dados .....	38
4.1.3.	Distribuição de Produtos por Categoria .....	39
4.1.4.	Distribuição histórica do preço médio do milho em Moçambique .....	40
	O Pico do Preço em 2017 .....	40
	Preço Médio Anual em Tete .....	40
	Variação dos Preços entre os Mercados em 2017 .....	41
4.2.	Engenharia de Atributos .....	43
4.2.1.	Seleção de Dados Pós-2000 .....	43
4.2.2.	Valores Ausentes .....	43
4.2.3.	Outras Variáveis com Valores Ausentes .....	44
4.2.4.	Tratamento de Duplicados .....	44
4.2.5.	Análise de Correlação .....	46
4.3.	Configuração dos Experimentos .....	47
4.3.1.	Primeira Série de Experimentos – Modelos Univariados .....	47
4.3.2.	Segunda Série de Experimentos – Modelos Multivariados .....	48
4.3.3.	Resultados e Avaliação .....	49
4.4.	Discussão dos Resultados .....	52
4.4.1.	Impacto das Variáveis Externas .....	52
4.4.2.	Comparação com Modelos Tradicionais .....	53
4.4.3.	Conclusão dos Experimentos .....	53
4.4.4.	Gráfico de Dispersão .....	54
4.4.5.	Preços Previstos vs Preços Reais .....	55
4.5.	Implementação da Plataforma Web .....	55
5.	Conclusão .....	59
5.1.	Contribuições e Implicações Práticas .....	60

5.2. Recomendações e Resultados Esperados.....	60
5.2.1. Estratégias de Disseminação e Adopção no Sector Agrícola .....	61
5.2.2. Trabalhos Futuros .....	61
Referências Bibliográficas .....	62

## Lista de Figuras

FIGURA 1. PRODUÇÃO DE MILHO EM MOÇAMBIQUE .....	15
FIGURA 2. USO DE SOLUÇÕES TECNOLÓGICAS NA AGRICULTURA.....	17
FIGURA 3. FLUXO DE TRABALHO DE UM MODELO ML.....	26
FIGURA 4. PLANTAÇÃO DE MILHO.....	31
FIGURA 5. COLHEITA DE MILHO.....	33
FIGURA 6. ESTRUTURA DE UMA RANDOM FOREST .....	39
FIGURA 7. CÉLULA E FÓRMULA DO MODELO LSTM .....	41
FIGURA 8. FÓRMULA DO MODELO DE RLM .....	42
FIGURA 9. ERRO QUADRÁTICO MÉDIO (FÓRMULA) .....	45
FIGURA 10. RAIZ DO ERRO QUADRÁTICO MÉDIO (FÓRMULA) .....	46
FIGURA 11. COEFICIENTE DE DETERMINAÇÃO (FÓRMULA) .....	46
FIGURA 12. DIAGRAMA DE CLASSES.....	50
FIGURA 13. DIAGRAMA DE SEQUÊNCIA DE EVENTOS .....	50
FIGURA 14. TABELA DO CONJUNTO DE DADOS ORIGINAL.....	52
FIGURA 15. DICIONÁRIO DE DADOS.....	53
FIGURA 16. AMOSTRA DE CONTAGEM DE VALORES.....	54
FIGURA 17. DISTRIBUIÇÃO DE REGISTOS POR CATEGORIA.....	54
FIGURA 18. EVOLUÇÃO DO PREÇO MÉDIO DO MILHO EM MOÇAMBIQUE .....	55
FIGURA 19. EVOLUÇÃO DO PREÇO MÉDIO DO MILHO EM TETE.....	56
FIGURA 20. DISTRIBUIÇÃO DA ÁREA E PRODUÇÃO DO MILHO, 2020 .....	56
FIGURA 21. TABELA DAS DIFERENÇAS DE PREÇO ENTRE AS PROVÍNCIAS.....	57
FIGURA 22. TABELA D DE DADOS EM FALTA ENTRE AS PROVÍNCIAS .....	59
FIGURA 23. TABELA DOS DADOS (TRATADA).....	60
FIGURA 24. TABELA DOS DADOS (COMPLETA).....	60

FIGURA 25. MAPA DE CORRELAÇÃO DOS DADOS .....	61
FIGURA 26. TABELA DE AVALIAÇÃO DO PRIMEIRO EXPERIMENTO .....	64
FIGURA 27. TABELA DE AVALIAÇÃO DO SEGUNDO EXPERIMENTO.....	66
FIGURA 28. GRÁFICO DE DISPERSÃO DO PREÇO REAL E PREVISTO .....	69
FIGURA 29. TABELA COM OS DADOS ORIGINAIS (VISTO DO EXCEL) .....	70
FIGURA 30. TABELA COM OS PREÇOS PREVISTOS (VISTO DO ANACONDA) .....	70
FIGURA 31. UI DO MVP.....	71
FIGURA 32. UI DO MVP.....	72
FIGURA 33. MILHO E SEUS GRÃOS.....	75

# 1. Introdução

Este capítulo tem como objectivo introduzir o tema da previsão do preço do milho em Moçambique, contextualizando a importância deste estudo no cenário económico e agrícola do país. As ideias fundamentais em torno da aplicação da Aprendizagem Automática (ML) para previsões económicas serão discutidas e ilustradas em relação ao agronegócio. Além disso, será apresentada uma breve reflexão acerca da importância do milho enquanto cultura estratégica para a segurança alimentar em Moçambique. Serão igualmente discutidas as razões pelas quais desafios enfrentados pelos agricultores, nomeadamente a volatilidade dos mercados e a incerteza dos preços, reforçam a pertinência e a necessidade de desenvolver ferramentas de previsão mais rigorosas e precisas.

## 1.1. Contextualização

Em Moçambique, a actividade económica é impulsionada principalmente pela agricultura, e esta representa também o sustento da maioria da população (Marcos et al., 2022b). Entre os produtos agrícolas mais significativos cultivados em suas terras férteis, o milho destaca-se como uma cultura fundamental.

Entre as culturas agrícolas produzidas em Moçambique, o milho ocupa um lugar de destaque (MADER, 2020). Este, faz parte da dieta estável da maioria das famílias e fornece sustento nutricional e económico a milhões de pessoas em Moçambique. No entanto, a volatilidade dos preços do milho ainda constitui uma questão sensível e representa um desafio significativo para agricultores, comerciantes e consumidores.

As flutuações nos preços do milho podem ter impactos consideráveis nas comunidades rurais e urbanas, afectando não apenas a disponibilidade de alimentos, mas também a viabilidade económica das famílias e a capacidade do país de garantir a segurança alimentar para sua população.

Neste contexto, existe uma necessidade real de criar um modelo de Machine Learning que consiga prever os preços do milho em Moçambique. Este modelo procurará alavancar os dados históricos de forma eficiente na previsão das tendências futuras para os preços do milho. Através deste esforço, o modelo fornecerá informação a vários stakeholders (partes interessadas) envolvidos na cadeia de

produção e distribuição de milho, o que fundamentará a sua tomada de decisões mais estratégicas e esclarecidas.



*Figura 1. Produção de Milho em Moçambique (fonte: unsplash)*

Estudos científicos demonstram que modelos de Machine Learning são altamente eficazes na previsão de preços de alimentos, como o milho.

Silva et al (2020) propuseram um sistema de previsão baseado em modelos de Machine Learning que utiliza dados de preços e análise de sentimento de *tweets* para prever o preço do milho. Este estudo veio a destacar as vantagens de integrar dados tradicionais com informações de redes sociais para melhorar a precisão das previsões em questão.

Por outro lado, Rai et al (2019) comparou o desempenho de diferentes algoritmos de Machine Learning, incluindo redes neurais artificiais, regressão linear e Support Vector Machines (SVM), na previsão do preço do milho na Índia.

Ademais, várias técnicas de deep learning, como redes neurais convolucionais (CNNs) e redes neurais recorrentes (LSTMs) foram exploradas para poder prever (com maior potencial) preços de produtos agrícolas, incluindo o milho (Mehta et al, 2018).

Esses estudos procuram ilustrar como técnicas avançadas de Machine Learning e deep learning conseguem revolucionar a previsão de preços agrícolas, provendo ferramentas e soluções mais precisas e confiáveis para auxiliar na tomada de decisões no sector agrícola.

O presente estudo não apenas responde a uma necessidade crítica no contexto agrícola e económico de Moçambique, como também demonstra de que forma essa melhoria pode ser alcançada: através do desenvolvimento e aplicação de modelos de ML capazes de analisar séries temporais de preços do milho e identificar padrões que permitam gerar previsões mais precisas. Ao fornecer estimativas fiáveis sobre a evolução futura dos preços, o estudo contribui para uma maior estabilidade no planeamento agrícola, na gestão de stocks e na definição de estratégias de mercado, promovendo assim um acesso mais previsível e seguro a este alimento básico para milhões de moçambicanos.

### 1.1.1. Delimitação do Tema

O presente estudo, é delimitado na previsão dos preços do milho em Moçambique, utilizando modelos avançados de Machine Learning. Está geograficamente limitado a Moçambique, considerando suas particularidades no mercado de milho, e utiliza dados históricos de preços do milho como variáveis de entrada, juntamente com outros factores relevantes como condições climáticas e indicadores económicos, optando por modelos de regressão, devido à sua capacidade de modelar sequências temporais e capturar dependências de longo prazo.

O objectivo deste estudo é fornecer previsões de curto e médio prazo, avaliadas por medidas para orientar decisões estratégicas no sector agrícola de Moçambique, com ênfase na construção e avaliação dos modelos de previsão.

O milho é o produto escolhido como foco do modelo devido à sua importância vital para a segurança alimentar e económica de Moçambique. O milho é o principal cereal consumido no país, representando boa parte da ingestão calórica da população (MADER, 2023). A sua produção, é uma actividade económica essencial para muitas famílias rurais, fornecendo rendimento e sustento (FAO, 2022).

## 1.2. Definição do Problema

O milho é um dos principais produtos agrícolas cultivados e consumidos pela população em Moçambique (MADER, 2020). Assim sendo, dada sua importância como fonte primária de sustento alimentar para várias famílias moçambicanas, **a previsão precisa e antecipada do preço do milho torna-se fundamental** para várias tomadas de decisão relacionadas à produção, distribuição, investimento e políticas governamentais.

O problema chave deste estudo tem como base a **necessidade de desenvolver um modelo preditivo capaz de antecipar com precisão as flutuações nos preços do milho em Moçambique**. Estas

variações de preço, muitas vezes imprevisíveis, são influenciadas por vários factores, como o clima, a oferta e procura, e os custos de produção, e podem afectar significativamente a segurança alimentar, a estabilidade económica e o rendimento das famílias que dependem deste produto agrícola essencial.

Ao prever os preços do milho com antecedência, as partes interessadas serão capacitadas a **tomar decisões informadas e estratégicas**. Por exemplo, os agricultores poderão planejar suas colheitas com base nas expectativas de preço, os comerciantes poderão ajustar suas estratégias de compra e venda, os distribuidores poderão gerir seus stocks de forma eficaz e os formuladores de políticas poderão levar a cabo intervenções pertinentes **para mitigar qualquer impacto adverso sobre a segurança alimentar** e a estabilidade económica.



*Figura 2. Uso de Soluções Tecnológicas na Agricultura (fonte: unsplash)*

### **1.3. Importância do Estudo e suas Soluções**

#### **1.3.1. Impacto Social e Económico**

Conseguir prever antecipadamente os preços do milho usando modelos de Machine Learning é um grande passo na inovação e carrega o potencial de transformar significativamente vários sectores da sociedade moçambicana.

Para os produtores rurais, esse passo oferece oportunidades de **planeamento agrícola mais eficiente**, e auxilia na tomada de decisões informadas sobre os momentos ideais para plantar, colher e vender. Ao mesmo tempo, possibilita maior acesso a **informações de mercado**, fortalece sua posição em negociações e **umenta sua rentabilidade**. Para os consumidores, a previsão antecipada serve de base

na estabilidade nos preços de consumo e melhoria no acesso a alimentos básicos, promovendo a **segurança alimentar** e reduzindo os impactos das flutuações de mercado.

Do ponto de vista governamental, esse passo promete ser um instrumento inovador e valioso para o **planeamento estratégico**. Com o acesso antecipado às tendências de preços, torna-se mais viável implementar políticas governamentais eficazes, como subsídios ou stocks reguladores, que estabilizem o mercado. Além disso, a previsão antecipada de flutuações auxilia em uma gestão mais eficiente de crises económicas e contribui para a resiliência do sector agrícola e da economia como um todo.

### **1.3.2. Contribuições Tecnológicas e Acadêmicas**

O presente trabalho também desempenha um papel crucial no avanço inovador das aplicações tecnológicas e acadêmicas em Moçambique. Com a construção de modelos de Machine Learning ajustados ao contexto local, conseguimos destacar o grande potencial de soluções personalizadas em países em desenvolvimento, onde problemas reais frequentemente requerem abordagens inovadoras. Uma das principais inovações deste estudo é a integração de dados históricos de diferentes fontes (climáticas, agrícolas e económicas) em modelos preditivos avançados para “olhar para o futuro” da agroeconomia de Moçambique.

Considerando o âmbito acadêmico, observa-se que a presente investigação serve de base para futuros estudos, tanto em Moçambique quanto em outros países com condições semelhantes. A metodologia desenvolvida pode ser replicada e adaptada para outros produtos agrícolas como o arroz e a mandioca, ampliando o impacto da investigação, ou ainda, para outros estudos científicos.

Assim, este trabalho não apenas **propõe soluções de problemas reais**, mas também contribui significativamente para o corpo de conhecimento em ciência de dados aplicada à agricultura e economia.

## **1.4. Objectivos**

### **1.4.1. Objectivo Geral**

Desenvolver um sistema de previsão do preço do milho em Moçambique baseado em Machine Learning.

### **1.4.2. Objectivos Específicos**

- Estudar a evolução histórica dos preços do milho em Moçambique, através da aplicação de técnicas de análise exploratória de séries temporais, medindo padrões sazonais, tendências e variações ao longo do tempo por meio de métricas estatísticas como autocorrelação e amplitude de variação.
- Identificar e quantificar os factores que mais influenciam o comportamento do preço do milho em Moçambique, integrando variáveis relevantes nas análises.
- Desenvolver e aplicar modelos de Machine Learning para prever os preços futuros do milho, utilizando como base os dados históricos disponíveis.
- Avaliar e comparar o desempenho final dos modelos de Machine Learning de modo a escolher o mais adequado para o sistema final.

## **1.5. Estrutura do Trabalho**

O presente relatório está estruturado em seis (6) capítulos, a saber: Introdução, Revisão de Literatura, Material e Métodos, Experimentos e Resultados, Conclusão e Referências Bibliográficas.

A seguir, apresenta-se um resumo de cada capítulo:

### **Capítulo 1: Introdução**

Neste capítulo, realiza-se a contextualização do tema, onde destaca-se a importância do milho no cenário agrícola e económico de Moçambique e os desafios relacionados à volatilidade dos seus preços. Enuncia-se o problema que motivou o desenvolvimento do trabalho e apresenta-se os objectivos gerais e específicos, assim como a relevância desta investigação para a inovadora modernização do sector agrícola. Por fim, descreve-se a estrutura do relatório.

## **Capítulo 2: Revisão de Literatura**

Este capítulo aborda os principais conceitos e estudos relacionados à previsão de preços agrícolas com base em técnicas de Machine Learning. Aqui, discute-se a relevância de modelos como LSTM, Random Forest e Regressão Multilinear, explorando casos reais de aplicação em contextos semelhantes. Por fim, são apresentados também estudos que fundamentam a utilização de variáveis climáticas, económicas e históricas para prever flutuações de preços.

## **Capítulo 3: Material e Métodos**

Neste capítulo, detalha-se a metodologia aplicada para atingir os objectivos propostos. Apresenta-se o processo de colecta, limpeza e preparação dos dados, assim como a engenharia de atributos e a análise exploratória. Descrevem-se as técnicas de modelagem preditiva, os critérios para ajuste de parâmetros e validação dos modelos, e as ferramentas tecnológicas utilizadas para implementar o sistema de previsão.

## **Capítulo 4: Experimentos e Resultados**

Este capítulo descreve os experimentos realizados para avaliar o desempenho dos diferentes algoritmos seleccionados na previsão do preço do milho. Também apresentam-se as medidas de avaliação utilizadas, como o RMSE, e os resultados obtidos com os mesmos modelos. Destaca-se o modelo LSTM como o mais eficaz e robusto, com explicações detalhadas sobre os seus resultados em relação aos outros modelos.

## **Capítulo 5: Conclusão**

Neste capítulo, sintetizam-se as principais contribuições do trabalho, enfatizando o avanço representado pela aplicação de Machine Learning no sector agrícola de Moçambique. Apresentam-se as limitações identificadas e recomendações para trabalhos futuros, incluindo a exploração de novos modelos e ampliação do impacto da solução proposta.

## **Capítulo 6: Referências Bibliográficas**

Neste capítulo, são listadas todas as obras, artigos científicos, relatórios técnicos e documentos consultados para a elaboração do trabalho. As referências seguem as normas de formatação adoptadas garantindo a rastreabilidade e credibilidade das fontes utilizadas. Incluem-se tanto estudos internacionais

que exploram técnicas de Machine Learning aplicadas à agricultura como documentos que fornecem informações sobre a produção e o preço do milho em Moçambique.

## 2

## **2. Revisão de Literatura**

Este capítulo tem como objectivo apresentar e discutir os principais conceitos e estudos relacionados com a previsão do preço de alimentos, como o milho, baseada em Machine Learning (ML). A análise centrar-se-á na aplicação de técnicas de ML como ferramenta central para a previsão de preços agrícolas, assim como nos factores que influenciam as variações de preço do milho, incluindo factores económicos, ambientais e de mercado.

### **2.1. Segurança Alimentar**

A segurança alimentar é definida como o "acesso de todas as pessoas, em todo o momento, a alimentos suficientes, seguros e nutritivos para atender às suas necessidades alimentares e preferências alimentares, permitindo-lhes levar uma vida activa e saudável" (FAO, 2020).

Ela envolve não apenas a disponibilidade de alimentos, mas também o acesso, a utilização e a estabilidade desses recursos ao longo do tempo.

Vamos explorar cada um desses aspectos.

#### **2.1.1. Disponibilidade de Alimentos**

A disponibilidade de alimentos refere-se à existência de alimentos em quantidade e qualidade suficientes para responder às necessidades da população (FAO, 2020). Assim, adoptar práticas agrícolas sustentáveis e gerir de forma eficaz os recursos do sector é fundamental para assegurar que a oferta alimentar acompanhe o ritmo do aumento populacional.

### **2.1.2. Acesso a Alimentos**

Segundo Maxwell e Smith (1992), a segurança alimentar é alcançada quando todas as pessoas têm, em todo momento, acesso físico, social e económico a alimentos suficientes, seguros e nutritivos para satisfazer as suas necessidades dietéticas e preferências alimentares para uma vida activa e saudável. O acesso envolve a capacidade das pessoas de obter alimentos. Isso pode ser afectado por factores económicos, como o rendimento e os preços dos alimentos, mas também por questões sociais, como a discriminação e a exclusão.

### **2.1.3. Utilização de Alimentos**

A utilização de alimentos refere-se ao processo pelo qual os alimentos são utilizados para suprir as necessidades nutricionais e de consumo. Essa dimensão da segurança alimentar está directamente ligada à qualidade dos alimentos, à dieta equilibrada e à capacidade do organismo de absorver os nutrientes. De acordo com Smith e Haddad (2000), a utilização adequada depende não apenas da disponibilidade de alimentos nutritivos, mas também de factores como educação nutricional, condições de saneamento e acesso a serviços de saúde. Esses factores garantem que os alimentos consumidos contribuam efectivamente para o bem-estar físico e mental.

### **2.1.4. Estabilidade**

A estabilidade na segurança alimentar refere-se à capacidade de garantir o acesso consistente a alimentos ao longo do tempo. Segundo FAO (2006), a segurança alimentar não pode ser atingida sem estabilidade, pois variações no acesso à comida devido a crises económicas, conflitos, mudanças climáticas ou pandemias comprometem a resiliência dos sistemas alimentares. Para assegurar essa estabilidade, é necessário o desenvolvimento de sistemas resilientes que possam absorver choques e se adaptar às condições mutáveis (Ericksen, 2008).

### **2.1.5. Desafios da Segurança Alimentar**

Os desafios da segurança alimentar são complexos e interligados, de tal modo que requerem soluções integradas. As mudanças climáticas representam um dos maiores desafios, porque afectam directamente a produção agrícola e a disponibilidade de água (IPCC, 2014). O crescimento populacional também pressiona os sistemas de produção alimentar, aumentando a demanda por alimentos (United Nations, 2019). Além disso, a desigualdade económica dificulta o acesso a alimentos saudáveis para comunidades em situação de pobreza (Sen, 1981). Os conflitos armados, por sua vez, comprometem tanto a produção

quanto a distribuição de alimentos, intensificando assim a insegurança alimentar nas regiões vulneráveis (FAO, 2020).

### **2.1.6. Importância da Segurança Alimentar**

Garantir a segurança alimentar é crucial para a saúde pública, a redução da pobreza e o desenvolvimento económico. Alimentos adequados e nutritivos são essenciais para prevenir doenças relacionadas à desnutrição e melhorar a produtividade das populações. Segundo Fanzo et al. (2013), a segurança alimentar contribui directamente para a estabilidade social e o crescimento económico sustentável. Além disso, é um pilar fundamental para garantir a qualidade de vida e promover sociedades mais saudáveis e equitativas.

## **2.2. Aprendizagem de Máquina**

O aprendizado de máquina, ou Machine Learning, é definido como um campo da inteligência artificial que capacita sistemas computacionais a aprenderem e melhorarem automaticamente a partir de dados e experiências, sem a necessidade de serem programados explicitamente para cada tarefa. Essa abordagem utiliza algoritmos para identificar padrões e generalizar comportamentos a partir de conjuntos de dados, permitindo previsões e decisões baseadas em evidências empíricas (Mitchell, 1997).

Esta abordagem inovadora tem transformado vários sectores da economia e impactado significativamente a forma como vivemos e interagimos com a tecnologia. Por causa da sua flexibilidade e eficiência, a aprendizagem de máquina é amplamente aplicada em diversas áreas, como descrito a seguir.

### **i. Saúde**

No sector da saúde, sistemas de ML têm revolucionado o diagnóstico médico, personalizado tratamentos e permitido a previsão de epidemias. Um exemplo notável é a utilização de algoritmos para análise de imagens radiológicas, como raios-X e ressonâncias magnéticas. Estudos demonstram que modelos como o ResNet50 têm alcançado precisão comparável à de especialistas humanos na detecção de patologias torácicas (Wang et al., 2017). Além disso, modelos preditivos são usados para identificar precocemente doenças crónicas, como diabetes, permitindo cuidados mais eficazes e a redução de custos (Obermeyer et al., 2016).

## **ii. Finanças**

No sector financeiro, o aprendizado de máquina tem sido utilizado para prever tendências de mercado, detectar fraudes e melhorar a análise de crédito. Por exemplo, um sistema antifraude pode utilizar algoritmos de redes neurais para analisar milhões de transações em tempo real, reduzindo significativamente os riscos financeiros. Além disso, algoritmos baseados em boosting, como o XGBoost, têm sido aplicados para prever falências empresariais com alta precisão (Chen & Guestrin, 2016).

## **iii. Entretenimento**

Várias plataformas de streaming (como a Netflix), utilizam o aprendizado de máquina para personalizar recomendações de conteúdo com base no comportamento dos utilizadores. O modelo de recomendação da Netflix combina algoritmos de filtragem colaborativa e análise de conteúdo com o objectivo de melhorar continuamente com as interações dos utilizadores (Gomez-Uribe & Hunt, 2015). Essa abordagem ajuda a maximizar o engajamento e a satisfação do cliente.

## **iv. Transportes e Logística**

O aprendizado de máquina desempenha um papel crucial em sistemas de transporte inteligente, como veículos autónomos e aplicações de rotas optimizadas. Como exemplo popular temos o sistema de condução autónoma da Tesla, que utiliza redes neurais profundas para processar dados em tempo real e tomar decisões seguras na condução (Bojarski et al., 2016). Enquanto isso, no sector logístico, empresas como a DHL utilizam o aprendizado de máquina para prever a demanda e optimizar a alocação de recursos, melhorando a eficiência operacional (DHL, 2018).

## **v. Indústria e Automação**

Na indústria, o aprendizado de máquina tem sido aplicado para prever falhas em equipamentos, promovendo a manutenção preditiva. Por exemplo, a Siemens implementou sistemas baseados em ML para monitorar turbinas, reduzindo assim o tempo de inatividade (Siemens, 2019). Além disso, robôs industriais treinados com o aprendizado de máquina realizam tarefas repetitivas ou perigosas com alta precisão, aumentando a produtividade (LeCun et al., 2015).

Essa capacidade de adaptação e aprendizado contínuo dos algoritmos de ML é o que os torna tão poderosos e relevantes. Por meio de técnicas como Aprendizado Supervisionado, Aprendizado Não Supervisionado e Aprendizado por Reforço, esses sistemas podem ser aplicados a uma ampla gama de problemas reais, desde tarefas simples de classificação até decisões complexas em ambientes dinâmicos.

Entretanto, o avanço do aprendizado de máquina também traz consigo desafios importantes, como a necessidade de garantir a privacidade e segurança dos dados, a transparência dos algoritmos (para evitar decisões enviesadas ou injustas) e o impacto ético do uso indiscriminado dessas tecnologias.

Deste modo, o aprendizado de máquina não é apenas uma tendência passageira, mas sim um componente essencial do futuro da tecnologia e vai desempenhando um papel crítico na criação de soluções mais inteligentes, eficientes e acessíveis em todas as áreas da sociedade.

### **2.2.1. Modelos de Machine Learning**

Os modelos de Machine Learning são ferramentas poderosas que utilizam estruturas matemáticas para capturar e representar o conhecimento extraído de grandes volumes de dados (Toolify, 2023). O seu processo de treino é fundamental nesse contexto, onde algoritmos específicos aprendem a partir de um conjunto de dados, ajustando seus parâmetros com o objectivo de minimizar um erro ou otimizar uma função específica. Este processo é muitas vezes iterativo e permite que o modelo refine de forma contínua as suas previsões/saídas conforme mais dados são disponibilizados.

Uma vez que o modelo esteja treinado, pode ser aplicado a novos dados para fazer previsões, tomar decisões ou realizar tarefas específicas. Isso permite que organizações automatizem processos, identifiquem padrões e ganhem insights valiosos, onde otimizam operações e aumentam a eficiência.

Por outro lado, a qualidade dos dados utilizados no treino de modelos de Machine Learning é crucial para seu desempenho. Geralmente quando se trabalha com dados imprecisos ou enviesados, estes costumam resultar em previsões insatisfatórias e decisões erradas, por isso o pré-processamento de dados é um passo fundamental. Além disso, a interpretação dos resultados é essencial, uma vez que entender como um modelo chega a suas conclusões é tão importante quanto a precisão das previsões, especialmente em áreas sensíveis como saúde e finanças. Assim, o campo de Machine Learning exige não apenas habilidades técnicas, mas também uma consideração ética e prática das suas implicações no mundo real.



Figura 3. Fluxo de Trabalho de um Modelo ML

Os algoritmos de Aprendizagem de Máquina podem ser categorizados em várias classes, incluindo:

### Aprendizagem Supervisionada

O aprendizado supervisionado utiliza um conjunto de dados rotulado, no qual os pares de entrada (características) e saída (rótulos) são conhecidos. O objectivo é ensinar o modelo a generalizar a relação entre entradas e saídas, para que possa fazer previsões com precisão em dados novos (Alpaydin, 2020).

### Modelos Comuns de Aprendizagem Supervisionada

#### Regressão Linear

A regressão linear é usada para prever uma variável contínua com base em uma ou mais variáveis independentes. O modelo tenta ajustar uma linha recta que minimize o erro entre os valores previstos e os reais.

**Exemplo:** Estudos mostraram que a regressão linear é eficaz na previsão de preços de imóveis com base em características como tamanho, localização e número de quartos (Draper & Smith, 1998).

### **Máquina de Vetores de Suporte (SVM - Support Vector Machine)**

O SVM é usado tanto para classificação quanto para regressão. Ele encontra um hiperplano ótimo que separa os dados em diferentes classes no espaço multidimensional.

**Exemplo:** O SVM é amplamente utilizado na classificação de e-mails como spam ou não spam, utilizando características como a frequência de palavras (Joachims, 1998).

### **Regressão Logística**

A regressão logística é usada para tarefas de classificação binária. Ela modela a probabilidade de uma observação pertencer a uma classe específica, utilizando uma função sigmoide para transformar os valores preditos em probabilidades.

**Exemplo:** Estudos mostraram que a regressão logística é eficiente no diagnóstico de doenças como diabetes com base em dados clínicos (Hosmer et al., 2013).

### **Árvores de Decisão**

Este modelo segmenta os dados em diferentes ramos com base em condições que maximizam a separação entre classes ou explicam a variância.

**Exemplo:** Em sistemas bancários, árvores de decisão são usadas para determinar a elegibilidade de clientes para empréstimos com base no histórico de crédito e rendimento (Breiman et al., 1984).

### **Floresta Aleatória (Random Forest)**

Conjunto de árvores de decisão construídas com amostras aleatórias do conjunto de dados. A previsão final é feita pela agregação (votação ou média) dos resultados das árvores.

**Exemplo:** A Random Forest tem sido usada para prever os preços de ações com base em factores económicos e históricos financeiros (Liaw & Wiener, 2002).

## **Redes Neurais Artificiais (ANN - Artificial Neural Networks)**

Inspiradas no cérebro humano, essas redes processam informações através de camadas de neurônios para aprender padrões complexos.

**Exemplo:** As ANNs são amplamente usadas em reconhecimento de imagens para identificar objectos em fotos, como demonstrado em estudos de LeCun et al. (1998).

## **Aprendizagem Não Supervisionada**

Nesse método, os dados não possuem rótulos. O modelo busca identificar padrões, agrupamentos ou estruturas nos dados. Ele é particularmente útil para análise exploratória ou pré-processamento de dados (Bishop, 2006).

### **Modelos Comuns de Aprendizagem Não Supervisionada:**

#### **K-Means Clustering**

O K-Means é um algoritmo de agrupamento que particiona os dados em k grupos distintos. Ele minimiza a distância entre os pontos de um grupo e o centroide do grupo correspondente.

**Exemplo:** Este algoritmo é utilizado na segmentação de clientes para criar campanhas de marketing personalizadas com base em comportamentos de compra (MacQueen, 1967).

#### **Análise de Componentes Principais (PCA - Principal Component Analysis)**

O PCA é usado para redução de dimensionalidade, transformando os dados para um espaço de menor dimensão, enquanto preserva o máximo de variação possível.

**Exemplo:** O PCA é utilizado em compressão de imagens, como na redução de características para processamento eficiente (Jolliffe, 2002).

#### **Hierarchical Clustering**

Forma grupos criando uma hierarquia de aglomerados com base nas similaridades entre os dados.

**Exemplo:** O clustering hierárquico tem sido usado em bibliotecas digitais para agrupar documentos por tópicos semelhantes (Johnson, 1967).

## **Aprendizagem por Reforço**

O aprendizado por reforço envolve um agente que interage com um ambiente para aprender uma política de acções que maximize recompensas acumuladas. É particularmente útil em cenários onde o agente precisa tomar decisões sequenciais (Sutton & Barto, 2018).

## **Modelos Comuns de Aprendizagem por Reforço**

### **Q-Learning**

Algoritmo baseado em tabelas que utiliza uma função para avaliar a qualidade de executar uma acção em um estado. O agente aprende com base nas recompensas recebidas e actualiza os valores iterativamente.

**Exemplo:** O Q-Learning tem sido usado na navegação de robôs em ambientes desconhecidos (Watkins & Dayan, 1992).

### **Proximal Policy Optimization (PPO)**

PPO é um método de aprendizagem por reforço baseado em gradiente que ajusta as políticas directamente. Ele é projectado para equilibrar exploração e estabilidade durante o treino.

**Exemplo:** Treinamento de agentes em jogos digitais complexos, como mostrado nos estudos da OpenAI (Schulman et al., 2017).

### **Deep Q-Learning (DQN)**

Variante do Q-Learning que utiliza redes neurais profundas para estimar funções de valor em ambientes complexos.

**Exemplo:** O DQN foi usado para ensinar agentes a jogar jogos clássicos da Atari com desempenho humano (Mnih et al., 2015).

### **Monte Carlo Methods**

Algoritmos baseados em simulação que avaliam recompensas acumuladas através de trajectórias completas, usados para aprendizagem em ambientes estocásticos.

**Exemplo:** Modelagem de comportamento de traders no mercado financeiro (Metropolis & Ulam, 1949).

### **SARSA (State-Action-Reward-State-Action)**

Algoritmo de aprendizagem on-policy, ou seja, aprende a política com base nas acções tomadas pelo próprio agente.

**Exemplo:** Controle de temperatura em sistemas de aquecimento inteligentes (Rummery & Niranjan, 1994).

### **Policy Gradient Methods**

Métodos que optimizam directamente a política do agente ajustando probabilidades de acções com base em gradientes de recompensa esperada.

**Exemplo:** Desenvolvimento de bots para jogos multiplayer online (Silver et al., 2014).

### **Actor-Critic Methods**

Combina abordagens de política (actor) e valor (critic) para aprender de forma eficiente tanto a política quanto o valor das acções.

**Exemplo:** Controle de drones para missões de busca e resgate (Konda & Tsitsiklis, 2000).

Em suma, o aprendizagem de máquina continua a expandir as suas aplicações nas áreas como agricultura, finanças, transporte e entretenimento. Cada abordagem - supervisionada, não supervisionada e por reforço - atende a diferentes tipos de problemas e proporciona soluções personalizadas para demandas específicas. Os modelos exemplificados neste relatório mostram a diversidade de ferramentas disponíveis.

## **2.3. Previsão do Preço do Milho**

Quando se fala da previsão dos preços dos alimentos, é necessário entender que esta desempenha um papel essencial na gestão da segurança alimentar, uma vez que permite aos governos e às empresas planearem a oferta e procura, com o objectivo de prevenir instabilidade no mercado e assegurar que a população tenha acesso aos alimentos (Scodro L, Corso LL., 2023).



*Figura 4. Plantação de Milho (fonte: unsplash)*

Ainda, a previsão do preço do milho é um tema que tem atraído a atenção de vários investigadores, especialmente à luz da crescente importância do milho como uma cultura fundamental para a segurança alimentar não só nacional, assim como global.

Nos últimos anos, várias investigações propuseram implementar esquemas de previsão de preços do milho utilizando diferentes algoritmos de aprendizagem de máquina e suas metodologias como a identificação de eventos estatisticamente significativos, a descoberta da estrutura inerente e a realização de previsões numéricas precisas. Entre as abordagens populares estão as Long Short Term Memory (LSTM) e Support Vector Machine (SVM), que têm sido amplamente debatidas devido à sua capacidade de modelar padrões não lineares e relações causais complexas, além de sua eficácia em grandes conjuntos de dados históricos.

No entanto, o desempenho da aprendizagem de máquina na previsão dos preços futuros do milho em Moçambique em específico ainda é pouco explorado.

### **2.3.1. Estudos de Previsão em Machine Learning**

Um estudo significativo realizado por Hoffman et al. (2021) explorou a previsão dos preços globais do milho a partir das produções regionais, utilizando algoritmos de Machine Learning baseados em árvores de regressão. Neste trabalho, foram testados três algoritmos e comparados entre si, onde o seu desempenho foi avaliado com métodos tradicionais de séries temporais, mostrando assim a eficácia das abordagens de Machine Learning na identificação de regiões influentes na formação de preços.

Outra investigação inovadora foi realizada por Pukrongta et al. (2024), que apresentou o modelo PEnsemble 4. Este apresentou ser um sistema de previsão que integra múltiplos modelos de Machine Learning para prever rendimentos de milho. Neste mesmo modelo foi feita a utilização de dados de imagens de veículos aéreos não tripulados (UAV) e dados ambientais obtidos através da Internet das Coisas (IoT) e mostrou uma taxa de precisão notável de 91%. Por fim, este estudo conseguiu destacar a importância de prever não apenas os rendimentos, mas também o estresse hídrico e doenças nas culturas.

Além disso, Asamoah et al. (2024) desenvolveu um modelo de Machine Learning utilizando a técnica Random Forest para prever os rendimentos e a eficiência agronómica do milho em Gana. O estudo analisou uma vasta gama de factores, como clima, solo e práticas de manejo, e encontrou que as variáveis climáticas eram menos importantes do que as variáveis de solo para a previsão de rendimentos. Deste modo, enfatizando a necessidade de modelos preditivos que considerem a complexidade dos factores ambientais. Estes e outros estudos visam abordar a versatilidade e a eficácia das abordagens de Machine Learning na previsão do preço e rendimento do milho, oferecendo insights valiosos para agricultores, formuladores de políticas e investigadores.

Embora alguns pesquisadores tenham se concentrado na aplicação de redes neurais para a previsão do preço do milho, a maior parte dos estudos foca na previsão de intervalos, enquanto a previsão pontual tem recebido menos atenção, e isso mantém uma limitação no contexto de observação diária dos preços. É necessário mais esforço para abordar uma análise preditiva mais aprofundada dos modelos de Machine Learning.



*Figura 5. Colheita de Milho (fonte: unsplash)*

### **3. Material e Métodos**

O objectivo deste capítulo é descrever detalhadamente os procedimentos metodológicos utilizados para a investigação na previsão do preço do milho em Moçambique com base em técnicas de Machine Learning. Serão apresentadas as ferramentas tecnológicas empregues. Adicionalmente, serão discutidos os métodos de tratamento e análise dos dados, as técnicas de modelagem preditiva aplicadas, e os critérios utilizados para a avaliação dos modelos, com o intuito de garantir a robustez e a precisão das previsões.

#### **3.1. Metodologia de Pesquisa**

A metodologia adoptada neste trabalho fundamenta-se numa abordagem exploratória, dado o carácter inovador do estudo que procura propor um modelo de previsão do preço do milho em Moçambique baseado em técnicas de Machine Learning. A base deste tipo de pesquisa é permitir que se obtenha um entendimento aprofundado do problema, fornecendo subsídios teóricos e práticos para o desenvolvimento do modelo proposto. Segundo Creswell (2014), a pesquisa exploratória é apropriada para estudos iniciais em áreas pouco exploradas, permitindo o levantamento de dados e a construção de hipóteses.

Neste contexto, a análise incluiu a revisão de literatura científica e técnica sobre modelos preditivos e sua aplicabilidade na agricultura, assim como a recolha de dados históricos relacionados aos preços do milho e factores económicos, sociais e climáticos que influenciam essas variações.

Adicionalmente, a pesquisa envolveu o desenvolvimento de um modelo computacional, utilizando algoritmos de Machine Learning, que foram treinados e validados com base nos dados recolhidos. Este modelo foi concebido para fornecer previsões que possam apoiar agricultores, comerciantes e decisores políticos na gestão e planeamento das actividades relacionadas ao mercado de milho. Assim, a metodologia combinou aspectos teóricos e práticos, utilizando ferramentas computacionais para consolidar os objectivos do estudo.

## **3.2. Classificação da Pesquisa**

A presente pesquisa é classificada sob diversas perspectivas, conforme se segue.

### **3.2.1. Quanto à abordagem**

De acordo com Bryman (2012), a abordagem de uma pesquisa pode ser qualitativa, quantitativa ou mista, sendo escolhida de acordo com a natureza do problema investigado e os métodos de análise necessários. A pesquisa qualitativa foca na compreensão profunda do fenómeno estudado, enquanto a quantitativa prioriza a análise numérica de dados e o estabelecimento de relações estatísticas. Este trabalho adoptou uma abordagem mista, considerando que o problema exige uma análise integrada de dados qualitativos e quantitativos.

A dimensão qualitativa permitiu compreender os desafios enfrentados pelos agricultores moçambicanos e as dinâmicas do mercado de milho, enquanto a quantitativa viabilizou a modelagem matemática, utilizando dados numéricos históricos para avaliar o desempenho dos algoritmos de Machine Learning. Este tipo de abordagem mista foi crucial para alinhar a complexidade do problema com a proposta de soluções, assegurando resultados robustos e relevantes para a realidade local.

### **3.2.2. Quanto aos objectivos**

Seguindo a classificação proposta por Lakatos e Marconi (2003), os estudos podem ser exploratórios, descritivos ou explicativos. Este trabalho enquadra-se como exploratório, pois busca oferecer uma compreensão inicial sobre a aplicação de Machine Learning no contexto da previsão de preços agrícolas, especialmente do milho, em Moçambique. O carácter exploratório é reforçado pela necessidade de identificar padrões e relações entre os factores que influenciam os preços, construindo hipóteses e fundamentações que possibilitem a implementação do modelo preditivo.

Embora o foco principal seja exploratório, o trabalho também apresenta elementos descritivos ao detalhar os processos de recolha e análise dos dados, bem como as etapas de construção e validação do modelo preditivo. Ao relacionar os factores económicos e climáticos com os preços históricos, o estudo contribui para a compreensão das dinâmicas do mercado agrícola e oferece subsídios para futuras pesquisas e intervenções no sector.

### **3.2.3. Quanto à natureza**

De acordo com Gil (2002), a classificação da pesquisa quanto à natureza pode ser teórica, aplicada ou metodológica. Este trabalho enquadra-se como uma pesquisa aplicada, pois visa desenvolver um modelo

tecnológico prático que utiliza algoritmos de Machine Learning para prever o preço do milho em Moçambique, e o seu foco principal é resolver um problema concreto e específico, relacionado às necessidades de agricultores, comerciantes e gestores no sector agrícola. A aplicação prática dos resultados obtidos pelo modelo proposto demonstra o carácter pragmático da pesquisa, destacando sua contribuição directa para a melhoria da tomada de decisão no mercado agrícola.

#### **3.2.4. Quanto ao método**

Conforme defendido por Prodanov e Freitas (2013), os métodos de pesquisa podem ser classificados como experimental, histórico, comparativo, estudo de caso ou levantamento. O método utilizado neste trabalho é predominantemente de estudo de caso, uma vez que a pesquisa se debruça sobre o contexto específico do mercado de milho em Moçambique. Esta abordagem permite explorar profundamente as dinâmicas locais, como os factores económicos, sociais e climáticos que afectam os preços do milho, além de avaliar a eficácia dos algoritmos de Machine Learning aplicados a esse cenário particular.

A escolha do estudo de caso é adequada, pois oferece uma visão contextualizada e detalhada do problema, proporcionando subsídios para a construção de um modelo preditivo que reflecte as características únicas do ambiente moçambicano. A combinação dessa abordagem com técnicas exploratórias reforça o alinhamento entre os objectivos da pesquisa e os métodos empregados, garantindo a relevância e aplicabilidade dos resultados no contexto estudado.

#### **3.2.5. Material e Ficha Técnica**

Os dados históricos sobre o preço do milho em Moçambique foram obtidos a partir de fontes públicas online fornecidas pela Organização das Nações Unidas para Alimentação e Agricultura (FAO - Food and Agriculture Organization) sob a **licença** “*Creative Commons Attribution for Intergovernmental Organisations (CC BY-IGO)*”. Estes dados abrangem um período extenso de 32 anos, de 1992 a 2024, permitindo uma análise abrangente da evolução dos preços ao longo de várias décadas.

Além dos preços históricos do milho, foram incluídos outros factores económicos e ambientais que influenciam o comportamento dos preços. Entre eles, destacam-se:

**1. Taxa de Inflação:** Disponibilizada publicamente no site oficial do Banco Central de Moçambique. Estes dados fornecem um indicador macroeconómico essencial para entender a variação do poder de compra ao longo do tempo.

**2. Nível de Precipitação Regional:** Este dados (também disponibilizados pela FAO) são cruciais para avaliar o impacto das condições climáticas sazonais e anuais na produção agrícola.

**3. Nível de Vegetação Regional:** Estes (também disponibilizados pela FAO), permitem analisar o estado da vegetação em várias regiões, influenciando a disponibilidade e a qualidade do milho.

Entre as **variáveis** destacadas na análise, observou-se que:

- O preço do milho apresentou uma correlação positiva forte com o **ano** ( $r \approx 0,85$ ), evidenciando uma tendência de crescimento estrutural dos preços ao longo das décadas.
- O **nível de precipitação** revelou uma correlação positiva moderada com o preço ( $r \approx 0,62$ ), sugerindo que maiores volumes de chuva podem estar associados a variações de custo, evidentemente pela relação entre excesso de precipitação e perdas de produção.
- O **índice de vegetação (NDVI)** exibiu uma correlação negativa baixa com os preços ( $r \approx -0,29$ ), o que indica que melhores condições vegetativas tendem a estar associadas a preços mais baixos, mas com impacto limitado.
- A **taxa de inflação (IPC)** apresentou uma **correlação fraca positiva** com o preço ( $r \approx 0,09$ ), mostrando que a sua influência directa é menos expressiva em comparação com factores climáticos e sazonais.

Para o tratamento, análise e modelação dos dados, recorreu-se ao **software Anaconda**, que integra um ecossistema robusto para ciência de dados. Dentro dele, destacaram-se as seguintes ferramentas: **Python**, com bibliotecas como *pandas* para manipulação de dados, *matplotlib* e *seaborn* para visualização gráfica, e *scikit-learn* e *TensorFlow* para modelagem e experimentação com algoritmos de *Machine Learning*.

### 3.3. Procedimentos Técnicos

#### 3.3.1. Recolha e Análise Exploratória de Dados

Para operacionalizar a pesquisa, foi realizada a recolha de dados históricos sobre os preços do milho em Moçambique, bem como variáveis correlacionadas, como climáticas e económicas. Esses dados foram obtidos em bases públicas web, incluindo instituições nacionais e organizações internacionais, **respeitando suas respectivas licenças**.

O pré-processamento de dados para modelagem envolve várias etapas essenciais. Primeiro, a uma observação geral dos dados, realiza-se a limpeza destes para remover valores ausentes, duplicados ou inválidos, garantindo a sua qualidade. Para os preços históricos do milho, podem ser aplicadas técnicas de imputação, como a média ou interpolação. Em seguida, os dados brutos, como os preços diários, são normalizados ou escalonados para assegurar que as variáveis estejam na mesma escala, o que é crucial para algoritmos como LSTM.

Além disso, é feita a engenharia de atributos, onde variáveis adicionais que influenciam o preço, como condições meteorológicas, custos adicionais, etc, são incorporadas para melhorar a precisão do modelo.

Para representação dos dados, foi utilizada a análise gráfica de tendência, explorando séries temporais e diagramas de dispersão que permitiram identificar padrões sazonais, flutuações cíclicas e possíveis outliers ao longo do tempo. Foram também gerados gráficos comparativos entre os preços do milho e variáveis correlacionadas, como precipitação, índice de vegetação e taxa de inflação, possibilitando uma visualização clara da relação entre factores climáticos, económicos e o comportamento do mercado. Essa abordagem visual foi adoptada para facilitar a compreensão preliminar da dinâmica dos dados, como também servir de suporte para a formulação de hipóteses e para a seleção das variáveis mais relevantes a serem incluídas no processo de modelação.

### 3.3.2. Escolha dos Modelos de Machine Learning

As características dos dados e a natureza do problema a ser abordado neste estudo envolvem alta variabilidade e múltiplos factores influenciadores, logo, a selecção dos modelos de Machine Learning para este trabalho foi feita tendo em conta essa complexidade.

Este estudo considera a combinação de robustez, capacidade de generalização e desempenho preditivo como critérios fundamentais para garantir que os modelos escolhidos sejam adequados ao contexto da previsão de preços do milho em Moçambique. Com base em pesquisas existentes e práticas recomendadas, foram seleccionados os modelos **Random Forest**, **LSTM (Long Short-Term Memory)** e **Regressão Multilinear**, cada um com capacidades específicas que se complementam para lidar com os desafios inerentes ao problema.

#### **Random Forest**

O modelo **Random Forest** foi seleccionado para este estudo pela sua eficácia em lidar com conjuntos de dados complexos e não lineares, uma característica essencial num domínio como a agricultura, onde

variáveis como condições climáticas, factores económicos e práticas agrícolas podem interagir de formas imprevisíveis.

Segundo Breiman (2001), o modelo de Random Forest combina a criação de múltiplas árvores de decisão independentes, de modo que permite alcançar uma precisão robusta, mesmo em cenários com dados ruidosos ou incompletos. Esta robustez é particularmente relevante em contextos onde a qualidade dos dados pode ser afectada por lacunas ou inconsistências nas medições, como frequentemente ocorre em sistemas agrícolas. Além disso, o Random Forest oferece a capacidade de medir a importância relativa das variáveis independentes, um recurso valioso para identificar os factores que mais influenciam a precificação do milho.

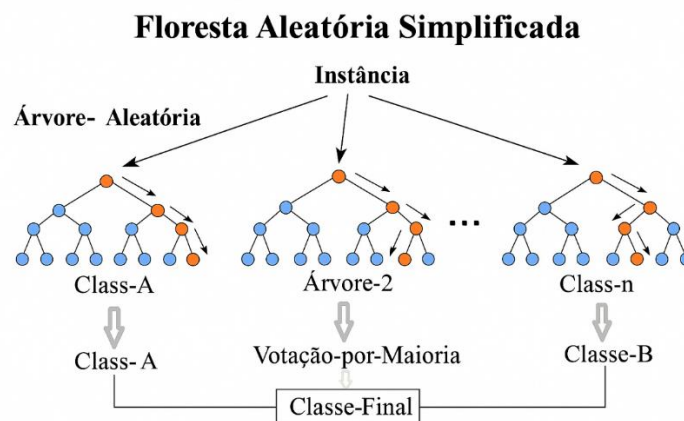


Figura 6. Estrutura de uma Random Forest (fonte: IBM)

## Descrição do Funcionamento

O Random Forest opera construindo múltiplas árvores de decisão durante a fase de treino e produzindo a classe que é o modo das classes (classificação) ou a média das previsões (regressão) das árvores individuais. Cada árvore é construída a partir de uma amostra aleatória com reposição dos dados de treino (bootstrap) e, em cada nó de decisão, uma seleção aleatória de características é considerada para determinar a melhor divisão. Isso introduz diversidade entre as árvores, resultando em um modelo mais robusto e menos propenso ao overfitting (IBM, 2023).

## Principais Parâmetros

- **Número de Árvores (n\_estimators):** Quantidade de árvores na floresta.

- **Número Máximo de Características (max\_features):** Número de características a serem consideradas para determinar a melhor divisão em cada nó.
- **Profundidade Máxima da Árvore (max\_depth):** Profundidade máxima permitida para cada árvore.
- **Tamanho Mínimo da Folha (min\_samples\_leaf):** Número mínimo de amostras necessárias para formar uma folha.
- **Critério de Divisão (criterio\_n):** Função utilizada para medir a qualidade de uma divisão (por exemplo, Gini ou entropia para classificação).

**Vantagens:** O Random Forest é robusto e capaz de capturar interações não-lineares entre variáveis, sendo aplicável a uma ampla variedade de problemas de previsão, incluindo séries temporais e dados multivariados. É menos sensível a outliers e é eficaz para problemas de classificação e regressão.

**Desvantagens:** Apesar de sua flexibilidade, este modelo não possui uma memória interna de longo prazo, como o LSTM, o que o torna menos indicado para dados com dependências temporais complexas. Ele também pode ser computacionalmente pesado em conjuntos de dados grandes e menos interpretável do que modelos lineares.

## LSTM

Por outro lado, o modelo **LSTM**, uma variante das redes neurais recorrentes, foi seleccionado devido à sua capacidade superior de modelar dependências temporais em séries cronológicas.

Conforme Hochreiter e Schmidhuber (1997), a LSTM é projectada para lidar com a longa dependência temporal, preservando informações relevantes ao longo de períodos extensos. Essa característica é crucial para o problema em questão, pois os preços do milho frequentemente exibem padrões sazonais e flutuações associadas a factores climáticos e económicos que se desenvolvem ao longo do tempo. Ao capturar essas dinâmicas complexas, a LSTM pode fornecer previsões precisas e bem fundamentadas, adaptando-se a variações temporais e tendências emergentes. Essa capacidade torna este modelo uma ferramenta indispensável para compreender e prever alterações nos preços do milho no mercado moçambicano.

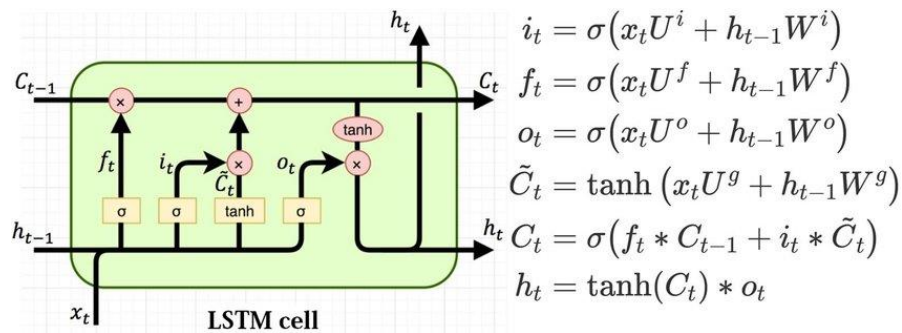


Figura 7. Célula e fórmula do modelo LSTM (fonte: IBM)

### Descrição do Funcionamento:

Segundo Hochreiter & Schmidhuber (1997), as LSTMs são projetadas para modelar dependências de longo prazo em seqüências de dados. Elas utilizam uma estrutura de portas que controlam o fluxo de informações, permitindo que a rede retenha ou descarte informações conforme necessário. Isso as torna eficazes em tarefas como processamento de linguagem natural e previsão de séries temporais, onde a memória de eventos passados é crucial.

### Principais Parâmetros:

- **Dimensão do Estado Oculto (hidden\_size):** Número de unidades na camada oculta.
- **Número de Camadas (num\_layers):** Quantidade de camadas LSTM empilhadas.
- **Taxa de Dropout (dropout):** Probabilidade de zerar as conexões para evitar overfitting.
- **Direcionalidade (bidirectional):** Indica se a LSTM é unidirecional ou bidirecional.
- **Função de Ativação:** Funções como *tanh* e *sigmoid* são comumente usadas nas portas.

**Vantagens:** O LSTM é capaz de capturar padrões sazonais, tendências e ciclos de longo prazo, sendo especialmente útil em séries temporais onde há dependências temporais complexas. É amplamente utilizado em previsões de dados financeiros, meteorológicos e de demanda.

**Desvantagens:** É um modelo computacionalmente intensivo, exigindo um volume significativo de dados para ser treinado de maneira eficaz. Isso o torna menos eficiente em comparação a modelos mais simples em casos com séries curtas ou onde o relacionamento entre variáveis é linear.

## Regressão Multilinear

Por fim, a **Regressão Multilinear** foi incluída neste estudo devido à sua simplicidade, interpretabilidade e capacidade de fornecer uma base para análise comparativa.

De acordo com Montgomery et al. (2012), este modelo estatístico é amplamente utilizado para explorar relações lineares entre variáveis dependentes e independentes, proporcionando uma visão clara sobre o peso de cada factor na explicação das variações observadas. No contexto da previsão de preços do milho, a Regressão Multilinear é especialmente útil para estabelecer uma relação directa entre variáveis explicativas, como produção agrícola, custo de transporte e condições macroeconómicas. Além disso, a sua inclusão permite uma análise contrastante com os resultados obtidos pelos modelos mais complexos, como o Random Forest e a LSTM, assegurando uma abordagem compreensiva e diversificada.

The diagram shows the formula  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \epsilon$  with arrows pointing to each part from descriptive labels. 'Dependent Variable (Response Variable)' points to 'Y'. 'Independent Variables (Predictors)' points to the 'X' terms. 'Y intercept' points to ' $\beta_0$ '. 'Slope Coefficient' points to ' $\beta_1$ ' and ' $\beta_2$ '. 'Error Term' points to ' $\epsilon$ '.

Figura 8. Fórmula do modelo de R.L.M (fonte: IBM)

### Descrição do Funcionamento:

Segundo o IBM, a Regressão Multilinear modela a relação entre uma variável dependente e múltiplas variáveis independentes. O objectivo é ajustar uma linha (ou hiperplano em dimensões superiores) que minimize a soma dos quadrados dos resíduos entre os valores observados e os valores previstos pelo modelo. Isso permite prever o valor da variável dependente com base nos valores das variáveis independentes.

### Principais Parâmetros:

- **Coefficientes ( $\beta_1, \beta_2, \dots, \beta_n$ ):** Os valores que indicam a magnitude e direção do impacto de cada variável independente na variável dependente.
- **Intercepto ( $\beta_0$ ):** O valor esperado da variável dependente quando todas as variáveis independentes são zero.
- **Erro ( $\epsilon$ ):** O desvio entre os valores observados e previstos pelo modelo, assumindo que segue uma distribuição normal.

- **Método de Estimação:** Normalmente, os coeficientes são determinados utilizando o método dos mínimos quadrados ordinários (OLS), que minimiza a soma dos quadrados dos erros.

**Vantagens:** A simplicidade do modelo torna-o fácil de interpretar e de implementar, sendo útil em problemas onde as variáveis apresentam relações aproximadamente lineares. A Regressão Multilinear é amplamente utilizada como uma linha de base em previsões de curto prazo e para análises exploratórias.

**Desvantagens:** Este modelo não captura bem as interações complexas e não-lineares entre variáveis. Quando os dados têm comportamento não-linear ou efeitos de dependência complexos, como sazonalidade ou ciclicidade, a regressão multilinear tende a ter um desempenho inferior a modelos como o LSTM ou o Random Forest.

Em suma, a combinação desses modelos foi estrategicamente adoptada para maximizar a capacidade de previsão e proporcionar uma análise rica e abrangente. Enquanto o Random Forest destaca-se na identificação de interações complexas entre variáveis, a LSTM captura padrões temporais de longo prazo, e a Regressão Multilinear fornece uma base interpretativa clara.

A presente abordagem integrada garante um reforço da precisão do estudo e assegura que os resultados obtidos sejam confiáveis, práticos e aplicáveis às necessidades reais do mercado agrícola em Moçambique.

### 3.3.3. Divisão da Amostra de Treino e Teste

Para avaliar de forma rigorosa a capacidade de generalização dos modelos de previsão, o conjunto de dados foi dividido em dois subconjuntos distintos: **treino (80%) e teste (20%)**. O subconjunto de treino foi utilizado para ajustar os parâmetros internos dos algoritmos, permitindo que estes aprendessem os padrões históricos do preço do milho e as relações existentes com as variáveis explicativas. Já o subconjunto de teste foi reservado para validar o desempenho em exemplos completamente novos, que não foram apresentados durante o processo de treino, garantindo assim uma avaliação imparcial.

A divisão da amostra seguiu práticas recomendadas na literatura (Bishop, 2006), que sugerem a utilização de percentagens equilibradas entre treino e teste, de modo a assegurar que ambos os subconjuntos sejam representativos. Este procedimento é fundamental para reduzir o risco de viés e

para identificar situações de sobreajuste (overfitting) ou subajuste (underfitting). Assim, a avaliação com dados de teste fornece uma medida realista da precisão dos modelos quando aplicados em cenários futuros, reforçando a sua robustez e aplicabilidade prática.

#### **3.3.4. Processo de Ajuste de Parâmetros e Métricas**

O ajuste dos parâmetros é uma fase essencial na construção de modelos preditivos eficazes, como os utilizados neste trabalho. Especificamente, este procedimento centra-se na definição dos hiperparâmetros, que são estabelecidos antes da etapa de treino do modelo. Estes hiperparâmetros têm um papel crucial e influenciam directamente na capacidade do modelo de realizar previsões precisas e generalizar adequadamente para novos dados. De acordo com Goodfellow, Bengio e Courville (2016), os hiperparâmetros devem ser cuidadosamente ajustados para evitar problemas como subajuste ou sobreajuste, garantindo que o modelo consiga capturar os padrões complexos presentes nos dados sem perder a capacidade de generalização.

Para determinar os valores ideais dos hiperparâmetros, foram utilizadas técnicas de optimização, com destaque para a pesquisa em grelha (grid search) e a pesquisa aleatória (random search). A pesquisa em grelha permite explorar sistematicamente combinações de valores dentro de intervalos pré-definidos, enquanto a pesquisa aleatória selecciona valores dentro de intervalos especificados de forma estocástica, aumentando a eficiência ao reduzir o espaço de busca. Ambas as abordagens foram aplicadas no contexto deste trabalho, tendo em conta as características específicas de cada modelo utilizado, como Random Forest, LSTM e Regressão Multilinear.

Além disso, a validação cruzada foi empregue para avaliar a performance dos modelos durante o processo de ajuste. Nesta técnica, os dados são subdivididos em múltiplos subconjuntos, e o treino é realizado iterativamente em diferentes combinações desses subconjuntos, permitindo uma avaliação mais robusta do desempenho do modelo.

#### **3.3.5. Medidas de Avaliação de Desempenho do Modelo**

Neste processo, foram adoptadas medidas quantitativas, como o **RMSE** (Raiz do Erro Quadrático Médio) e o **R<sup>2</sup>** (Coeficiente de Determinação), para guiar o processo de ajuste dos modelos tendo como base a sua simplicidade de interpretação e capacidade de fornecer uma medida clara e confiável da diferença entre os valores previstos pelo modelo e os valores reais.

O RMSE é amplamente reconhecido como uma das medidas mais informativas para avaliar a precisão de modelos de previsão, sobretudo em contextos onde a magnitude do erro tem implicações práticas significativas. De acordo com Hyndman e Athanasopoulos (2018), o RMSE, por ser derivado directamente do Mean Squared Error (MSE), herda a capacidade de penalizar de forma mais severa os desvios de maior magnitude, característica essencial em domínios sensíveis a grandes erros, como é o caso da previsão de preços agrícolas. Neste tipo de aplicação, erros acentuados podem conduzir a decisões económicas erradas, comprometer o planeamento da produção ou prejudicar a alocação de recursos.

Uma das principais vantagens do RMSE em relação ao MSE é o facto de este apresentar os erros na mesma unidade dos dados observados, o que torna a sua interpretação mais directa e intuitiva. Esta propriedade facilita a comunicação dos resultados a partes interessadas não técnicas e permite uma avaliação mais realista do desempenho do modelo (Chai & Draxler, 2014). Por exemplo, ao prever o preço de uma tonelada de milho, um RMSE de 150 meticais indica que, em média, os desvios das previsões em relação aos valores reais rondam esse valor, o que é mais compreensível do que um erro ao quadrado.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Figura 9. Erro Quadrático Médio (fórmula; fonte: kevinma, medium.com)

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Figura 10. Raiz do Erro Quadrático Médio (fórmula; fonte: kevinma, medium.com)

Por outro lado, o **coeficiente de determinação (R<sup>2</sup>)** é uma das medidas estatísticas mais utilizadas na avaliação do desempenho de modelos de regressão, sobretudo devido à sua capacidade de quantificar a proporção da variância total da variável dependente que é explicada pelo modelo. Segundo Hyndman e Athanasopoulos (2018), o R<sup>2</sup> fornece uma medida normalizada da qualidade do ajuste do modelo,

variando entre 0 e 1, onde valores mais próximos de 1 indicam um modelo com elevada capacidade explicativa.

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Figura 11. Coeficiente de Determinação (fórmula; fonte: kevinma, medium.com)

Portanto, ter essas medidas em consideração no presente trabalho não apenas visa seguir práticas consolidadas na área de Machine Learning, mas também reflecte a necessidade de garantir um desempenho de previsão confiável e preciso para o modelo desenvolvido. A sua utilização como métrica principal assegura que os resultados obtidos sejam sólidos e adequados às exigências práticas do mercado de milho em Moçambique e contribui para a confiabilidade e relevância do sistema de previsão.

### 3.3.6. Processo de Melhoramento de Modelo

Após o ajuste sistemático dos hiperparâmetros, recorreu-se a procedimentos baseados em normas consolidadas, como a utilização de pesquisa em grelha (grid search) e pesquisa aleatória (random search), complementadas por validação cruzada, para garantir a selecção dos valores que maximizassem o desempenho preditivo dos modelos. Este processo envolveu a definição de intervalos adequados para parâmetros críticos como a taxa de aprendizagem, profundidade máxima das árvores (no caso do RF) e número de estimadores, seguido de múltiplas iterações que permitiram avaliar a robustez das configurações testadas.

Durante o treino, os modelos analisaram os padrões históricos de variação de preços e as relações entre as variáveis explicativas, aprendendo a identificar dependências complexas. A implementação foi realizada em frameworks avançados de Machine Learning (TensorFlow e Scikit-learn), que oferecem mecanismos otimizados para treino eficiente, monitorização do erro e prevenção de problemas como sobreajuste (overfitting) e subajuste (underfitting). Esta etapa assegurou que os modelos retivessem apenas os padrões mais relevantes, mantendo a capacidade de generalizar para novos dados.

### 3.3.7. Processo de Tomada de Decisão

Com base nos resultados obtidos nas fases anteriores, foi realizada a selecção do modelo final que apresentasse o melhor desempenho no contexto específico da previsão de preços do milho. O critério principal para a escolha foi a capacidade de generalização demonstrada pelos modelos nos dados de teste. Entre os modelos testados, o **LSTM** destacou-se pela sua robustez em lidar com a volatilidade e a não linearidade associadas às séries temporais de preços agrícolas.

A escolha do LSTM como modelo principal deve-se à sua capacidade de capturar padrões complexos em dados sequenciais, conforme evidenciado por Hochreiter e Schmidhuber (1997). Este modelo mostrou-se particularmente eficaz em identificar flutuações sazonais e tendências emergentes, factores essenciais para prever preços futuros no mercado agrícola moçambicano. Além disso, a sua capacidade de preservar informações relevantes ao longo do tempo reforçou sua adequação ao problema em questão.

Após a selecção, o modelo LSTM foi integrado a um sistema web interactivo, que permite aos utilizadores aceder às previsões de forma prática e eficiente. Este sistema foi projectado para atender às necessidades de diferentes partes interessadas, como agricultores, comerciantes e decisores políticos, fornecendo informações cruciais para o planeamento e gestão das actividades relacionadas ao mercado de milho.

## 3.4. Métodos de Implementação

A implementação do sistema de previsão de preços do milho envolveu a utilização de ferramentas e técnicas modernas, tanto no front-end como no back-end. Este ponto descreve a prática de implementação do sistema, abrangendo desde as linguagens de programação e frameworks até a infraestrutura necessária para suportar a previsão dos preços de milho em larga escala.

### 3.4.1. Linguagem de Programação e Frameworks

A linguagem central escolhida para o desenvolvimento do sistema foi **Python**, devido ao seu vasto ecossistema de bibliotecas voltadas para o Machine Learning. Entre as bibliotecas utilizadas, destacam-se o **TensorFlow** e o **Scikit-learn**, que foram fundamentais na implementação e treino dos modelos de previsão, especialmente o LSTM, considerado o mais adequado para lidar com previsões de longo prazo.

Para a construção da plataforma web, o framework **Django** foi o seleccionado. Baseado em Python e seguindo a arquitectura Model-Template-View (MTV), o Django facilitou a integração dos algoritmos de previsão com a aplicação. Além disso, foi utilizado para desenvolver uma API RESTful para permitir que as previsões dos preços do milho fossem oferecidas como serviços web, que **podem ser consumidos por outros sistemas** ou interfaces

No que diz respeito ao desenvolvimento da interface gráfica (frontend), a biblioteca **React.js** foi a escolhida por proporcionar uma experiência interactiva e eficiente para os utilizadores, esta permite a visualização e interacção com os dados de forma dinâmica e intuitiva.

### 3.4.2. Integração de Modelos e Dados

A integração entre os modelos de previsão e o sistema web foi implementada utilizando uma arquitectura modular. Isso permitiu uma fácil manutenção e actualização dos algoritmos sempre que necessário. Esse tipo de arquitectura garante que os diferentes componentes do sistema, como os modelos de previsão e as interfaces, possam ser modificados ou melhorados de forma independente, sem impactar negativamente no funcionamento geral.

No que respeita à base de dados, foi escolhido o **SQLite3** para armazenar os dados históricos dos preços do milho, assim como outros dados essenciais, tais como condições climáticas e variáveis de mercado. A escolha do SQLite3 deve-se à sua robustez e à capacidade de lidar com operações de forma simples, além de oferecer flexibilidade no armazenamento de dados em sistemas django, o que foi importante para a variedade de informações utilizadas neste estudo.

O processo de treino dos modelos é realizado em segundo plano, através de uma fila de tarefas assíncronas. Esta abordagem assegura que o sistema web continue a funcionar de forma eficiente, respondendo rapidamente às solicitações dos utilizadores, mesmo durante o treino ou actualização dos modelos com novos dados. Desta forma, o desempenho da aplicação não é comprometido, e a experiência do utilizador mantém-se fluida.

Por fim, as previsões geradas pelos modelos são disponibilizadas através de uma **API RESTful**, que permite aos utilizadores acederem às previsões de preços do milho de maneira prática e em tempo real. Além disso, essa API facilita a integração com outras plataformas ou sistemas que possam necessitar dessas previsões, ampliando o alcance e a utilidade dos dados gerados pela aplicação.

### 3.4.3. Implementação da Interface Gráfica

A interface gráfica (UI) foi desenhada com ênfase na simplicidade e facilidade de uso, **especialmente para utilizadores com pouca ou nenhuma experiência técnica**. O objectivo principal foi criar uma experiência intuitiva que permitisse aos utilizadores acederem rapidamente às previsões e entenderem os dados apresentados.

Esta, proporciona aos utilizadores uma visão detalhada sobre o mercado do milho e inclui explicações sobre os factores que influenciam os preços, tendências históricas e previsões futuras para que os utilizadores compreendam melhor as previsões geradas pelo sistema, permitindo uma tomada de decisão mais informada.

### 3.4.4. Testes e Validação

A fase de testes foi crucial para assegurar que o sistema estivesse a funcionar correctamente. Aqui, foram aplicados diversos tipos de testes, de modo a garantir a fiabilidade e o desempenho da aplicação.

Foram executados **Testes Unitários** para avaliar cada componente individualmente, como os modelos de previsão, a API e as funcionalidades da interface. O objectivo destes testes foi verificar se cada parte do sistema operava de forma isolada e garantir o seu correcto funcionamento antes da integração com os demais componentes.

Além disso, foram realizados **Testes de Integração**, focando na interacção entre os diferentes módulos do sistema. Esses testes avaliaram a comunicação entre o backend e o frontend, bem como a interacção com a base de dados. O intuito foi garantir que a troca de informações entre os diferentes componentes fosse fluida e sem falhas, assegurando a coesão do sistema como um todo.

Por fim, foram realizados **Testes de Desempenho** a fim de avaliar a capacidade do sistema de lidar com um grande número de utilizadores. Estes testes garantiram que, mesmo sob alta demanda, o sistema continuasse a fornecer previsões de forma eficiente, sem comprometer a experiência do utilizador.

### 3.4.5. Diagramas

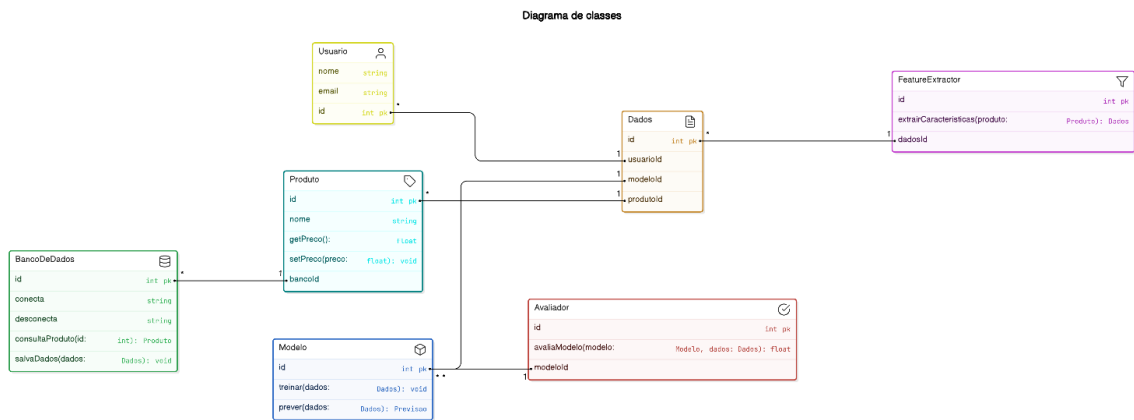


Figura 12. Diagrama de Classes

**Diagrama de Use Case**

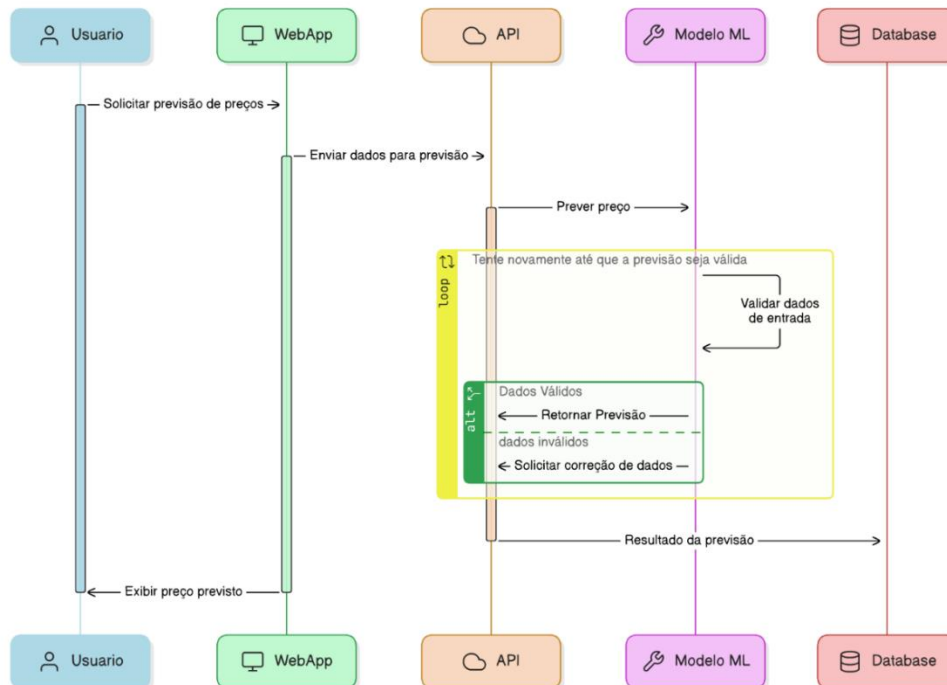


Figura 13. Diagrama de Sequência de Eventos

## 4. Experimentos e Resultados

Nesta secção, são detalhados os experimentos realizados para avaliar o desempenho dos modelos de Machine Learning na previsão do preço do milho em Moçambique. A análise foi conduzida com dados históricos e múltiplos factores externos, como as condições climáticas e a taxa de inflação no mercado, para garantir a precisão e a robustez das previsões. Os resultados são apresentados com base em medidas de avaliação e na capacidade dos modelos de generalizar para novos dados.

### 4.1. Estatística Analítica Descritiva (EAD)

A tabela abaixo ilustra a primeira impressão que temos dos dados brutos antes de qualquer tratamento.

date	admin1	admin2	market	latitude	longitude	category	commodity	unit	priceflag	pricetype	currency
#date	#adm1+name	#adm2+name	#loc+market+name	#geo+lat	#geo+lon	#item+type	#item+name	#item+unit	#item+price+flag	#item+price+type	#currency
1992-11-15	Maputo City	Cidade_De_Maputo	Maputo	-25.965278	32.589167	cereals and tubers	Maize (white)	KG	actual	Retail	MZN
1992-12-15	Gaza	Chokwe	Chokwe	-24.533333	32.983333	cereals and tubers	Maize (white)	KG	actual	Retail	MZN
1992-12-15	Inhambane	Maxixe	Maxixe	-23.859722	35.347222	cereals and tubers	Maize (white)	KG	actual	Retail	MZN
1992-12-15	Maputo City	Cidade_De_Maputo	Maputo	-25.965278	32.589167	cereals and tubers	Maize (white)	KG	actual	Retail	MZN

Figura 14. Tabela do Conjunto de Dados Original

#### 4.1.1. Tratamento Preliminar dos Dados

Na fase inicial de tratamento dos dados, foram realizadas as seguintes etapas:

- Remoção das linhas de cabeçalho com comentários.
- Verificação da ausência de valores duplicados.
- Inspeção das colunas para confirmar a inexistência de valores em falta; planeamento de estratégias para imputação. Criação de um dicionário de dados com a descrição detalhada das variáveis.

Estas etapas serviram de preparo ao o conjunto de dados para exploração detalhada (EDA), engenharia de atributos e modelagem preditiva.

Coluna	Descrição
date	Data da observação
admin1	Nome da província (admin1)
admin2	Nome do distrito (admin2)
market	Nome do mercado
latitude	Latitude do mercado
longitude	Longitude do mercado
category	Categoria do alimento
commodity	Produto alimentício
unit	Unidade de medida
priceflag	Indicador de validação do preço
pricetype	Tipo de preço (por exemplo, atacado ou varejo)
currency	Moeda em que o preço foi registrado
price	Preço na moeda local (MZN)
usdprice	Preço convertido para dólares americanos (USD)

Figura 15. Dicionário de Dados

#### 4.1.2. Compreensão Inicial dos Dados

A análise inicial dos dados nos revela um total de **63.834** registros no conjunto.

A variável data (date) apresenta 370 datas únicas. Geograficamente, o conjunto de dados inclui 11 províncias (admin1), com maior representação de Inhambane (9.276 registros). A nível distrital (admin2), destacam-se 81 distritos, sendo a Cidade de Maputo o mais representado (4.096 registros). Os dados abrangem 99 mercados distintos, destacando-se Maputo com 2.913 registros.

A categoria de produtos mais frequente é dos cereais e tubérculos (cereals and tubers) (25.851 registros), e o produto predominante é o milho (5.830 registros). Este destaque do milho como o produto com mais dados disponíveis agrega valor na escolha deste conjunto para o estudo de previsão do preço do milho, uma vez que a abundância de observações aumenta o potencial de construir modelos preditivos robustos e fiáveis.

A unidade de medida mais comum é o quilograma (KG, 54.019 registos), com preços consistentemente registados como tipo “actual” e “Retail” em moeda local (MZN).

O milho destaca-se entre as amostras, com “Maize (white)” a somar 5 830 registos e várias farinhas de milho (primeira qualidade, com/sem farelo) ultrapassando juntas 6 000 registos, evidenciando a sua importância como alimento básico e a diversidade de produtos ajustada às necessidades dos consumidores.

Maize (white)	5830
Sugar (brown, local)	4270
Rice (imported)	4260
Oil (vegetable, local)	4222
Cowpeas	3395
Wheat flour (local)	2990
Rice	2936
Maize meal (white, first grade)	2921
Beans (butter)	2667
Groundnuts (small, shelled)	2518

Figura 16. Amostra de Contagem de Valores

#### 4.1.3. Distribuição de Produtos por Categoria

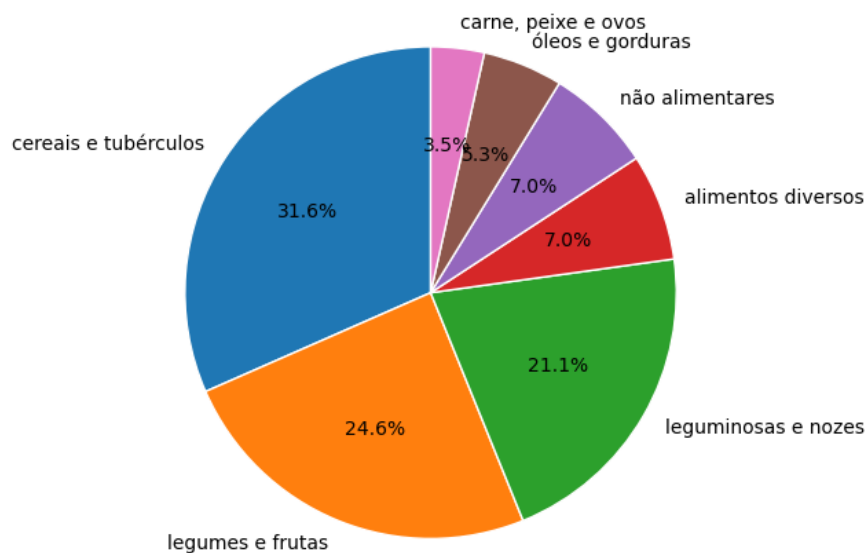


Figura 17. Distribuição de Registos por Categoria

#### 4.1.4. Distribuição histórica do preço médio do milho em Moçambique

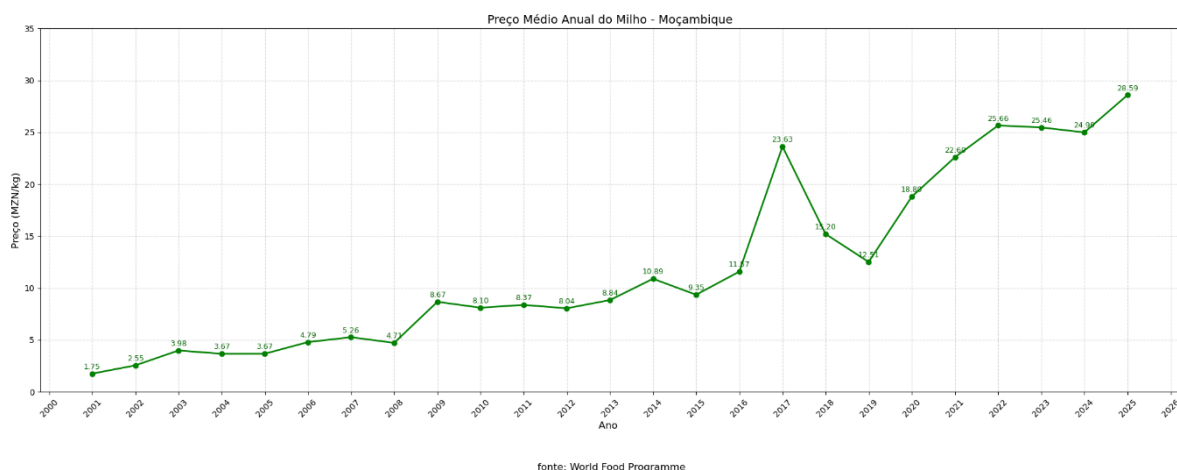


Figura 18. Evolução Histórica do Preço Médio do Milho em Moçambique

#### O Pico do Preço em 2017

No gráfico da distribuição histórica dos preços médios é percebido um fenômeno particular. O pico isolado no preço médio do milho em Moçambique durante o ano de 2017 foi amplamente documentado pelo Ministério da Agricultura e Desenvolvimento Rural (MADER). Segundo relatórios oficiais, esse aumento foi particularmente acentuado nas províncias do sul do país, como Inhambane, Gaza e Maputo, onde os preços ultrapassaram os 25 meticais por quilograma.

#### Preço Médio Anual em Tete

A província de **Tete é a maior produtora de milho em Moçambique**. De acordo com dados do Instituto Nacional de Estatística (INE), em 2023, Tete liderou a produção nacional com 501.080 toneladas, seguida pela província de Manica, que produziu 398.619 toneladas.

Além disso, o Inquérito Agrário Integrado de 2020 indicou que Tete contribuiu com aproximadamente 28% da produção nacional de milho, destacando-se como a principal região produtora do país.

fonte: Inquérito Agrário Integrado, IAI (MADER, 2020)



Figura 19. Evolução Histórica do Preço Médio do Milho em Tete

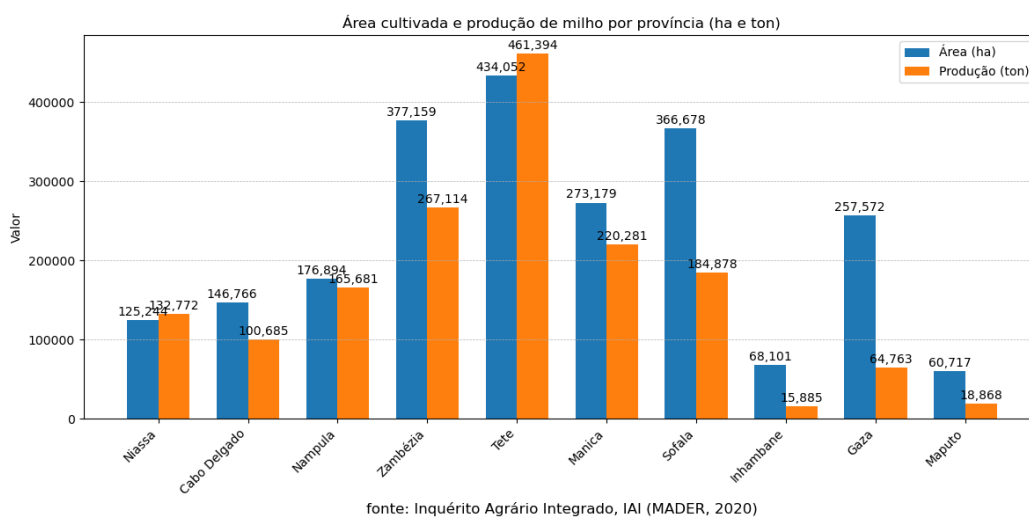


Figura 20. Distribuição da Área e Produção do Milho, 2020

### Varição dos Preços entre os Mercados em 2017

De modo a compreender como cada mercado/província reagiu ao pico de 2017, observamos a variação dos preços entre os mercados.

Resumo das diferenças por província:				
admin1	min	max	range	std
Zambezia	11.622500	17.917500	6.295000	2.833378
Gaza	18.931250	25.018333	6.087083	4.304218
Sofala	11.072500	15.619167	4.546667	2.314385
Tete	11.647500	15.725000	4.077500	2.213467
Cabo_Delgado	11.430000	14.534167	3.104167	2.194977
Manica	14.718333	17.443333	2.725000	1.926866
Inhambane	18.912222	21.382222	2.470000	1.746554
Niassa	6.035000	7.322500	1.287500	0.910400
Nampula	12.370909	13.316364	0.945455	0.668537
Maputo City	21.391667	21.391667	0.000000	0.000000

Figura 21. Tabela das Diferenças de Preço entre as Províncias

A tabela resume a variação dos preços médios do milho entre os diferentes mercados dentro das províncias para o ano de 2017.

### Destaques principais:

- **Zambézia e Gaza** têm as maiores disparidades de preço entre mercados locais:
  - **Zambézia:** faixa ampla (6,29 MZN/kg), indicando mercados significativamente mais caros e mais baratos dentro da mesma província.
  - **Gaza:** também mostra uma alta amplitude (6,08 MZN/kg) e o maior desvio-padrão (4,30), indicando muita heterogeneidade entre mercados.
- **Maputo Cidade** é homogénea:
  - Todos os mercados registam o mesmo preço médio, resultando em amplitude zero. Isso indica uma uniformidade perfeita, porém não podemos descartar o fato de que esses são os registros oficiais e, na realidade, os preços costumam sempre variar, ainda que minimamente.
- **Províncias com menor variação:**
  - **Nampula e Niassa** possuem a menor amplitude, sugerindo maior uniformidade dos preços do milho entre seus mercados locais.

## O que isto sugere em termos práticos:

- Províncias com alta amplitude (**Zambézia, Gaza, Sofala, Tete**) indicam diferenças significativas nas condições locais (acessibilidade, oferta local, infraestrutura ou níveis de procura).
- Províncias com baixa amplitude (**Nampula, Niassa**) sugerem uma distribuição mais uniforme dos preços e condições semelhantes entre os mercados locais.

## 4.2. Engenharia de Atributos

Segundo Hyndman e Athanasopoulos (2018), a engenharia de features — ou engenharia de atributos — é essencial num estudo de previsão de preços, pois permite extrair e transformar variáveis brutas em indicadores que capturam sazonalidade, tendências e choques exógenos, melhorando significativamente o sinal disponível para os modelos. Por exemplo, ao criar atributos como grupos temporais, médias móveis ou indicadores de clima e festividades, reduz-se o ruído e potencializa-se a capacidade preditiva dos algoritmos de Machine Learning.

Kuhn e Johnson (2013) reforçam que, sem um conjunto rico de atributos relevantes, mesmo os modelos mais sofisticados tendem a subestimar padrões complexos nos dados. Além disso, Zheng e Casari (2018) demonstram que a seleção e a construção criteriosa de características são determinantes para evitar overfitting e facilitar a interpretabilidade dos resultados, especialmente em séries temporais agrícolas sujeitas a variações climáticas e económicas.

### 4.2.1. Seleção de Dados Pós-2000

Foi inicializada a engenharia com um recorte temporal a partir do ano de 2000, visando garantir consistência e relevância analítica, eliminando períodos históricos com baixa qualidade de dados.

### 4.2.2. Valores Ausentes

Foram observados valores em falta no conjunto de dados para cada província com base no ano e mês. Com isso, foi gerado o índice completo de combinações esperadas de modo a garantir assim um painel balanceado onde cada província tem observações para todos os períodos.

<b>Província</b>	<b>Num. Faltantes</b>
Cabo Delgado	24
Gaza	54
Inhambane	12
Manica	31
Maputo	291
Maputo City	12
Nampula	30
Niassa	74
Sofala	62
Tete	20
Zambézia	84
<b>Total</b>	<b>694</b>

*Figura 22. Tabela da Distribuição de Dados em falta entre as Províncias*

“A construção de um painel balanceado, com todas as combinações esperadas de unidades  $\times$  períodos, simplifica a estimação de efeitos fixos e modelos dinâmicos, evitando problemas de endogeneidade introduzidos por ausências irregulares.” - Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*.

#### **4.2.3. Outras Variáveis com Valores Ausentes**

Para lidar com valores em falta nas variáveis de NDVI, precipitação e IPC, empregaram-se as abordagens de imputação “interpolação linear” e “médias históricas”. A interpolação linear garante continuidade temporal ao estimar pontos intermédios de acordo com as tendências imediatas dos dados (Shumway; Stoffer, 2017), ao passo que a utilização de médias sazonais históricas pode reduzir o viés em séries com forte componente cíclico (Van Buuren, 2018).

#### **4.2.4. Tratamento de Duplicados**

Nos dados, foram encontradas variações de preços dentro das mesmas províncias, meses e anos, devido a variação dos mercados nas mesmas, criando deste modo registros duplicados quando observados por província. De modo a normalizar esses casos, os valores foram consolidados pela média, para manter um ponto mensal único por província. Essa abordagem não só simplificou o quadro de dados como também reduziu o ruído provocado pelas variações locais excessivas sem descaracterizar a tendência global.

Segundo estudos sobre agregação de dados agrícolas, otimizar conjuntos de dados geográficos por média regional é uma prática validada. Por exemplo, quando se consolida séries temporais de preços por áreas com várias estações de medição, a agregação por média representa adequadamente o valor regional sem comprometer a precisão. Lehtonen et al. (2021) demonstraram que “usar uma série média única para representar dados geográficos múltiplos minimiza perdas informacionais e preserva a validade estatística dos dados”.

provincia	ano	mes	NDVI	precipitacao (mm)	IPC	preco
Cabo_Delgado	2000	1	0.612490	28.421009	0.604579	1.831944
Cabo_Delgado	2000	2	0.613449	26.641566	4.448624	1.140000
Cabo_Delgado	2000	3	0.613587	26.384400	2.213949	1.040000
Cabo_Delgado	2000	4	0.613449	26.641566	0.773196	1.140000
Cabo_Delgado	2000	5	0.613518	26.512983	0.797799	1.090000

Figura 23. Tabela dos Dados (Tratada)

ano	mes	NDVI	precipitacao (mm)	IPC	preco	provincia_Gaza	provincia_Inhambane	provincia_Manica	provincia_Maputo	provincia_Maputo City
2000	1	0.612490	28.421009	0.604579	1.831944	0	0	0	0	0
2000	2	0.613449	26.641566	4.448624	1.140000	0	0	0	0	0
2000	3	0.613587	26.384400	2.213949	1.040000	0	0	0	0	0
2000	4	0.613449	26.641566	0.773196	1.140000	0	0	0	0	0
2000	5	0.613518	26.512983	0.797799	1.090000	0	0	0	0	0

Figura 24. Tabela dos Dados (Completa)

#### 4.2.5. Análise de Correlação

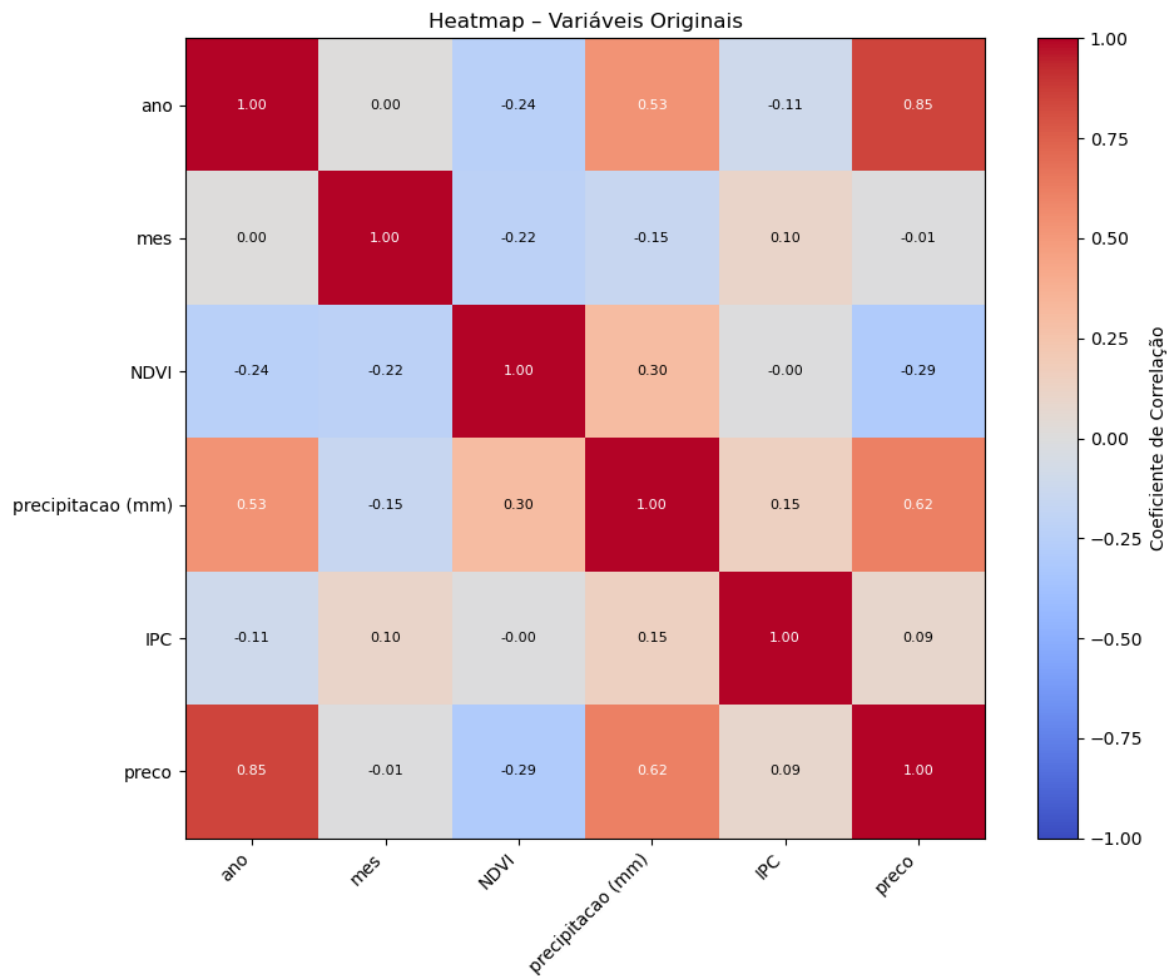


Figura 25. Mapa de Correlação dos Dados

Quando observado o mapa de correlação, percebemos que entre o ano e o preço existe uma tendência de aumento sustentado dos valores ao longo do tempo, enquanto a precipitação mostra relação moderada positiva com o preço, sugerindo que anos mais chuvosos podem elevar os preços, possivelmente pela qualidade das colheitas.

Por outro lado, o NDVI apresenta correlação negativa com o preço, o que indica que maiores índices de vegetação, refletindo oferta mais abundante, pressionam os preços para baixo. Esses padrões reforçam a necessidade de incluir termos de tendência temporal e variáveis climáticas nos modelos de previsão, em linha com boas práticas de análise de séries temporais (Hyndman e Athanasopoulos, 2018).

### **4.3. Configuração dos Experimentos**

Para avaliar a eficácia de diferentes algoritmos de previsão no contexto dos preços do milho em Moçambique, foram realizadas duas séries de experimentos, onde cada conjunto de modelos foi testado e comparado com base em sua capacidade de prever os preços futuros.

#### **4.3.1. Primeira Série de Experimentos – Modelos Univariados**

Na primeira série de experimentos, o objectivo principal foi avaliar o desempenho dos modelos com base apenas nos dados históricos de preços, sem a inclusão de variáveis externas, como inflação ou dados climáticos. O enfoque na previsão univariada permitiu testar até que ponto a série temporal do preço do milho contém padrões consistentes para suportar uma previsão precisa.

#### **Modelos Testados**

##### **LSTM (Long Short-Term Memory)**

O modelo foi configurado com camadas LSTM, dropout e densidade, projetado para capturar padrões temporais. A camada de saída foi configurada para prever um único passo à frente, enquanto a função de custo utilizada foi o RMSE para quantificar a precisão das previsões.

##### **Regressão Multilinear**

A Regressão Multilinear foi incluída como um modelo de base para comparar o desempenho dos algoritmos de Machine Learning. Embora este modelo seja menos complexo, ele forneceu uma linha de base para compreender a relação linear entre valores passados e futuros, e captar tendências gerais sem considerar dependências não-lineares mais profundas.

##### **Random Forest**

Este modelo baseado em árvores de decisão foi configurado com estimadores, níveis de profundidade e `min_samples_leaf` para capturar variações e padrões locais nos dados de forma não-linear. O Random Forest é robusto a ruídos nos dados, o que é útil em séries temporais como as de preços agrícolas, onde oscilações bruscas podem ocorrer devido a factores externos imprevistos.

Esses experimentos iniciais forneceram uma base sólida para a segunda série, onde os modelos passaram a incorporar variáveis adicionais para enriquecer o contexto e a precisão das previsões.

### **4.3.2. Segunda Série de Experimentos – Modelos Multivariados**

Na segunda série de experimentos, o foco foi integrar variáveis adicionais – taxa de inflação, precipitação regional e nível de vegetação – para avaliar se esses factores externos melhorariam a capacidade dos modelos de prever os preços do milho em Moçambique. A inclusão destas variáveis foi motivada pelo entendimento de que os preços agrícolas são influenciados por uma combinação de condições económicas e ambientais. Assim, essa abordagem multivariada permite que os modelos capturem não apenas padrões temporais, mas também efeitos externos que podem amplificar a precisão das previsões.

#### **Modelos Testados**

##### **LSTM (Long Short-Term Memory)**

Para esta série multivariada, o modelo LSTM foi ajustado para receber múltiplas entradas, incluindo todas as variáveis externas. Esta configuração multivariada do LSTM permite que o modelo capture interdependências complexas entre os diferentes factores. O modelo foi configurado com mais camadas de LSTM, e uma camada densa final para produzir uma única previsão de preço à frente, com uma função de custo baseada no RMSE, adequada para séries temporais contínuas. A arquitectura multivariada do LSTM é especialmente útil para identificar relações entre o preço do milho e variáveis como mudanças sazonais na vegetação ou variações na inflação.

##### **Regressão Multilinear**

A Regressão Multilinear foi novamente usada como modelo de comparação, mas desta vez incorporando as variáveis adicionais como entradas. Este modelo identificou a relação linear entre os preços do milho e os factores externos, fornecendo uma linha de base para verificar o valor de adicionar estas variáveis na previsão multivariada. Apesar de não capturar dependências não-lineares complexas, a Regressão Multilinear conseguiu indicar a contribuição básica de cada variável externa.

##### **Random Forest**

O modelo Random Forest foi expandido para incluir as variáveis de inflação, precipitação e vegetação. Este, foi ajustado para explorar as relações não-lineares entre os diferentes factores e o preço do milho, e mostrou ser ideal para detectar padrões complexos e interações entre variáveis. O modelo também

mostrou ser particularmente robusto contra ruídos e mudanças bruscas, permitindo explorar com precisão os impactos de factores voláteis, como a precipitação.

### 4.3.3. Resultados e Avaliação

Os três modelos foram avaliados com as mesmas medidas usadas na primeira série. Esperava-se que o LSTM, com sua capacidade de detectar interdependências, superasse os outros modelos em precisão. O Random Forest poderia ser mais eficaz em capturar relações não-lineares entre variáveis, especialmente em previsões de curto prazo. A Regressão Multilinear, embora mais simples, serviria como um referencial para avaliar a contribuição básica das variáveis adicionais.

Com esta abordagem multivariada, os resultados mostraram que a inclusão de variáveis económicas e ambientais é vantajosa e permitiu capturar flutuações externas que influenciam os preços do milho e melhorou a precisão global das previsões. Esta segunda série de experimentos evidenciou que modelos multivariados fornecem uma visão mais abrangente do sistema, com potencial para prever o comportamento de preços de forma mais robusta e alinhada ao contexto real do mercado agrícola em Moçambique.

#### 4.3.3.1. Primeira Série de Experimentos

Os resultados desta série mostraram que os modelos baseados em aprendizagem profunda, especialmente o **LSTM** (Long Short-Term Memory), tiveram desempenho superior em relação aos modelos tradicionais, como a Regressão Multilinear e o Random Forest. Isso deve-se principalmente à capacidade do LSTM de capturar padrões temporais de longo prazo, uma vantagem crítica quando se lida com séries temporais que podem ter dependências sazonais ou de longo alcance.

Algoritmo	RMSE	R <sup>2</sup>
LSTM	1,871	0,449
Random Forest	7,094	0,401
Regressão Multilinear	9,125	0,341

*Figura 26. Tabela de Avaliação do Primeiro Experimento*

**LSTM (Long Short-Term Memory):** O modelo LSTM apresentou o melhor desempenho entre os algoritmos testados, com um RMSE de 1,871. A estrutura em camadas do LSTM, permitiu ao modelo

capturar não apenas flutuações imediatas, mas também dependências prolongadas no histórico de preços. Assim, conseguimos perceber que a habilidade do LSTM de preservar informações ao longo do tempo foi especialmente eficaz em identificar ciclos sazonais e padrões que os modelos lineares e baseados em árvores de decisão não conseguiram captar.

**Random Forest:** Embora o Random Forest tenha demonstrado bom desempenho ao capturar variações não-lineares nos dados, ele não obteve resultados tão precisos quanto o LSTM. O modelo apresentou um RMSE maior do que o LSTM, refletindo sua limitação em captar dependências temporais de longo prazo. No entanto, o Random Forest mostrou-se robusto em períodos de volatilidade acentuada nos dados, uma característica valiosa, pois o modelo não é tão sensível a outliers quanto os modelos de aprendizagem profunda.

**Regressão Multilinear:** Como modelo de referência, a Regressão Multilinear destaca-se por sua simplicidade e velocidade de execução. No entanto, devido à sua natureza linear, o modelo não conseguiu capturar a complexidade das mudanças nos preços ao longo do tempo, refletindo limitações na previsão de padrões não-lineares e sazonalidades presentes na série temporal.

## **Conclusão da Primeira Série de Experimentos**

Com os resultados da primeira série, tornou-se evidente que o LSTM é particularmente adequado para a previsão de séries temporais em contextos univariados. A sua capacidade de capturar dependências de longo prazo torna-o um modelo robusto para previsões de preços, onde flutuações sazonais e tendências a longo prazo são críticas. Através desses experimentos, ficou claro que o LSTM pode servir como uma base sólida para cenários mais complexos, como na segunda série de experimentos, onde variáveis adicionais serão incorporadas para testar modelos multivariados. Esta abordagem permitirá verificar se a inclusão de factores externos poderá ainda mais melhorar a precisão, adicionando contexto às previsões de preços.

### **4.3.3.2. Segunda Série de Experimentos**

Na segunda série de experimentos, foi realizada uma análise multivariada incluindo variáveis adicionais, como taxa de inflação, precipitação e vegetação regional, para verificar se factores económicos e climáticos contribuiriam para uma previsão mais precisa dos preços do milho. Essa abordagem

multivariada proporcionou uma visão mais abrangente dos elementos que influenciam as variações de preço, aproveitando a interação entre esses factores externos e os dados históricos de preço.

A inclusão dessas variáveis resultou em uma melhora significativa na precisão das previsões para todos os modelos, com destaque especial para o desempenho aprimorado do LSTM:

<b>Algoritmo</b>	<b>RMSE</b>	<b>R<sup>2</sup></b>
LSTM	0,348	0,874
Random Forest	5,578	0,662
Regressão Multilinear	4,972	0,731

Figura 27. Tabela de Avaliação do Segundo Experimento

**LSTM (Long Short-Term Memory):** Com a adição de variáveis económicas e climáticas, o desempenho do LSTM foi significativamente aprimorado. Para alimentar o modelo, foram normalizadas todas as variáveis com recurso ao *StandardScaler* e criadas janelas deslizantes de 24 meses — ou seja, foram usados os dados dos 24 meses anteriores para prever o preço do mês seguinte, analisando cada província de forma independente.

O processo de treino foi realizado em até 4 épocas, com a técnica de *EarlyStopping* a monitorizar a perda na validação (*val\_loss*) e com uma tolerância de 2 épocas sem melhorias.

No conjunto de teste, o modelo atingiu um RMSE de **0,348** e um coeficiente de determinação ( $R^2$ ) de **0,874**. Estes resultados indicam que o modelo consegue explicar cerca de **87%** da variação nos preços do milho, mantendo um erro médio de previsão relativamente baixo — o que demonstra uma boa capacidade de ajustamento aos dados.

**Random Forest:** A inclusão das variáveis adicionais também melhorou o desempenho do Random Forest, embora o modelo ainda ficasse atrás do LSTM em precisão. Com o aumento de variáveis, o Random Forest conseguiu capturar algumas interações não-lineares que influenciam os preços do milho, especialmente em períodos de volatilidade climática e económica. No entanto, a ausência de uma

memória interna limita sua capacidade de lidar com dependências temporais profundas, tornando-o mais apropriado para previsões de curto prazo.

**Regressão Multilinear:** A Regressão Multilinear também apresentou uma melhora no RMSE ao incorporar as novas variáveis, evidenciando a contribuição das variáveis externas, ainda que de forma linear. O modelo continuou a ter limitações em comparação com o LSTM e o Random Forest, uma vez que não consegue capturar interações complexas entre variáveis nem dependências não-lineares. No entanto, foi útil como uma linha de base para avaliar a importância individual de cada variável adicional.

## Conclusão da Segunda Série de Experimentos

Os resultados da segunda série de experimentos mostraram que a inclusão de variáveis externas, como condições económicas e climáticas, é fundamental para melhorar a precisão das previsões. O LSTM, em especial, demonstrou um ganho expressivo de desempenho com dados multivariados mostrando que é adequado para cenários onde o preço do milho é influenciado por uma diversidade de factores interligados.

Ainda, essa série evidenciou que, ao adicionar contexto ao modelo, é possível obter previsões mais robustas e alinhadas com as dinâmicas reais do mercado.

## 4.4. Discussão dos Resultados

Os experimentos realizados destacaram que modelos baseados em redes neurais profundas, especialmente o **LSTM** (Long Short-Term Memory), são altamente eficazes para prever o preço do milho em Moçambique, principalmente devido à sua capacidade de lidar com dados temporais complexos e multivariados. A análise dos resultados reforça a importância de duas abordagens principais: a utilização de modelos de aprendizagem profunda para captar padrões de longo prazo em séries temporais e a inclusão de variáveis adicionais, como condições climáticas e económicas, que influenciam o comportamento dos preços.

### 4.4.1. Impacto das Variáveis Externas

A inclusão de variáveis externas – como taxa de inflação, nível de precipitação e vegetação – mostrou-se decisiva para aumentar a precisão das previsões. Essas variáveis proporcionaram um contexto mais

amplo para o comportamento dos preços do milho, já que estes são influenciados não apenas por tendências de demanda e oferta, mas também por factores climáticos e custos económicos.

A influência das variáveis climáticas, por exemplo, é particularmente importante em regiões agrícolas, onde a produção pode ser drasticamente afectada por condições como secas ou chuvas excessivas. A taxa de inflação também exerce uma influência significativa ao afectar o poder de compra dos consumidores e o custo dos insumos agrícolas. Esses factores são essenciais para um modelo que visa prever preços em um ambiente economicamente dinâmico e climaticamente volátil, como o de Moçambique.

#### **4.4.2. Comparação com Modelos Tradicionais**

Em comparação com a **Regressão Multilinear** e o **Random Forest**, o **LSTM** mostrou-se mais robusto na captura de padrões temporais e interações complexas entre variáveis. O Random Forest conseguiu captar padrões não-lineares, especialmente em períodos de maior volatilidade de preços, mas demonstrou limitações devido à falta de memória interna, essencial para previsões de longo prazo. A Regressão Multilinear, embora útil como uma linha de base, não conseguiu captar adequadamente as complexidades das interações entre os factores externos e o preço do milho, reforçando a necessidade de modelos mais sofisticados para previsões em ambientes agrícolas complexos.

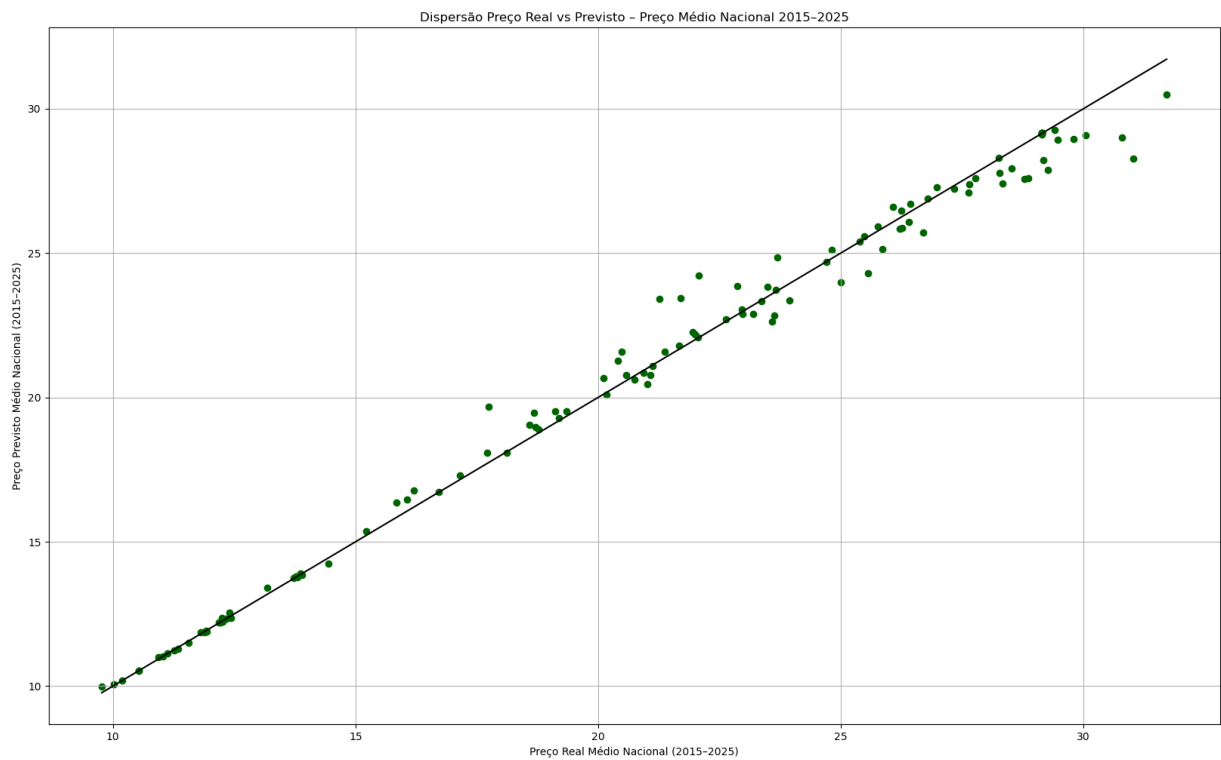
Conforme Zhang et al. (2024), no estudo “Short term forecasting of vegetable prices based on LSTM model — Evidence from Beijing’s vegetable data”, ao prever preços agrícolas, os modelos LSTM superaram claramente métodos tradicionais como modelos autoregressivos e integrados de médias móveis similares e regressão.

#### **4.4.3. Conclusão dos Experimentos**

Os resultados indicam que a utilização de redes neurais profundas, como o LSTM, combinado com uma abordagem multivariada que inclui variáveis externas, é a estratégia mais eficaz para previsão de preços do milho em Moçambique. Este modelo não apenas fornece uma previsão mais precisa, mas também permite uma interpretação mais abrangente dos factores que afectam o mercado de milho. A inclusão de dados económicos e climáticos reflete um contexto mais realista e robusto, melhorando a resiliência e a aplicabilidade das previsões em um ambiente onde eventos externos podem ter impactos consideráveis no preço final. Esses insights podem, portanto, servir de base para estratégias mais informadas e eficazes, beneficiando os agricultores, decisores políticos e outros actores envolvidos no sector agrícola moçambicano.

#### 4.4.4. Gráfico de Dispersão

Após o desenvolvimento do modelo, foi observado, por meio do gráfico de dispersão, uma forte correlação entre os preços reais e os preços previstos do milho para os últimos 10 anos. Os pontos concentram-se ao longo da linha de identidade, o que indica que as previsões estão próximas dos valores observados. Isto sugere que o modelo possui uma grande capacidade preditiva e consegue explicar de forma satisfatória a variação do preço médio nacional do milho, com erros sistemáticos reduzidos.



*Figura 28. Gráfico de Dispersão do Preço Real e Previsto*

#### 4.4.5. Preços Previstos vs Preços Reais

Por fim, entre as várias validações realizadas, destacamos uma comparação com uma amostra de dados históricos do preço médio da província de Inhambane para o período de outubro de 2023 e março de 2024 (6 meses) com resultados positivos e uma média de 1,6 MZN de diferença entre os preços.

date	admin1	commodity	unit	price	usdprice	Preço Médio Real	Preço Médio Previsto	Diferença
15/10/2023	Inhambane	Maize (white)	KG	22,86	0,36			
15/10/2023	Inhambane	Maize (white)	KG	27,14	0,43			
15/10/2023	Inhambane	Maize (white)	KG	20,00	0,32			
15/10/2023	Inhambane	Maize (white)	KG	24,00	0,38			
15/10/2023	Inhambane	Maize (white)	KG	27,43	0,43	24,29	28,36	4,07
15/11/2023	Inhambane	Maize (white)	KG	28,57	0,45			
15/11/2023	Inhambane	Maize (white)	KG	25,71	0,41	27,14	27,92	0,78
15/12/2023	Inhambane	Maize (white)	KG	25,71	0,41	25,71	27,6	1,89
15/01/2024	Inhambane	Maize (white)	KG	27,43	0,43	27,43	27,31	0,12
15/02/2024	Inhambane	Maize (white)	KG	22,86	0,36			
15/02/2024	Inhambane	Maize (white)	KG	34,52	0,55	28,69	27,03	1,66
15/03/2024	Inhambane	Maize (white)	KG	28,00	0,44	28,00	26,55	1,45

Figura 29. Tabela com os Dados Originais (Visto do Excel)

provincia	Gaza	Inhambane	Manica	Maputo	Maputo City	Nampula	Niassa	Sofala	Tete	Zambezia
Mês										
2023-10-01	27.16	28.36	27.23	26.59	25.15	25.37	25.34	26.27	25.26	25.79
2023-11-01	26.66	27.92	26.44	25.82	24.01	24.56	24.21	25.92	24.05	25.12
2023-12-01	26.74	27.60	25.98	25.60	23.32	24.40	23.64	26.55	23.40	24.62
2024-01-01	25.91	27.31	25.49	24.95	22.34	23.71	22.78	25.82	22.56	24.18
2024-02-01	25.27	27.03	25.14	24.24	21.54	23.05	21.93	25.14	21.76	23.78
2024-03-01	24.38	26.55	24.43	23.13	20.47	22.27	20.66	24.01	20.52	23.16

Figura 30. Tabela com os Preços Previstos (Visto do Anaconda)

#### 4.5. Implementação da Plataforma Web

O sistema desenvolvido como resultado do modelo, consiste numa plataforma web interativa de monitoria e previsão de preços do milho em Moçambique. O seu desenho foi pensado para ser acessível e intuitivo, permitindo que diferentes perfis de utilizadores, incluindo agricultores e famílias, consigam beneficiar das funcionalidades sem necessidade de conhecimentos técnicos avançados.



Figura 31. UI do MVP

#### 4.5.1. Acessibilidade e Público-Alvo

A plataforma foi construída com foco em **simplicidade de uso**. A navegação é clara e direta, com menus laterais que organizam a informação por província, facilitando a consulta rápida de dados específicos. Além disso, a interface gráfica utiliza cores, ícones e representações visuais amigáveis, garantindo que até utilizadores com baixa literacia digital consigam interpretar os resultados. Assim, o sistema posiciona-se como uma ferramenta inclusiva para agricultores, comerciantes locais e decisores comunitários.

#### 4.5.2. Funcionalidades Principais

- **Previsão de Preços do Milho:** a funcionalidade central do sistema é a geração de previsões de preços por província, apresentadas através de gráficos que distinguem claramente o período histórico do período previsto. Isso permite que os agricultores planeiem o melhor momento para comprar ou vender a sua produção.
- **Indicadores Agroclimáticos:** a plataforma disponibiliza variáveis de apoio, como índice de vegetação, precipitação e índice de consumo, que ajudam a interpretar as oscilações de preços e antecipar tendências de mercado.
- **Assistente Virtual:** integrado na plataforma, o assistente virtual responde a perguntas em linguagem natural, orientando o utilizador de forma simples. Por exemplo, é possível perguntar “Qual é o

melhor mês para comprar milho em Manica em 2025?” e receber uma resposta direta e compreensível.

- **Dúvidas Frequentes:** a secção de ajuda explica, em linguagem acessível, o funcionamento da plataforma, a origem dos dados e como interpretar os gráficos, promovendo a autonomia dos utilizadores.

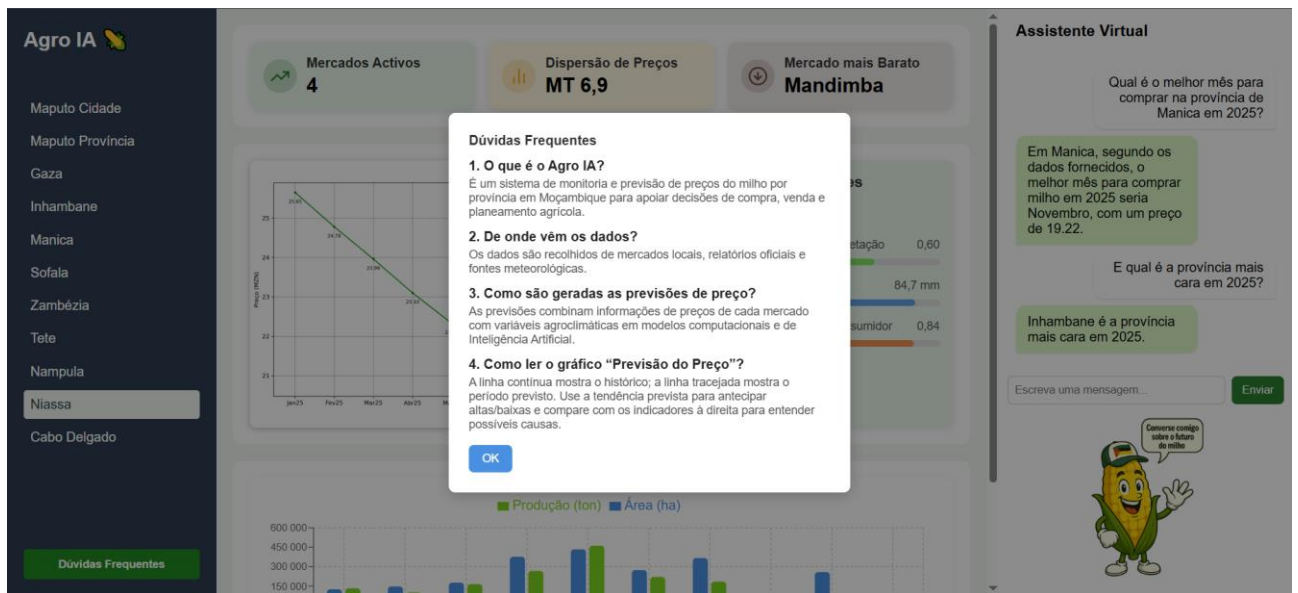


Figura 32. UI do MVP

#### 4.5.3. Usabilidade e Experiência do Utilizador

O Agro IA aposta numa **experiência de utilização agradável e educativa**. O sistema não apenas fornece previsões, mas também explica as informações exibidas, tornando-se uma ferramenta de apoio à tomada de decisão. O uso de elementos visuais — gráficos de preços, barras de produção e áreas de cultivo — ajuda na compreensão rápida da informação, mesmo para quem não está habituado a relatórios técnicos.

#### 4.5.4. Impacto e Benefícios

O impacto esperado da plataforma é significativo:

- **Para os agricultores:** maior capacidade de planear o momento ideal de venda e compra, reduzindo riscos financeiros.

- **Para as famílias de baixo rendimento:** acesso a informação que pode ajudar a otimizar os gastos alimentares.
- **Para a economia agrícola:** promoção de maior transparência nos mercados e apoio a políticas públicas baseadas em evidências.

#### 4.5.5. Imagem e Potencial Futuro

Combinando **Inteligência Artificial, dados agroclimáticos e acessibilidade digital**, o sistema representa uma inovação no sector agrícola moçambicano. A plataforma se posiciona como uma ponte entre a tecnologia e a vida rural de modo a reforçar a visão de que a transformação digital pode servir directamente comunidades agrícolas.

## 5. Conclusão

O presente trabalho apresentou uma solução inovadora para um desafio crítico no sector agrícola moçambicano: **prever com precisão o preço do milho**, um alimento básico essencial para a segurança alimentar e económica do país. Através da utilização de modelos avançados de Machine Learning, especificamente o LSTM, foi possível demonstrar que essas tecnologias oferecem uma abordagem robusta e eficiente para enfrentar as complexidades das séries temporais multivariadas. O modelo LSTM destacou-se por sua capacidade de capturar padrões temporais de longo prazo e integrar factores externos que afectam o preço do milho, como variáveis económicas e climáticas.

Os resultados obtidos reforçam a relevância da utilização de Machine Learning para lidar com as flutuações de preços em mercados agrícolas voláteis. O desempenho superior do modelo LSTM, particularmente em cenários multivariados, evidenciou que a inclusão de factores contextuais, como a taxa de inflação, nível de precipitação e indicadores de vegetação regional, enriquece significativamente a qualidade das previsões. Além disso, o modelo demonstrou ser resiliente a mudanças no comportamento do mercado, um atributo essencial num ambiente onde condições económicas e climáticas frequentemente se alteram.

Quanto a integração do modelo preditivo numa plataforma web, esta permite que agricultores, comerciantes, decisores políticos e outros intervenientes tenham acesso a previsões detalhadas e personalizáveis, possibilitando uma tomada de decisão mais informada e estratégica. Por exemplo, agricultores podem planejar suas colheitas com maior precisão, comerciantes podem ajustar suas estratégias de compra e venda, e formuladores de políticas podem levar a cabo intervenções direccionadas para mitigar os impactos negativos das flutuações de preços.

## 5.1. Contribuições e Implicações Práticas

O presente estudo sublinha a relevância de integrar a ciência de dados com o conhecimento contextual local, evidenciando como a inclusão de dados específicos de Moçambique permite adaptar modelos preditivos às particularidades do mercado nacional. A aplicação de redes neurais profundas, como o LSTM, no contexto moçambicano, preenche uma lacuna na literatura existente, oferecendo insights valiosos para futuros estudos e aplicações.



*Figura 33. Milho e seus Grãos (fonte: unsplash)*

Ao fornecer previsões mais precisas e confiáveis, o sistema auxilia na mitigação dos efeitos da volatilidade dos preços, promovendo maior estabilidade económica e segurança alimentar (World Bank, 2022).

## 5.2. Recomendações e Resultados Esperados

O sistema de previsão desenvolvido demonstra um potencial significativo para transformar o sector agrícola de Moçambique. A integração de tecnologias avançadas de previsão pode servir como catalisador para a modernização do sector, proporcionando informações valiosas para toda a cadeia de valor do milho – desde agricultores até consumidores finais (FAO, 2020). Este trabalho exemplifica como a aplicação de ciência de dados pode responder a desafios sociais e económicos complexos, criando impacto positivo e duradouro.

### **5.2.1. Estratégias de Disseminação e Adopção no Sector Agrícola**

Para maximizar os benefícios do modelo preditivo baseado no LSTM, é imperativo investir em estratégias de disseminação que assegurem a sua adopção ampla e sustentável por parte dos principais intervenientes do sector agrícola moçambicano. A integração de ferramentas tecnológicas com programas de extensão agrícola e treinos comunitários é fundamental para garantir que agricultores e comerciantes compreendam e confiem nas previsões geradas (Thornton et al., 2018).

Parcerias com instituições locais, como associações de agricultores e organizações governamentais, podem facilitar a implementação de políticas que incentivem o uso da plataforma web preditiva, promovendo a segurança alimentar e a estabilidade económica. Esta abordagem colaborativa ajuda a mitigar barreiras de acesso tecnológico e favorece a criação de um ecossistema integrado de partilha de dados, onde os utilizadores se tornam simultaneamente beneficiários e contribuidores do sistema (Chivale, 2021).

### **5.2.2. Trabalhos Futuros**

Que este trabalho sirva de referência a trabalhos futuros de modo a desenvolver novos modelos ou expandir o sistema de previsão desenvolvido para abranger outros produtos agrícolas estratégicos em Moçambique, tais como arroz, feijão, sorgo e mandioca, que também desempenham um papel crucial na segurança alimentar nacional devido às flutuações significativas nos seus preços. Segundo Chivale (2021) e FAO (2020), diversificar o portefólio de culturas analisadas é essencial para fornecer informações robustas e específicas, capazes de orientar melhor decisões económicas e políticas agrícolas no país.

Estudos recentes, como os apresentados por Thornton et al. (2018), têm demonstrado eficácia em prever a procura e os preços de culturas como arroz e hortícolas, através de modelos clássicos como o ARIMA e outras técnicas de séries temporais. Trabalhos como estes, reforçam a relevância de estender a abordagem baseada em redes neurais profundas, já validada neste estudo, para outros produtos agrícolas. Desta forma, será possível fortalecer ainda mais o impacto económico e social das ferramentas preditivas implementadas, contribuindo para uma maior estabilidade nos mercados agrícolas moçambicanos.

## Referências Bibliográficas

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., & Zheng, X. (2016). **TensorFlow: A System for Large-Scale Machine Learning**.

Agropecuáriamoz. (2016). **A agricultura e pecuária em Moçambique**.  
<http://agropecuáriamoz.blogspot.com/2016/09/agricultura-epecuaria-mocambique.html>

Alpaydin, E. (2020). **Introduction to Machine Learning**. MIT Press.

Asamoah, R., Agyeman, E. A., & Opoku, F. (2024). **Impact of Agricultural Extension Services on the Adoption of Improved Agricultural Practices in Rural Communities**. Wageningen University & Research. <https://www.wur.nl/en/publication-details.htm?publicationId=e7159e79-abde-451d-95d7-caa06403c922>

Bill, P., Smith, R., & Johnson, T. (2020). **Machine Learning in Agriculture: Strategies and Impacts**. Springer.

Bishop, C. M. (2006). **Pattern Recognition and Machine Learning**. Springer.

Bojarski, M., Testa, D., Dworakowski, D., et al. (2016). **End-to-End Deep Learning for Self-Driving Cars**. *arXiv:1604.07316*.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). **Time Series Analysis: Forecasting and Control** (5th ed.). Wiley.

Breiman, L. (2001). **Random forests**. *Machine Learning*, 45(1), 5–32.  
<https://doi.org/10.1023/A:1010933404324>

Bryman, A. (2012). **Social Research Methods**. Oxford University Press.

Chai, T., & Draxler, R. R. (2014). **Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE**. *Geoscientific Model Development*, 7(3), 1247–1250.  
<https://doi.org/10.5194/gmd-7-1247-2014>

Chen, T., & Guestrin, C. (2016). **XGBoost: A Scalable Tree Boosting System**. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Chivale, A. (2021). **Transformação digital no sector agrícola em Moçambique: perspectivas e desafios**. *Revista de Agricultura e Desenvolvimento Rural*, 34(2), 45–60.

Creswell, J. W. (2014). **Research Design: Qualitative, Quantitative, and Mixed Methods Approaches**. SAGE Publications.

DHL. (2018). **Artificial Intelligence in Logistics**. DHL Trend Research.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). **Dermatologist-level classification of skin cancer with deep neural networks**. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>

- FAO. (2009, Outubro 13). **How to Feed the World in 2050**. Retrieved Abril 29, 2019, from [http://www.fao.org/fileadmin/templates/wsfs/docs/expert\\_paper/How to Feed the World in 2050.pdf](http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf)
- FAO. (2020). **Digital Agriculture: Harnessing Technology for Sustainable Development**. Food and Agriculture Organization.
- FAO. (2020). **Food Outlook – Biannual Report on Global Food Markets**. Organização das Nações Unidas para a Alimentação e Agricultura.
- FAO. (s.d.). **Food Balance Sheets / FAOSTAT**. Organização das Nações Unidas para a Alimentação e Agricultura.
- Gomez-Uribe, C. A., & Hunt, N. (2015). **The Netflix recommender system: Algorithms, business value, and innovation**. *ACM Transactions on Management Information Systems*, 6(4), Article 13.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). **Deep Learning**. MIT Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). **The Elements of Statistical Learning: Data Mining, Inference, and Prediction** (2nd ed.). Springer.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., et al. (2012). **Deep neural networks for acoustic modeling in speech recognition**. *IEEE Signal Processing Magazine*, 29(6), 82–97.
- Hoffman, B., Gil, J., Loboguerrero, A. M., Asseng, S., Hertel, T., & Ewert, F. (2021). **Enhancing Climate-Smart Agriculture: A review of how economics and policy research addresses sustainability**. *Frontiers in Sustainable Food Systems*, 5, 836437. <https://doi.org/10.3389/fsufs.2022.836437>
- Hyndman, R. J., & Athanasopoulos, G. (2018). **Forecasting: Principles and Practice** (2nd ed.). OTexts.
- Jordan, M. I., & Mitchell, T. M. (2015). **Machine learning: Trends, perspectives, and prospects**. *Science*, 349(6245), 255–260.
- Kuhn, M., & Johnson, K. (2013). **Applied Predictive Modeling**. Springer.
- Lakatos, E. M., & Marconi, M. A. (2003). **Fundamentos de Metodologia Científica** (5ª ed.). Atlas.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). **Deep Learning**. *Nature*, 521(7553), 436–444.
- Liaw, A., & Wiener, M. (2002). **Classification and regression by randomForest**. *R News*, 2(3), 18–22.
- MADER/SIMA. (vários anos). **Boletins de Preços de Mercados Agrícolas**. Ministério da Agricultura e Desenvolvimento Rural, Moçambique.
- Ministério da Agricultura. (2011). **Plano Estratégico de Desenvolvimento do Sector Agrário (PEDSA) 2011–2020**. Governo de Moçambique.

- Ministério da Agricultura. (2013). **Plano Nacional de Investimento do Sector Agrário (PNISA) 2013–2017**. Governo de Moçambique.
- Mitchell, T. M. (1997). **Machine Learning**. McGraw-Hill Education.
- Mundial, B. (2018, Outubro 17). **Banco Mundial: quase metade da população global vive abaixo da linha da pobreza**. Retrieved April 17, 2019, from <https://oglobo.globo.com/economia/onu-divulga-dados-mundiais-de-pobrezasem-informacoes-sobre-brasil-23085580>
- Murphy, K. P. (2012). **Machine Learning: A Probabilistic Perspective**. MIT Press.
- Obermeyer, Z., & Emanuel, E. J. (2016). **Predicting the Future — Big Data, Machine Learning, and Clinical Medicine**. *The New England Journal of Medicine*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). **Scikit-learn: Machine Learning in Python**. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pukrongta, P., Thanapirom, P., & Wongpinyochit, T. (2024). **Enhancing Agricultural Production through Precision Farming: A review of advanced technologies in sustainable agriculture**. *Applied Sciences*, 14(8), 3313. <https://doi.org/10.3390/app14083313>
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., et al. (2017). **CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning**. *arXiv:1711.05225*.
- Schulman, J., et al. (2017). **Proximal Policy Optimization Algorithms**.
- Shumway, R. H., & Stoffer, D. S. (2017). **Time Series Analysis and Its Applications** (4th ed.). Springer.
- Siemens. (2019). **Predictive Maintenance Solutions**. Siemens Industrial Automation.
- Sutton, R. S., & Barto, A. G. (2018). **Reinforcement Learning: An Introduction**. MIT Press.
- Toolify. (2023). **Ferramentas de Automação para a Agricultura Sustentável**. Toolify Technologies.
- Van Buuren, S. (2018). **Flexible Imputation of Missing Data** (2nd ed.). CRC Press.
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). **ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases**. *arXiv:1705.02315*.
- World Bank. (2022). **Agricultural Innovations in Sub-Saharan Africa**. World Bank Group.
- Wooldridge, J. M. (2010). **Econometric Analysis of Cross Section and Panel Data** (2nd ed.). MIT Press.
- Zheng, A., & Casari, A. (2018). **Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists**. O'Reilly.