Global Temperature Prediction using Machine Learning & Deep Learning

A technical exploration into forecasting future climate patterns with advanced predictive analytics.

Submitted by: Jordan Manoj

Presentation Roadmap

This presentation outlines our approach to building a robust climate forecasting system using both classical and deep learning techniques.

Overview

Defining the scope and importance of the project.

Problem Statement

Identifying the complexity of non-linear climate prediction.

Dataset & Preprocessing

Source, attributes, and critical data preparation steps.

Methodology & Models

Implementation of ML and LSTM architectures.

Model Comparison & Results

Evaluating performance metrics across all models.

Conclusion & Outcomes

Key takeaways and the path forward for climate analytics.

Project Overview: Predicting the Future Climate

> Climate change is arguably the biggest challenge of our time. Accurate predictive modeling is essential for informed policy and mitigation strategies.

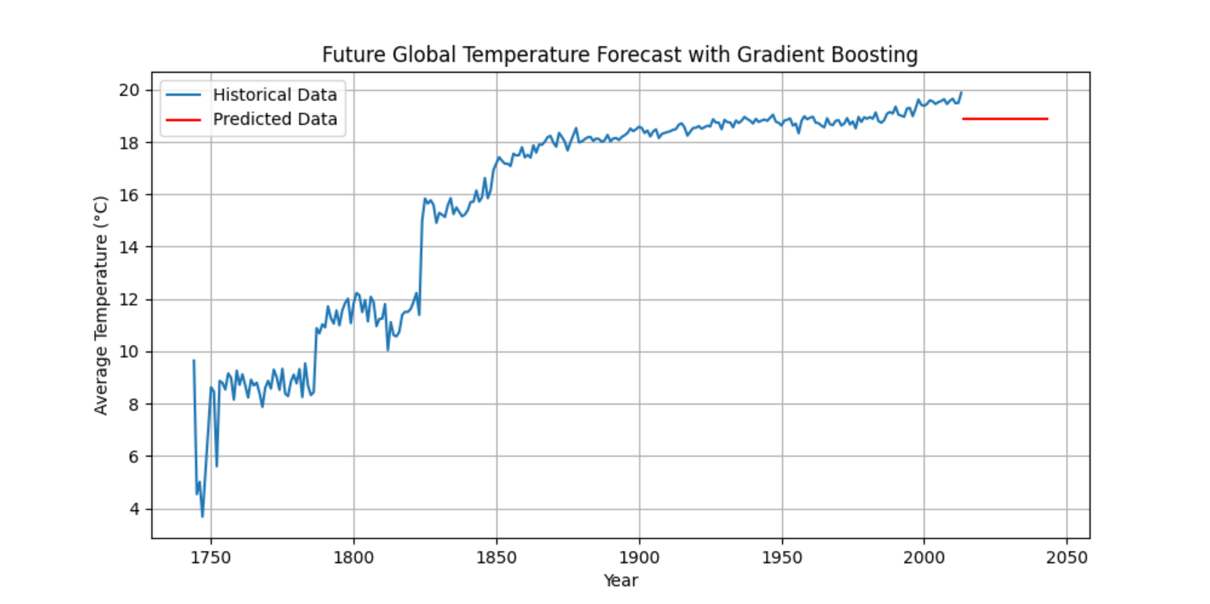
This project aims to analyze complex historical global temperature data to accurately forecast future temperature trends, providing a clear picture of the long-term warming trajectory.

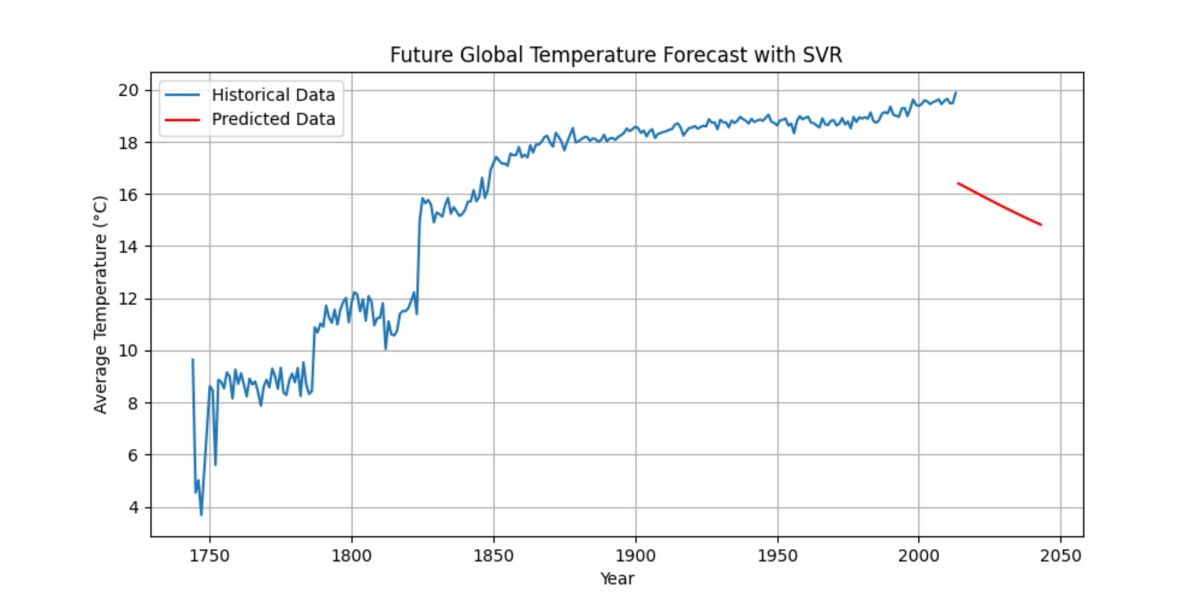
Linear Regression

Random Forest Regressor

Support Vector Regressor (SVR)

LSTM Neural Network



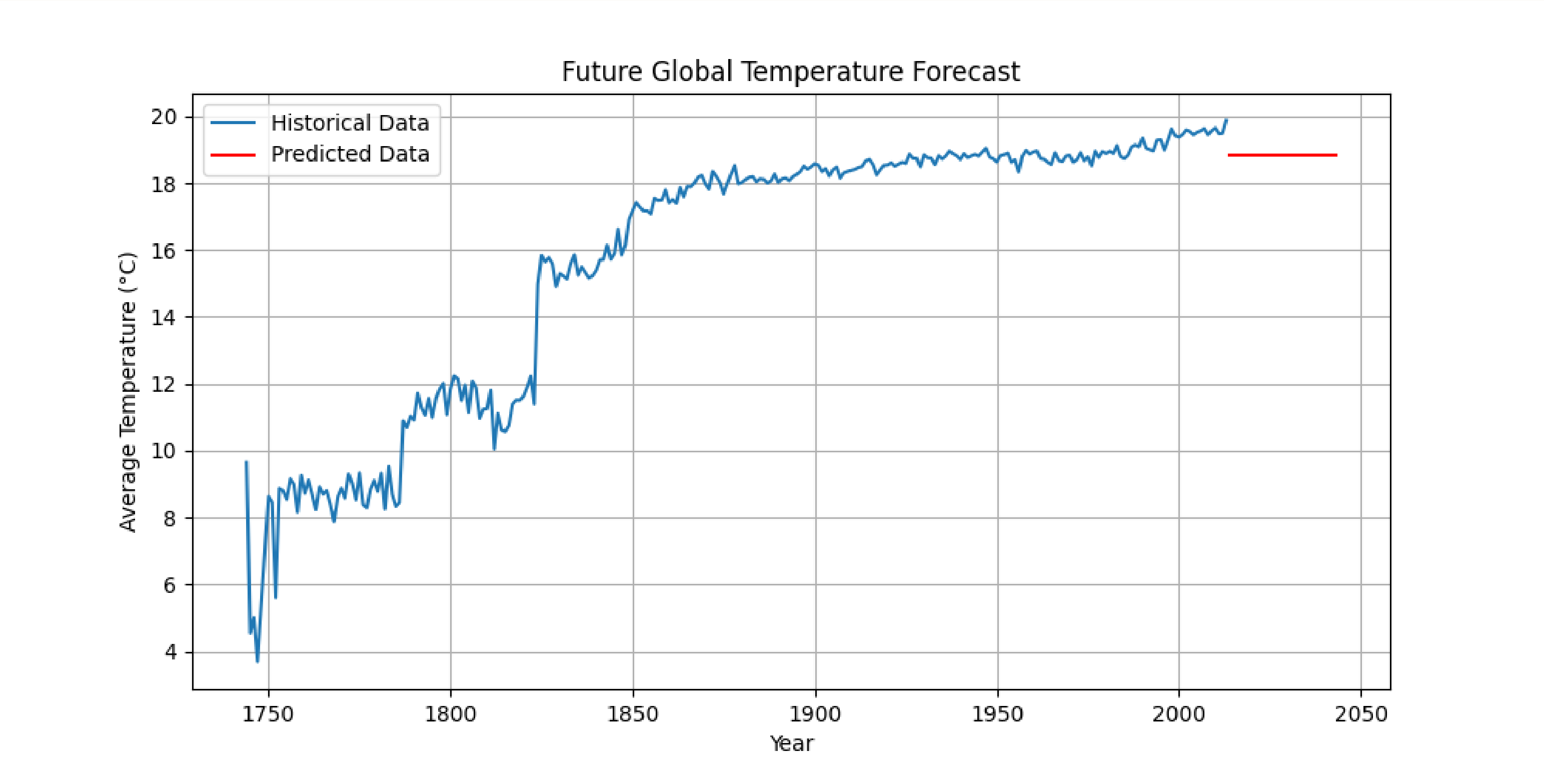


The Challenge: Modeling Non-Linear Climate Behavior

Complexities in Temperature Forecasting

While historical data clearly indicates a long-term warming trend driven by greenhouse gases, predicting precise future patterns is complicated by several technical factors:

- The sheer volume and size of global climate data.
- Pronounced seasonal and short-term temperature fluctuations.
- The inherently non-linear and chaotic nature of climate systems.



Primary Objective

The main goal is to develop a highly **accurate and robust predictive model** using ML and DL that can overcome these non-linear challenges to forecast future global temperature trends.

Dataset and Preparation for Climate Modeling

We utilized the publicly available `GlobalLandTemperaturesByCountry.csv` dataset, which provides a rich historical record necessary for training complex sequential models.

Key Attributes

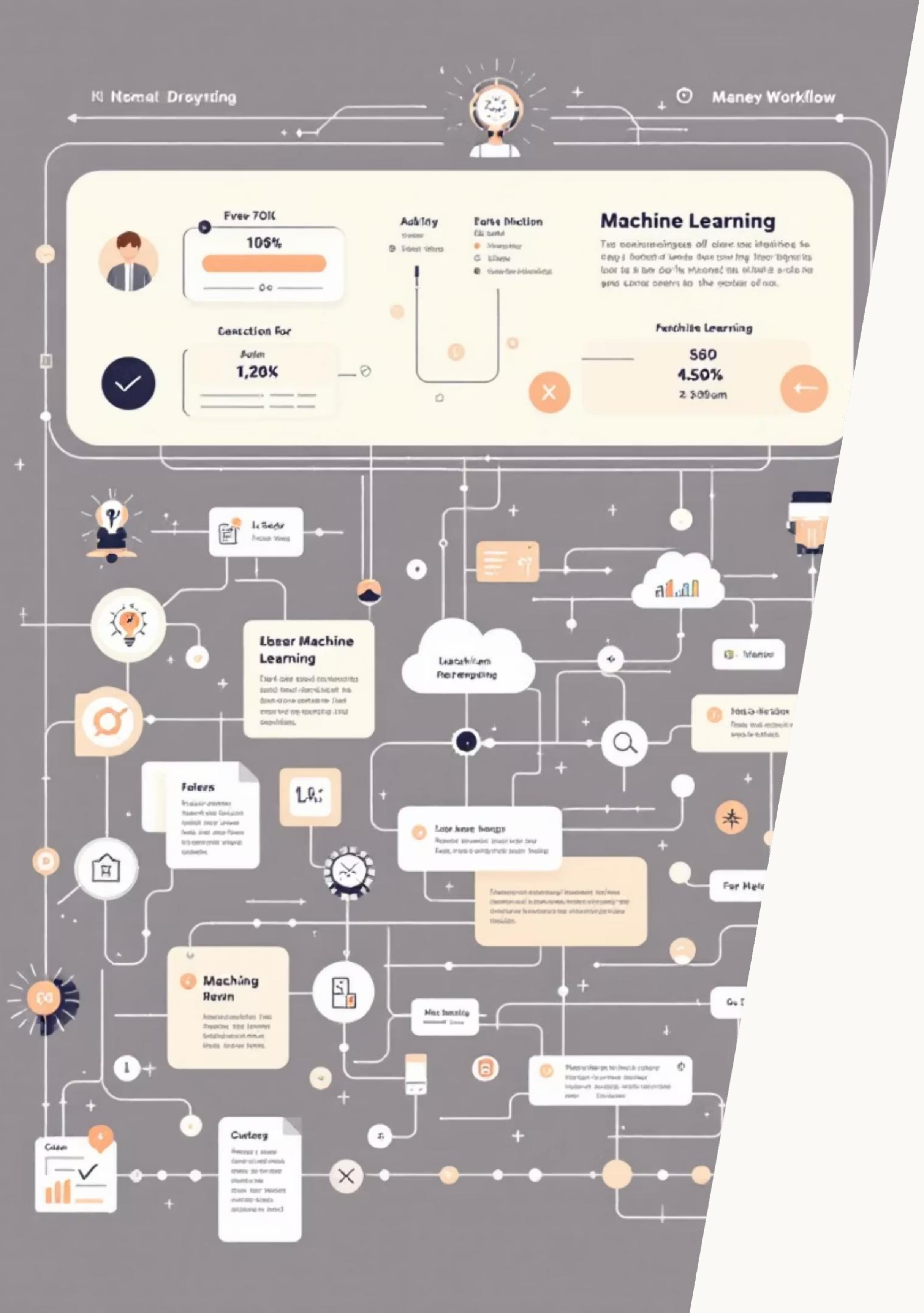
dt: Date of recording; **AverageTemperature:** The recorded mean temperature; **Country:** Geographic identifier.

Data Processing

- Linear interpolation to handle missing values.
- Extraction and conversion of date fields (Year, Month).

Aggregation & Cleaning

Data was aggregated to stable monthly and yearly averages to smooth noise, followed by z-score based outlier removal.



Machine Learning Methodology: Classical Approaches

Our initial methodology focused on rigorous data preparation to ensure model stability, followed by the implementation of three standard regression techniques.

Preprocessing Essentials

- Handling Missing Values (Interpolation)
- Outlier Identification & Removal
- Feature Scaling (`MinMaxScaler`)

Feature Engineering

- Rolling Mean & Standard Deviation (for trend/volatility)
- Cyclical Encoding of Months (Sin/Cos transformation)

Models Implemented

Linear Regression (Baseline), Random Forest Regressor (Ensemble), and Support Vector Regressor (SVR).

Deep Learning: Sequential Forecasting with LSTM

To capture the temporal dependencies inherent in climate data, we employed a Long Short-Term Memory (LSTM) network a powerful recurrent neural network architecture.

Model Architecture

A sequential model incorporating 128 LSTM units, specifically designed to process the time-series nature of temperature data.

Training Configuration

Input sequences of 24 months, utilizing the `Adam` optimizer and Mean Squared Error (MSE) as the loss function.

Preventing Overfitting

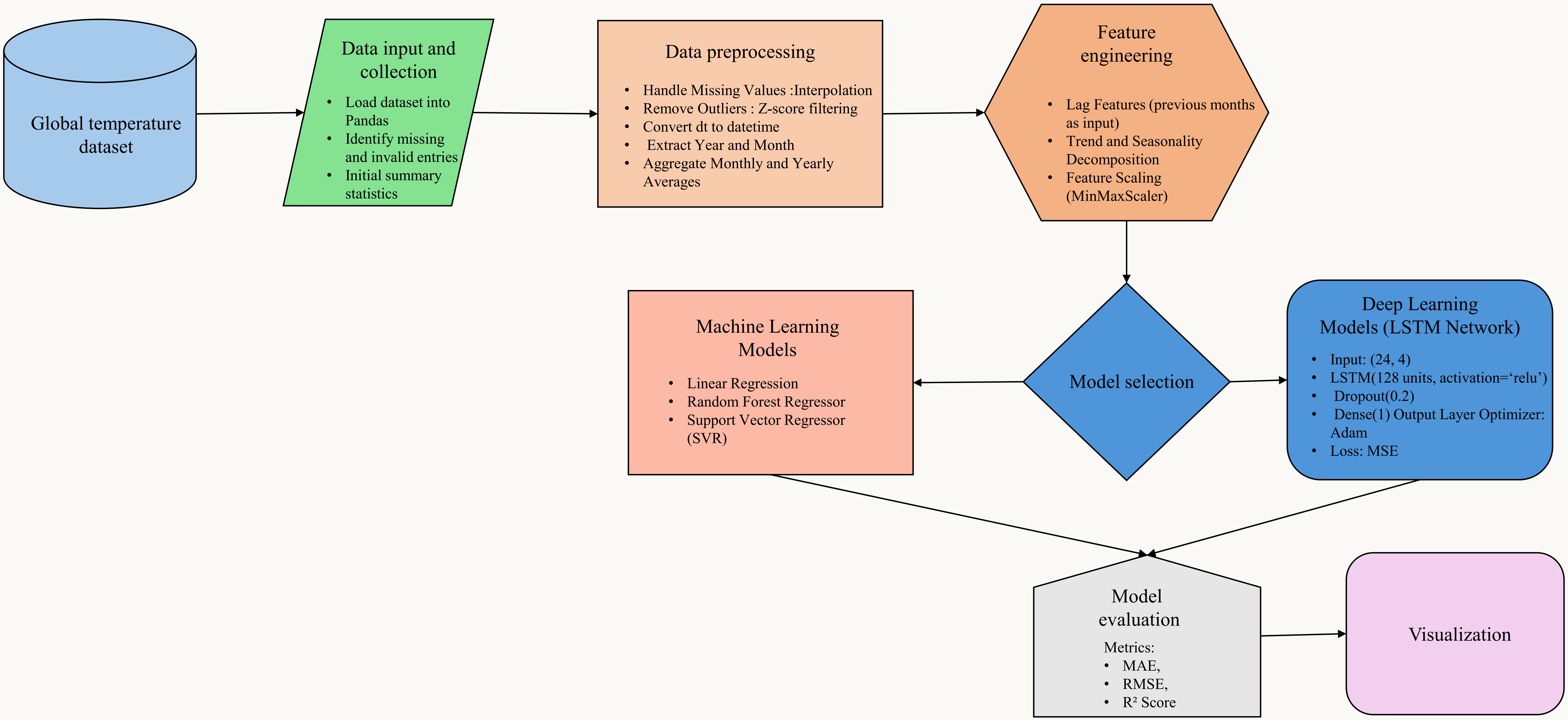
A `Dropout` layer was included to enhance model generalization and robustness across different time periods.

Model evaluation metrics used

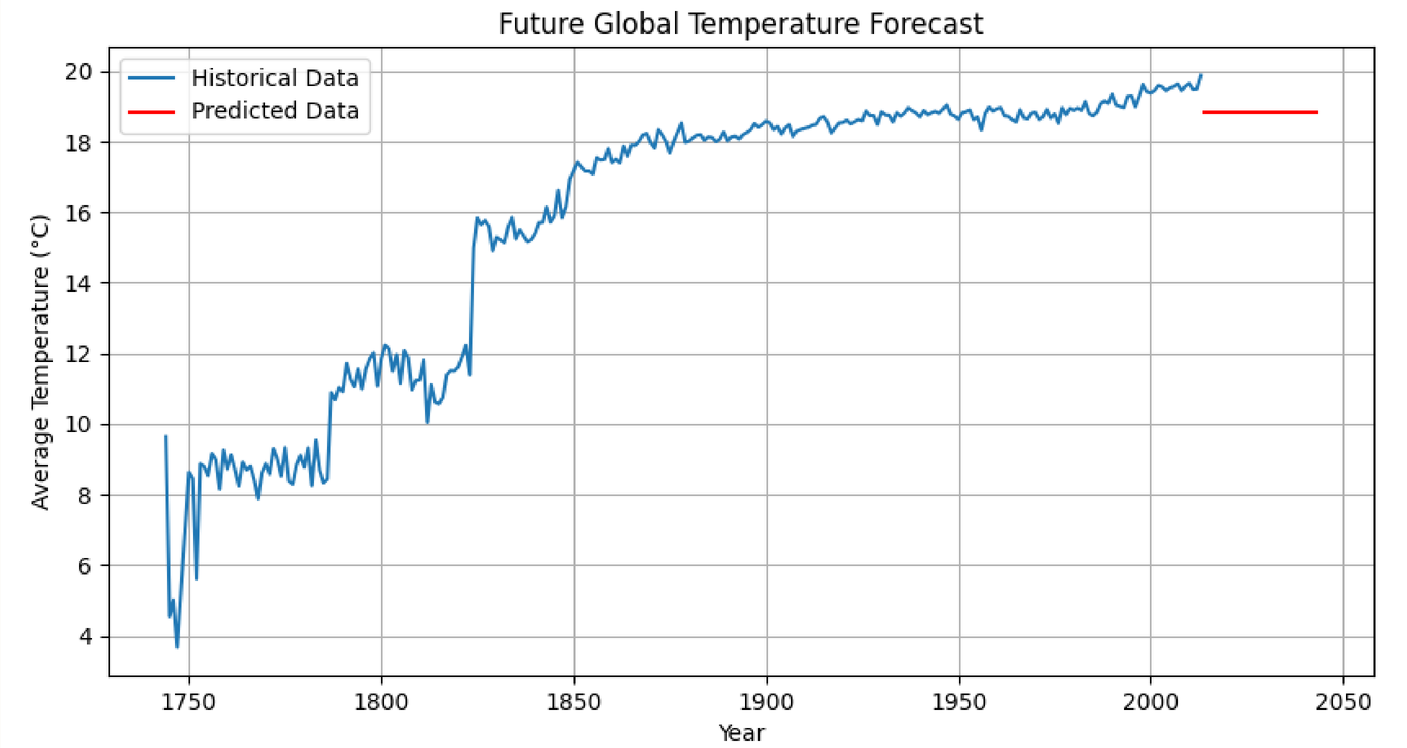
- MASE: Measures the average size of the errors in the predictions without considering whether they're positive or negative.
- RMSE: Measures the square root of the average squared difference between predicted and actual values.
- R^2 Score: R^2 tells you how well your model's predictions match the real data , it shows how much of the actual temperature changes your model can explain.

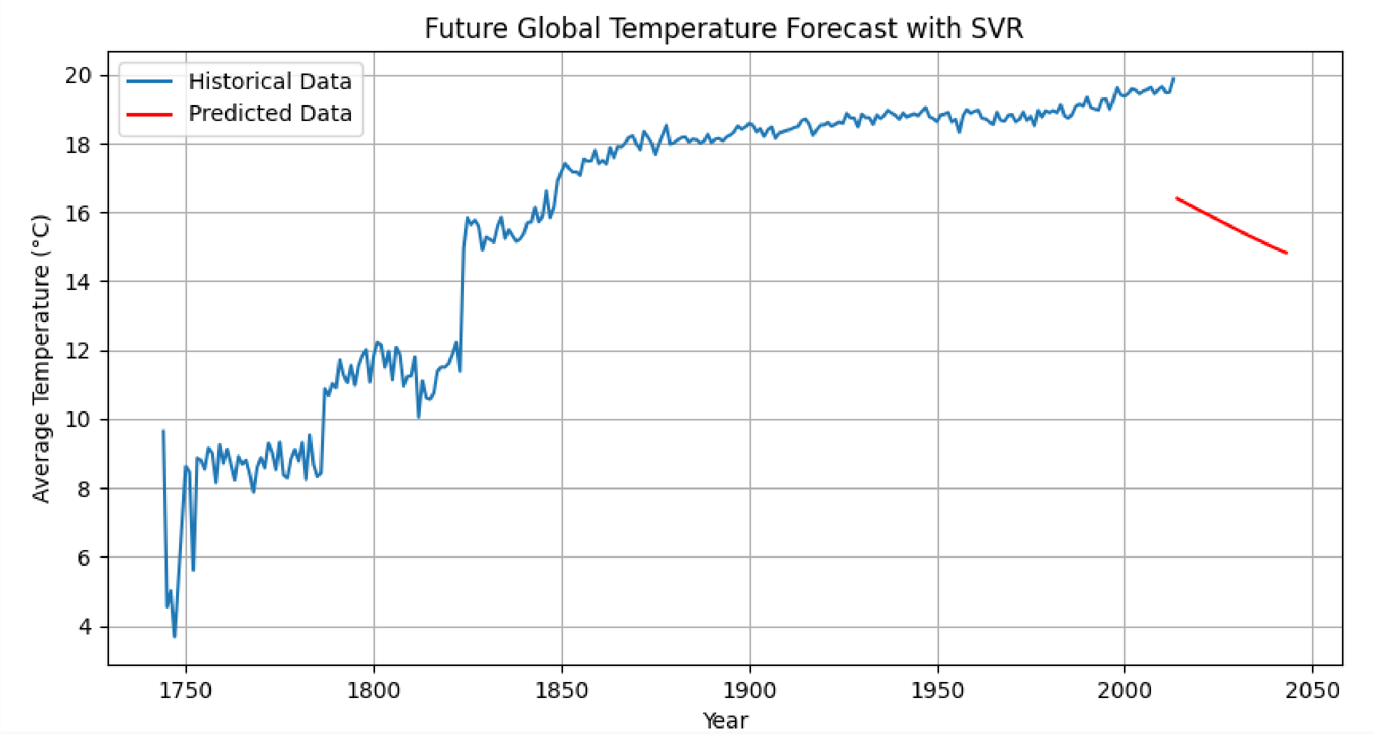
Note:
$R^2 = 1$ (Perfect fit)
$R^2 = 0$ (No predictive power)
$R^2 < 0$ (Worse than random guess)

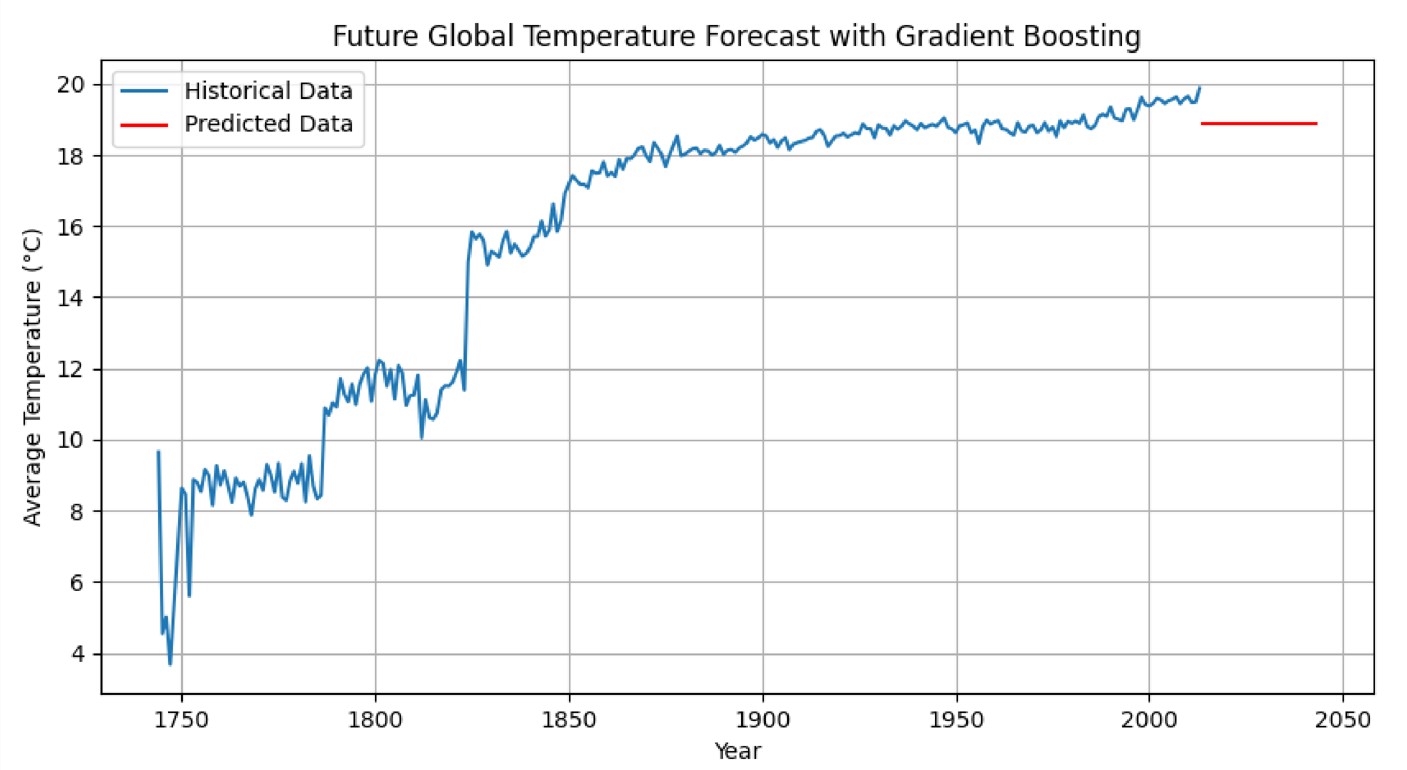
Flowchart

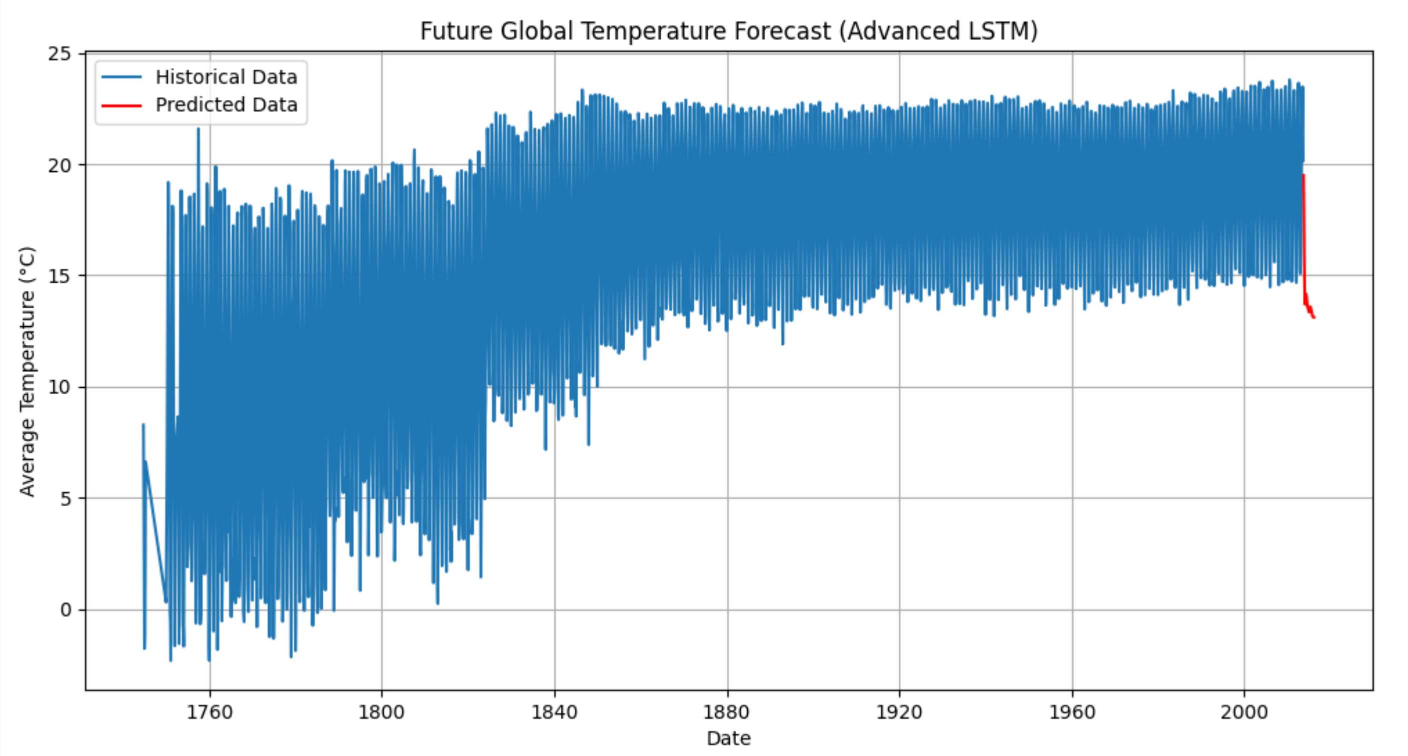

Global temperature dataset
Data input and collection
• Load dataset into Pandas
• Identify missing and invalid entries
• Initial summary statistics
Data preprocessing
• Handle Missing Values :Interpolation
• Remove Outliers : Z-score filtering
• Convert dt to datetime
• Extract Year and Month
• Aggregate Monthly and Yearly Averages
Feature engineering
• Lag Features (previous months as input)
• Trend and Seasonality Decomposition
• Feature Scaling (MinMaxScaler)
Machine Learning Models
• Linear Regression
• Random Forest Regressor
• Support Vector Regressor (SVR)
Model selection
Deep Learning Models (LSTM Network)
• Input: (24, 4)
• LSTM(128 units, activation='relu')
• Dropout(0.2)
• Dense(1) Output Layer Optimizer: Adam
• Loss: MSE
Model evaluation
Metrics:
• MAE,
• RMSE,
• R² Score
Visualization

Visualization









Model Comparison: LSTM Achieves Superior Performance

The performance metrics clearly demonstrate the advantage of using a Deep Learning approach (LSTM) for capturing the complex, sequential relationships in global temperature time series data.

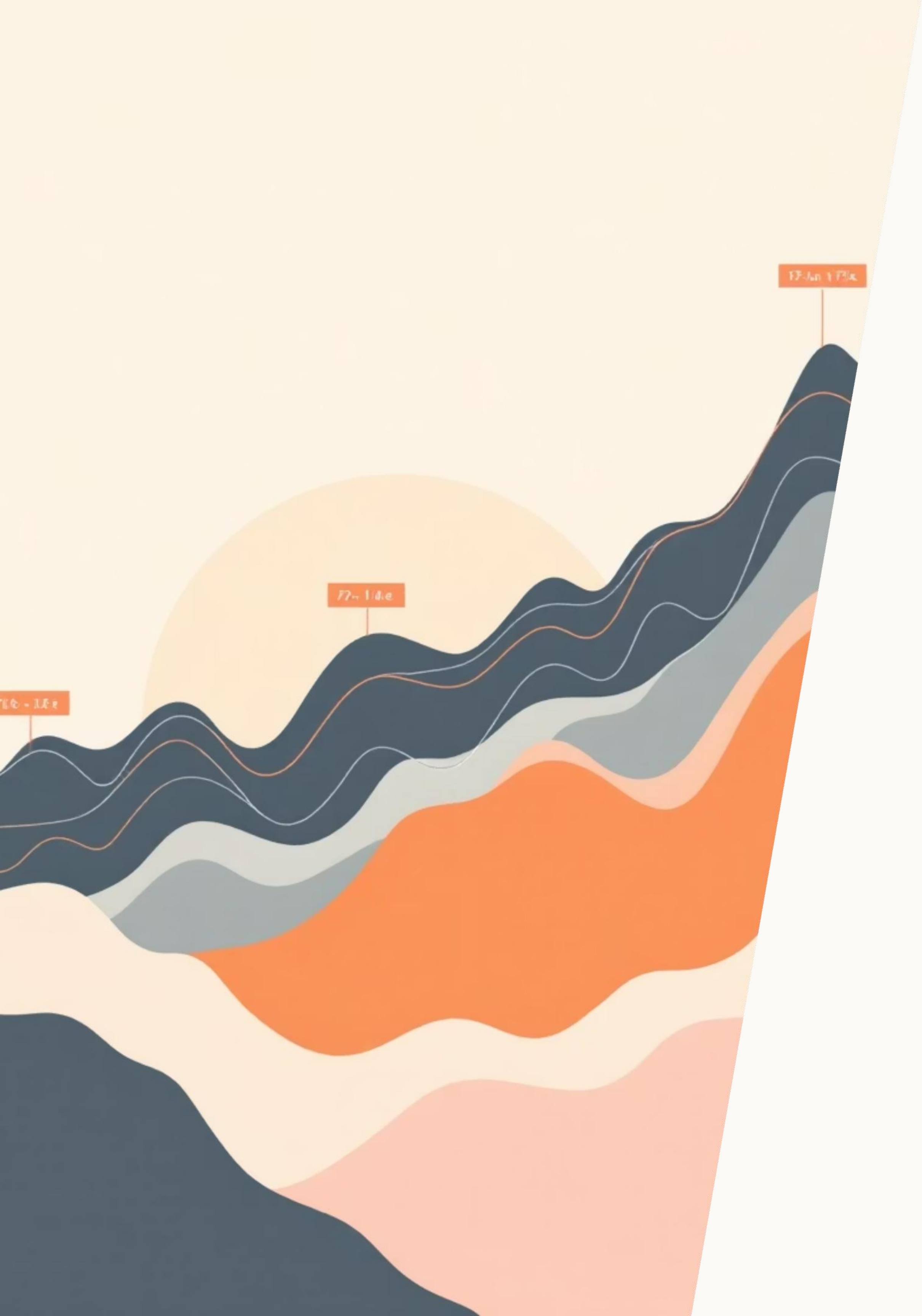
Models	MAE	RMSE	R^2 Score
Gradient Boosting Regressor	0.311	0.407	-0.354
Random Forest Performance	0.327	0.422	-0.452
Support Vector Regressor Performance	1.214	1.571	-19.156
LSTM	0.328	0.414	0.981

Classical ML Limitations

Linear Regression and SVR performed poorly, unable to model non-linear variations. Random Forest was moderate but inferior to LSTM.

LSTM Superiority

With the lowest MAE and RMSE, and the only positive R^2 score, LSTM effectively captured complex temporal patterns and dependencies.



Insights & Future Warming Trajectory

The analysis confirms a critical global trend, with the LSTM model providing the most reliable forecast for the coming years.

Consistent Upward Trend

Historical data and model predictions both show an undeniable and accelerating increase in global average temperature.

Deep Learning Validation

The LSTM's success demonstrates the value of sequential models in handling time-series data with dependencies and memory requirements.

Future Forecasts

Projections from our best model indicate a continued, significant warming trend, underscoring the urgency of climate action.

Conclusion: Expected Outcomes and Future Work

This project successfully validated the application of Deep Learning for climate forecasting, delivering a highly performant predictive system.

Operational Predictive System

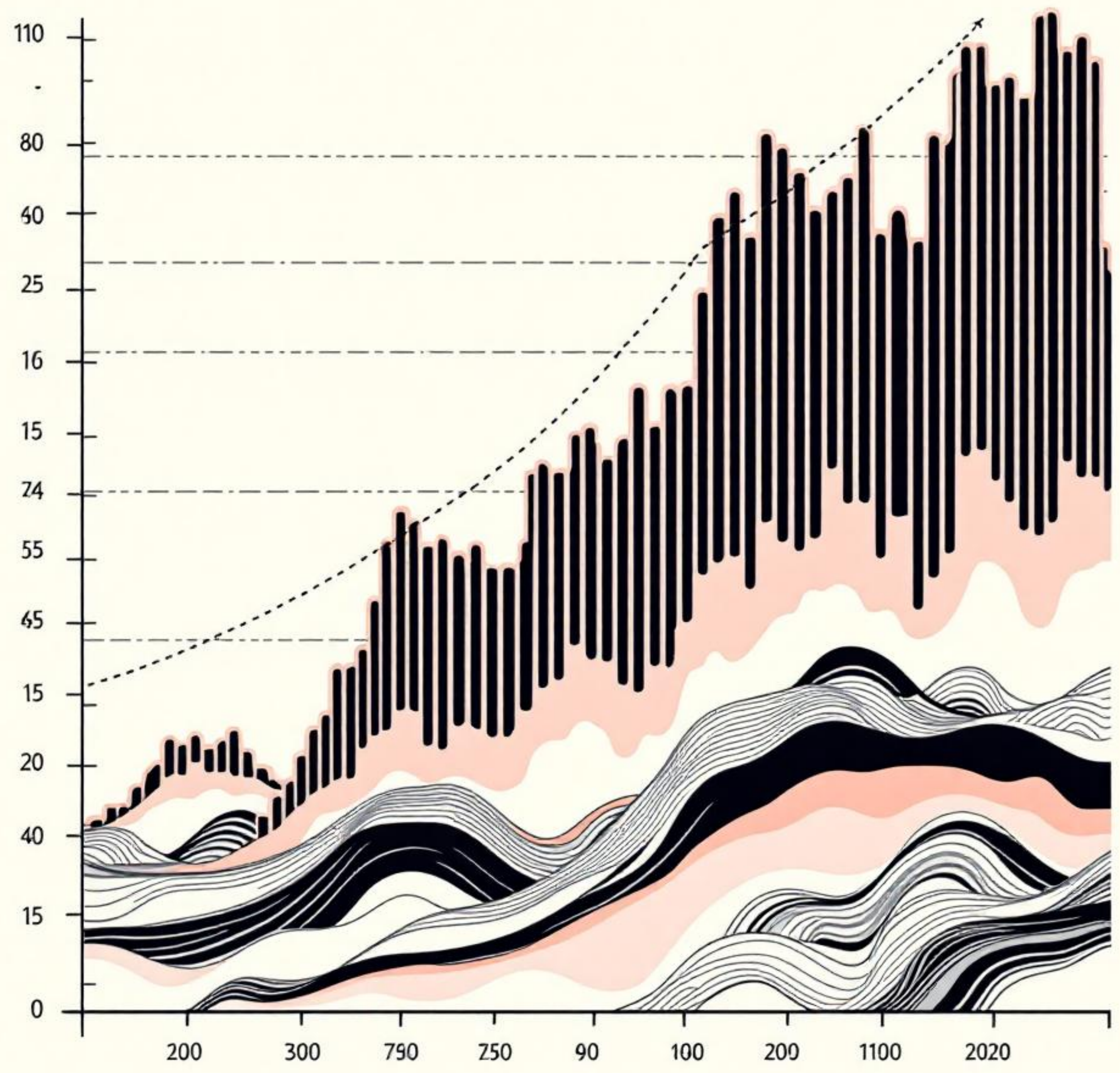
A functional system capable of accurately forecasting temperature trends for upcoming years.

Rate of Global Warming

Deeper, quantitative insights into the current and projected acceleration rate of global warming.

ML/DL in Climate Science

Demonstration of how advanced ML and DL techniques can significantly complement traditional climate research.



Thank you

Jordan Manoj

Email: jordi.manoj@gmail.com