



FACULDADE DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA
LICENCIATURA EM ENGENHARIA INFORMÁTICA

Desenvolvimento de um Modelo de Previsão da Seca com Recurso à Inteligência Artificial

Caso de Estudo: Instituto Nacional de Meteorologia (INAM)

Autora:

MUCAVELE, Ludmila Emília Lourenço

Supervisor:

Mestre Rúben Manhiça, Eng.º

Maputo, Novembro de 2024



FACULDADE DE ENGENHARIA

DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA

LICENCIATURA EM ENGENHARIA INFORMÁTICA

Desenvolvimento de um Modelo de Previsão da Seca com Recurso à Inteligência Artificial

Caso de Estudo: Instituto Nacional de Meteorologia (INAM)

Autor:

MUCAVELE, Ludmila Emília Lourenço

Supervisor:

Mestre Rúben Manhiça, Eng.º

Maputo, Novembro de 2024



FACULDADE DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA

TERMO DE ENTREGA DO RELATÓRIO DE TRABALHO DE LICENCIATURA

Declaro que a estudante Ludmila Emília Lourenço Mucavele entregou no dia _____, as _____ cópias do relatório do seu Trabalho de Licenciatura com referência _____, intitulado: Desenvolvimento de um Modelo Previsão da Seca com Recurso à Inteligência Artificial.

Maputo, _____ de _____ de 2024

A Chefe da Secretaria



FACULDADE DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA

DECLARAÇÃO DE HONRA

Declaro sob compromisso de honra que o presente trabalho é resultado da minha investigação e que foi concebido para ser submetido apenas para a obtenção do grau de Licenciatura em Engenharia Informática na Faculdade de Engenharia da Universidade Eduardo Mondlane.

Maputo, ____ de _____ de 2024

A Autora

(Ludmila Emília Lourenço Mucavele)

Dedicatória

À minha Mãe, Emília Timane

Ao meu pai, Agnelo Jorge

Aos meus irmãos, Gerson, Shemy e Ákilah

Aos meus avós, David Timane e Anastácia Macuacua

Aos demais familiares e amigos.

Agradecimentos

Primeiramente, agradeço a Allah (Deus), o Todo Poderoso, por este feito e por me permitir chegar até esta fase importante da minha vida e carreira.

Agradeço à minha mãe, que sempre me guiou e auxiliou-me sempre e em qualquer situação.

Agradeço ao meu pai, por sua paciência e gentileza ao cuidar de mim.

Agradeço aos meus irmãos, pela paciência e por me manterem sempre motivada na luta para o alcance dos meus objectivos.

Agradeço à minha prima (Karen Timane) e às minhas amigas (Umaima Aly, Nayra Aly e Ana Tiffany Machovo, Amarilda Chihepe), por sempre mostrarem-se disponíveis para me ajudar em tudo.

Agradeço também aos meus docentes, que pacientemente partilharam os seus conhecimentos comigo e com todos os meus colegas da faculdade no geral e da turma em particular.

Um especial agradecimento aos Mestres Eng^o Rúben Manhiça, dr. Vali Issufo, Eng^a Leila Omar e Eng^o Felizardo Munguambe, pela atenção extra dispensada e pelos valiosos conselhos que ajudaram-me muito a chegar à esta fase.

Epígrafe

“Acertamos naquilo para que miramos e, se não mirarmos para nada, acertamos sempre em nada.”

Zig Ziglar

Resumo

O presente trabalho de pesquisa aborda a problemática das secas meteorológicas em Moçambique, focando-se na aplicação de técnicas modernas de previsão de secas e na utilização do Índice de Precipitação Padronizada (SPI) como metodologia de análise. A seca é um fenómeno que afecta directamente a vida de todos os seres vivos, limitando a disponibilidade de água e alimentos, sendo especialmente crítica nas zonas sul e centro do país, onde a limitada capacidade do monitoramento climático adequado agrava os impactos nas comunidades e nos ecossistemas.

Esta pesquisa tem enfoque nas dificuldades enfrentadas pelo Instituto Nacional de Meteorologia (INAM) em colectar dados climáticos amplos, devido a limitações infraestruturais, e propõe a integração de tecnologias de *Machine Learning* para melhor precisão das previsões.

Para testar esta abordagem, foi desenvolvido um modelo de previsão LSTM (Long Short-Term Memory) utilizando o SPI para intervalos de 1, 3 e 6 meses, com o objectivo de fornecer uma visão clara da severidade e duração das secas.

A implementação prática dessas soluções enfrentou certos desafios, tais como a necessidade de grandes volumes de dados históricos e a optimização de modelos para garantir menor margem de erro nos resultados. Contudo, a pesquisa oferece uma base sólida para a modernização das práticas de previsão climática em Moçambique.

Palavras-chave: Seca Meteorológica, SPI, Machine Learning, LSTM, Previsão de Secas.

Abstract

This research paper addresses the problem of meteorological droughts in Mozambique, focusing on the application of modern drought forecasting techniques and the use of the Standardized Precipitation Index (SPI) as an analysis methodology. Drought is a phenomenon that directly affects the lives of all living beings, limiting the availability of water and food, and is especially critical in the southern and central areas of the country, where the limited capacity for adequate climate monitoring exacerbates the impacts on communities and ecosystems.

This research focuses on the difficulties faced by the National Meteorological Institute (INAM) in collecting comprehensive climate data, due to infrastructural limitations, and proposes the integration of Machine Learning technologies to improve forecast accuracy.

To test this approach, an LSTM (Long Short-Term Memory) forecast model was developed using the SPI for intervals of 1, 3 and 6 months, with the aim of providing a clear view of the severity and duration of droughts.

The practical implementation of these solutions faced certain challenges, such as the need for large volumes of historical data and the optimization of models to ensure a smaller margin of error in the results. However, the research provides a solid basis for modernizing climate forecasting practices in Mozambique.

Keywords: Meteorological Drought, SPI, Machine Learning, LSTM, Drought Forecasting.

Índice

CAPÍTULO I.....	1
1. Introdução.....	1
1.1. Contextualização.....	1
1.2. Definição do Problema.....	2
1.3. Objectivos.....	3
1.3.1. Objectivo Geral	3
1.3.2. Objectivos específicos	3
1.4. Metodologia	4
1.4.1. Etapas da Pesquisa	4
1.4.2. Classificação da Metodologia.....	5
1.4.3. Técnicas de Colecta de Dados	7
1.4.4. Metodologia de Desenvolvimento do Trabalho Prático.....	8
1.4.5. Paradigma de Programação.....	8
1.4.6. Ambiente de Treinamento do Modelo de Aprendizagem.....	8
1.4.7. Linguagens de Programação.....	9
1.4.8. Testagem de Resultados	9
1.4.9. Validação de Resultados.....	9
CAPÍTULO II.....	10
2. Revisão de Literatura.....	10
2.1. A Seca Como Fenómeno Natural	10
2.1.1. Tipos de Secas	10
2.1.2. Indicadores de Detecção das Secas.....	11
2.1.3. A Precipitação	12
2.1.4. Standardized Precipitation Index - SPI	12
2.2. Impacto das Secas	14
2.3. Moçambique e a Vulnerabilidade à Seca	15
2.3.1. Desafios na Região Sul de Moçambique em Termos de Mitigação à Seca	16
2.4. Previsão de Secas: Tecnologias e Ferramentas	17
2.4.1. Tecnologias de Previsão de Secas em Moçambique.....	18
2.5. Inteligência Artificial e Aprendizagem de Máquina.....	19
2.5.1. O Modelo Genérico de Aprendizagem de Máquina.....	20
2.5.2. Redes Neurais Artificiais	23

2.5.3. Abordagens de Aplicação da Aprendizagem de Máquina na Previsão de Secas	24
2.6. Desafios na Implementação de ML	27
CAPÍTULO III	29
3. Caso de Estudo (INAM)	29
3.1. Local de Estudo	29
3.1.1. Resultados das Entrevistas	29
3.1.2. Problemas Enfrentados no Local de Estudo	33
3.1.3. Comparação das Abordagens de IA para os Problemas Identificados	33
3.2. Proposta de Solução	36
3.2.1. Modelo de IA para Previsão do SPI	37
3.2.2. Sistema para Previsão de Secas	37
CAPÍTULO IV	39
4. Construção dos Modelos de Previsão de secas	39
4.1. Definição do Objectivo da Pesquisa	39
4.2. Colecta de Dados	39
4.3. Pré-Processamento dos Dados	40
4.4. Arquitectura dos Modelos de Inteligência Artificial	41
4.5. Treinamento e Teste dos Modelos	42
4.5.1. Resultados da Fase de Treinamento dos Modelos	43
4.5.2. Resultados da Fase de Teste dos Modelos	44
4.6. Iteracção Contínua com o Modelo	45
CAPÍTULO V	47
5. Desenvolvimento de Protótipo da Solução Proposta	47
5.1. Modelagem do Sistema	47
5.1.1. Requisitos de Negócio	48
5.1.2. Requisitos de Usuários	48
5.1.3. Requisitos Funcionais e Requisitos Não Funcionais	49
5.2. Restrições de Desenvolvimento	50
5.3. Arquitectura do Sistema	51
5.4. Mapa de Navegação do Sistema	51
5.4.1. Testes Funcionais Realizados no Sistema	52
5.5. Implantação do Sistema	54
CAPÍTULO VI	56
6. Discussão de Resultados	56

6.1. Revisão de Literatura	56
6.2. Proposta de Solução	57
CAPÍTULO VII	58
7. Considerações Finais	58
7.1. Conclusões	58
7.2. Constrangimentos.....	59
7.3. Recomendações	59
Referências Bibliográficas.....	61
Anexos.....	64
Anexo 1 - Guião de Entrevista dirigida à equipa de Monitoramento de Seca no INAM. A1	
Anexo 1.1 - Resultados da Entrevista 1	A1.3
Anexo 1.2 - Resultados da Entrevista 2.....	A1.6
Anexo 1.3 - Resultados da Entrevista 3.....	A1.9
Anexo 2 - Fluxo de Desenvolvimento do Modelo de Proposto.....	A2.1
Anexo 3 - Descrição dos Dados de Históricos Climáticos.....	A3.1
Anexo 4 – Arquitectura e Hyperparâmetros do Modelo Proposto.....	A4.1
Anexo 5 – Resultados do Processo de Treinamento do Modelo de Proposto.....	A5.1
Anexo 6 – Resultados do Processo de Teste do Modelo de Proposto	A6.1
Anexo 7 - Especificação dos Casos de Uso	A7.1
Anexo 8 - Diagrama de sequências	A8.1
Anexo 9 - Diagrama de Classes	A9.1

Lista de Figuras

Figura 1. Componentes de um modelo de ML genérico. Fonte: Jafazr Alzubi et al., <i>Machine Learning from Theory to Algorithms: An Overview</i> , 2018.	20
Figura 2. Um neurônio artificial. Fonte: Zakaria, AL-Shebany, & Sarhan, <i>Artificial Neural Network : A Brief Overview</i> , 2014.	23
Figura 3. Estrutura geral de uma rede neural artificial com camadas ocultas. Fonte: Zakaria, AL-Shebany, & Sarhan, <i>Artificial Neural Network: A Brief Overview</i> , 2014.	24
Figura 4. Como o modelo de Gradient Boosting é construído. Fonte: Internet, <i>Understanding Gradient Boosting</i> , <i>Hemashreekilari</i> , 2023.	25
Figura 5. Como o modelo de Random Forest é construído. Fonte: Internet, <i>Bagging and Random Forest</i> , 2021.	26
Figura 6. Categorias de Redes Neurais Recorrentes. Fonte: Internet.	27
Figura 7. Processo de previsão de secas pelo INAM.	33
Figura 8. Visão geral da solução proposta.	36
Figura 9. Funcionamento dos módulos da proposta de solução.	38
Figura 10. Valores de SPI ao longo dos anos em função da escala de 1, 3 e 6 meses, na bacia hidrográfica do rio Umbelúzi.	40
Figura 11. Divisão de Dados para Treinamento e Teste.	41
Figura 12. Resultados do teste do modelo LSTM em escala de 1.	44
Figura 13. Resultados do teste do modelo LSTM em escala de 3.	44
Figura 14. Resultados do teste do modelo LSTM em escala de 6.	45
Figura 15. Diagrama de casos de uso.	48
Figura 16. Arquitectura do sistema.	51
Figura 17. Mapa de navegação do sistema proposto.	52
Figura 18. Dados climáticos usados no primeiro cenário de teste.	53
Figura 19. Resultado obtido do primeiro cenário de teste.	53
Figura 20. Dados climáticos usados no segundo cenário de teste.	53
Figura 21. Resultado obtido do segundo cenário de teste.	54
Figura 22. Diagrama ilustrando a sincronização de actualizações no sistema.	55
Figura A2 - 1. Fluxo de desenvolvimento do modelo.	A2.1
Figura A4 - 1. Arquitectura e hiperparâmetros dos modelos LSTM. Fonte: Jupyter Notebook.	A4.1
Figura A5 - 1. Gráfico de perda no processo de treinamento do modelo LSTM para previsão de SPI na escala de 1. Fonte: Jupyter Notebook.	A5.1
Figura A5 - 2. Gráfico de perda no processo de treinamento do modelo LSTM para previsão de SPI na escala de 3. Fonte: Jupyter Notebook.	A5.1
Figura A5 - 3. Gráfico de perda no processo de treinamento do modelo LSTM para previsão de SPI na escala de 6. Fonte: Jupyter Notebook.	A5.2

Figura A6 - 1. Resultado das métricas de desempenho do modelo LSTM em escala de 1.	A6.1
Figura A6 - 2. Resultado das métricas de desempenho do modelo LSTM em escala de 3.	A6.2
Figura A6 - 3. Resultado das métricas de desempenho do modelo LSTM em escala de 6.	A6.2
Figura A8 - 1. Diagrama de sequência: Autenticação de usuário.	A8.1
Figura A8 - 2. Diagrama de sequência: Carregar dados.	A8,1
Figura A8 - 3. Diagrama de sequência: Realizar e visualizar previsão.	A8.2
Figura A8 - 4. Diagrama de sequência: Exportar resultado.	A8.2
Figura A8 - 5. Diagrama de sequência: Exportar relatório de últimas previsões.	A8.3
Figura A9 - 1. Diagrama de Classes da proposta de solução.	A9.1

Lista de Tabelas

Tabela 1. Tabela com as definições de valores e classes do SPI.	13
Tabela 2. Resultados das entrevistas realizadas no caso de estudo - divisão em temas.	30
Tabela 3. Abordagens de implementação de ML como solução.	35
Tabela 4. Arquitetura do LSTM para previsão de SPI nas escalas de 1, 3 e 6.	42
Tabela 5. Configurações de parâmetros de treinamento do modelo LSTM.	42
Tabela 6. Resultados de treino de cada modelo em função da escala de SPI.	43
Tabela 7. Configurações de parâmetros de teste dos modelos.	43
Tabela 8. Resultados de perda/erro da previsão de dados de teste no modelo LSTM.	45
Tabela 9. Requisitos Funcionais.	49
Tabela 10. Requisitos Não Funcionais.	50
Tabela A3 - 1. Descrição dos dados climáticos.	A3.1
Tabela A7 - 1. Especificação de caso de uso: Autenticação de Usuário.	A7.1
Tabela A7 - 2. Especificação de caso de uso: carregar dados meteorológicos.	A7.2
Tabela A7 - 3. Especificação de caso de uso: Realizar previsão.	A7.3
Tabela A7 - 4. Especificação de caso de uso: Visualizar previsão.	A7.4
Tabela A7 - 5. Especificação de caso de uso: Exportar última previsão.	A7.4
Tabela A7 - 6. Especificação de caso de uso: Exportar últimas previsões.	A7.5

Lista de Abreviaturas e Siglas

CHIRP	Climate Hazards Group Infra-Red Precipitation
IA	Inteligência Artificial
INAM	Instituto Nacional de Meteorologia
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectroradiometer
MSE	Mean Squared Error
MVC	Model View Controller
NASA	National Aeronautics and Space Administration
OMM	Organização Mundial de Meteorologia
ONU	Organização das Nações Unidas
RNA	Rede Neural Artificial
SADC	Comunidade para o Desenvolvimento da África Austral
SPI	Índice de Precipitação Padronizado

Glossário de Termos

Batch	Subconjunto de dados processado pelo modelo em cada interação durante o treinamento.
Época (Epoch)	Uma rodada completa em que o modelo processa todos os dados de treinamento.
Função de Perda (Loss function)	Função que mede o erro entre as previsões do modelo e os valores reais
Hiperparâmetros	Parâmetros definidos antes do treinamento.
Jupyter Notebook	Ambiente interativo para escrever e executar código, usado especialmente em ciência de dados.
Python	Linguagem de programação amplamente usada em aprendizagem de máquina e ciência de dados.
Script	Conjunto de instruções escritas em uma linguagem de programação, para automatizar tarefas ou executar comandos de forma sequencial.

CAPÍTULO I

1. Introdução

Neste capítulo é estabelecido o cenário para o trabalho, apresentando a importância da previsão de secas e sua relevância para Moçambique. Assim, em jeito de contextualização, se aborda o problema ou a problemática da vulnerabilidade de Moçambique às secas, destacando-se a necessidade de identificação e desenvolvimento de métodos ou mecanismos cada vez mais eficazes para lidar com este desafio; é formulada uma questão central para orientar a investigação sobre previsão de secas em Moçambique, seguida da apresentação dos objectivos geral e específicos do trabalho; e por fim, a abordagem metodológica que foi seguida para o alcance dos objectivos estabelecidos.

1.1. Contextualização

A água é o bem mais precioso da humanidade, essencial para a vida, para saúde e para o desenvolvimento de todas as nações. A crescente escassez de recursos hídricos é um desafio global que afecta inúmeras regiões, impactando negativamente o fornecimento seguro e sustentável de água para diversos sectores, incluindo o consumo humano, a agricultura e a indústria.

Estima-se que entre 1900 e 2019, as secas tiveram impacto em 2,7 bilhões de pessoas no mundo (ONU News, 2022), destacando que o total de mortes durante esse período chegou a 11,7 milhões. Adicionalmente, o Relatório da ONU de Maio de 2022, indica que actualmente mais de 2,3 bilhões de pessoas enfrentam *stress* hídrico em todo o mundo (ONU, 2022), prevendo-se que até 2050, a seca poderá afectar mais de 75% da população global (World Economic Forum, 2023).

Segundo um relatório elaborado pelas Nações Unidas, a agricultura é fundamental para o desenvolvimento económico de Moçambique (Nações Unidas, 2015), sendo que, de acordo com a FAO, aproximadamente 80% da população moçambicana depende da agricultura de subsistência (FAO, 2019). O Programa Mundial de Alimentos (World Food Programme, 2020) também refere que as secas recorrentes em Moçambique têm agravado a insegurança alimentar, especialmente nas regiões sul e central, aumentando os desafios socioeconómicos.

As informações relativas a ocorrência deste cenário, foram reforçadas por um canal radiofónico de Moçambique, que noticiou que as regiões sul e centro do país poderiam ser afectadas por seca na época chuvosa 2023-2024, e que o facto resultaria da passagem do fenómeno El Niño, que por sua vez provocaria escassez de precipitação nessas zonas (RM, 2023).

O abastecimento de água às cidades de Maputo e Matola, depende da barragem dos Pequenos Limbobos, alimentada pelo rio Umbelúzi, sendo a agricultura irrigada o seu maior consumidor (Wils, et al., 2001). Segundo Wils, a capacidade projectada na barragem não poderá atender à demanda por mais uma década, tendo em conta a dependência que esta barragem tem das chuvas, que por sinal, são altamente incertas na região.

Isto torna o fenómeno da seca um assunto preocupante, o que motiva o uso de mecanismos para prevenir os impactos ou consequências negativas que podem por este serem causados. Neste contexto, diante das emergências climáticas e enorme volume de dados históricos, usar a IA poderá contribuir para uma melhor estimativa futura destes eventos climáticos extremos (Barbosa, 2024). Barbosa adiciona que a aplicação da IA é um avanço tecnológico importante, se considerado que os modelos climáticos actuais não possuem habilidade para detectar secas-relâmpagos.

1.2. Definição do Problema

A seca evidencia-se um fenómeno recorrente em Moçambique abrangendo extensas regiões do país, e o seu monitoramento bem como as tentativas de aplicação de previsões como medidas de prevenção, já têm sido realizados de tempos em tempos, reconhecem-se os desafios específicos enfrentados em regiões propensas à seca, como a região sul de Moçambique, nas previsões emitidas e difundidas pelo Instituto Nacional de Meteorologia (INAM). Estes desafios assentam no seguinte:

- Carência de infraestrutura para uma adequada colecta de dados, o que limita a capacidade de coltar dados climáticos e hidrológicos em tempo real;
- Limitados modelos matemáticos climáticos, que são intrinsecamente complexos e, muitas vezes, baseados em suposições simplificadoras, podendo ou não capturar totalmente a complexidade dos sistemas climáticos, resultando em previsões de secas imprecisas;

- Modelos matemáticos que dependem fortemente de dados históricos, o que pode ser problemático em regiões onde a disponibilidade de dados climáticos é limitada ou inconsistente;
- Modelos matemáticos tradicionais que podem não ser facilmente adaptados para incluir novos dados ou variáveis em tempo real, tornando-os menos ágeis na previsão de secas;
- A actualização e calibração de modelos matemáticos que requerem esforço e *expertise* constantes, o que pode ser caro e demorado.

Nesse contexto, foi identificado como problema a ser resolvido, a previsão **eficiente das secas**, com foco na seca meteorológica e recurso à Inteligência Artificial, o que se torna uma prioridade crucial para enfrentar os efeitos adversos das condições climáticas extremas.

A previsão de secas com recurso à Inteligência Artificial pode ser considerada eficiente se os seguintes critérios forem observados:

- Capacidade de fornecer informações actualizadas sobre as condições climáticas de secas;
- Previsão das secas com adaptação fácil dados novos;
- Amenização de custos excessivos ou esforço de manutenção dos sistemas.

1.3. Objectivos

1.3.1. Objectivo Geral

Propor um modelo de previsão da seca para prevenção de seus impactos com base em algoritmos de Inteligência Artificial.

1.3.2. Objectivos específicos

- Descrever a seca como um fenómeno natural e o seu impacto em Moçambique.
- Caracterizar o sistema de previsão actualmente em uso em Moçambique com foco na disponibilidade e qualidade dos dados climáticos, e na capacidade de colecta, processamento e análise de dados na região de Maputo.
- Descrever as abordagens de implementação de Inteligência Artificial para os problemas de previsão de seca.

- Construir um modelo de aprendizagem de máquina para a previsão de secas na região da bacia hidrográfica do rio Umbelúzi, em Maputo, com base nas abordagens de implementação de tecnologias de Inteligência Artificial seleccionada.
- Desenvolver um protótipo para a previsão de seca com base no modelo construído.

1.4. Metodologia

1.4.1. Etapas da Pesquisa

A pesquisa visa propor um modelo para previsão de secas em Moçambique, com o objectivo de prevenir os impactos causados por esse fenómeno. Assim sendo, a metodologia adoptada para o desenvolvimento do modelo baseia-se na pesquisa qualitativa e exploratória, a qual, na visão de Creswell & Clark (2007), inclui diversas etapas a destacar:

- Inicialmente, foi realizada uma pesquisa sobre o sistema de previsão actual, descrevendo o seu funcionamento, por meio de entrevistas com especialistas e análise de documentos, para identificar limitações na colecta, processamento e análise de dados climáticos, conforme destacado por Braun & Clarke (2006), proporcionando uma base sólida para as etapas subsequentes da pesquisa;
- Em seguida, foi realizada uma descrição das tecnologias de inteligência artificial aplicáveis, por meio de revisão da literatura e entrevistas com especialistas, com o objetivo de entender suas capacidades e limitações, tal como defende Stebbins (2001). Essa análise aprofundada permitiu a identificação das melhores práticas e abordagens para a implementação de soluções inovadoras no contexto da previsão de secas em Moçambique;
- Posteriormente, foram comparadas abordagens de implementação de tecnologias de inteligência artificial, considerando custo, eficácia e escalabilidade, para identificar a melhor solução para o contexto de previsão de secas em Moçambique, procedimento este defendido por Creswell & Clark (2007);

- Após essa fase, foi construído um modelo de aprendizagem de máquina utilizando as técnicas identificadas na etapa de investigação, treinando-o com dados climáticos históricos da região da bacia hidrográfica do rio Umbelúzi, na região de Maputo; e
- Por fim, foi desenvolvido um protótipo funcional de previsão de secas, integrando o modelo de aprendizagem de máquina e realizando testes e validações, seguindo as directrizes de Stebbins (2001). Essa etapa envolveu a colecta e análise de dados quantitativos para avaliar o desempenho do sistema, proporcionando insights valiosos para a implementação de soluções práticas e eficazes na previsão de secas em Moçambique.

1.4.2. Classificação da Metodologia

- ***Quanto à abordagem***

A pesquisa adopta uma abordagem mista, pois combina elementos de pesquisa qualitativa e quantitativa.

- Abordagem Qualitativa: A pesquisa qualitativa é evidenciada pela realização de entrevistas com especialistas e análise de documentos para avaliar a infraestrutura de previsão actual. Esses métodos qualitativos visam obter uma compreensão mais profunda das questões relacionadas à previsão de secas em Moçambique, explorando experiências e perspectivas dos envolvidos. Recorreu-se à metodologia de análise temática qualitativa de Braun & Clarke, como um método sistemático para identificar, analisar e relatar padrões ou temas significativos em dados qualitativos, iniciando pela familiarização com os dados, seguindo a geração de códigos, a partir dos quais são identificadas e agrupadas ideias semelhantes em temas. Esses temas são revisados e refinados para garantir sua consistência e relevância em relação aos objectivos da pesquisa;
- Abordagem Quantitativa: A pesquisa quantitativa está implícita na implementação do modelo de aprendizagem de máquina, que envolve o treinamento do modelo com dados climáticos históricos da região da bacia

hidrográfica do rio Umbelúzi, na região de Maputo. Essa etapa envolve a análise de dados quantitativos para desenvolver um modelo preditivo eficaz para prever secas. Além disso, o desenvolvimento do protótipo funcional também implica a realização de testes e validações, que incluem a colecta e análise de dados quantitativos para avaliar o desempenho do sistema.

- ***Quanto aos objectivos***

Seguindo a classificação de Gil (2002) sobre os objectivos de um trabalho, este estudo é considerado exploratório. O foco inicial foi fornecer uma base sólida de conceitos sobre o problema, abrindo caminho para investigações mais detalhadas no futuro.

- ***Quanto aos procedimentos***

Quanto aos procedimentos metodológicos adoptados, o trabalho se caracteriza por uma abordagem híbrida, combinando pesquisa bibliográfica e documental. A pesquisa bibliográfica foi realizada através do levantamento de referências teóricas em artigos científicos online, com o intuito de embasar teoricamente e praticamente os conceitos abordados no estudo. Por sua vez, a pesquisa documental envolveu a busca e análise de informações em fontes diversas e dispersas, como sites de discussões e fóruns especializados em programação, visando encontrar soluções para questões específicas levantadas em relação ao desenvolvimento do protótipo.

Assim, o procedimento de criação do modelo de previsão de secas, segue as seguintes etapas, conforme descrito por Cielen, Meysman & Ali (2016):

1. Definição do objectivo da pesquisa: claramente estabelecido, determinando quais dados serão colectados e quais métodos serão utilizados;
2. Colecta de dados: que são obtidos através da solicitação de dados relevantes para a pesquisa, garantindo que sejam precisos, confiáveis e relevantes para o objetivo da pesquisa;
3. Pré-processamento dos dados: os dados colectados são limpos e pré-processados, removendo valores inconsistentes ou ausentes, normalizando os dados e transformando-os num formato adequado para o modelo;

4. Seleção do modelo: diversos modelos de previsão de secas são utilizados e comparados, recomendando-se o que apresenta melhores resultados em termos de precisão e acurácia;
5. Treinamento do modelo: o modelo é treinado com os dados colectados e pré-processados, sendo um processo que pode ser demorado, dependendo da complexidade do modelo e da quantidade de dados;
6. Avaliação do modelo: o desempenho do modelo é avaliado utilizando um conjunto de dados de teste, verificando se é capaz de prever secas com precisão;
7. Implementação do modelo: o modelo final é implementado em um sistema operativo, permitindo que seja utilizado para prever secas; e
8. Monitoramento e actualização do modelo: o desempenho do modelo é monitorado periodicamente, podendo ser actualizado com novos dados ou métodos para melhorar sua precisão.

É necessário ter em conta que ao realizar um projecto de ciência de dados, nem sempre é possível planear todos os detalhes do processo com antecedência. Muitas vezes, será necessário iterar entre as diferentes etapas do processo, o que implica rever e ajustar as etapas à medida que o projecto avança.

1.4.3. Técnicas de Colecta de Dados

As técnicas de colecta de dados incluem entrevistas com especialistas, análise de documentos, revisão da literatura e investigação de tecnologias de inteligência artificial.

Em relação ao desenvolvimento do modelo de previsão de secas, a técnica de colecta de dados inclui a requisição de dados históricos colectados pelo INAM. O processo para solicitar dados históricos do INAM pode ser dividido em seguintes etapas:

1. Identificar o período de interesse, as variáveis climáticas relevantes e a resolução espacial desejada;
2. O contacto com o INAM foi feito através do email disponível no site oficial, onde foi solicitado o preenchimento de um formulário de requisição detalhado; e
3. Os dados foram disponibilizados por email ou por meio de um link para download, em diferentes formatos como CSV, Excel ou NetCDF.

Alternativamente, os dados foram colectados à partir do Projecto POWER do Centro de Pesquisa Langley da NASA financiado pelo Programa de Ciência Aplicada da Diretoria

de Ciências da Terra da NASA. Os produtos de dados foram desenvolvidos como parte do projecto Climate Adaptation Science Investigators Phase 2 (CASI2) da NASA e baseiam-se nas projecções do modelo WCRP (Coupled Model Intercomparison Project) - Fase 6 (CMIP6) do clima futuro.

As séries cronológicas foram retiradas e as quantidades derivadas calculadas a partir de dados de modelos reduzidos (conjunto de dados NEX-GDDP-CMIP6). Ao clicar em qualquer ponto do mapa no aplicativo, os dados da instalação mais próxima do local estarão disponíveis para a transferência em formato CSV.

1.4.4. Metodologia de Desenvolvimento do Trabalho Prático

Para o desenvolvimento do protótipo, foi adoptada a metodologia Kanban Lean, que se concentra na optimização do fluxo de trabalho e na eliminação de desperdícios, especialmente para projectos conduzidos por uma única pessoa (De Oliveira, 2023). O Kanban Lean é dividido em várias etapas principais, destacando-se as seguintes:

1. Visualização do Fluxo de Trabalho: mapear todo o processo de trabalho em um quadro Kanban, onde as tarefas são representadas como cartões que se movem através de colunas que representam diferentes etapas do processo;
2. Limitação do Trabalho em Progresso (WIP): estabelecendo limites para o número de tarefas a executar simultaneamente em cada etapa;
3. Gestão do Fluxo: O foco está em manter um fluxo de trabalho constante e equilibrado, ajustando continuamente o processo para melhorar a eficiência; e
4. Melhoria Contínua: A metodologia incentiva a avaliação constante e a melhoria dos processos.

1.4.5. Paradigma de Programação

O paradigma de programação utilizado inclui paradigmas como orientação a objectos (para a implementação do protótipo) e programação funcional (para a implementação do modelo de previsão).

1.4.6. Ambiente de Treinamento do Modelo de Aprendizagem

Para treinar o algoritmo de aprendizagem, foi utilizado o Jupyter Lab, uma plataforma que possibilita a escrita, execução e gravação de código, além de permitir compartilhar

análises. Optou-se por essa ferramenta devido à fácil integração com bibliotecas populares de aprendizagem de máquina.

1.4.7. Linguagens de Programação

A linguagem de programação para criação do protótipo foi o Python. O Python foi escolhido devido ao seu amplo suporte na comunidade de desenvolvimento de sistemas de inteligência artificial, que servirá para visualização de previsões realizadas.

1.4.8. Testagem de Resultados

Os resultados serão testados por meio de análise quantitativa e qualitativa, incluindo a comparação de métricas de desempenho do modelo de previsão de secas e a avaliação do protótipo funcional.

1.4.9. Validação de Resultados

A validação dos resultados será realizada por meio de testes e validações, seguindo diretrizes específicas para garantir a eficácia e a precisão do modelo proposto. Os resultados do modelo são confirmados através do uso de um conjunto separado de dados chamado conjunto de teste ou validação. Este conjunto consiste em dados que o algoritmo de aprendizagem não viu durante o treinamento, garantindo uma avaliação imparcial do desempenho do modelo.

CAPÍTULO II

2. Revisão de Literatura

Neste capítulo, são explorados os conceitos básicos sobre secas, incluindo seus tipos e características, conectando-se à contextualização fornecida no Capítulo I. Também serão descritos os principais indicadores utilizados para detectar secas, com destaque para o Índice Padronizado de Precipitação (SPI), os desafios específicos enfrentados por Moçambique em relação à seca e suas implicações e ferramentas actualmente disponíveis para a sua previsão. É também feita uma introdução aos conceitos fundamentais de aprendizagem de máquina e suas aplicações na previsão de secas.

2.1. A Seca Como Fenómeno Natural

A seca é uma das ocorrências mais comuns em todo o mundo e pode manifestar-se em qualquer lugar e sob qualquer regime climático (Bryant et al., 2005; Sheffield e Wood, 2012). No entanto, ela continua a ser um dos eventos naturais menos compreendidos devido à sua complexidade e às dificuldades na quantificação (Hagenlocher et al., 2019). Os efeitos das secas, que podem ser sentidos em sectores como produção de energia, alimentos, e abastecimento de água, geralmente são percebidos quando as consequências se tornam difíceis de mitigar, tornando as secas o desastre que causa as maiores perdas socioeconômicas globalmente (WMO, 2006). Espera-se que esses efeitos se agravem ao longo do tempo devido às mudanças climáticas globais.

Os períodos de seca, conforme apontado por Mishra e Singh (2011), estão intrinsecamente ligados à redução prolongada na quantidade de precipitação recebida, resultando em impactos significativos na humanidade e no meio ambiente. Mishra e Singh destacam ainda que as mudanças nos padrões espaço-temporais de precipitação e temperaturas extremas, provavelmente tornarão as secas mais frequentes.

2.1.1. Tipos de Secas

A seca pode ser compreendida pela análise de certas características como a frequência, intensidade, duração e extensão de área. No âmbito da pesquisa, é crucial compreender que a seca pode ser categorizada de acordo com seus impactos. Pontes Filho et al. (2019) e Doesken et al. (1991) ressaltam que essa análise requer uma ampla informação actual e histórica das condições meteorológicas e hidrológicas detalhadas de uma região

específica. As secas podem ser classificadas em quatro categorias, cada uma correspondendo ao tipo de impacto ocasionado:

- **Seca meteorológica:** decorrente de uma redução significativa na precipitação e aumento expressivo da evapotranspiração, afetando o sistema hidrológico e sua utilização (Pontes Filho et al., 2019; Pessini, 2017; Van Loo, 2015; Farajzadeh et al., 2009);
- **Seca agrícola:** ocorre quando o *déficit* de precipitação e humidade do solo impactam a agricultura, prejudicando o plantio e a colheita (Mishra e Singh, 2010; Farajzadeh et al., 2009; Kentayeshe e Dracup, 2002);
- **Seca hidrológica:** descreve uma redução significativa nos níveis de água superficiais e subterrâneos, causando distúrbios no abastecimento hídrico e irrigação (Pessini, 2017; Van Loo, 2015; Mishra e Singh, 2010; Farajzadeh et al., 2009); e
- **Seca socioeconômica:** resulta da combinação dos três tipos anteriores e está relacionada à oferta e demanda por bens como água, alimentos e energia, afectando recursos económicos (Pessini, 2017; Ferreira et al., 2015; Van Loo, 2015; Mishra e Singh, 2001; Farajzadeh et al., 2009).

Tomando em consideração que esta pesquisa tem como foco a seca meteorológica sob justificativa de que o INAM enfrenta desafios significativos na colecta abrangente de dados relacionados à ela, devido à limitações na infraestrutura em uso, que não se evidencia adequada para o contexto de frequentes alterações climáticas e ambientais, bem como ao facto de o tema sobre a seca se demonstrar uma questão vasta e abrangente, a análise de todos os tipos de secas em todas as áreas de Moçambique é uma tarefa desafiadora enfrentada pelo país devido à limitação de recursos, tanto financeiros quanto técnicos.

2.1.2. Indicadores de Detecção das Secas

O processo de adaptação à escassez hídrica requer o monitoramento da seca, que pode ser realizado a partir de índices, que tem por finalidade medir a intensidade, duração e frequência desses eventos. As informações fornecidas pelos índices são úteis para planejar e desenvolver ferramentas de apoio à decisão.

Os índices são representações numéricas que expressam e quantificam, por meio de valores, a severidade e magnitude das secas, facilitando a determinação e interpretação

de um evento ou característica. Para isso, os índices utilizam dados provenientes de indicadores, assim como, informações climáticas e hidrometeorológicas (Yihdegol et al., 2019; Svoboda e Fuchs, 2016; Sobral et al., 2011; Jannuzzi, 2004; Valipour, 2013; Jang et al., 2017).

2.1.3. A Precipitação

Das inúmeras variáveis climáticas que podem ser utilizadas para analisar secas, a precipitação é reconhecida como o indicador chave de detecção e caracterização do fenômeno e da sua persistência (Lloyd-Hughes, 2002), sendo também a variável essencial na maioria dos modelos hidrológicos (Santos et al., 2005). Até o momento, não existe um único indicador ou índice que seja aplicável para monitorar ou prever todos os tipos de seca, cada um possui a sua especificidade. Praticamente, todos os índices utilizam como variável de entrada a precipitação, seja ela de forma individual ou combinada com outros elementos. Essa combinação vai depender de qual tipo de seca se deseja monitorar (Suellen Teixeira Nobre Gonçalves, et al., 2021).

Variáveis como precipitação, vegetação, humidade do solo e armazenamento hídrico podem ser usadas como dados para calcular diversos índices e monitorar eventos de seca em uma região (Svoboda & Fuchs, 2016). Com foco na precipitação, diversos índices foram desenvolvidos para identificar seu déficit, incluindo o Índice de Precipitação Padronizado (SPI), Índice de Precipitação e Evapotranspiração Padronizado (SPEI), Índice da Anomalia da Precipitação e Índice Chinês Z (China Z Index), entre outros. Destaca-se o SPI, amplamente utilizado e reconhecido pela Organização Mundial de Meteorologia (WMO) como referência global para estudos de seca meteorológica (Ribeiro Neto, 2017; Hayes et al., 2011; Fernandes et al., 2009).

2.1.4. Standardized Precipitation Index - SPI

Desenvolvido em 1993 por Mckee, Doesken e Kleist como alternativa ao Índice de Seca de Palmer (PDSI), o SPI superou dificuldades de cálculo e limitações de uso. Citado por Gonçalves (2021), ele quantifica o *déficit* de precipitação em diversas escalas de tempo, avaliando a severidade da seca (Ribeiro Neto, 2017; Svoboda and Fuchs, 2016; Fernandes et al., 2009; Patel et al., 2007; Guttman, 1998). Gonçalves adiciona que o SPI calcula o déficit de precipitação com base nos dados mensais acumulados para um intervalo de tempo de n meses, aplicando funções gama e probabilidades de ocorrência

para encontrar desvios em relação à média (Santos, 2020; Martins et al., 2015; Patel et al., 2007; Mckee et al., 1993).

$$SPI = \frac{-(t - C_0 + C_1t + C_2t^2)}{1 + d_1t + d_2t^2 + d_3t^3} \quad \text{Se } 0 < H(x) < 0,5$$

$$SPI = \frac{(t - C_0 + C_1t + C_2t^2)}{1 + d_1t + d_2t^2 + d_3t^3} \quad \text{Se } 0,5 < H(x) < 1$$

Onde:

- $t = \sqrt{\ln [1 / (H(x))^2]}$ para $0 < H(x) > 0,5$ e $t = \sqrt{\ln [1 / (1 - (H(x))^2)]}$ para $0,5 < H(x) < 1$;
- $H(x)$ = distribuição de probabilidade acumulada;
- C_0, C_1, C_2, d_1, d_2 e d_3 são constantes e equivalem respectivamente aos valores: 2,515517; 0,802853; 0,010328; 1,432788; 0,189269; 0,001308.

Utiliza-se o SPI que foi acumulado em intervalos relativos a três e quatro meses para identificar possíveis secas de curto prazo, e para identificação de secas de longo prazo adoptam-se os intervalos de doze, dezoito e vinte e quatro meses, assim como, para o intervalo de seis meses objectivando a verificação da transição entre os períodos de curto e longo prazo (Svoboda e Fuchs, 2016; Martins et al., 2015; Fernandes et al., 2009; Mckee et al., 1993). Em suma, o SPI está disponível nas escalas de tempo de 1, 3, 6, 12, 24 e 48 meses, significando o período durante o qual se acumula o valor de precipitação: o SPI 1 corresponde à precipitação mensal, o SPI 3 corresponde à precipitação acumulada em 3 meses, e assim por diante. Este indicador fornece informações cruciais para o planeamento e desenvolvimento de ferramentas de apoio à decisão diante da complexidade das secas.

A tabela a seguir apresenta a escala padrão de classificação do SPI, com variação de valores positivos a valores negativos, os quais representam suas respectivas seguintes condições climáticas (McKee et al., 1993; Byun; Wilhite, 1999).

Tabela 1. Tabela com as definições de valores e classes do SPI.

Valores do SPI	Classes
≥ 2	Extremamente Chuvoso
1,50 a 1,99	Severamente Chuvoso

1,00 a 1,49	Moderadamente Chuvoso
-0,99 a 0,99	Normal
-1,00 a -1,49	Moderadamente Seco
-1,50 a -1,99	Severamente Seco
≤ -2	Extremamente Seco

2.2. Impacto das Secas

As secas, assim como a deterioração do solo, afectam quase todas as regiões do globo, mesmo as áreas húmidas. Componentes ambientais, como terra e vegetação, são recursos críticos para a produção de alimentos e outros bens e serviços ecossistêmicos, como conservação da biodiversidade, regulação do regime hidrológico, ciclagem de nutrientes do solo e armazenamento de carbono, entre outros (Assan, E. et al., 2018). O recurso natural mais importante é, sem dúvida, a terra produtiva e o solo fértil (Liu, Y. et al., 1998).

A seca tem uma variedade de efeitos no ambiente. Plantas, animais e seres humanos dependem da água. Quando há uma seca, sua fonte de alimento pode ser reduzida, e seu *habitat* pode ser prejudicado. A redução nas fontes de alimentos e água resultou no aumento de doenças em animais selvagens (Kala et al., 2017). Os serviços ecossistêmicos desempenham um papel crítico na gestão ambiental e no desenvolvimento de políticas. A Avaliação do Milênio dos Ecossistemas define serviços ecossistêmicos como os benefícios que os seres humanos obtêm dos ecossistemas, incluindo serviços de provisão, regulação, culturais e de suporte (Li, P. et al., 2017).

A disponibilidade de água doce é crucial para o bem-estar humano e uma economia saudável. O fornecimento de água doce é um serviço de provisão vital que descreve a alteração ecológica da água utilizada para diversas finalidades, incluindo consumo, irrigação, produção de energia hidrelétrica, recreação e pesca (Dennedy-Frank et al., 2016). Durante períodos de seca, o déficit hídrico se propaga pelo ciclo hidrológico, reduzindo o fluxo dos rios, os níveis dos lagos e os lençóis freáticos, sendo às vezes chamado de seca hidrológica (Narasimhan et al., 2005). Além da água doce, a seca pode

reduzir o crescimento vegetal e a produtividade agrícola devido à escassez de precipitação e humidade do solo, afectando potencialmente os serviços de provisão de alimentos (Narasimhan et al., 2005).

Num clima mais quente, a seca pode ter um impacto significativo no ciclo global e regional do carbono, além de causar danos irreparáveis à função do ecossistema (Seddon, 2016). Pesquisas recentes indicam que a seca combinada com temperaturas extremamente altas tem uma influência negativa na absorção de carbono, reduzindo os benefícios da fotossíntese e da fertilização por dióxido de carbono e nitrogênio nas plantas terrestres dos ecossistemas (Yuan et al., 2019). Descobriu-se um crescente impacto do *déficit* de pressão de vapor induzido pela seca no desenvolvimento da vegetação nas últimas três décadas. O impacto da seca nos ecossistemas é complexo, e muitas ambiguidades e questões permanecem sem resposta (Jiao, W. et al., 2021).

Assim, foram destacadas a complexidade e a gravidade dos impactos das secas, o que enfatiza a necessidade da definição de políticas e medidas eficazes de mitigação e adaptação para lidar com esses desafios e proteger as comunidades vulneráveis.

2.3. Moçambique e a Vulnerabilidade à Seca

Moçambique, localizado na costa leste da África Austral, é um dos países menos desenvolvidos da região. Com sua costa de 2.700 km, o país já está experimentando os efeitos devastadores do aumento da frequência de ciclones, inundações e secas nas formas de subsistência agrícola em áreas rurais (FAO, 2012). Moçambique sofre, em média, um desastre por ano e ocupa o terceiro lugar em danos relacionados ao clima, seguindo Bangladesh e Etiópia (Buys et al., 2007; Artur, 2011).

De fato, o país é reconhecido como propenso a desastres e entre os mais vulneráveis às mudanças climáticas e a desastres relacionados ao clima (Osbahr et al., 2008; Patt and Schröter, 2008; Eriksen and Silva, 2009). Por essas razões, Moçambique recebeu considerável atenção e apoio para a implementação de um programa de adaptação às mudanças climáticas (World Bank, 2000; Foley, 2007; World Food Program, 2007; FAO, 2012). Assim, fornece uma área útil para explorar perdas e danos no contexto de vulnerabilidades sociais e resiliência.

Moçambique pode ser dividido em três zonas principais vulneráveis a diferentes riscos naturais, resultando em diferentes impactos. Evidências observacionais sugerem um aumento no número de desastres naturais no país nas últimas três décadas (INGC,

2009). O centro do país é mais propenso a inundações, ciclones tropicais e epidemias, seguido pelo sul e pelo norte. O sul do país, com seu clima de savana tropical seca, é mais propenso à seca do que o centro ou norte, que são dominados, respectivamente, por um clima tropical chuvoso e um clima moderadamente húmido modificado pela altitude. Em termos do número de pessoas afectadas, as inundações e as secas são os desastres naturais mais graves no país (Ange-Benjamin Brida and Tom Owiyo, 2013).

Moçambique é um dos países africanos mais expostos a riscos relacionados ao clima, que são e continuarão a ser exacerbados pelas mudanças climáticas. Eventos extremos perigosos e destrutivos têm sido notáveis e associados à ocorrência de desastres de grandes impactos socioeconômicos (INGC, 2009). Sua população era de 13 milhões em 1990, atingiu 27,9 milhões em 2017 e 29,5 milhões em 2018, com uma taxa de crescimento de 2,9% ao ano (INE, 2018; Banco Mundial, 2019). A sub-região da África Austral (África do Sul e Moçambique) é, após o Norte da África, a mais urbanizada do continente e espera-se que atinja uma maioria urbana em toda a região por volta do final da década actual.

2.3.1. Desafios na Região Sul de Moçambique em Termos de Mitigação à Seca

As entidades ligadas ao M&E (Monitoramento e Avaliação) em Moçambique, enfrentam vários desafios que incluem a falta de harmonização entre planos sectoriais, metas ambiciosas demais para sistemas de M&E, confiabilidade de dados e, conseqüentemente, o uso da informação. Existem também problemas relacionados à capacidade ou nível de conhecimento do pessoal ou ainda de treinamento em matérias relacionadas. Em relação às mudanças climáticas, estudos científicos foram realizados, mas ainda existem lacunas nos dados locais que podem resultar em problemas para estabelecer uma linha de base para a vulnerabilidade, e os dados colectados estão em grande parte em formato de projecto dissimilar e não foram racionalizados ou consolidados. As intervenções de adaptação às mudanças climáticas são relativamente novas em Moçambique, e historicamente houve uma abordagem descoordenada em relação às mudanças climáticas no planeamento do desenvolvimento. Embora tenha havido vários projectos relacionados às mudanças climáticas em Moçambique, não parece ter havido uma aprendizagem consolidada sobre o M&E desses projectos nesse contexto (TAMD, 2013).

Acções Estratégicas Implementadas:

a. Reforçar o Sistema de Aviso Prévio:

Verificar a Taxa de acesso à informação hidrometeorológica (% da população vulnerável informada a tempo sobre a ocorrência de desastre - através da fonte de dados do INAM).

b. Redução de Riscos Climáticos:

- Reforçar (Capacitação) os sistemas de Aviso prévio: rádios comunitárias, estações meteorológicas, infraestrutura para compartilhar informações com setor público e privado, alertas de incêndio, uso de idiomas locais - através do INGC e INAM.
- Reforçar (Capacitação) capacidades de Preparação e Resposta a Desastres: capacidades de evacuação, coordenação do INGC, redução da vulnerabilidade à seca, melhor trabalho dos Centros de Recursos de Uso Múltiplo (CERUM) apoiando comunidades locais na gestão de recursos naturais e mapeamento de vulnerabilidades.

c. Pesquisa e Observação Sistemática:

Criar um Sistema Integrado de Gerenciamento de Informações Climáticas, integrado por INAM, DNA, IIAM e INAHINA.

2.4. Previsão de Secas: Tecnologias e Ferramentas

A previsão frequente de secas é crucial para reconhecer a gravidade, duração e extensão do fenômeno. Isso permite a implementação de preparação e mitigação adequadas para os possíveis efeitos negativos no meio ambiente, na agricultura e na economia (Israel R. Orimoloye, 2022).

Anteriormente, eventos de seca eram monitorados e previstos usando dados de observação adquiridos de estações meteorológicas. No entanto, a distribuição desigual e esparsa dessas estações em países em desenvolvimento aumenta a ambiguidade na pesquisa relacionada à avaliação de secas (Stampoulis, D., Anagnostou, E.N., 2012). Esse problema foi resolvido com a tecnologia de sensoriamento remoto por satélite, a partir da qual dados climáticos em alta resolução em grande escala podem ser facilmente obtidos (Ghajarnia, N., Liaghat, A., 2015; Khan, S.I.; Hong, Y., 2014). Plataformas de satélite possibilitam a aquisição de dados espaciais frequentes, atuais, quase em tempo

real e em alta resolução (Zhang, X, 2014). Nesse sentido, inovações nos campos de sensoriamento remoto (SR) e Sistema de Informação Geográfica (SIG) têm ajudado profundamente na avaliação de riscos de seca nas últimas três décadas.

Segundo Omolola M. Adisa, et al. (2020), as ferramentas de análise evoluíram de estatísticas simples para o uso de ferramentas geoespaciais, como Sensoriamento Remoto (SR) e Modelos de Sistema de Informação Geográfica (SIG), e mais recentemente, aprendizagem de Máquina (ML). Adisa acrescentou que o ML, em particular, contribuiu com cerca de 11% das publicações científicas totais, enquanto os modelos RS e SIG e análises estatísticas básicas respondem por cerca de 44%, 20% e 25%, respectivamente. Com isso, a integração de tecnologias espaciais e ML é fundamental para o desenvolvimento de sistemas robustos de monitoramento e previsão de secas, especialmente na África, considerada um continente propenso a secas.

Embora as plataformas de satélite tenham aprimorado a capacidade de obtenção de dados, é crucial destacar que a eficácia desses métodos ainda enfrenta desafios. A citação de Omolola M. Adisa, et al. (2020) destaca a evolução das ferramentas de análise, mas ressalta que apenas cerca de 11% das publicações científicas estão relacionadas ao aprendizagem de máquina (ML), indicando uma possível lacuna na adoção generalizada dessa tecnologia.

2.4.1. Tecnologias de Previsão de Secas em Moçambique

O Boletim Climático Provincial do INAM emerge como uma ferramenta abrangente para a vigilância de secas em Moçambique. Ao combinar dados de estações meteorológicas e satélites, como o CHIRP e a plataforma MODIS, oferece uma visão robusta dos padrões de precipitação no país. No entanto, é crucial contextualizar essa iniciativa no cenário das limitações existentes na infraestrutura de monitoramento climático em Moçambique.

Apesar dos esforços, a infraestrutura actual enfrenta desafios significativos. Modelos obsoletos, que demandam investimentos consideráveis, destacam-se como uma barreira. A complexidade desses modelos torna imperativo canalizar recursos para treinamento e compreensão, especialmente para gerações mais jovens. A necessidade de explicar tendências e falhas na previsão ressalta a urgência de modernizar e fortalecer a infraestrutura existente.

2.5. Inteligência Artificial e Aprendizagem de Máquina

Tal como foi referido no capítulo I, o estudo realizado foca-se no uso de Inteligência Artificial (IA) para a criação de um modelo que torna eficiente a problemática de previsão de secas meteorológicas, e para isso, irá fazer o uso específico da área de *Machine Learning* (ML) para a análise e mineração de dados.

A IA e o ML não são conceitos novos. Ao longo de mais de 60 anos, cientistas da computação, engenheiros, pesquisadores, estudantes e profissionais da indústria têm investigado, aplicado e reinventado essas disciplinas fundamentadas em álgebra, estatísticas e probabilidade. A era significativa do desenvolvimento sério de ML e IA iniciou-se nas décadas de 1950 e 1960, marcada por contribuições notáveis de pesquisadores como Alan Turing, John McCarthy, Arthur Samuels, Alan Newell e Frank Rosenblatt (Jafar Alzubi, 2018).

De acordo com o Russell e Norvig (2016), a Inteligência artificial é definida como a ciência e engenharia de fazer máquinas inteligentes, especialmente programas de computador inteligentes. Essa definição enfatiza a capacidade das máquinas de realizar tarefas que normalmente requerem inteligência humana, como aprendizagem, raciocínio, solução de problemas, compreensão de linguagem natural e percepção visual. A IA busca desenvolver algoritmos e sistemas que possam aprender com dados, adaptar-se a novas situações e realizar tarefas de forma autônoma, simulando a inteligência humana. Tendo a IA várias áreas de aplicação, importa referir que somente será abordado sobre o Aprendizagem de Máquina ou *Machine Learning* como área, por ser a aplicada neste projecto.

Ao longo do tempo, os pesquisadores formalizaram a definição de ML. O termo foi cunhado por Arthur Samuel em 1959, que o definiu como um campo de estudo que confere capacidade de aprendizagem a computadores sem serem explicitamente programados. Tom Mitchell (1997) proporcionou uma definição mais aplicável ao cenário de engenharia: "Um programa de computador aprende com a experiência E em relação a alguma tarefa T e alguma medida de desempenho P, se seu desempenho em T, medido por P, melhora com a experiência E".

Alzubi destaca que o ML é essencial para capacitar computadores a realizar tarefas sofisticadas sem intervenção humana, baseando-se na aprendizagem e experiência contínua, adaptando-se à complexidade dos problemas. Ao enfrentar desafios em ciência de dados, conforme definido por Hayashi Chikio (1998), que busca unificar

estatísticas, análise de dados, ML e métodos relacionados, a escolha apropriada de algoritmos de ML é crucial para resolver problemas de maneira eficaz.

2.5.1. O Modelo Genérico de Aprendizagem de Máquina

Os algoritmos de Aprendizagem de Máquina (ML) geralmente se enquadram em duas categorias principais: Aprendizagem Supervisionada e Aprendizagem Não Supervisionada. Na Aprendizagem Supervisionada, utiliza-se dados rotulados para treinamento, resultando em uma produção de dados com base em experiências anteriores. Já na Aprendizagem Não Supervisionada, a tarefa é mais desafiadora, pois o sistema lida apenas com dados não rotulados, trabalhando autonomamente para descobrir informações. O projecto segue as directrizes da aprendizagem supervisionada.

O ML é aplicado para solucionar uma variedade de problemas que demandam aprendizagem por parte da máquina. Um problema de aprendizagem possui três características principais:

- **Classes de Tarefas:** refere-se à natureza da tarefa que a máquina está aprendendo;
- **Medida de Desempenho a Ser Aprimorada:** indica o critério usado para avaliar a eficácia da aprendizagem; e
- **Processo de Aquisição de Experiência:** representa como a máquina acumula conhecimento ao longo do tempo.

O modelo genérico de ML consiste em seis componentes independentes do algoritmo adotado. A Figura a seguir ilustra esses componentes primários.

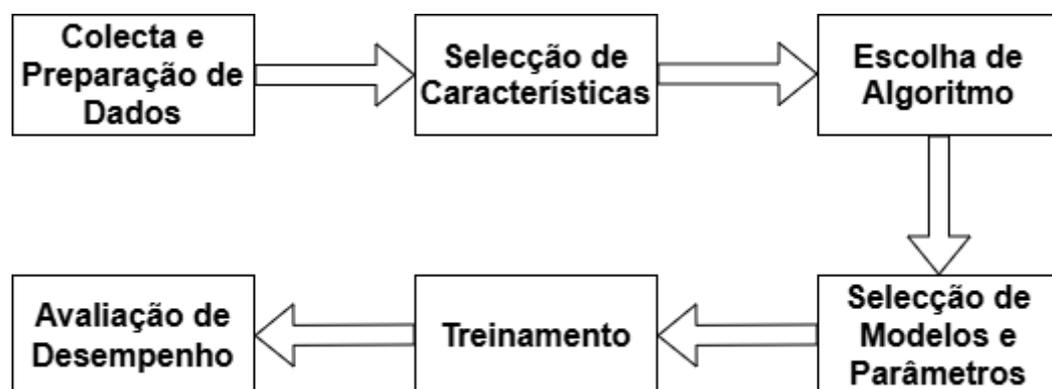


Figura 1. Componentes de um modelo de ML genérico. Fonte: Jafazr Alzubi et al., *Machine Learning from Theory to Algorithms: An Overview*, 2018.

Cada componente do modelo possui uma tarefa específica, conforme descrito a seguir:

1. Colecta e Preparação de Dados:

A tarefa principal do processo de aprendizagem de máquina é colectar e preparar dados em um formato que possa ser fornecido como entrada para o algoritmo.

2. Selecção de Características:

Os dados obtidos na etapa anterior podem conter inúmeras características, e sendo que nem todas são relevantes para o processo de aprendizagem, é necessário remover essas características e obter um subconjunto das mais importantes.

3. Escolha do Algoritmo:

Escolher o melhor algoritmo de aprendizagem de máquina para o problema em questão é crucial para obter os melhores resultados possíveis.

4. Selecção de Modelos e Parâmetros:

A maioria dos algoritmos de aprendizagem de máquina requer alguma intervenção manual inicial para definir os valores mais apropriados de vários parâmetros.

5. Treinamento:

Após seleccionar o algoritmo apropriado e os valores de parâmetros adequados, o modelo precisa ser treinado usando uma parte do conjunto de dados como dados de treinamento.

6. Avaliação de Desempenho:

Antes da implementação em tempo real do sistema, o modelo deve ser testado contra dados não vistos para avaliar quanto foi aprendido, utilizando vários parâmetros de desempenho, como precisão, acurácia ou erro.

- ***Parâmetros de Avaliação de desempenho***

Para avaliação de desempenho em modelos de aprendizagem supervisionada, usam-se como parâmetros as **Funções de Perda (Loss Functions)**. Estas funções quantificam a discrepância entre as previsões do modelo e os valores reais, permitindo que o modelo aprenda ajustando seus parâmetros para minimizar essa discrepância. Algumas das funções de perda mais comuns são:

○ Mean Squared Error (MSE)

Amplamente utilizada em problemas de regressão, sendo a função de erro mais comum para medir a diferença entre os valores reais e previstos. Esta é definida como a média das diferenças ao quadrado entre as previsões e os valores reais:

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

Onde:

- y_i = valores reais.
- \hat{y}_i = valores previstos pelo modelo.
- n = número total de observações.

Interpretação: quanto menor o MSE, melhor é o modelo. Um MSE alto indica que as previsões estão muito distantes dos valores reais.

O MSE é uma função de perda útil quando se deseja penalizar erros grandes, pois ao elevar ao quadrado a diferença, erros maiores têm um impacto mais significativo na função de perda (Goodfellow, Bengio, & Courville, 2016). A desvantagem do MSE é que ele é sensível a outliers, que podem influenciar de forma desproporcional o ajuste do modelo (Montgomery, Peck, & Vining, 2012).

○ Mean Absolute Error (MAE)

Função que mede a média das diferenças absolutas entre os valores previstos e os valores reais:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Interpretação: A MAE reflete o erro médio absoluto entre previsões e valores reais. Quanto menor o valor da MAE, mais preciso é o modelo.

Ao contrário do MSE, a MAE não eleva as diferenças ao quadrado, o que significa que ela é menos sensível a *outliers*. Isso a torna uma escolha preferida quando erros grandes são aceitáveis e se deseja um ajuste que trate todos os erros de maneira uniforme (Willmott & Matsuura, 2005). No entanto, uma desvantagem da MAE é que ela pode ser menos eficiente em lidar com erros que devem ser

penalizados mais fortemente, como em tarefas onde grandes erros podem ser críticos.

2.5.2. Redes Neurais Artificiais

De acordo com Carvalho (2023) com o desenvolvimento dos sistemas computacionais, em particular o GPU, possibilitaram o crescimento exponencial das técnicas de ML, que culminou com novas abordagens e metodologias que agora formam uma nova categoria de ML, o *Deep Learning*, que se refere ao uso de redes neurais artificiais que contêm muitas camadas de processamento.

As redes neurais artificiais são modelos de computação inspirados biologicamente. Geralmente, uma rede neural consiste em um conjunto de neurônios artificiais, comumente referidos como neurões, e um conjunto de arestas direcionadas entre eles, que representam intuitivamente as *sinapses* em uma rede neural biológica.

Segundo Zakaria, AL-Shebany, & Sarhan (2014), a RNA pode ser definida com base nas seguintes três características:

- A arquitetura, indicando o número de camadas e o número de nós em cada uma das camadas.
- O mecanismo de aprendizagem aplicado para actualizar os pesos das conexões.
- As funções de activação utilizadas em várias camadas.

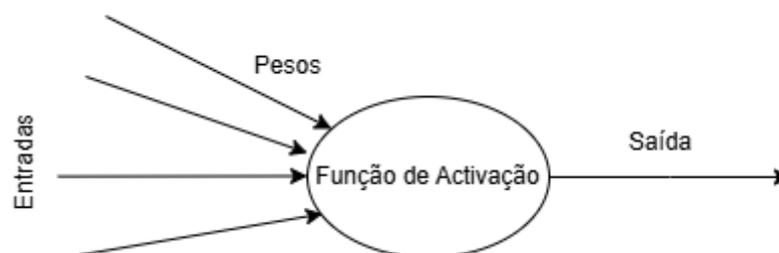


Figura 2. Um neurônio artificial. Fonte: Zakaria, AL-Shebany, & Sarhan, *Artificial Neural Network : A Brief Overview*, 2014.

Perceptrons simples e multicamada são tipos fundamentais de RNAs. O *perceptron* simples é a forma mais básica, com uma única camada de saída, enquanto a multicamada, como o nome sugere, possui múltiplas camadas, incluindo camadas ocultas.

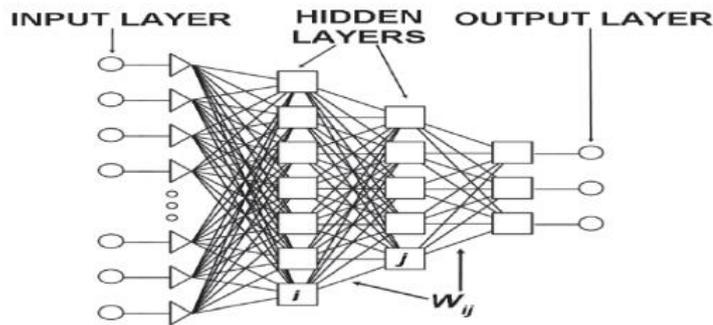


Figura 3. Estrutura geral de uma rede neural artificial com camadas ocultas. Fonte: Zakaria, AL-Shebany, & Sarhan, *Artificial Neural Network: A Brief Overview*, 2014.

A fórmula básica do *perceptron* para a actualização de pesos é:

$$W_{ij}(t + 1) = W_{ij}(t) + \eta \cdot (d_j - y_j) \cdot x_i$$

Onde:

- W_{ij} é o peso da conexão entre i e j .
- t denota o instante de tempo
- η é a taxa de aprendizagem
- d_j é o valor desejado da saída
- y_j é a saída real do neurônio
- x_i é a entrada do neurônio.

2.5.3. Abordagens de Aplicação da Aprendizagem de Máquina na Previsão de Secas

Os dados climáticos passados são um grande tesouro, portanto, com esses dados passados e modelos de ML, os cientistas de dados podem alcançar previsões climáticas quase 100% precisas (Karpagam Sundararajan, et al., 2021). A seguir, os modelos que podem ser aplicados para previsão de secas, segundo (Brieman, 2001) e (Feng, Liu, & Gong, 2015) :

- **Gradient Boosting**

O Gradient boosting é uma técnica de aprendizagem automática para problemas de regressão e classificação, que produz um modelo de previsão sob a forma de um conjunto de modelos de previsão fracos, normalmente árvores de decisão

(Fafalios, Charonyktakis & Tsamardinos, 2020). Esta técnica constrói o modelo por fases, tal como outros métodos de reforço, e generaliza-os ao permitir a optimização de uma função de perda diferenciável arbitrária. A ideia do Gradient Boosting é que o reforço pode ser interpretado como um algoritmo de optimização de uma função de custo adequada (Leo Breiman, 1997).

Como funciona o Gradient Boosting:

- Inicialização: começa com um modelo inicial, geralmente simples como por exemplo, para prever a média da variável alvo.
- Melhoria iterativa: em cada iteração, um novo modelo é treinado para prever os erros residuais (as diferenças entre os valores reais e as previsões do modelo actual).
- Combinação: o novo modelo é adicionado ao conjunto e as previsões são actualizadas. O processo é repetido por um número especificado de iterações ou até que os resíduos sejam minimizados.
- Modelo final: o modelo final é uma soma ponderada de todos os modelos no conjunto, o que normalmente resulta em um modelo de previsão forte.

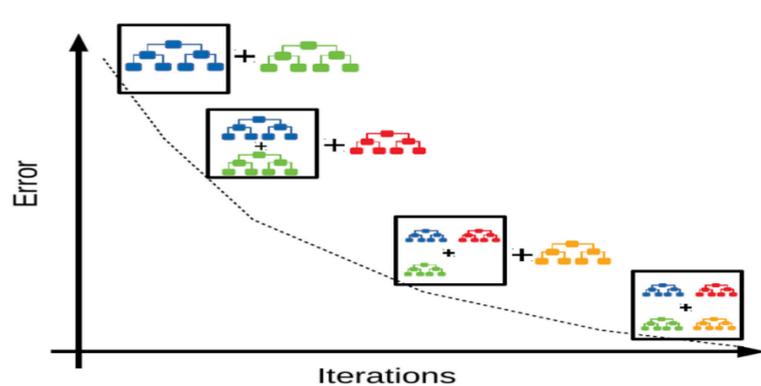


Figura 4. Como o modelo de Gradient Boosting é construído. Fonte: Internet, Understanding Gradient Boosting, *Hemashreekilari, 2023*.

• **Random Forest**

O Random Forest (Floresta Aleatória) é um conjunto de árvores em que cada árvore depende de um conjunto de variáveis aleatórias (Cutler, Cutler, & Stevens, 2011). Cutler, Cutler, & Stevens adicionam que as florestas aleatórias são uma combinação de árvores preditoras de tal forma que cada árvore depende dos valores de um vector aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores da floresta.

Os Random Forests podem ser usados tanto para uma variável de resposta categórica, chamada de “classificação”, quanto para uma resposta contínua, chamada de “regressão”. Da mesma forma, as variáveis preditoras podem ser categóricas ou contínuas.

Como funciona o Random Forest:

- Conjunto de árvores de decisão: utilizando um subconjunto aleatório dos dados de treino e características, são construídas árvores de decisão durante o treino.
- Bagging (Bootstrap Aggregating): são criados subconjuntos do conjunto de dados original por amostragem com substituição para treinar uma árvore de decisão diferente.
- Aleatoriedade das características: em cada divisão da árvore de decisão, é considerado um subconjunto aleatório de características, para criar árvores diversas e a reduzir a correlação entre elas.
- Agregação: em tarefas de classificação, a previsão final é feita por votação majoritária entre todas as árvores, e para tarefas de regressão, a média das previsões de todas as árvores.

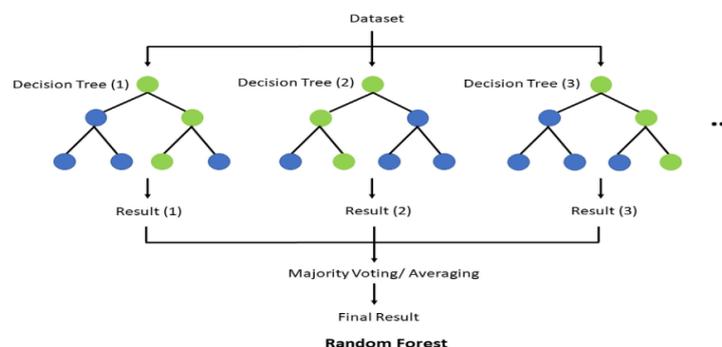


Figura 5. Como o modelo de Random Forest é construído. Fonte: Internet, Bagging and Random Forest, 2021.

- **Redes Neurais Recorrentes**

As redes neurais padrão têm limitações, pois dependem da suposição de independência entre os exemplos de treinamento e teste. Após o processamento de cada exemplo (ponto de dados), todo o estado da rede é perdido (Lipton, Berkowitz, & Elkan, 2015).

Segundo Lipton, Berkowitz, & Elkan, Redes Neurais Recorrentes (RNNs) são modelos conexionistas com a capacidade de passar selectivamente informações através de etapas sequenciais, processando dados sequenciais um elemento de cada vez. Assim, podem modelar entradas e/ou saídas consistindo em seqüências de elementos que não são independentes. Lipton acrescenta que as redes neurais recorrentes podem modelar simultaneamente dependências sequenciais e temporais em várias escalas.

Diferentes tipos de RNNs foram desenvolvidos para melhorar a capacidade de memória das redes neurais. De acordo com Chen (2020), o LSTM, uma melhoria em relação à RNN, adiciona um componente chamado "estado de célula de memória" à sua camada oculta para determinar se a informação é útil ou não (Hu, 2019). Também possuem designação de RNN, modelos como o TLRN, as redes NARX, a rede neural de Elman, e outros.

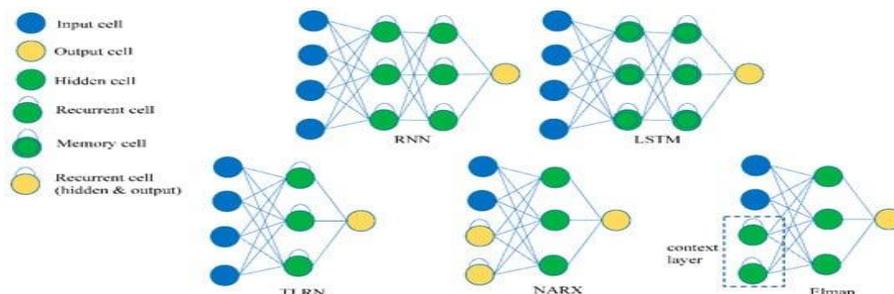


Figura 6. Categorias de Redes Neurais Recorrentes. Fonte: Internet.

2.6. Desafios na Implementação de ML

Segundo Karpagam Sundararajan, et al., (2021), constituem desafios na implementação da Aprendizagem de Máquina como modelo os seguintes:

- A necessidade de uma quantidade massiva de dados históricos para treinar e desenvolver o modelo de ML. A falta de disponibilidade de dados reduzirá a precisão das previsões;
- A necessidade de imagens de alta resolução sem contaminação por nuvens para os modelos que colectam dados de imagens de satélite; e

- Supercomputadores são necessários para o processo de previsão do tempo, pelo facto de a Terra ser um sistema vasto e complexo, exigindo um sistema de monitoramento ou previsão de alta resolução.

CAPÍTULO III

3. Caso de Estudo (INAM)

Este capítulo, incide sobre os resultados da pesquisa realizada junto ao Instituto Nacional de Meteorologia (INAM), onde serão apresentados os resultados das entrevistas e os problemas enfrentados na sua busca. É também neste capítulo onde estes resultados serão conectados à revisão de literatura apresentada no Capítulo II, com o objectivo único de produzir ou elaborar uma proposta de solução para os problemas identificados em matéria de previsão de secas, com base nas abordagens de inteligência artificial discutidas anteriormente.

3.1. Local de Estudo

O INAM é uma instituição pública de carácter técnico-científico que realiza e fornece serviços na área de meteorologia em Moçambique. O seu objectivo é fornecer informações meteorológicas necessárias para garantir o desenvolvimento sustentável da economia nacional e para mitigar os impactos negativos relacionados com o clima, sobre o bem-estar humano e ambiente natural, e cumprir com as responsabilidades nacionais e internacionais na partilha de dados. O INAM tem a sua sede na cidade de Maputo e desenvolve a sua actividade em todo o território nacional - funcionando com 3 Centros Regionais de Previsão nas províncias de Nampula, Sofala e Gaza, mas com dependência técnica, financeira e administrativa da Sede.

3.1.1. Resultados das Entrevistas

Foram conduzidas entrevistas, com base em um guião estruturado (**Anexo 1**), abordando métodos e desafios do monitoramento de secas, com vista a cumprir com os objectivos deste projecto mencionados acima, tendo os participantes compartilhado suas experiências e sugestões de melhorias;

Após a colecta de informação, as entrevistas foram categorizadas e sumarizadas seguindo as fases de análise qualitativa de Braun & Clarke, permitindo identificar padrões e insights relevantes, tal como se demonstra na tabela 2 abaixo.

Tabela 2. Resultados das entrevistas realizadas no caso de estudo - divisão em temas.

	Entrevista 1	Entrevista 2	Entrevista 3
Documentos e Variáveis	Entrevistado 1	Entrevistado 2	Entrevistado 3
Seca	- A seca é um fenómeno climático que não tem controle. É um fenómeno muito vasto, que pode ser medido de várias formas.	- A seca é um fenómeno que se dá por variados motivos, daí que é classificado em vários tipos e é monitorado de vários meios.	- A seca climatológica é causada por déficit prolongado de precipitação em uma determinada região, em comparação com as condições climáticas normais.
Instituição	- O INAM está expandindo seus métodos de monitoramento para incluir a previsão de secas. - O monitoramento de secas atualmente é considerado obsoleto.	- Para fazer a monitoria de precipitação, é possível usar dados observados ou previstos, sendo que o segundo método é mais complexo. - A previsão de secas é uma área nova e em processo de avaliação e aprimoramento do modelo. - Este ano foi o primeiro em que foi possível realizar uma previsão e será avaliado o grau de acerto ao final do ano.	- O INAM realizou previsão de probabilidade de precipitação no ano passado com resultados positivos recentemente. - O INAM é uma instituição que faz parte da OMM e deve estar actualizada com tecnologias e métodos usados em outras instituições.
Origem de dados	- São usados dados de satélite e observados de precipitação para monitoramento - Os dados são fornecidos por outras entidades ou recolhidos das estações que monitoram a precipitação em cada região.	- Dados como humidade do solo são difíceis de colectar devido à falta de recursos.	- A capacidade de prever secas não é influenciada pelos dados, pois há procedimentos para colecta e processamento. - São usados sistemas europeus e de satélite CHIRPS para colecta de dados. - Os dados do INAM são integrados com outros para garantir maior precisão.
Sistema de previsão de secas	- O Sistema de Previsão de Secas acaba de ser implementado, estando em fase de teste. - Utiliza-se o SPI para monitorar a seca meteorológica no país. - É um modelo estatístico que é executado através de um script em R. - Os dados colectados são alimentados no modelo estatístico que gera novos dados. - Os resultados são lidos em um dashboard que mostra a categorização da seca em	- O Sistema de Previsão de Secas utiliza o índice SPI para categorizar e monitorar secas. - Existem outros índices que também permitem o monitoramento de secas. - O SPI é usado para monitoramento e previsão de secas, sendo que a diferença está no uso de dados de precipitação observada ou prevista. - O processo de monitoramento através da previsão é mais complexo.	- Utiliza-se um modelo estatístico para prever secas, que calcula o SPI com base em dados processados. - Os dados são colectados por estações e complementados pelo sistema Europeu ESA, que fornece informações sobre precipitação. - O sistema de satélite CHIRPS também é utilizado para estimar a chuva com base na densidade das nuvens.

	<p>cores.</p> <p>- O processo está em fase de teste e só foi realizada uma previsão até agora.</p> <p>- O objectivo é melhorar a previsão de secas, já que o método atual é baseado no controle de dados de 3 a 6 meses.</p>		
Desafios	- Eficácia e performance do modelo é um obstáculo.	- Escassez de dados limita o uso de índices. - Seca é um fenómeno complexo e de progressão lenta.	- Previsão de secas é um campo recente em África. - INAM é pioneiro na aplicação de métodos de previsão de secas. - Demora na realização de actividades e necessidade de intervenção humana e experiência são desafios.
Eficiência	- O modelo usado é lento e leva tempo para fazer o processamento de dados.		- Representa um desafio o facto de todos os procedimentos precisarem de intervenção humana.
Uso de Inteligência Artificial	- Apesar de não ter muito conhecimento sobre Inteligência Artificial, seria interessante ver como poderia automatizar alguns dos nossos processos.	- Seria vantajoso fazer o uso de sistemas desses que ajudassem a colectar os restantes índices de seca para monitoramento com maior precisão.	- O uso da inteligência artificial seria útil para prever alguns processos de análise de secas climatológica que requerem tempo e esforço no trabalho.

Das entrevistas, foram recolhidos pontos essenciais para o tema em estudo, conforme se destaca em seguida:

- A seca é um fenómeno climático complexo e difícil de controlar, que pode ser classificado de diferentes maneiras dependendo da causa e do impacto que provoca. Em Moçambique, a seca representa uma ameaça significativa, especialmente para a segurança alimentar e a disponibilidade de recursos hídricos. Os entrevistados destacaram que a seca pode ser monitorada por meio de diversos métodos, mas que o fenómeno é vasto e de progressão lenta, exigindo uma compreensão aprofundada para uma resposta eficaz;
- O sistema de previsão de secas em Moçambique, embora ainda em fase inicial, utiliza o Índice Padronizado de Precipitação (SPI) como principal ferramenta para monitoramento e previsão. Os dados para esse sistema são colectados a partir de várias fontes, incluindo estações meteorológicas locais e sistemas de satélite como o CHIRPS e o ESA. Apesar da implementação desse sistema, os entrevistados indicaram que a eficácia e a performance do

modelo são desafios significativos. Além disso, a escassez de dados e a necessidade de intervenção humana no processamento limitam a flexibilidade e a eficiência do sistema. O INAM, sendo pioneiro na aplicação desses métodos em Moçambique, ainda está avaliando a acurácia das previsões realizadas; e

- Embora os entrevistados tenham expressado conhecimento limitado sobre Inteligência Artificial (IA), reconheceram o potencial da tecnologia para otimizar os processos actuais de previsão de secas.

O processo de previsão de secas pelo INAM é dividido em quatro etapas:

1. **Colecta e Preparação de dados:** O INAM colecta dados de precipitação de estações meteorológicas em todo o país, e complementa com dados de outras origens (sistema europeu ESA e Satellite CHIRPS). Para que os dados sejam utilizáveis, estes são preparados de seguida, através de procedimentos de limpeza e padronização dos mesmos;
2. **Cálculo do SPI:** O INAM calcula o SPI para cada estação meteorológica;
3. **Análise do SPI:** O INAM analisa o SPI para identificar áreas onde a precipitação está abaixo da média climatológica (dependendo do que é considerado como clima normal em determinada região); e
4. **Divulgação dos resultados:** O INAM divulga os resultados de previsão de secas para as restantes instituições que esperam receber estes dados para tomada de decisões ou acções.

A seguinte figura mostra o diagrama de fluxo do processo de previsão de secas, através do cálculo do SPI utilizando pelo INAM:

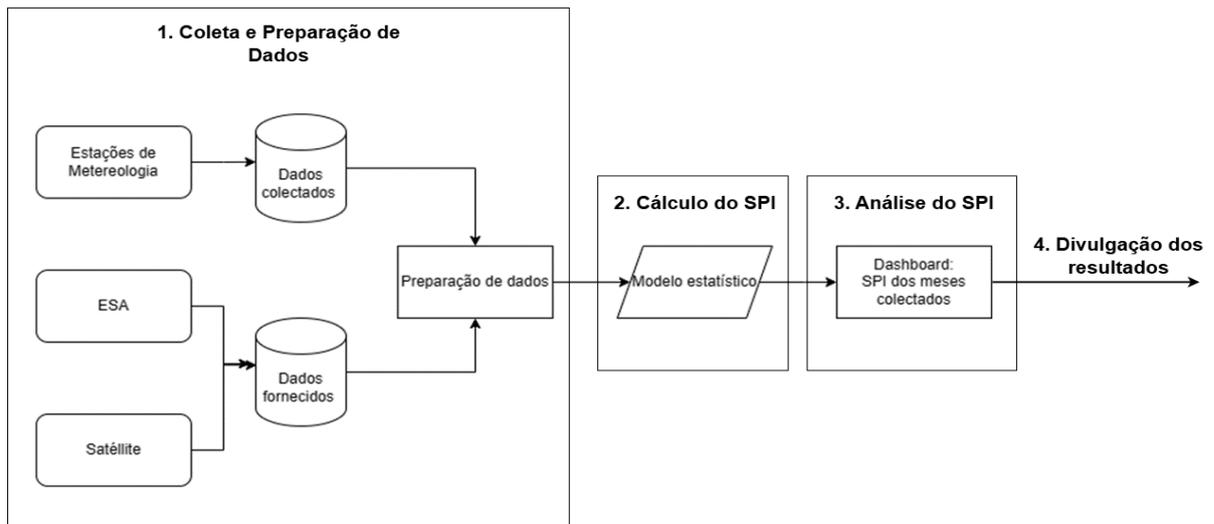


Figura 7. Processo de previsão de secas pelo INAM.

3.1.2. Problemas Enfrentados no Local de Estudo

Com base nos resultados das entrevistas foram listados os problemas abaixo, sobre os quais há uma necessidade de serem identificadas soluções:

- A infraestrutura actual de monitoramento de secas em Moçambique é baseada somente no índice de precipitação acumulada (SPI). Este índice é relativamente simples de calcular, mas tem algumas limitações, como a sensibilidade à variabilidade espacial da precipitação; e
- O sistema actual de monitoramento de secas é lento e pouco flexível. Os dados de precipitação são colectados de estações meteorológicas terrestres e complementados por dados de origem do sistema europeu ESA e dados do satélite CHIRPS, mas o processamento dos dados e a geração dos resultados levam dias.

3.1.3. Comparação das Abordagens de IA para os Problemas Identificados

Os resultados das entrevistas mostram que a aplicação de IA pode ser conveniente para a problemática da previsão da seca. Para tal, recorre-se aos dados históricos para poder alimentar o modelo de ML com o objectivo de monitorar e prever eventos de seca na região da bacia hidrográfica do rio Umbelúzi, em Maputo. No entanto, existem algumas considerações gerais a tomar para escolher a melhor alternativa de implementação de ML:

- A disponibilidade de dados, onde são necessários para o treinamento e o teste do modelo devem ser garantidos, sob risco de a implementação do projecto ficar prejudicada;
- O custo de colectar e preparar os dados, isto é, se o custo for muito elevado, o projecto pode não ser viável;
- A complexidade de integrar a IA ao processo de monitoramento de secas, podendo tornar o projecto difícil de implementar e manter; e
- A precisão da categorização da seca, onde a IA deve ser capaz de gerar categorizações precisas para que o projecto seja eficaz.

Existem várias abordagens do uso do ML, que podem ser aplicadas para resolver variados problemas. Foram definidos critérios de comparação com respectiva pontuação, que são essenciais para avaliar as abordagens como modelos de previsão de secas. Cada critério tem um impacto significativo no desempenho, na eficiência e na capacidade de o modelo prever com precisão eventos de seca. Abaixo estão os critérios selecionados e a justificativa para sua inclusão:

- Capacidade de Processamento: A capacidade do modelo em lidar com grandes volumes de dados e realizar cálculos complexos de forma eficiente,
- Tratamento de Dados Temporais: A capacidade de o modelo lidar com dados temporais, pois séries temporais são comuns em previsão de secas;
- Complexidade do Modelo: A complexidade do modelo em relação à sua arquitectura e número de parâmetros;
- Interpretabilidade dos Resultados: O quão fácil é interpretar e entender os resultados produzidos pelo modelo;
- Tendência à *Overfitting*: A propensão do modelo em se ajustar demais aos dados de treinamento, prejudicando sua capacidade de generalização para dados não vistos;
- Robustez e Estabilidade: A capacidade do modelo de manter um bom desempenho em diferentes condições e cenários, bem como sua estabilidade ao longo do tempo;
- Facilidade de Implementação: A facilidade com que o modelo pode ser implementado e integrado em sistemas existentes, levando em consideração factores como disponibilidade de bibliotecas de código aberto e recursos de computação necessários, e

- Flexibilidade: A capacidade do modelo de se adaptar a diferentes tipos de dados e cenários de previsão.

Para a busca de soluções para o problema de pesquisa, foram destacadas como opções as seguintes abordagens (a pontuação de cada uma com base nos valores numéricos: Alto = 3; Médio = 2; Baixo = 1):

Tabela 3. Abordagens de implementação de ML como solução.

Critérios de Seleção	Random Forest	LSTM	Gradient Boosting
Capacidade de Processamento	Médio	Alto	Médio
Tratamento de Dados Temporais	Baixo	Alto	Baixo
Complexidade do Modelo	Baixo	Alto	Médio
Interpretabilidade dos Resultados	Alto	Baixo	Médio
Tendência a Overfitting	Baixo	Médio	Médio
Robustez e Estabilidade	Alto	Médio	Alto
Facilidade de Implementação	Alto	Médio	Médio
Flexibilidade	Médio	Alto	Médio
Capacidade de Capturar Padrões Longos	Baixo	Alto	Baixo
Total	18 / 27	22 / 27	17 / 27

Esta avaliação foi realizada com base em critérios geralmente considerados na literatura de aprendizagem de máquina e previsão de séries temporais. Não foi baseada em uma pesquisa específica ou em um autor em particular, mas sim em conceitos amplamente reconhecidos e discutidos em diversas fontes acadêmicas e técnicas sobre o assunto.

Com base na análise comparativa dos critérios de seleção para os modelos Random Forest, Gradient Boosting e LSTM, nota-se que o Random Forest é um método que combina múltiplas árvores de decisão, sendo eficiente em termos de processamento e robusto a ruídos nos dados; contudo, não captura explicitamente dependências temporais inerentes a séries temporais, o que pode limitar sua eficácia em previsões que dependem de padrões temporais complexos. Por outro lado, o Gradient Boosting constrói modelos sequencialmente, cada novo modelo tentando corrigir os erros do conjunto anterior, podendo ser computacionalmente intensivo e mais suscetível ao

overfitting; além disso, não é projectado especificamente para capturar dependências temporais.

Fica claro que a LSTM foi a opção mais adequada para a previsão de secas. Com uma capacidade de processamento muito alta, tratamento eficaz de dados temporais e uma tendência baixa a *overfitting*, a LSTM demonstra robustez e estabilidade, além de oferecer alta flexibilidade e uma complexidade de modelo moderada. Embora a interpretabilidade dos resultados seja baixa (por ser uma rede neural), a LSTM equilibra essa característica com uma facilidade moderada de implementação. Portanto, conclui-se que a LSTM foi a escolha mais equilibrada e eficaz para desenvolver um modelo de previsão de secas, atendendo aos requisitos técnicos necessários para lidar com os desafios complexos desse problema.

Usando a IA com dados históricos colectados, estes são utilizados especificamente para a previsão do SPI num período de 6 meses (SPI de cada mês) na bacia hidrográfica do rio Umbeluzi, em Maputo, assim calculando o índice de secas na região de forma mais flexível. O diagrama na figura abaixo ilustra o que foi feito.

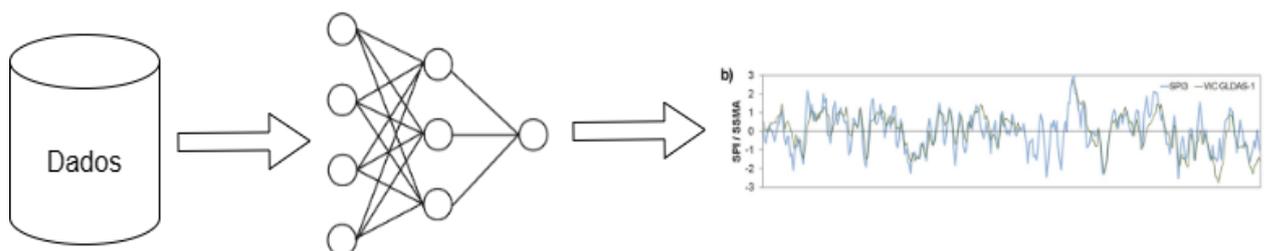


Figura 8. Visão geral da solução proposta.

3.2. Proposta de Solução

Para além de se determinar o modelo a ser usado, esta proposta de implementação se insere nos domínios da Engenharia de Software e Engenharia de Inteligência Artificial, buscando sinergias entre ambas as partes e incorporando metodologias para garantir eficiência, precisão e adaptabilidade.

Sendo a engenharia de IA o ramo da engenharia que se concentra no desenvolvimento de sistemas inteligentes, onde os sistemas são capazes de aprender e se adaptar, e podem ser usados para uma variedade de aplicações, a engenharia de software vem

com o objectivo de apoiar o desenvolvimento profissional de software, mais do que a programação individual. Esta inclui técnicas que apoiam a especificação, projecto e evolução de programas, que normalmente não são relevantes para o desenvolvimento de software pessoal (Sommerville, 2011).

Nesse contexto, a solução foi desenvolvida em dois módulos: Sistema para Previsão de Secas e Modelo de IA para Previsão do SPI e a forma como estes módulos funcionam e se integram é descrita a seguir:

3.2.1. Modelo de IA para Previsão do SPI

Este módulo concentra-se no **uso de algoritmos de IA**, mais especificamente o LSTM, que foram treinados com dados climáticos históricos **para prever os valores do SPI dos próximos 6 meses**. Este módulo funcionará da seguinte forma:

- Entrada de dados climáticos: O modelo de IA recebe como entrada dados climáticos brutos, a partir do sistema principal;
- Processamento de dados: O modelo de IA processa esses dados, realizando os cálculos necessários para definir o valor de SPI com maior probabilidade de estar correcto; e
- Saída de previsões: as previsões são geradas em valores numéricos que representam os índices de SPI e são enviados de volta ao sistema principal para serem exibidas em gráficos de série temporal.

3.2.2. Sistema para Previsão de Secas

Este módulo serve como a *interface* central para os usuários, onde **aloja o modelo de IA como um dos seus componentes** e torna possível visualizar os resultados das previsões de seca. As principais funcionalidades são:

- Recepção e visualização de dados: O sistema recebe os dados climáticos actuais do usuário, integra-os no modelo de previsão, e exhibe os resultados em gráficos de série temporal; e

- Geração de relatórios: os usuários extraem relatórios detalhados sobre as condições de seca previstas, baseados nas previsões do SPI realizadas sobre o modelo.

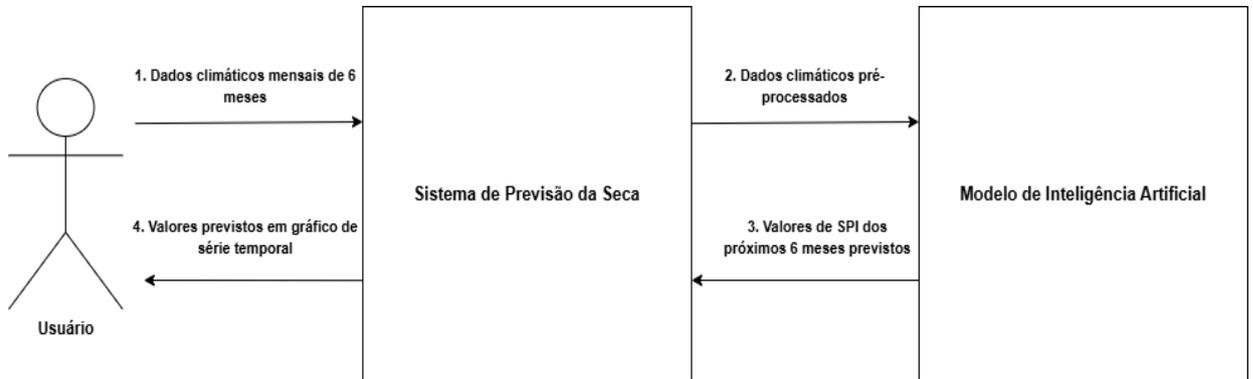


Figura 9. Funcionamento dos módulos da proposta de solução.

CAPÍTULO IV

4. Construção dos Modelos de Previsão de secas

Este capítulo aborda o processo de construção dos modelos de previsão de secas para diferentes escalas de SPI. Foram implementadas metodologias de LSTM com base nos dados climatológicos observados na bacia hidrográfica do rio Umbelúzi. A partir da análise e tratamento dos dados, foram ajustados os hiperparâmetros para otimizar a precisão dos modelos. Este capítulo conecta-se com as soluções tecnológicas mencionadas no Capítulo III, pois é demonstrado como a inteligência artificial pode ser aplicada para prever secas com maior confiabilidade.

4.1. Definição do Objectivo da Pesquisa

O Objectivo da pesquisa se assenta na construção do modelo de inteligência artificial focado no cálculo e previsão do Índice Padronizado de Precipitação (SPI). Este modelo será criado e treinado com dados históricos relevantes para automatizar o processo de monitoramento e previsão de secas e posteriormente integrado ao software. Para a concretização desse objectivo, a seguir se descrevem os procedimentos a aplicar. O Fluxo de criação do modelo pode ser verificado no **Anexo 2**.

4.2. Colecta de Dados

Os dados colectados são referentes ao período de 1981 a 2022 e foram obtidos do Projecto POWER do Centro de Pesquisa Langley da NASA financiado pelo Programa de Ciência Aplicada da Directoria de Ciências da Terra da NASA.

Para cada histórico climático, os dados que constam são: ano, mês, irradiação de ondas curtas do mês, quantidade de névem, precipitação, pressão atmosférica, humidade específica, humidade relativa, temperatura média, temperatura máxima, temperatura mínima, direcção do vento, velocidade do vento e SPI. Para mais detalhes sobre o conjunto de dados, vide o **Anexo 3**.

Assim, o gráfico abaixo, mostra a distribuição dos valores de SPI ao longo do período em análise na bacia hidrográfica do rio Umbeluzi.

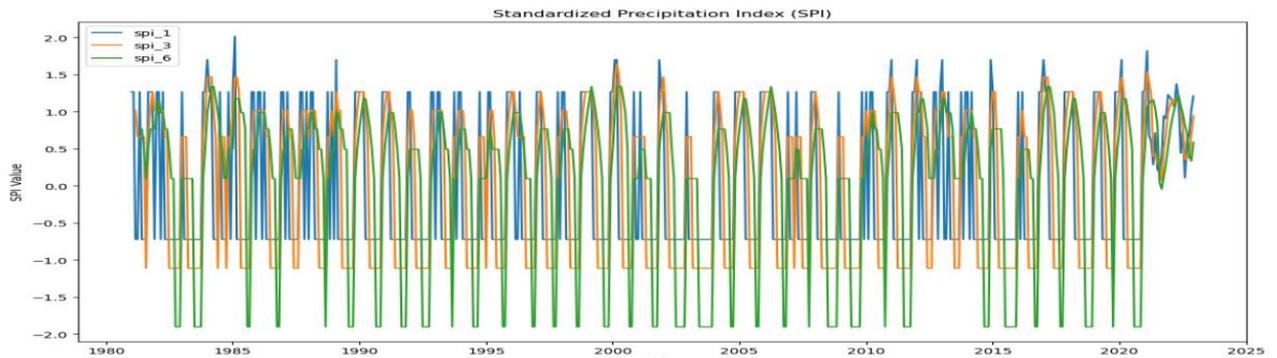


Figura 10. Valores de SPI ao longo dos anos em função da escala de 1, 3 e 6 meses, na bacia hidrográfica do rio Umbelúzi.

4.3. Pré-Processamento dos Dados

Para que o modelo treine e aprenda, será necessário dividir-se os dados preparados em duas partes. A divisão de dados de treino e teste é uma etapa fundamental no processo de aprendizagem de máquina, pois ela permite que o modelo seja avaliado em dados que ele não viu durante o treinamento, fornecendo uma medida mais precisa de seu desempenho no mundo real (Murphy, 2012). Para que se chegue à esta fase, foi necessário que os dados passassem por um processo de limpeza e pré-processamento antecipadamente.

Os métodos usados para a **limpeza** dos dados são:

1. Limpeza de variáveis altamente correlacionadas (para reduzir a multicolinearidade e simplificar o modelo); e
2. Limpeza de históricos com dados nulos.

Os métodos usados para o **pré-processamento** dos dados são:

1. Separação de colunas de Características e Alvos, para que o modelo possa aprender com base nos critérios de aprendizagem supervisionada. Os alvos são os resultados do SPI calculado nas escalas de 1, 3 e 6 para cada histórico temporal, e estes foram usados para criação de 3 modelos, correspondentes à previsão de SPI em cada escala;
2. Normalização dos Dados, para que estes estejam na escala, evitando que o modelo seja dominado por variáveis com maior magnitude e garantindo uma aprendizagem mais justa e precisa; e

3. Criação de Sequências Temporais e Rótulos correspondentes, que se refere ao comprimento de dados que serão alimentados ao modelo em função do número de 6 passos (6 dados históricos são alimentados ao modelo para que ele possa realizar a previsão) e o valor de SPI que o modelo deve prever.

Sobre a **divisão de dados em treinamento e teste**, é relevante saber que:

- Os dados de treinamento são usados para treinar o modelo, enquanto que os dados de teste são usados para avaliar o modelo. Os dados de treino e teste são representativos do conjunto de dados real;
- A divisão de dados de treinamento e teste geralmente é feita em uma proporção de 80% para treino e 20% para teste; e
- Dos dados de treinamento, 80% são divididos em para o treinamento do modelo e 20% para a validação do próprio modelo durante o treinamento.

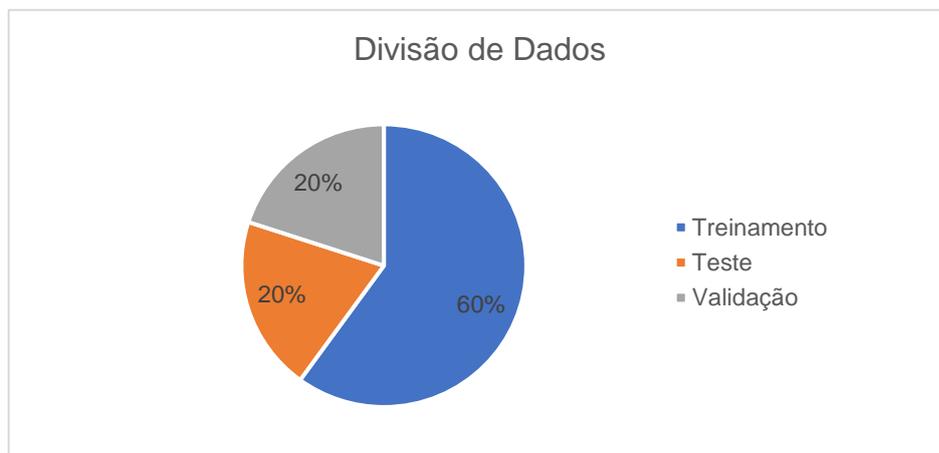


Figura 11. Divisão de Dados para Treinamento e Teste.

4.4. Arquitectura dos Modelos de Inteligência Artificial

Com dados pré-processados, estes serão alimentados ao modelo de inteligência artificial, tendo sido criados três modelos de LSTM separadamente, para a previsão de SPI nas escalas 1, 3 e 6. As especificações dos modelos mostram-se a seguir:

Tabela 4. Arquitectura do LSTM para previsão de SPI nas escalas de 1, 3 e 6.

Nome da Arquitectura	LSTM
Entrada	Dados mensais climáticos observados no período de 6 meses
Camada oculta 1	256 neurônios
Camada oculta 2	128 neurônios
Camada oculta 3	64 neurônios
Camada oculta 4	32 neurônios
Camada oculta 5	16 neurônios
Camada de saída	6 neurônios
Saída	Valores de SPI dos próximos 6 meses

A biblioteca geralmente usada para desenvolver modelos LSTM é o TensorFlow e Keras, que oferece uma variedade de recursos que o tornam ideal para o desenvolvimento de modelos LSTM. Por essa razão, foi a preferida para este projecto.

A arquitectura e hyperpâmetros do modelo podem ser visualizadas no **Anexo 4**.

4.5. Treinamento e Teste dos Modelos

Para a previsão de SPI, a tarefa principal que os modelos utilizaram é a de **Regressão**. Para estas tarefas, métricas como precisão e acurácia não são normalmente utilizadas, pois estes termos estão geralmente associados às tarefas de classificação. As métricas usadas nestas tarefas foram funções para cálculos de erros que quantificam a diferença entre os valores previstos pelo modelo e os reais observados nos dados.

Os parâmetros para o treinamento e optimização do desempenho de aprendizagem do modelo são:

Tabela 5. Configurações de parâmetros de treinamento do modelo LSTM.

Conjunto de dados	80%
Taxa de aprendizagem	0.00005
Tamanho de batch	32

Epochs	500
Função de perda (loss function)	MSE
Dropout	0.2
Optimizador	Adam
Early stopping	False

4.5.1. Resultados da Fase de Treinamento dos Modelos

Os modelos foram treinados para previsão de SPI em escalas de 1 mês, 3 meses e 6 meses, separadamente, para evitar o risco de comprometer a precisão de uma escala para otimizar outra. Ao treinar as categorias separadamente, foi possível maximizar a performance de cada modelo para o seu respectivo horizonte de previsão. A seguir, pode-se visualizar os resultados de treino de cada modelo em função da escala de SPI. Os gráficos relacionados com os valores de perda podem ser visualizados no **Anexo 5**.

Tabela 6. Resultados de treino de cada modelo em função da escala de SPI.

Modelo	Escala de SPI	Perda	Perda de Validação
LSTM	1 mês	0.0751	0.0938
LSTM	3 meses	0.0535	0.0454
LSTM	6 meses	0.0457	0.0433

Normalmente procura-se saber o quão bom o algoritmo de aprendizagem é, em dados que não viu (dados que não foram usados para treinar o algoritmo) (Goodfellow, Bengio, & Courville, 2016). Sendo assim, os parâmetros para o teste do modelo, validando a sua aprendizagem são:

Tabela 7. Configurações de parâmetros de teste dos modelos.

Conjunto de dados	20%
Métricas de Avaliação	MSE MAE

4.5.2. Resultados da Fase de Teste dos Modelos

Nesta fase verifica-se os resultados das previsões de SPI geradas pelos modelos na fase de teste, onde são comparados os valores actuais (reais) com os valores previstos por cada um dos modelos LSTM, utilizando os dados que não foram expostos durante o treinamento.

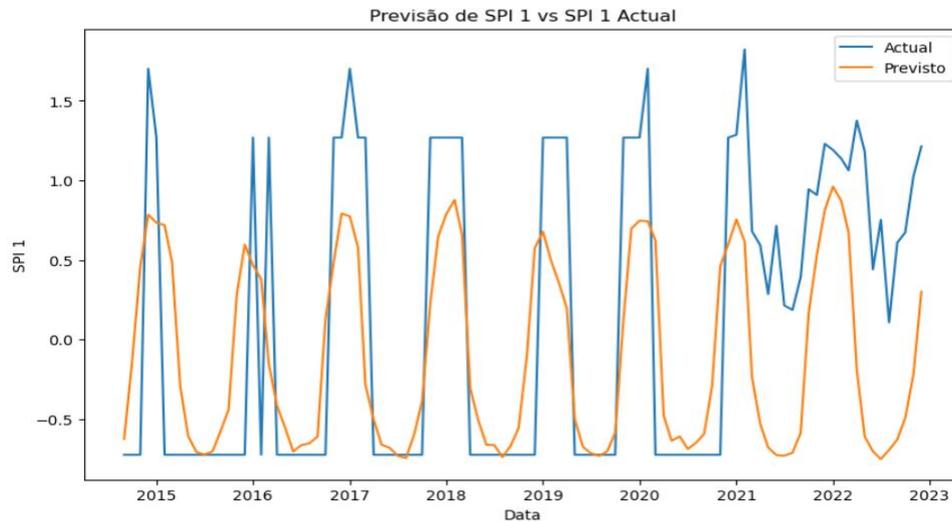


Figura 12. Resultados do teste do modelo LSTM em escala de 1.

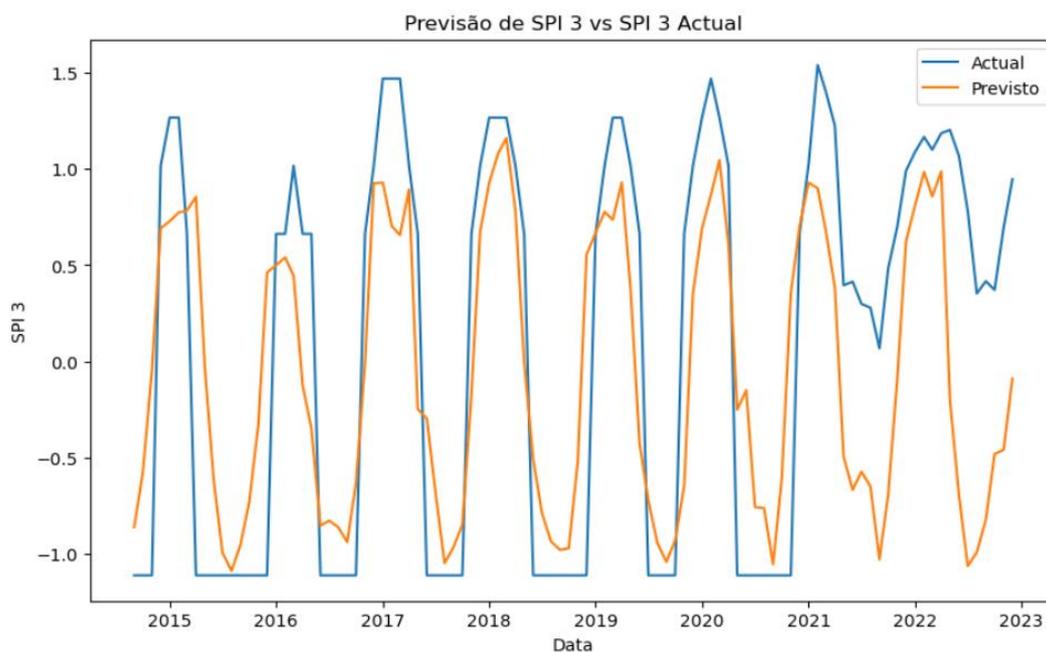


Figura 13. Resultados do teste do modelo LSTM em escala de 3.

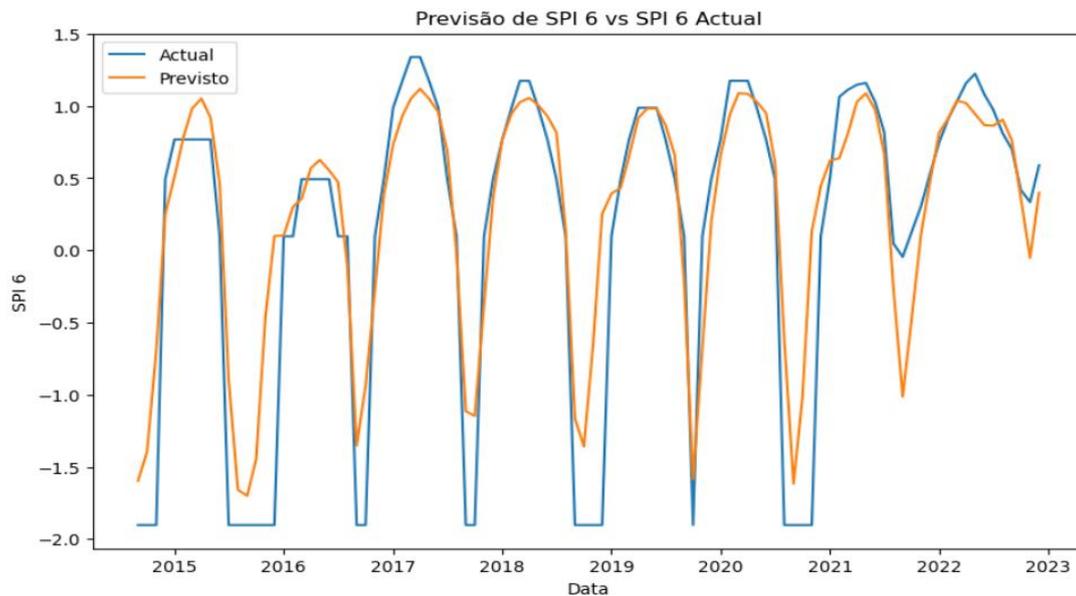


Figura 14. Resultados do teste do modelo LSTM em escala de 6.

Além dos gráficos de previsão, a tabela 8 a seguir, mostra os erros calculados para cada modelo durante a fase de teste para avaliar a precisão e a confiabilidade das previsões dos modelos:

Tabela 8. Resultados de perda/erro da previsão de dados de teste no modelo LSTM.

Modelo	Escala de SPI	MSE (quanto menor, melhor)	MAE (quanto menor, melhor)
LSTM	1 mês	0.0788	0.2180
LSTM	3 meses	0.0731	0.2143
LSTM	6 meses	0.0289	0.1064

Os resultados do processo de teste do modelo proposto também podem ser visualizados com mais detalhes no **Anexo 6**.

4.6. Iteração Contínua com o Modelo

A iteração contínua com o modelo é uma parte importante do processo de desenvolvimento de um modelo de aprendizagem de máquina. Essa iteração permite que se melhore o desempenho do modelo e o torne mais adequado para o problema específico. Após o treinamento inicial do modelo, foi importante realizar iterações

contínuas para melhorar o desempenho. Isso foi feito ajustando os parâmetros de treinamento e teste, ou adicionando novas características ao modelo.

CAPÍTULO V

5. Desenvolvimento de Protótipo da Solução Proposta

Neste Capítulo, detalha-se o processo de desenvolvimento do protótipo do sistema de previsão de secas baseado no modelo construído. Através de uma *interface* amigável, o protótipo permite que os usuários insiram dados climatológicos e visualizem as previsões geradas pelo modelo.

5.1. Modelagem do Sistema

De acordo com Pressman (2011), a análise de requisitos resulta na definição das características operacionais do software, estabelece sua *interface* com outros elementos do sistema e impõe as restrições que o *software* deve cumprir. Essa análise pode ser realizada por meio de modelos como:

- Modelos baseados em cenários (casos de uso);
- Modelos baseados em classes (diagrama de classes);
- Modelos comportamentais (diagrama de sequências); e
- Modelos de fluxo (modelos de dados).

Os requisitos, que representam as restrições que o sistema deve atender, são divididos em diferentes tipos. Para este trabalho, focaremos nos requisitos de negócio, requisitos de usuários, requisitos funcionais e requisitos não funcionais.

- Os requisitos de negócio delineiam o propósito do projecto e definem as necessidades do produto, muitas vezes associadas a regras de negócio;
- Os requisitos de usuários especificam as tarefas que os usuários finais poderão realizar com o produto;
- Os requisitos funcionais descrevem o comportamento que o produto deve ter ou suportar; e
- Os requisitos não funcionais complementam os funcionais, indicando quão bem o produto deve executar.

Adicionalmente, consideram-se as restrições de desenvolvimento, que abrangem a tecnologia de implementação, convenções, documentação e o processo de desenvolvimento do *software*. Esses elementos formam a base para a modelagem do

sistema, proporcionando uma visão abrangente das necessidades e limitações do projecto.

5.1.1. Requisitos de Negócio

Os requisitos de negócio são importantes para garantir que um sistema ou solução de informações atenda às necessidades do negócio. Neste trabalho, definiu-se o requisito de negócio como sendo a melhoria de eficiência na previsão de secas, trazendo mecanismos que facilitem o processo, com recurso a tecnologias de inteligência artificial.

5.1.2. Requisitos de Usuários

Os requisitos de usuários são declarações que descrevem as funcionalidades ou características que os usuários finais esperam do sistema. Esses requisitos são essenciais para entender como o software deve atender às necessidades dos usuários. O diagrama de casos de uso é uma ferramenta visual que representa as interações entre os usuários e o sistema. O diagrama de casos de uso, define somente o que existe fora do sistema (actores) e o que deve ser realizado pelo sistema (casos de uso) (Pressman, 2011), ou seja, ele mapeia os cenários de uso, identificando os diferentes casos em que os usuários interagem com o sistema para alcançar seus objectivos.

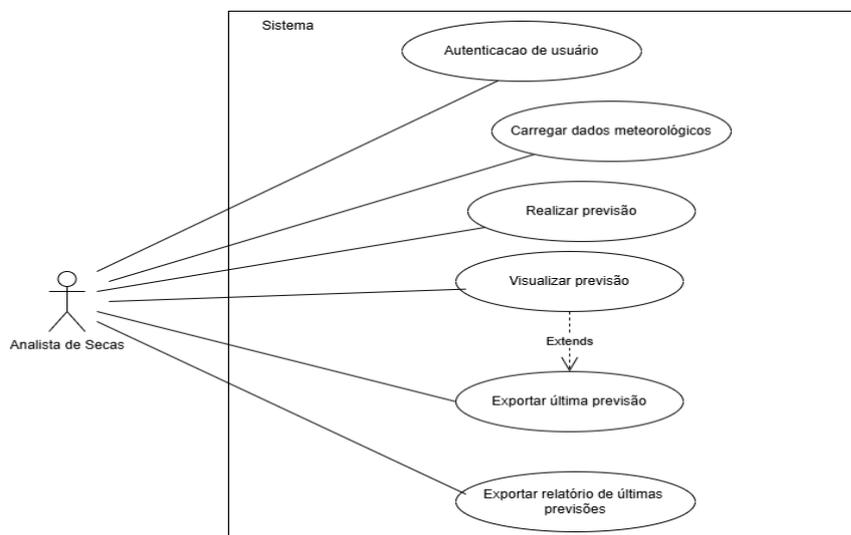


Figura 15. Diagrama de casos de uso.

As especificações dos casos de uso encontram-se no **Anexo 7**, juntamente com os diagramas de sequência para cada caso de uso, no **Anexo 8**.

5.1.3. Requisitos Funcionais e Requisitos Não Funcionais

Ao serem definidos requisitos, a priorização dos mesmos foi crucial no processo de desenvolvimento de *software*, pois permite concentrar esforços nos elementos mais essenciais para atender às necessidades do usuário e alcançar os objectivos do projecto. A categorização dos requisitos em essenciais, importantes e desejáveis ajuda a estabelecer uma hierarquia clara, facilitando a tomada de decisões durante o desenvolvimento.

- Essenciais: São requisitos fundamentais para o sucesso do projecto;
- Importantes: Requisitos significativos que embora não sejam vitais para a versão inicial, agregam valor substancial ao sistema; e
- Desejáveis: Reflectem requisitos que se implementados, proporcionam benefícios adicionais ou melhorias estéticas.

- **Requisitos Funcionais**

Tabela 9. Requisitos Funcionais.

Referência	Nome	Descrição	Prioridade
RF01	Autenticação de usuário	Autenticação do usuário ao sistema por meio de nome e senha	Importante
RF02	Carregar o modelo	Carregamento automático do modelo de aprendizagem de máquina ao inicializar o <i>software</i> .	Essencial
RF03	Carregar dados meteorológicos ao modelo	O <i>software</i> deve poder fazer o processamento e carregamento dos dados recolhidos em 6 meses para prever o SPI de 6 meses no <i>software</i> .	Essencial
RF04	Realizar previsão	O <i>software</i> deve ser capaz de realizar uma previsão do SPI do próximo mês, com base nos dados carregados.	Essencial
RF05	Visualizar resultado de previsão	O usuário deve poder visualizar o resultado do SPI da previsão	Essencial
RF06	Exportar relatório de última previsão	O usuário deve poder exportar o resultado da previsão para um arquivo	Importante

RF07	Exportar relatório de últimas previsões	O usuário deve poder visualizar um relatório contendo as 3 últimas previsões realizadas no sistema	Desejável
------	---	--	-----------

- **Requisitos Não Funcionais**

Tabela 10. Requisitos Não Funcionais.

Referência	Nome	Descrição	Prioridade
RNF01	Precisão	O intervalo de confiança do modelo deve ser de pelo menos 60%.	Essencial
RNF02	Uso de recursos computacionais	O uso de recursos computacionais deve ser mínimo	Importante
RNF03	Facilidade de uso	O sistema não deve ser complexo de se usar, permitindo que o usuário navegue sem dificuldades	Importante
RND04	Disponibilidade	O sistema deve manter-se disponível sem interrupções	Importante

5.2. Restrições de Desenvolvimento

Para a implementação do sistema, foram considerados os seguintes aspectos:

- **Linguagem de programação:**

Python – para a programação do modelo de aprendizagem de máquina e da aplicação *desktop*.

- **Ambiente de execução do sistema:**

- Jupyter Lab – ambiente interativo de desenvolvimento, para a criação do modelo de aprendizagem: e
- PyCharm – ambiente integrado de desenvolvimento para aplicações, incluindo aplicações *desktop*, quer para Windows ou macOS e Linux.

5.3. Arquitectura do Sistema

Para desenvolver o sistema optou-se pela arquitectura MVC (Model-View-Controller) devido à sua simplicidade de implementação e à capacidade de fornecer camadas bem definidas para separação de responsabilidades em cada componente, facilitando a manutenção do projecto, e permitindo uma gestão clara e eficiente das partes do sistema. Vide no **Anexo 9** o diagrama de classes do sistema.

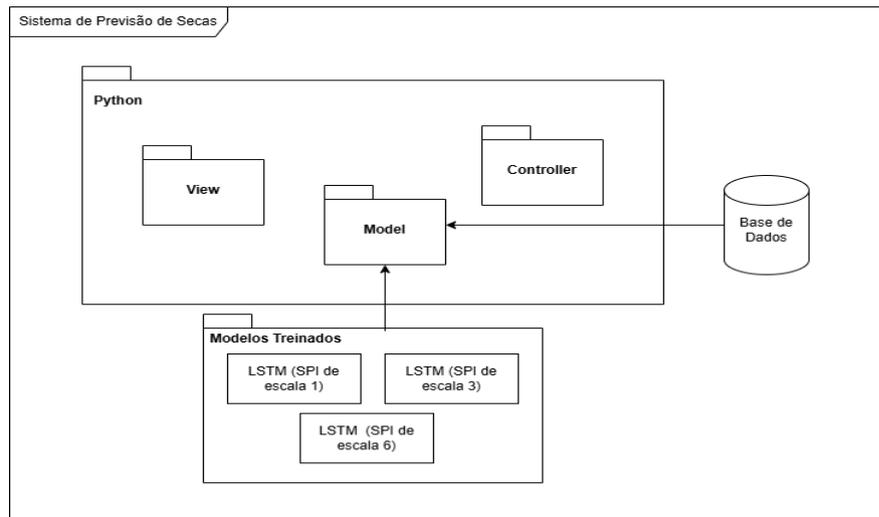


Figura 16. Arquitectura do sistema.

5.4. Mapa de Navegação do Sistema

O mapa de navegação do sistema faz a representação visual da estrutura e do fluxo de interação dentro da plataforma desenvolvida.

O fluxo de interacção com o sistema integra os seguintes passos:

1. O usuário autentica-se ao sistema;
2. O usuário clica em Previsão com Modelo de IA;
3. O usuário escolhe qual é a escala de SPI que pretende usar para previsão usando o modelo LSTM;
4. O usuário carrega dados climáticos mensais equivalentes a 6 meses (em formato *.csv ou *.xlsx) no modelo clica em “Previsão”;
5. O usuário visualiza os gráficos de SPI previstos na escala escolhida num período de 6 meses; e
6. Alternativamente, o usuário exporta o último resultado previsto ou as últimas 3 previsões em relatórios.

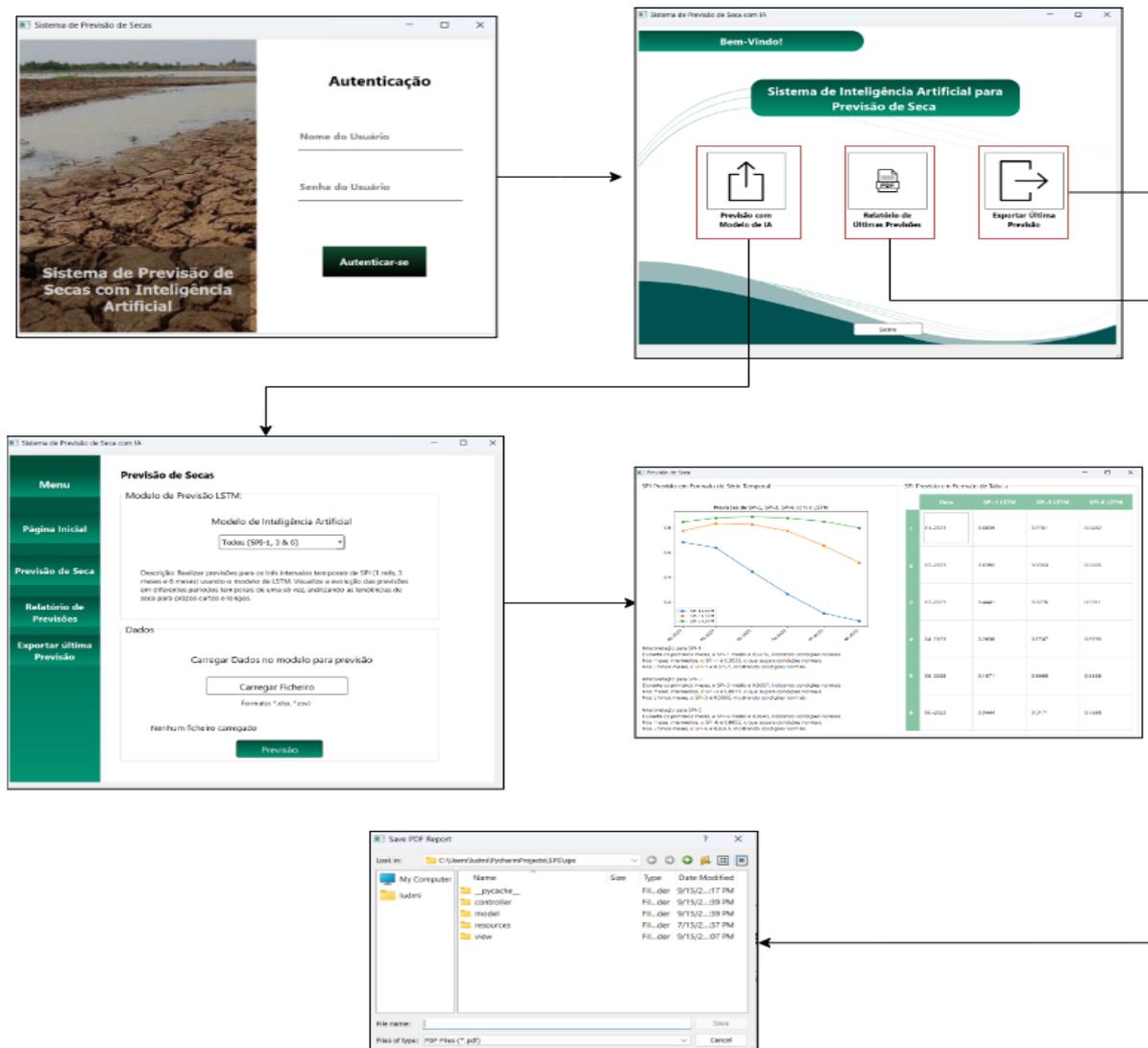


Figura 17. Mapa de navegação do sistema proposto.

5.4.1. Testes Funcionais Realizados no Sistema

Uma vez que o objectivo dos testes funcionais é verificar se as principais funções do sistema operam conforme os requisitos estabelecidos, recorreu-se a este exercício para verificar se este gera correctamente as previsões de SPI-1, SPI-3, e SPI-6. Os seguintes cenários cobrem as principais funcionalidades testadas (requisitos essenciais), com todos os resultados satisfatórios.

Cenário 1:

- **Descrição:** foram usados dados do ano de 2022 (Janeiro - Junho) para prever o SPI dos 6 meses seguintes (Julho - Dezembro) nas escalas de 1, 3 e 6 meses.

- **Passos:** inserir dados climáticos, selecionar o modelo LSTM e executar a previsão.

year	month	shortw_lrr	cloud_am	precip	surface_pr	specific_h	relative_h	temp	max_temp	min_temp	earth_tem	wind_dir	wind_spee	spi_1	spi_3	spi_6
2022	1	21.42	70.24	4.55	100.68	15.87	75.31	26.38	34.51	20.54	27.43	6.55	0.48	1.191447	1.088583	0.74831
2022	2	22.86	45.98	4.09	100.56	15.81	75.94	26.26	35.92	20.69	27.3	7.58	0.08	1.137721	1.166668	0.91515
2022	3	18.92	48.86	3.5	100.73	14.34	73.62	25.18	36.37	18.75	26.15	6.53	0.4	1.062749	1.099566	1.029303
2022	4	15.04	56.61	6.37	100.88	13.98	80.69	23.08	34.08	15.4	23.49	8.23	0.12	1.375573	1.18594	1.158531
2022	5	13.31	37.58	4.46	101.15	12.39	83.69	20.48	28.52	12.48	20.62	7.89	0.17	1.181219	1.202619	1.223671
2022	6	12.95	35.69	0.64	101.71	9.58	77.69	17.62	25.05	9.87	17.48	6.25	0.18	0.442037	1.065707	1.080299

Figura 18. Dados climáticos usados no primeiro cenário de teste.

- **Resultado Esperado:** Gráficos e valores de SPI para as três escalas.
- **Resultado Obtido:** Previsões geradas com sucesso, dentro dos parâmetros esperados.

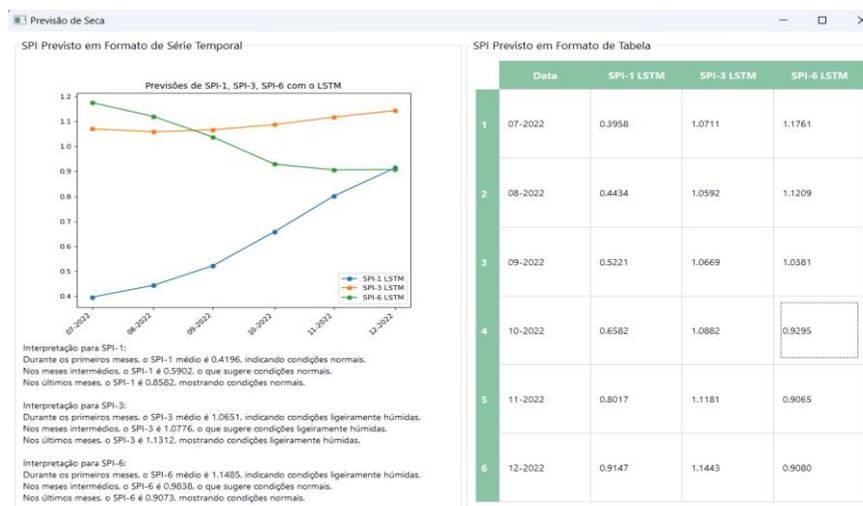


Figura 19. Resultado obtido do primeiro cenário de teste.

Cenário 2:

- **Descrição:** foram usados dados do ano de 2022 (Julho – Dezembro) para prever o SPI dos 6 meses seguintes (Janeiro – Junho de 2023 nas escalas de 1, 3 e 6 meses.
- **Passos:** inserir dados climáticos, selecionar o modelo LSTM e executar a previsão.

year	month	shortw_lrr	cloud_am	precip	surface_pr	specific_h	relative_h	temp	max_temp	min_temp	earth_tem	wind_dir	wind_spee	spi_1	spi_3	spi_6
2022	7	13.58	30.77	1.66	101.53	10.56	78.75	19.11	27.62	10.98	19.3	7.98	0.13	0.75261	0.780634	0.973525
2022	8	15.59	36.96	0.17	101.19	9.95	71.38	19.81	32.4	11.76	20.23	8.34	0.23	0.108633	0.35268	0.812257
2022	9	19.09	31.03	1.1	101.08	10.93	67.25	22.73	35.61	11.76	23.53	8.12	0.38	0.60895	0.416974	0.70108
2022	10	17.73	60.66	1.33	100.92	13.55	69.62	25.51	40.78	18.39	26.88	8.38	0.2	0.673202	0.371688	0.419125
2022	11	20.34	68.91	3.22	100.72	13.61	69.75	25.27	39.78	17.73	26.52	8.19	0.62	1.024247	0.694734	0.334276
2022	12	21.59	74.12	4.75	100.54	15.08	73.12	26.05	36.72	19	27.24	8.6	0.32	1.213723	0.946406	0.589469

Figura 20. Dados climáticos usados no segundo cenário de teste.

- **Resultado Esperado:** Gráficos e valores de SPI para as três escalas.
- **Resultado Obtido:** Previsões geradas com sucesso, dentro dos parâmetros esperados.

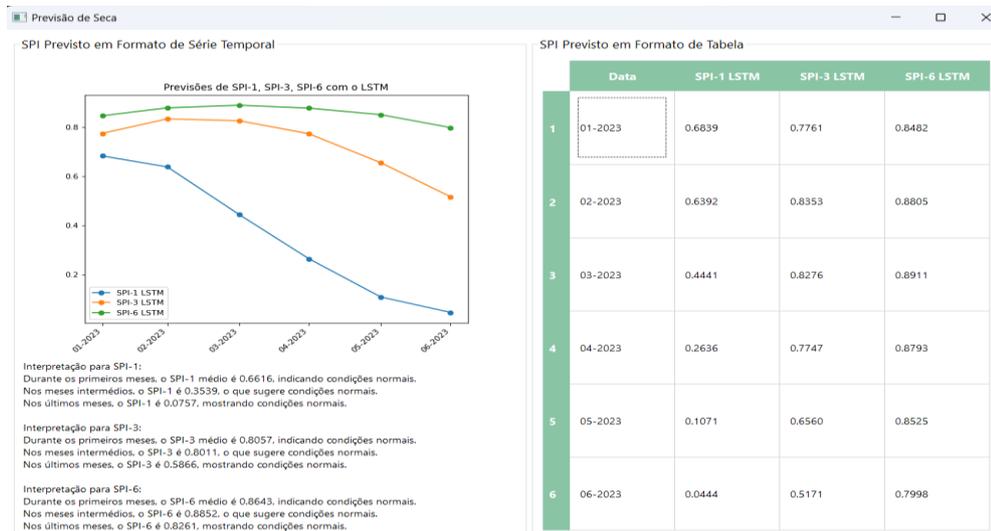


Figura 21. Resultado obtido do segundo cenário de teste.

5.5. Implantação do Sistema

O sistema deve ser implantado localmente, com instalação em cada desktop dos usuários finais. Este método evita o uso de servidores remotos ou nuvens, onde o sistema funciona em cada máquina, independentemente da conexão com a internet.

Processo de Implantação:

1. **Instalação Local:** cada desktop recebe uma instalação independente do aplicativo. Isso é realizado com um instalador que configura todas as dependências necessárias e verifica a compatibilidade do sistema com o ambiente de execução local.
2. **Sincronização Automática para Actualizações:** para aspectos de actualização do modelo e do aplicativo, foi implementado um sistema de sincronização automática que permite que sempre que houver uma actualização no modelo ou no *software*, as versões mais recentes sejam distribuídas automaticamente à todos os *desktops* conectados à rede interna, evitando a necessidade de actualizações manuais. A origem das versões mais recentes deve ser um servidor

central local (ou um dos computadores), especificamente configurado para armazenar e distribuir as actualizações. Este servidor actua como um repositório central onde pode-se carregar as novas versões do modelo e do software. A partir do servidor, essas actualizações serão distribuídas automaticamente à todos os *desktops que possuem o sistema e estão conectados à rede interna*.

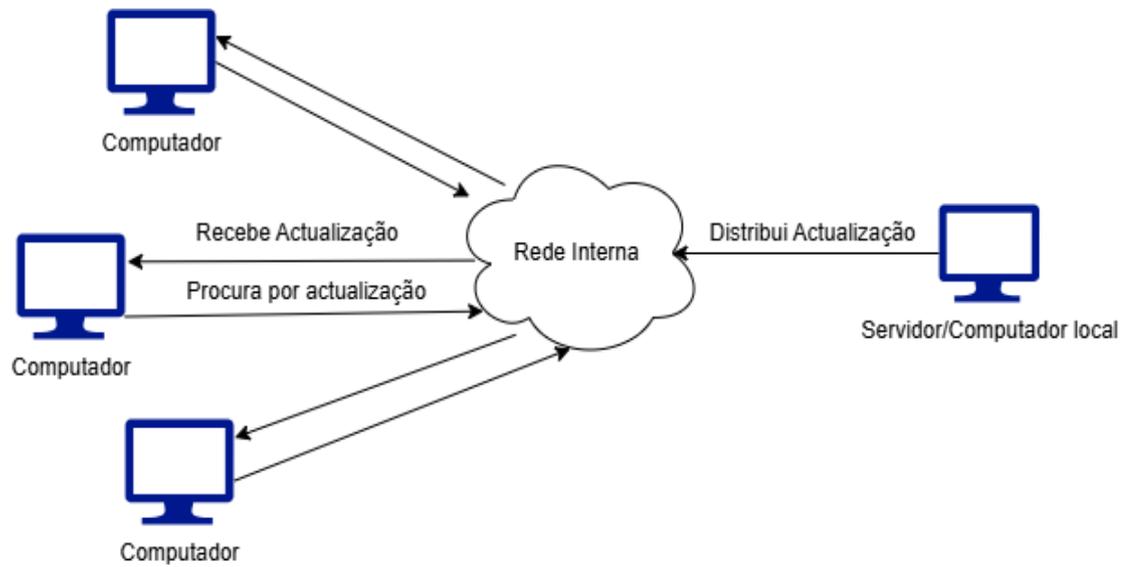


Figura 22. Diagrama ilustrando a sincronização de actualizações no sistema.

CAPÍTULO VI

6. Discussão de Resultados

Este capítulo revisa a literatura sobre os impactos das secas e os desafios da previsão em Moçambique. Em seguida, apresenta e discute a proposta de solução baseada nos modelos desenvolvidos, com destaque para a eficácia e limitações em termos de previsões.

6.1. Revisão de Literatura

A seca é um fenômeno natural que, apesar de sua frequência global e impactos significativos, continua sendo um dos eventos menos compreendidos devido à sua complexidade e à dificuldade de quantificação. Este fenômeno afecta diversos sectores, como a produção de energia, alimentos e abastecimento de água, resultando em grandes perdas socioeconômicas. O agravamento dos efeitos das secas é previsto devido às mudanças climáticas globais, que aumentam a frequência e a severidade desses eventos.

Em Moçambique, as secas têm um impacto significativo devido à vulnerabilidade regional a diferentes riscos naturais. A região do centro é mais propensa a inundações e ciclones, o sul enfrenta maior risco de secas, e o norte tem um clima mais moderado. As mudanças climáticas agravam esses riscos, com a seca sendo uma preocupação no sul, onde predomina um clima seco.

A gestão eficaz das secas em Moçambique enfrenta desafios significativos devido a limitações observadas na infraestrutura actualmente em uso e ainda à limitação de recursos para colecta e análise de dados. O Monitoramento e Avaliação (M&E) do país enfrenta problemas como a falta de harmonização entre planos sectoriais, metas ambiciosas demais e a confiabilidade dos dados. A integração de tecnologias avançadas, como o sensoriamento remoto e Sistemas de Informação Geográfica (SIG), surgiu como uma solução promissora para superar esses desafios.

O uso de sensoriamento remoto e SIG têm se mostrado essenciais para a avaliação de secas, especialmente em regiões com infraestrutura limitada, uma vez que permitem a colecta de dados em larga escala e quase em tempo real, proporcionando uma visão abrangente e actualizada das condições climáticas. No entanto, a eficácia dessa ferramenta é limitada pela infraestrutura obsoleta e a necessidade de modernização,

destacando a urgência de investimentos em tecnologia e capacitação para maximizar seu potencial.

O uso de *Machine Learning* e inteligência artificial para a previsão de secas apresenta uma abordagem inovadora e promissora. Essas tecnologias têm o potencial de melhorar significativamente a precisão das previsões e a capacidade de resposta a eventos climáticos. Porém, a adoção bem-sucedida de ML e IA enfrenta desafios relacionados à necessidade de grandes volumes de dados históricos, imagens de alta resolução e supercomputadores.

6.2. Proposta de Solução

A proposta de solução para a previsão de secas em Moçambique combina modelos avançados de LSTM com o Índice de Precipitação Padronizada (SPI) em escalas específicas, visando assegurar uma previsão eficiente e prática das condições de seca.

Foi notado que o modelo de LSTM tende a melhorar a precisão das previsões quanto maior for a escala de SPI a prever. Isto se deve ao facto de que, em escalas de tempo mais longas (como o SPI-3 e SPI-6), o modelo tem mais tempo para capturar padrões sazonais e tendências de longo prazo nos dados climáticos.

Por outro lado, a previsão de SPI em uma escala de 1 mês pode ser mais suscetível à ruídos e variações aleatórias nos dados, uma vez que os factores meteorológicos de curto prazo, como eventos isolados de chuva, podem influenciar fortemente a variabilidade dos resultados. Portanto, em escalas mais curtas, foi mais difícil para o LSTM capturar essas dinâmicas de curto prazo com a mesma precisão, o que explica o desempenho inferior nas previsões.

Adicionalmente, o uso dos modelos LSTM juntamente com dados climáticos adicionais possibilitaram que as previsões de SPI reflectissem as condições climáticas actuais e que pudessem ser ajustadas rapidamente com novas informações.

A flexibilidade dos modelos foi outro ponto que permitiu a adaptação fácil a novos dados. Além disso, a proposta buscou amenizar custos excessivos e esforços de manutenção dos sistemas. Os modelos escolhidos equilibraram precisão e eficiência computacional, uma vez que o LSTM também pode ser otimizado para lidar com grandes volumes de dados.

CAPÍTULO VII

7. Considerações Finais

Este Capítulo aborda as conclusões sobre o alcance dos objectivos do trabalho, os principais desafios enfrentados durante a pesquisa e recomendações para futuras melhorias e expansões do modelo de previsão de secas desenvolvido.

7.1. Conclusões

No concernente a busca de soluções para a problemática de limitações de eficiência relativamente a previsão de secas por parte do INAM, foi aplicada a metodologia de análise exploratória, documental e qualitativa, procedimento este que permitiu o alcance do objectivo geral da pesquisa, assente na apresentação de uma Proposta de modelo de previsão de secas para prevenção de impactos negativos por estas causados.

A clareza na definição dos objectivos específicos mostrou-se importante para o alcance do objectivo geral, na medida em que, estes serviram de guião para a limitação do âmbito e incidência na matéria relevante para o processo, que iniciou com a pesquisa documental e realização de entrevistas aos profissionais na área de climatologia no INAM, o que permitiu assim, filtrar a informação que diz respeito ao conceito de secas e à infraestrutura, bem como levantar os os procedimentos que são aplicados actualmente no monitoramento e previsão de secas.

Importa referir que durante o processo de recolha de informação para a pesquisa, mostraram-se como constrangimentos a superar, a escassez de dados não só na área de climatologia, com enfoque nas secas meteorológicas, mas também em outras áreas, observação esta, confirmada pelas respostas dos profissionais.

Esta situação, veio a reforçar a necessidade de buscar e apresentar uma proposta de mecanismos que pudessem ser úteis para o processos de previsão das secas em Moçambique, ou seja, identificar soluções que possam ajudar o INAM e outras entidades afins, a superar as limitações actuais, da forma mais conveniente e prática. Isto foi amplamente discutido no CAPITULO III do presente trabalho.

Foi nesta senda, que a construção de um modelo de aprendizagem de máquina voltado para a previsão de secas na região da bacia hidrográfica do rio Umbelúzi, em Maputo foi efectivada, através do uso de abordagens de Inteligência Artificial, tendo estas permitido a criação de um modelo LSTM, capaz de antecipar eventos de seca com base em dados

históricos climáticos, abrindo a possibilidade de se tornar numa ferramenta poderosa para auxiliar no planeamento e gestão de recursos hídricos da região.

O modelo de aprendizagem de máquina, referido acima, levou ao desenvolvimento de um protótipo de sistema para previsão de secas, proporcionando uma *interface* funcional para que instituições como o INAM, possam realizar as previsões de forma eficiente e tomar decisões preventivas e em tempo útil.

Para consolidar a contribuição deste estudo, é importante referir que as entrevistas realizadas demonstraram que a infraestrutura actual do local de estudo podem representar obstáculos na previsão das secas. O modelo LSTM proposto responde a esse desafio por ser flexível, adaptável à integração de novos dados e por não ser custoso em termos de recursos, numa fase inicial.

Assim, considera-se atingido o principal objectivo proposto neste trabalho, que vem reforçar a importância e relevância do aprimoramento das ferramentas de previsão de secas em Moçambique, no contexto das constantes e ameaçadoras mudanças climáticas a que o país está sujeito.

E por fim, espera-se que com este sistema possa ser possível contribuir para uma redução no tempo de resposta às crises hídricas, permitindo uma alocação mais eficiente dos recursos e uma mitigação dos impactos socioeconômicos da seca.

7.2. Constrangimentos

Durante a pesquisa representaram constrangimentos, as seguintes situações:

- Demora da resposta por parte do local de estudo (INAM), relativamente à requisição de dados;
- Disponibilização de dados requisitados ao INAM, sujeito a pagamento, cujo custo se evidenciava extremamente elevado; e
- Amostra de dados a fornecer que se evidenciaram inconsistentes e inadequados ao formato desejado para a pesquisa.

7.3. Recomendações

Este trabalho foi realizado com intuito de em jeito de proposta construir ou desenhar um Sistema para a previsão de secas com recurso à inteligência artificial, onde foi explorada

uma solução de implementação da tecnologia para o efeito. Contudo, existiram limitações que levam as seguintes recomendações:

- Que esta pesquisa não seja vista por si só como um produto acabado, podendo estar sujeita a um trabalho de aprofundamento da matéria no local de estudo, com o objectivo de enriquecimento do produto (sistema);
- Considerar o treinamento ou testagem do modelo orientado às outras áreas ou regiões a nível do país, para maior robustez do modelo;
- Considerar outros parâmetros contribuintes para a previsão de secas e aspectos que garantam maior interpretabilidade dos dados previstos pelo modelo no sistema; e
- Os modelos de inteligência artificial são de interpretação limitada, o que requer sempre conhecimento humano para melhor validação de resultados.

- Stefanos Fafalios, Pavlos Charonyktakis, Ioannis Tsamardinos. (Abril de 2020). Gradient Boosting Trees. 1-13.
- Suellen Teixeira Nobre Gonçalves, e. a. (2021). Índices e Metodologias de Monitoramento de Secas: Uma Revisão. Obtido de <https://www.scielo.br/j/rbmet/a/rGqGprMLXW5F66b6R98D98j/?lang=pt&format=pdf>
- Sundararajan, K. (2021). A Contemporary Review on Drought Modeling Using Machine Learning Approaches. Obtido em 2023, de <https://e-space.mmu.ac.uk/628387/1/TEch%20Science-%20Computer%20Modeling-%20Kathiravan.pdf>
- Tais de Moura Ariza Alpino, e. a. (2022). Os impactos das mudanças climáticas na Segurança Alimentar e Nutricional: uma revisão da literatura. Obtido de <https://www.scielo.br/j/csc/a/Rdr4LGpjWwGfmkgxMs6pLSL/?format=pdf&lang=pt>
- Ussalu, J. L., & Bassrei, A. (2021). CLIMATE DYNAMICS OF SOUTHERN REGION OF MOZAMBIQUE: STATISTICS AND FOURIER ANALYSIS. Obtido em 2023, de <https://revistas.ufpr.br/revistaabclima/article/download/75088/44938>
- Vaz, P. J. (2022). Hybrid Neural Network Based Models for Evapotranspiration Prediction Over Limited Weather Parameters. Obtido de <https://ieeexplore.ieee.org/document/10003192/metrics>
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 79-82. Obtido de Mendeley: <https://www.mendeley.com/catalogue/ce7e126c-75cd-383c-a833-867353f14a2b/#:~:text=RMSE%20varies%20with%20the%20variability%20within%20the%20distribution,of%20average%20error%2C%20and%20%28unlike%20RMSE%29%20is%20unambiguous.>
- World Economic Forum. (2023). *Secas mundiais podem afetar 75% da população até 2050, afirma Fórum Econômico Mundial*. Obtido de [exame.: https://exame.com/esg/secas-mundiais-podem-afetar-75-da-populacao-ate-2050-afirma-forum-economico-mundial/](https://exame.com/esg/secas-mundiais-podem-afetar-75-da-populacao-ate-2050-afirma-forum-economico-mundial/)
- Zakaria, M., AL-Shebany, M., & Sarhan, S. (2014). Artificial Neural Network : A Brief Overview. Obtido em 2024, de <https://d1wqtxts1xzle7.cloudfront.net/33278877/B42010712-libre.pdf?1395440006=&response-content-disposition=inline%3B+filename%3DB42010712.pdf&Expires=1705504944&Signature=Lh7jgXPUc8jVT6fGB2Ds0c05mHRDO~xV0uMG53TUWHN-j73z0TmJHrANLreo9F8vCEZnfGaB47zv5h3O66r>

Anexos

Anexo 1 - Guião de Entrevista dirigida à equipa de Monitoramento de Seca no INAM.

Qual é o papel do entrevistado no INAM?

Parte I: Descrição da seca em Moçambique

- a. Como se caracteriza a seca em Moçambique?
- b. Quais os impactos causados pela seca em Moçambique?

Parte II: Levantamento de Dados da Infraestrutura de Monitoramento com Profissionais

Sobre a infraestrutura actual:

- a. Quais padrões de medição de secas são actualmente utilizadas para o monitoramento de secas?
- b. Descreva brevemente os métodos empregados na colecta, processamento e análise de dados relacionados às condições de seca.
- c. Existe uma variedade de sistemas em uso na previsão de secas? Quais são eles?

Eficiência e Eficácia:

- a. Como você avalia a eficácia do sistema actual de monitoramento/previsão de secas?
- b. Existem áreas específicas em que o sistema mostra limitações?

Integração de Tecnologias:

- a. Há planos para actualização ou expansão do sistema existente?

Parte III: Entrevistas aos Profissionais

Desafios Percebidos:

- a. Quais são os principais desafios que você percebe no monitoramento/previsão de secas?

- b. Como avaliaria a capacidade do INAM de prever secas com base nos dados disponíveis?
- c. Como avaliaria a capacidade atual do INAM de integrar tecnologias de inteligência artificial em seus processos de análise de dados para previsão de secas?
- d. Existem outros obstáculos específicos que afectam a eficácia do seu trabalho?

Expectativas Futuras:

Que outras melhorias ou tecnologias você espera ver no monitoramento/previsão de secas no INAM nos próximos anos?

Anexo 1.1 - Resultados da Entrevista 1

Esta entrevista foi feita ao Analista de Secas, com vista a perceber como funciona o mecanismo de monitoramento e previsão de secas e quais são os constrangimentos durante o processo.

Parte I: Descrição da seca em Moçambique

a. Como se caracteriza a seca em Moçambique?

R: A seca é um fenómeno climático que não tem controle. É um fenómeno muito vasto, que pode ser medido de várias formas.

b. Quais os impactos causados pela seca em Moçambique?

R: Sendo a seca de difícil controle, o impacto que causa no país também é severo. A seca impacta em vários sectores, desde agricultura, hidrologia e até em termos de finanças. Algumas áreas do país já sofrem com *déficit* de chuva, o que afecta a produção agrícola.

Parte II: Levantamento de Dados da Infraestrutura de Monitoramento de Secas com Profissionais

Sobre a infraestrutura actual:

a. Quais padrões de medição de secas são actualmente utilizadas para o monitoramento de secas?

R: Actualmente, fazemos o monitoramento de secas meteorológicas usando o SPI. Usamos um modelo estatístico, um algoritmo, que é corrido através de um script em R. Com isto, é possível fazer o processamento de dados para monitorar o índice de precipitação no país.

b. Descreva brevemente os métodos empregados na colecta de dados relacionados às condições de seca.

R: Usamos dados que são fornecidos por outras entidades, à nível global ou à nível do SADC, ou recolhemos dados das estações que verificam o nível de precipitação em cada região, especificamente dados relacionados com a monitoria de secas meteorológicas.

c. Existe uma variedade de sistemas em uso? Quais são eles?

R: Não são usados vários sistemas. Os dados recolhidos são alimentados ao script ou modelo estatístico que faz a geração de novos dados, e os resultados são depois lidos em um dashboard, onde podemos ver a categorização da seca em cores, em seca leve, moderada ou severa, em torno do mapa do país.

Eficiência e Eficácia:

a. Como você avalia a eficácia do sistema actual de monitoramento/previsão de secas?

R: Não é muito eficiente. Apesar de ser um mecanismo que usamos, possui algumas limitações.

a. Existem áreas específicas em que o sistema mostra limitações?

R: Sim, o modelo é muito lento e não é flexível, justamente devido ao tempo que leva para poder fazer o processamento. Costuma levar 1 a 2 dias, onde alimentamos os dados e o modelo fica a ser executado durante esse período de tempo para podermos ter os resultados.

Integração de Tecnologias:

a. Há planos para actualização ou expansão do sistema existente?

R: Sim, actualmente estamos a expandir os nossos métodos de monitoramento incluindo a previsão de secas também. É um processo que está em fase de teste, por isso, só realizamos uma previsão até agora, porque é um procedimento que não é facil de implementar e é demorado.

Parte III: Entrevistas aos Profissionais

Desafios Percebidos:

a. Quais são os principais desafios que você percebe no monitoramento de secas?

R: O monitoramento de secas em sí, é um processo que hoje-em-dia mostra-se obsoleto para a nossa realidade. Houve necessidade de se melhorar e podermos prever as secas, porque o que é implementado actualmente é a monitoria de secas através do controle de dados de 3 á 6 meses, e caso nós identificássemos um índice baixo de precipitação ou seca severa nesses dados, assumia-se que estamos em

período de seca. Este resultado não era tão preciso. Precisávamos achar um método para prever as secas.

a. Como avaliaria a capacidade do INAM de prever secas com base nos dados disponíveis?

R: Os dados existem, embora poderiam ser melhores.

b. Como avaliaria a capacidade atual do INAM de integrar tecnologias de inteligência artificial em seus processos de análise de dados para previsão de secas?

R: Estado actual, acho que seria custoso, devido aos recursos. Mas seria de grande benefício para automatizar os nossos processos.

c. Existem outros obstáculos específicos que afectam a eficácia do seu trabalho?

R: A eficácia e performance do modelo é um dos obstáculos que me vem em mente.

Expectativas Futuras:

a. Que outras melhorias ou tecnologias você espera ver no monitoramento/previsão de secas no INAM nos próximos anos?

R: Qualquer método que facilitasse o nosso trabalho seria bem-vindo. Inclusive a implementação de inteligência artificial se fosse aplicável.

Anexo 1.2 - Resultados da Entrevista 2

Esta entrevista foi feita ao Analista de secas, com vista a perceber como funciona o mecanismo de monitoramento e previsão de secas e quais são os constrangimentos durante o processo.

Parte I: Descrição da seca em Moçambique

a. Como se caracteriza a seca em Moçambique?

R: A seca é um fenómeno que se dá por variados motivos, daí que é classificado em vários tipos e é monitorado de vários meios, o mesmo se aplica para o caso de Moçambique.

b. Quais os impactos causados pela seca em Moçambique?

R: A Seca é um fenómeno complexo e de progressão lenta, o que existe em Moçambique não é fácil de combater. Ela influencia vários sectores, principalmente a agricultura, que conseqüentemente impacta outras áreas.

Parte II: Levantamento de Dados da Infraestrutura de Monitoramento de Secas com Profissionais

Sobre a infraestrutura actual:

a. Quais padrões de medição de secas são actualmente utilizadas para o monitoramento de secas?

R: Nós fazemos a categorização de secas usando o SPI, sendo que é o índice que depende da precipitação. Mas existem outros índices que permitem o monitoramento de secas. Usamos o SPI para fazer o monitoramento de secas, incluindo a previsão, a diferença é que se queremos fazer a monitoria sem previsão, usamos dados de precipitação observada, e se queremos fazer a monitoria através da previsão, usamos dados de precipitação previstos, é um processo mais complexo

b. Descreva brevemente os métodos empregados na colecta de dados relacionados às condições de seca.

R: Recolhemos os dados de satellite e dados observados de precipitação para poder fazer o monitoramento. Outro tipos de dados como humidade do solo, ainda são

díficeis de recolher devido aos recursos que estão em falta, nos permitindo por enquanto prever as secas meteorológicas.

c. Existe uma variedade de sistemas em uso? Quais são eles?

R: Não. Usamos um modelo estatístico.

Eficiência e Eficácia:

a. Como você avalia a eficácia do sistema actual de monitoramento/previsão de secas?

R: São procedimentos demorados.

b. Existem áreas específicas em que o sistema mostra limitações?

R: O facto do modelo em si não ser muito flexível é uma limitação.

Integração de Tecnologias:

b. Há planos para actualização ou expansão do sistema existente?

R: A previsão de secas é algo novo, então estamos no processo de avaliação também para afinar o próprio modelo. Este foi o primeiro ano em que conseguimos realizar uma previsão, e ao final do ano vamos ver até que ponto foi o nosso grau de acerto.

Parte III: Entrevistas aos Profissionais

Desafios Percebidos:

a. Quais são os principais desafios que você percebe no monitoramento/previsão de secas?

R: A limitação é de dados, existem vários índices, mas enfrentamos escassez de dados, por isso usamos somente o SPI.

b. Como avaliaria a capacidade do INAM de prever secas com base nos dados disponíveis?

R: A escassez de dados limita o uso de certos índices, porque há limitação de recursos para colectar esses dados. O que fazemos é colectar dados relacionados com a seca meteorológica somente, porque não há recursos para colecta de outros parâmetros relacionados com outros tipos de secas.

c. Como avaliaria a capacidade actual do INAM de integrar tecnologias de inteligência artificial em seus processos de análise de dados para previsão de secas?

R: Acho que o que ajudaria seria fazer o uso de inteligência artificial para melhorar a colecta de outros parâmetros que melhorariam a precisão de determinação da seca.

d. Existem obstáculos específicos que afectam a eficácia do seu trabalho?

R: A seca em si é um fenómeno bastante complexo, porque é de progressão lenta, então trata-se de algo que não é fácil de determinar.

Expectativas Futuras:

a. Que outras melhorias ou tecnologias você espera ver no monitoramento/previsão de secas no INAM nos próximos anos?

R: Realizar actividades para poder monitorar a secas de vários outros meios, com o uso de vários outros parâmetros.

Anexo 1.3 - Resultados da Entrevista 3

Esta entrevista foi feita ao Climatologista e Chefe de departamento de pesquisa, com vista a perceber como funciona o mecanismo de monitoramento e previsão de secas e quais são os constrangimentos durante o processo.

Parte I: Descrição da seca em Moçambique

a. Como se caracteriza a seca em Moçambique?

R: A seca climatológica é causada por déficit prolongado de precipitação em uma determinada região, em comparação com as condições climáticas normais. No caso de Moçambique, temos verificado áreas que sofrem desse déficit.

b. Quais os impactos causados pela seca em Moçambique?

R: A seca influencia em muitos aspectos no país, tanto que é crucial que os resultados que recolhemos sejam reportados para outras entidades, com vista a realizarem uma tomada de decisão adiantada para a situação climatológica que está por vir, como por exemplo, no que diz respeito à agricultura e colheita sazonal.

Parte II: Levantamento de Dados da Infraestrutura de Monitoramento com Profissionais

Sobre a infraestrutura actual:

a. Quais padrões de medição de secas são actualmente utilizadas para o monitoramento de secas?

R: O padrão de medição de secas que usamos é o SPI, calcula o déficit de precipitação em relação à média.

b. Descreva brevemente os métodos empregados na colecta, processamento e análise de dados relacionados às condições de seca.

R: Os dados do INAM são integrados com outros para garantir maior precisão. São usados sistemas europeus e de satélite CHIRPS para colecta de dados.

c. Existe uma variedade de sistemas em uso na previsão de secas? Quais são eles?

R: Utiliza-se um modelo estatístico para prever secas, que calcula o SPI com base em dados.

Eficiência e Eficácia:

a. Como você avalia a eficácia do sistema actual de monitoramento/previsão de secas?

R: Os métodos usados são eficientes, embora haja dependência da experiência humana para supervisão de resultados.

b. Existem áreas específicas em que o sistema mostra limitações?

R: Demora na realização de actividades e necessidade de intervenção humana e experiência são desafios. Há também limitações em certas áreas relacionadas com procedimentos antecedentes à previsão de secas em si.

Integração de Tecnologias:

a. Há planos para actualização ou expansão do sistema existente?

R: Neste momento, estamos em fase piloto de um mecanismo para previsão de secas, para além do monitoramento que temos feito.

Parte III: Entrevistas aos Profissionais

Desafios Percebidos:

b. Quais são os principais desafios que você percebe no monitoramento/previsão de secas?

R: A necessidade de intervenção humana e experiência para garantir acurácia nos dados que são gerados como resultados.

c. Como avaliaria a capacidade do INAM de prever secas com base nos dados disponíveis?

R: A capacidade de prever secas não é influenciada pelos dados, pois há procedimentos para colecta e processamento integrados.

d. Como avaliaria a capacidade actual do INAM de integrar tecnologias de inteligência artificial em seus processos de análise de dados para previsão de secas?

R: É possível e seria interessante para alguns procedimentos.

e. Existem outros obstáculos específicos que afectam a eficácia do seu trabalho?

R: Tirando o facto de se necessitar de experiência e expertise humana, não.

Expectativas Futuras:

b. Que outras melhorias ou tecnologias você espera ver no monitoramento de secas no INAM nos próximos anos?

R: A capacidade de integração de meios mais robustos para processamento de dados antes da análise.

Anexo 2 - Fluxo de Desenvolvimento do Modelo de Proposto.

Este fluxo representa os passos que são observados para o desenvolvimento do modelo proposto como solução do problema. Este fluxo mostra de forma breve como um modelo genérico é desenvolvido, incluindo o LSTM.

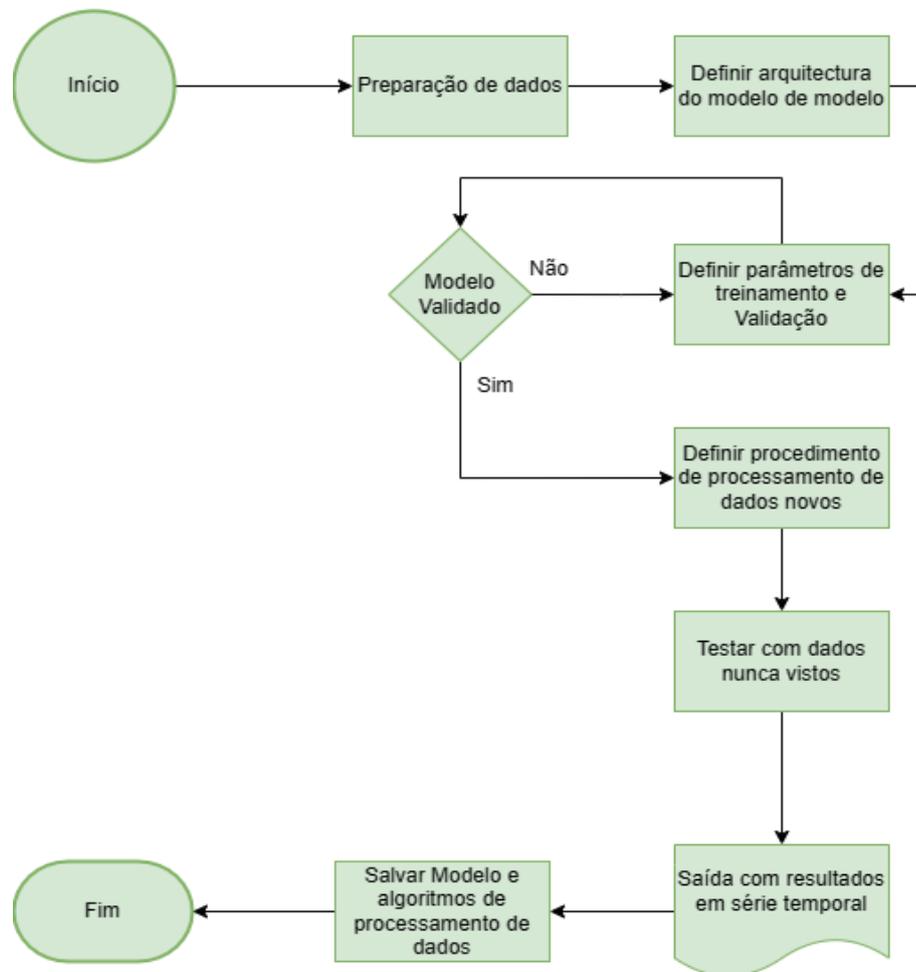


Figura A2- 1. Fluxo de desenvolvimento do modelo.

Anexo 3 - Descrição dos Dados de Históricos Climáticos

Foram obtidos dados históricos climáticos mensais dos anos de 1981 à 2022, com foco na região da bacia hidrográfica do rio Umbelúzi, na província de Maputo. Os dados provêm de projecto Climate Adaptation Science Investigators Phase 2 (CASI2) da NASA, formando em totalidade, 504 registos históricos. Cada registo histórico foi composto pelos seguintes parâmetros:

Tabela A3- 1. Descrição dos dados climáticos.

Parâmetro	Descrição	Unidade
PS	Pressão Superficial	kPa
TS	Temperatura da Superfície Terrestre	°C
T2M	Temperatura a 2 Metros	°C
QV2M	Humidade Específica a 2 Metros	g/kg
RH2M	Humidade Relativa a 2 Metros	%
WD10M	Direção do Vento a 10 Metros	Graus
WS10M	Velocidade do Vento a 10 Metros	m/s
T2M_MAX	Temperatura Máxima a 2 Metros	°C
T2M_MIN	Temperatura Mínima a 2 Metros	°C
CLOUD_AMT	Quantidade de Nuvens	%
PRECTOTCORR	Precipitação Corrigida	mm/dia
ALLSKY_SFC_SW_DWN	Irradiância de Onda Curta Superficial com Céu Claro	kW-h/m ² /dia

Anexo 4 – Arquitectura e Hyperparâmetros do Modelo Proposto

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 6, 256)	275,456
dropout (Dropout)	(None, 6, 256)	0
lstm_1 (LSTM)	(None, 6, 128)	197,120
dropout_1 (Dropout)	(None, 6, 128)	0
lstm_2 (LSTM)	(None, 6, 64)	49,408
dropout_2 (Dropout)	(None, 6, 64)	0
lstm_3 (LSTM)	(None, 6, 32)	12,416
lstm_4 (LSTM)	(None, 16)	3,136
dense (Dense)	(None, 1)	17

Total params: 537,553 (2.05 MB)

Trainable params: 537,553 (2.05 MB)

Non-trainable params: 0 (0.00 B)

Figura A4- 1. Arquitectura e hyperparâmetros dos modelos LSTM. Fonte: Jupyter Notebook.

Anexo 5 – Resultados do Processo de Treinamento do Modelo de Proposto

- Modelo LSTM para previsão de SPI em escala de 1.

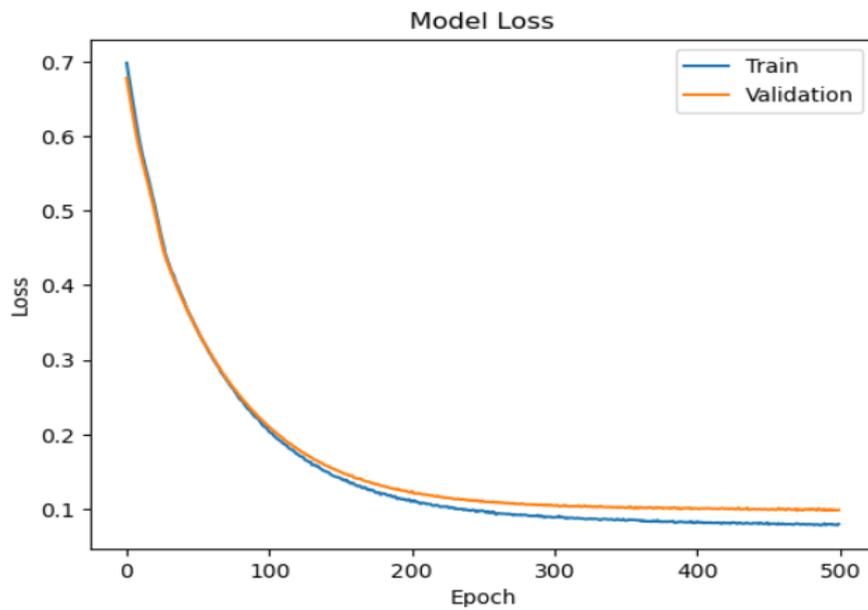


Figura A5- 1. Gráfico de perda no processo de treinamento do modelo LSTM para previsão de SPI na escala de 1. Fonte: Jupyter Notebook.

- Modelo LSTM para previsão de SPI em escala de 3.

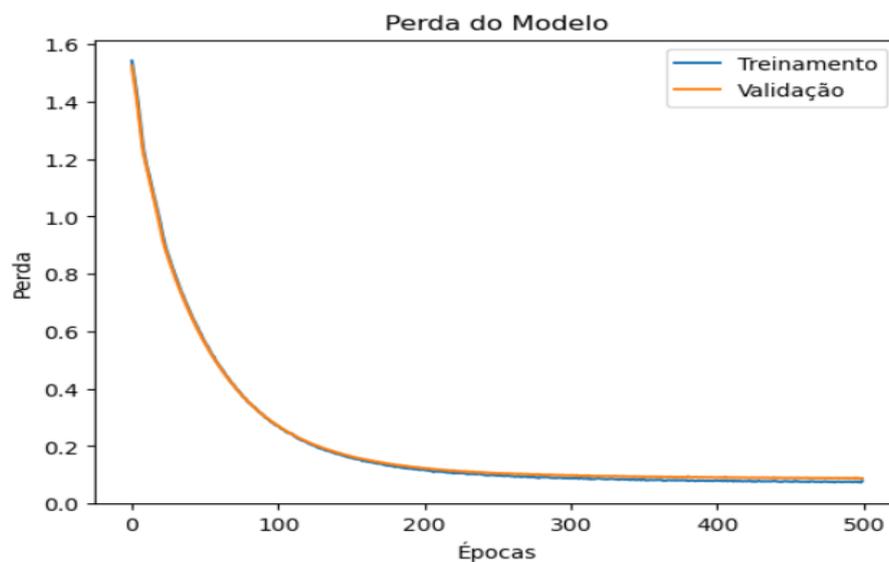


Figura A5- 2. Gráfico de perda no processo de treinamento do modelo LSTM para previsão de SPI na escala de 3. Fonte: Jupyter Notebook.

- Modelo LSTM para previsão de SPI em escala de 6.

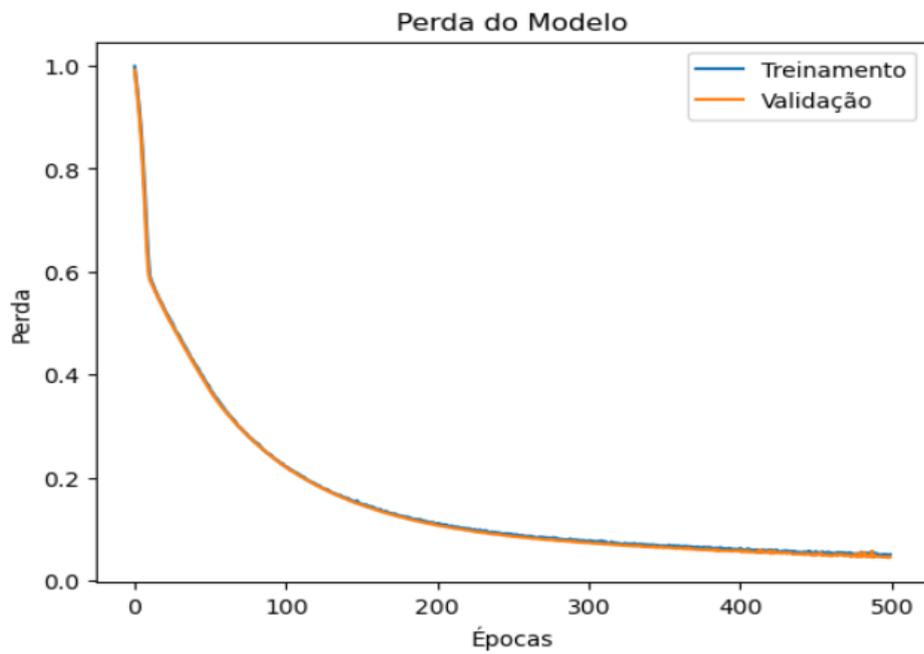


Figura A5- 3. Gráfico de perda no processo de treinamento do modelo LSTM para previsão de SPI na escala de 6. Fonte: Jupyter Notebook.

Anexo 6 – Resultados do Processo de Teste do Modelo de Proposto

- Modelo LSTM para previsão de SPI em escala de 1.

```
# Evaluate the best model on the test set
best_model = tf.keras.models.load_model('best_model_weights_spi_1.keras')
test_loss = best_model.evaluate(test_x, test_y)
print("Test Loss:", test_loss)
```

```
4/4 ————— 1s 10ms/step - loss: 0.0838
Test Loss: 0.0883130356669426
```

```
from sklearn.metrics import mean_absolute_error, mean_squared_error

# Predict temperatures using the trained model
predictions = best_model.predict(test_x)

# Calculate evaluation metrics
mae = mean_absolute_error(test_y, predictions)
mse = mean_squared_error(test_y, predictions)
rmse = np.sqrt(mse)

print("Mean Absolute Error (MAE):", mae)
print("Mean Squared Error (MSE):", mse)
```

```
4/4 ————— 0s 7ms/step
Mean Absolute Error (MAE): 0.2177651652629388
Mean Squared Error (MSE): 0.07875030765358482
```

Figura A6 - 1. Resultado das métricas de desempenho do modelo LSTM em escala de 1.

- Modelo LSTM para previsão de SPI em escala de 3.

```
# Evaluate the best model on the test set
best_model = tf.keras.models.load_model('best_model_weights_spi_3.keras')
test_loss = best_model.evaluate(test_x, test_y)
print("Test Loss:", test_loss)
```

```
4/4 ————— 1s 7ms/step - loss: 0.0827
Test Loss: 0.08920799940824509
```

```

from sklearn.metrics import mean_absolute_error, mean_squared_error

# Predict spi using the trained model
predictions = best_model.predict(test_x)

# Calculate evaluation metrics
mae = mean_absolute_error(test_y, predictions)
mse = mean_squared_error(test_y, predictions)
rmse = np.sqrt(mse)

print("Mean Absolute Error (MAE):", mae)
print("Mean Squared Error (MSE):", mse)

4/4 ————— 0s 7ms/step
Mean Absolute Error (MAE): 0.2143891387569289
Mean Squared Error (MSE): 0.07313757006683742

```

Figura A6 - 2. Resultado das métricas de desempenho do modelo LSTM em escala de 3.

- Modelo LSTM para previsão de SPI em escala de 6.

```

# Evaluate the best model on the test set
best_model = tf.keras.models.load_model('best_model_weights_spi_6.keras')
test_loss = best_model.evaluate(test_x, test_y)
print("Test Loss:", test_loss)

4/4 ————— 2s 9ms/step - loss: 0.0536
Test Loss: 0.051765184849500656

```

```

from sklearn.metrics import mean_absolute_error, mean_squared_error

# Predict spi using the trained model
best_model = tf.keras.models.load_model('best_model_weights_spi_6.keras')
predictions = best_model.predict(test_x)

# Calculate evaluation metrics
mae = mean_absolute_error(test_y, predictions)
mse = mean_squared_error(test_y, predictions)
rmse = np.sqrt(mse)

print("Mean Absolute Error (MAE):", mae)
print("Mean Squared Error (MSE):", mse)
print("Root Mean Squared Error (RMSE):", rmse)

4/4 ————— 2s 334ms/step
Mean Absolute Error (MAE): 0.1063671469141673
Mean Squared Error (MSE): 0.028881920025028655

```

Figura A6 - 3. Resultado das métricas de desempenho do modelo LSTM em escala de 6.

Anexo 7 - Especificação dos Casos de Uso

A seguir são apresentados os fluxos detalhados dos requisitos funcionais importantes para o funcionamento do sistema e que carecem de esclarecimento.

Tabela A7 - 1. Especificação de caso de uso: Autenticação de Usuário.

Nome	Autenticação de Usuário
Descrição	Processo de inserção de credenciais de usuário para poder aceder às funcionalidades do sistema.
Actor	Analista de secas
Prioridade	Importante
Pré-condição/Entrada	Sistema iniciado.
Pós-condição/Saída	Acesso à página de chegada do sistema.
Fluxo Principal	
Acção do Actor	Resposta do Sistema
1. Iniciar o sistema	
	2. Tela de Autenticação de usuário aberta.
3. Credencias do usuário inseridos	
	4. O sistema reconhece o usuário e aprova a sua autenticação
	5. O sistema abre a Tela de chegada com as funcionalidades do sistema
Fluxo Alternativo	
3. Credenciais de usuário não inseridos/reconhecidos e o bloqueando a entrada no sistema.	
5. Erro no processamento de dados, não permitindo a sua entrada no sistema.	

Tabela A7 - 2. Especificação de caso de uso: carregar dados meteorológicos.

Nome	Carregar dados meteorológicos
Descrição	Processamento e carregamento dos dados recolhidos em 6 meses para prever o SPI de 6 meses no software.
Actor	Analista de secas
Prioridade	Essencial
Pré-condição/Entrada	Sistema iniciado e modelo carregado.
Pós-condição/Saída	Dados processados e prontos para entrar ao modelo
Fluxo Principal	
Acção do Actor	Resposta do Sistema
1. Escolher o modelo LSTM a usar para previsão	
2. Escolher o ficheiro que contém os dados meteorológicos	
	3. Processar os dados e preparar para entrar ao modelo
	4. Notificação de dados carregados com êxito
Fluxo Alternativo	
1. Não escolher dados no formato correcto e não avançar com o procedimento	
2. Erro no processamento de dados, não permitindo a sua entrada no modelo.	

Tabela A7 - 3. Especificação de caso de uso: Realizar previsão.

Nome	Realizar previsão
Descrição	O software deve ser capaz de realizar uma previsão do SPI do próximo mês, com base nos dados carregados.
Actor	Analista de secas
Prioridade	Essencial
Pré-condição/Entrada	Dados carregados ao sistema.
Pós-condição/Saída	Previsão de SPI de 6 meses executada.
Fluxo Principal	
Acção do Actor	Resposta do Sistema
1.Clicar a opção para efectuar a previsão do SPI.	
	2. Introduzir os dados no modelo e realizar a previsão com o modelo.
	3. Realizar previsão dos dados inseridos
	4. Formatar os dados previstos para série temporal
Fluxo Alternativo	
2. O modelo não conseguir realizar a previsão por algum erro e encerrar o procedimento.	
3. Resultados que não podem ser convertidos ao formato de séries, encerrando o procedimento.	

Tabela A7 - 4. Especificação de caso de uso: Visualizar previsão.

Nome	Visualizar previsão
Descrição	Processo de visualização da previsão do SPI
Actor	Analista de secas
Prioridade	Essencial
Pré-condição/Entrada	Previsão de dados pelo e formatação para série temporal executada com êxito.
Pós-condição/Saída	Visualização de previsão do SPI de um período de 6 meses.
Fluxo Principal	
Acção do Actor	Resposta do Sistema
	1. Converter os dados de previsão para o gráfico
	2. Trazer os gráficos para a tela do usuário.
3. Visualização dos dados de previsão em gráficos de série temporal	
Fluxo Alternativo	
	1. Erro de conversão para o gráfico, cancelando o procedimento.

Tabela A7 - 5. Especificação de caso de uso: Exportar última previsão.

Nome	Exportar última previsão
Descrição	O usuário deve poder exportar o resultado da última previsão para um arquivo.
Actor	Analista de secas
Prioridade	Importante

Pré-condição/Entrada	Previsão visualizada na tela.
Pós-condição/Saída	Relatório da previsão criado.
Fluxo Principal	
Acção do Actor	Resposta do Sistema
1. Clicar a opção para poder exportar o resultado obtido.	
	2. Exportar os dados para um relatório.
	3. Transferência do relatório ao sistema do actor.
Fluxo Alternativo	
2. Erro de exportação, não podendo criar o relatório.	

Tabela A7 - 6. Especificação de caso de uso: Exportar últimas previsões.

Nome	Exportar últimas previsões
Descrição	O usuário deve poder exportar o resultado das últimas 3 previsões para um arquivo.
Actor	Analista de secas
Prioridade	Importante
Pré-condição/Entrada	Pelo menos 3 previsões realizadas no sistema.
Pós-condição/Saída	Relatório de previsões criado.
Fluxo Principal	
Acção do Actor	Resposta do Sistema

1. Clicar a opção para poder exportar as últimas três previsões realizadas.	
	2. Exportar os dados para um relatório.
	3. Transferência do relatório ao sistema do actor.
Fluxo Alternativo	
2. Erro de exportação, não podendo criar o relatório.	

Anexo 8 - Diagrama de seqüências

Estes diagramas de seqüências descrevem como e em qual ordem os objectos presentes numa iteração trabalham em conjunto para alcançar um determinado objectivo. Os diagramas de seqüência a seguir são: Autenticação do Usuário, Carregar dados, Realizar previsão, Exportar última previsão, Exportar últimas previsões.

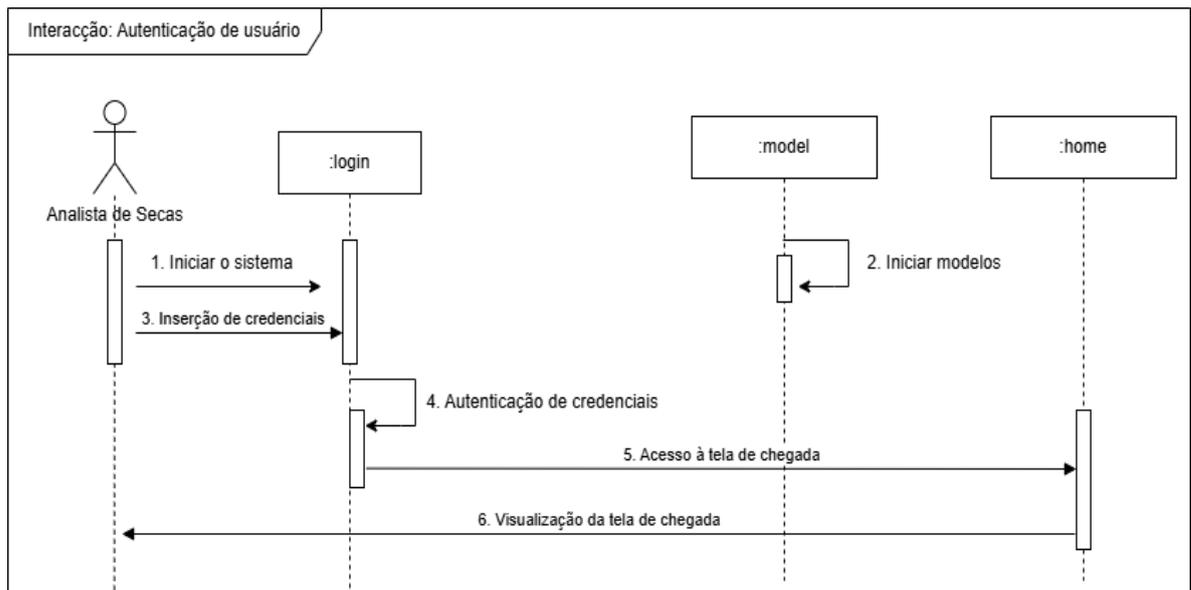


Figura A8- 1. Diagrama de seqüência: Autenticação de usuário.

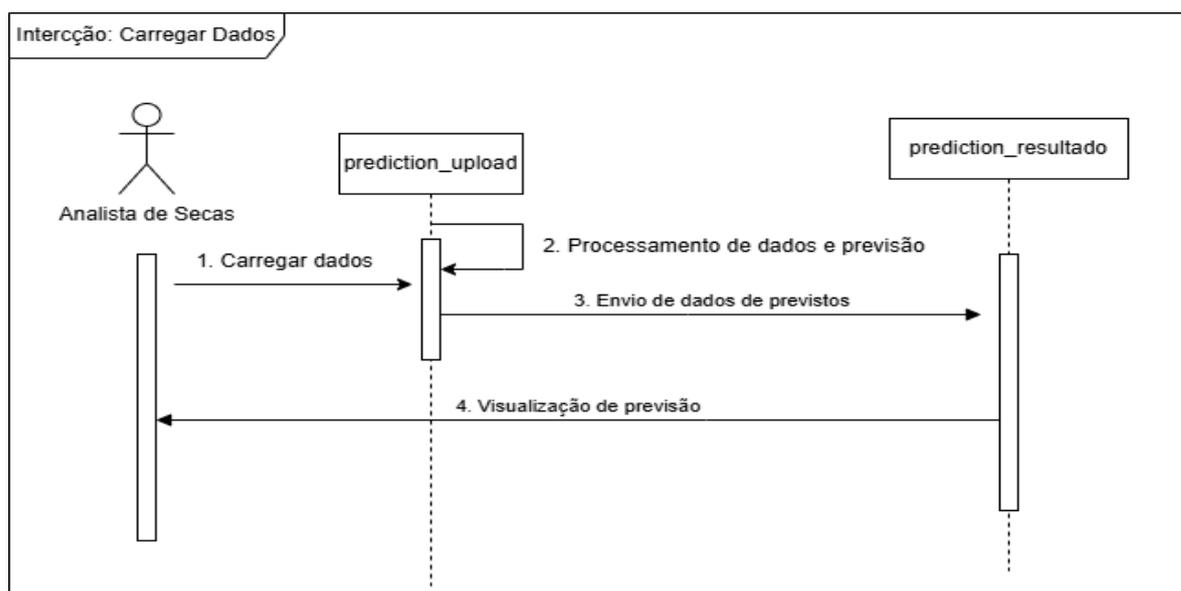


Figura A8- 2. Diagrama de seqüência: Carregar dados.

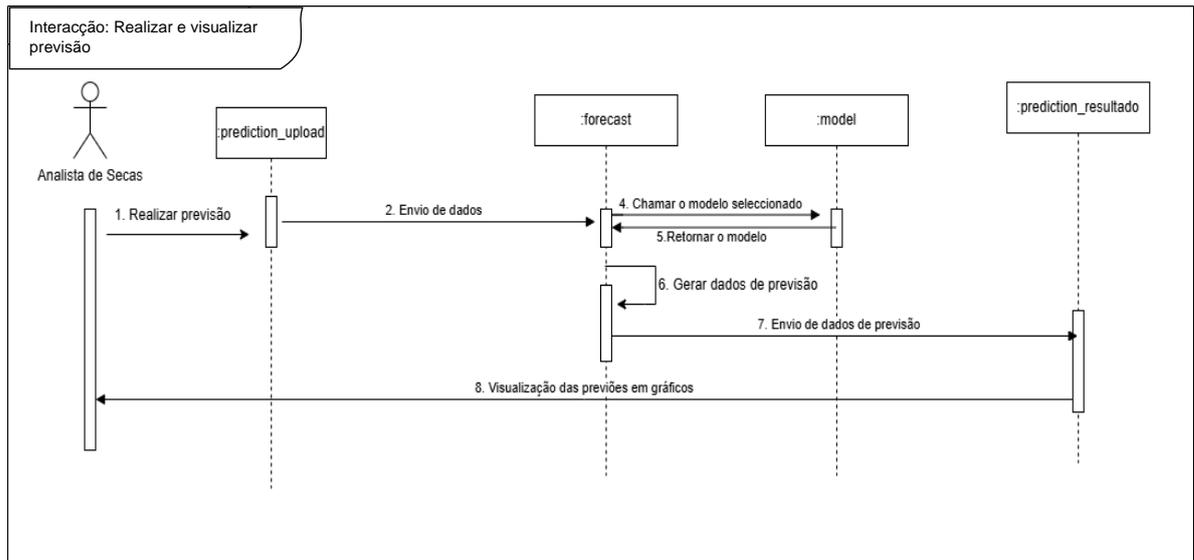


Figura A8- 3. Diagrama de seqüência: Realizar e visualizar previsão.

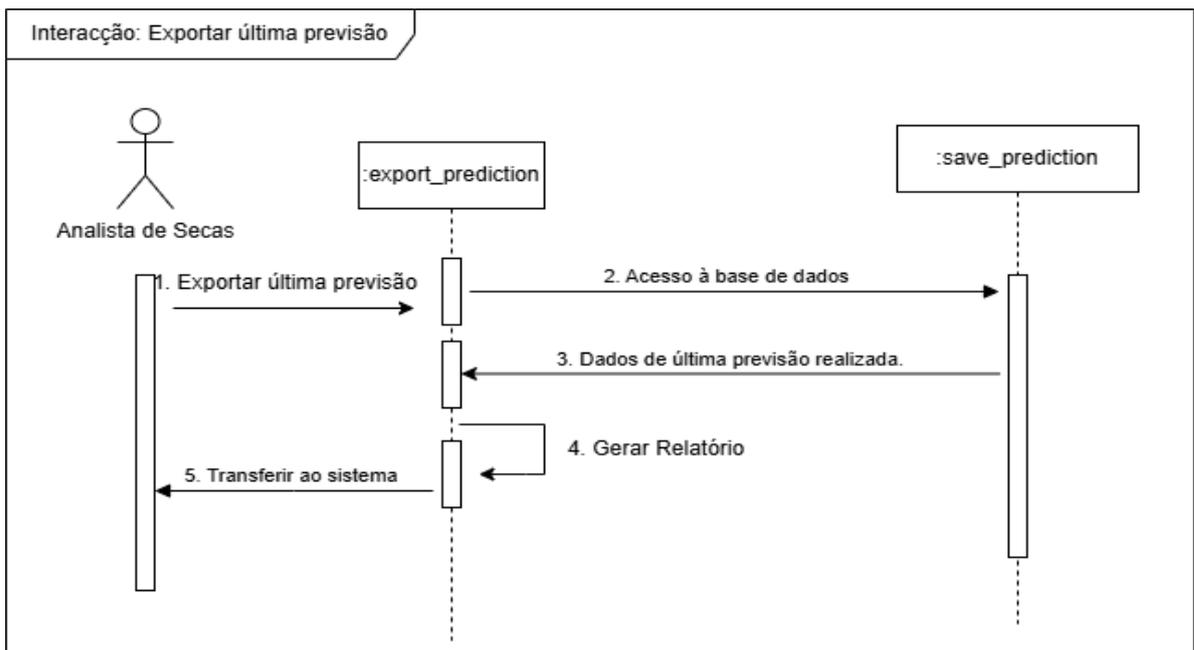


Figura A8- 4. Diagrama de seqüência: Exportar resultado.

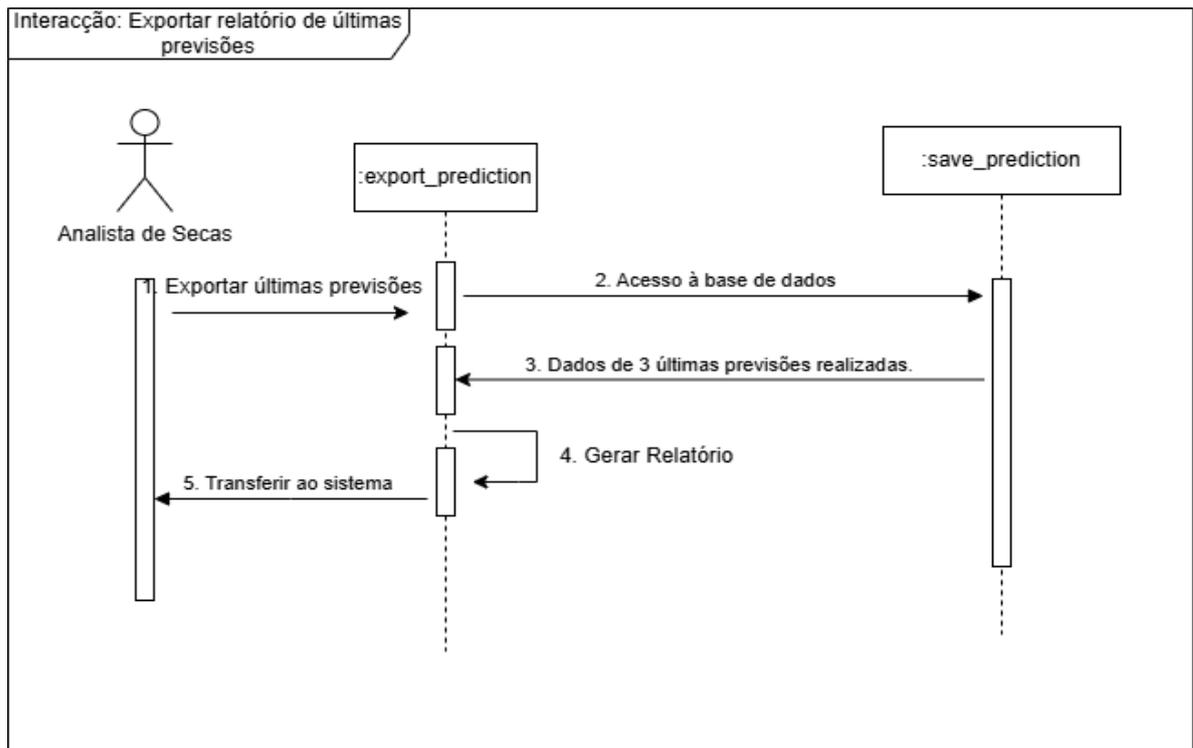


Figura A8- 5. Diagrama de sequência: Exportar relatório de últimas previsões.

Anexo 9 - Diagrama de Classes

O diagrama de classes a seguir apresenta uma visão geral do sistema proposto para a previsão de secas no INAM.

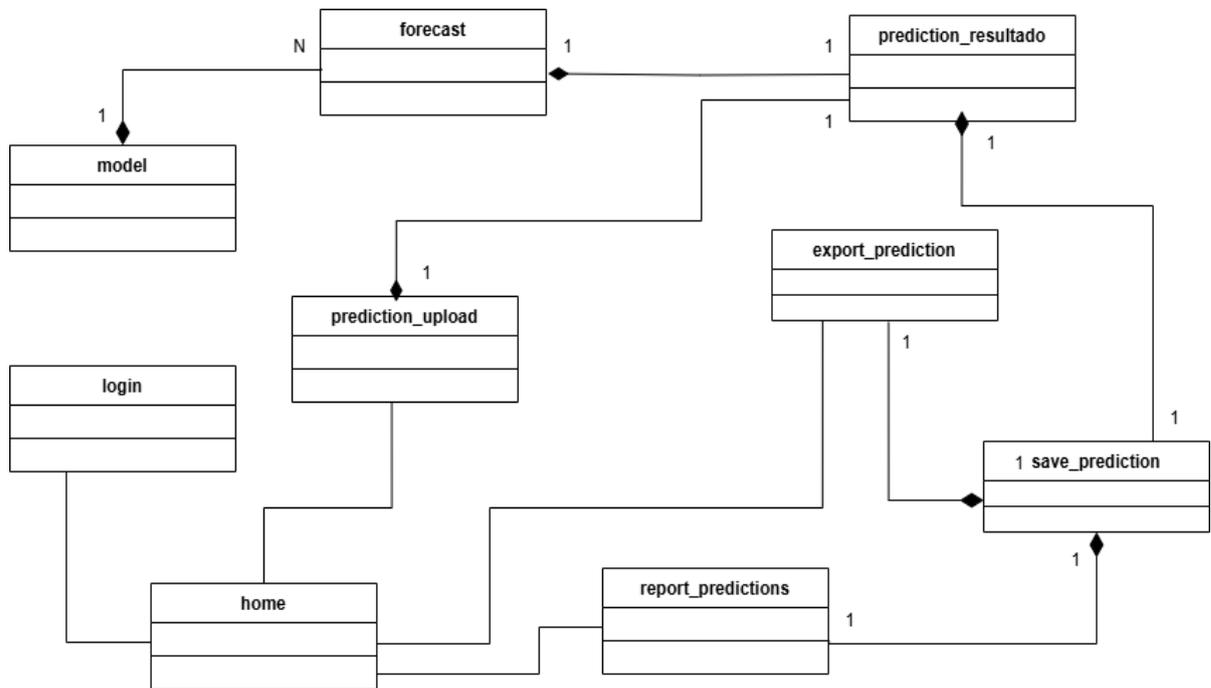


Figura A9 - 1. Diagrama de Classes da proposta de solução.