UNIVERSITY OF LONDON

Goldsmiths UNIVERSITY OF LONDON

2023

Final Project Presentation

Computer Science

Presented by Layal Ashi

Project Title: United Kingdom's Climate Change Analysis And Surface Temperature Forecasting

Abstract

In this project, we explored time series analysis for UK temperature forecasting. Our findings emphasize diverse model selection, data quality, and insights into monthly and yearly temperature variations. These discoveries support informed climate data analysis and temperature forecasting.

Overview

UNIVERSITY OF LONDON

Goldsmiths UNIVERSITY OF LONDON

Introduction

This project idea falls under the course "CM3015 Machine Learning and Neural Networks" and it's based on the Project Idea Title 1: Deep Learning on a Public Dataset. It aims to analyze UK surface temperature data and create predictive models for future temperature changes using various machine and deep learning models. It uses the "Climate Change: Earth Surface Temperature Data" from Kaggle, which provides global historical temperature records, making it suitable for studying temperature patterns and building predictive models.

Problems

— Problem 01

Understanding Climate Data and Time Series

Understanding climate data and time series involves analyzing historical temperature records, detecting patterns, and addressing climate-related questions. It requires knowledge of data analysis and time series to gain insights into climate dynamics.

— Problem 02

Building Time Series Predictive Models

Building predictive models for time series data is a challenging task, involving model selection, parameter tuning, and handling time-dependent information. It requires advanced modelling techniques for applications like climate forecasting and stock price prediction.

Project Plan

— WorkPlan

The project plan consisted of essential phases: initial research and literature review, project design, prototype development and evaluation, preliminary report creation, implementation and testing of forecasting models, and finally, the preparation of a comprehensive final academic report.

Data Overview

— Overview

For the public dataset, the choice was the "Climate Change: Earth Surface Temperature Data" dataset, available on Kaggle, which is a comprehensive collection of historical temperature records from around the world. It provides valuable insights into long-term climate trends and patterns, making it a valuable resource for climate analysis, research, and forecasting.

— Advantage

The Berkeley Earth Surface Temperature Study aggregates data from 1.6 billion temperature reports, sourced from 16 established archives. This dataset is thoughtfully organized and offers the flexibility to extract specific subsets, such as data for individual countries or states, making it a versatile resource for in-depth analysis and research.

Objectives

— Objectives 01

Conduct a comprehensive analysis of historical temperature data for the United Kingdom to identify trends, seasonality, and patterns in surface temperatures. This includes exploring variations over different time scales, such as monthly and yearly patterns.

— Objectives 02

Develop and evaluate predictive models that can forecast future temperature variations for the United Kingdom. Utilize a range of forecasting techniques, including traditional time series models like ARIMA and advanced machine learning models like neural networks, to achieve accurate predictions.

UNIVERSITY OF LONDON

Goldsmiths UNIVERSITY OF LONDON

2023

Literature Review

01 Insights from Time Series Analysis Techniques

02 Machine Learning Enhancements in Temperature Prediction

03 Diverse Forecasting Models for Climate Science

Literature Review (01)

Insights from Time Series Analysis Techniques

We've discovered that time series analysis techniques are essential for dissecting temperature data. They enable us to extract critical components like trends, seasonality, and noise, providing valuable insights for precise temperature forecasting.

Reference: Bartholomew, D. J. Review of time series analysis forecasting and control. Operational Research Quarterly (1970-1977) 22, 2 (1971), 199–201.

Literature Review (02)

Machine Learning Enhancements in Temperature Prediction

The literature reveals that machine learning techniques, including regression models and advanced neural networks such as CNNs and LSTMs, offer substantial improvements in temperature forecasting accuracy. By incorporating additional meteorological and environmental factors, we can further enhance the accuracy of our predictions.

Reference: Wu, F., Lu, S., Armando, L.-A., and She, J. Temperature prediction based on long short term memory networks. 312–317.

Literature Review (03)

Diverse Forecasting Models for Climate Science

Our review highlights a spectrum of forecasting models, ranging from basic methods to sophisticated models like ARIMA, SARIMA, and LSTM. These advanced models excel at capturing intricate temporal dependencies, seasonal patterns, and non-linear temperature variations, ultimately leading to more accurate and reliable temperature predictions.

Reference: Sarkar, P. P., Janardhan, P., and Roy, P. Prediction of sea surface temperatures using deep learning neural networks. SN Applied Sciences 2, 1458 (2020).

— Phase 01

Data Preprocessing

This step involves gathering the required datasets from the public source and performing initial data cleaning tasks to ensure data integrity. Any missing values or outliers are addressed to prepare the data for further analysis. And finally preparing two sets of data one for Yearly Data and the other for Monthly Data

— Phase 02

Time Series Analysis

The second phase revolves around time series analysis. This involves checking for stationarity in the temperature data and conducting autocorrelation analysis to understand temporal dependencies. Additionally, the impact of these findings on prediction models is assessed to determine the suitability of various modeling techniques.

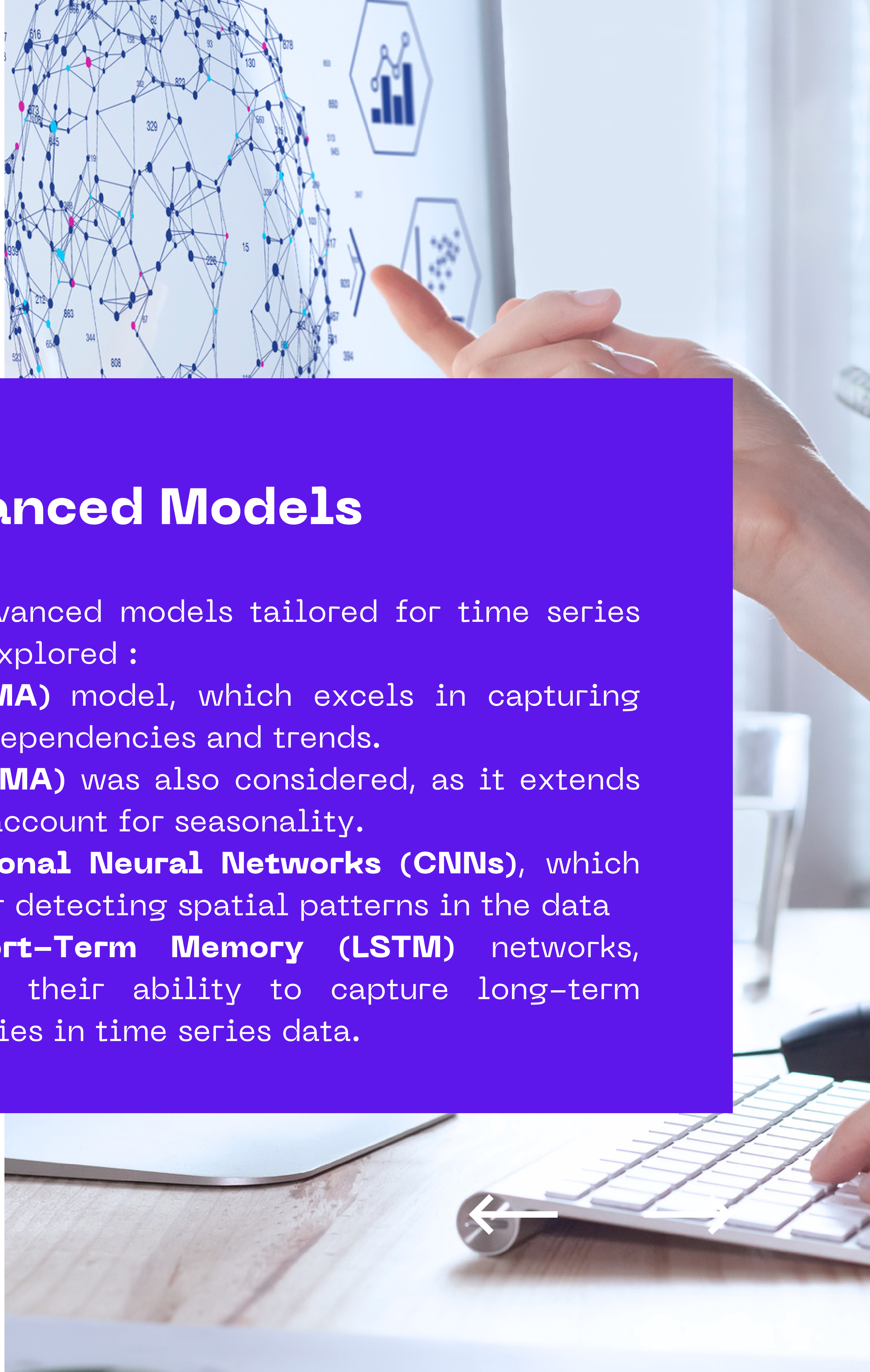
— Phase 03

Models Implementation

The final phase encompasses the actual implementation of prediction models. It begins with the selection of appropriate models based on the insights gained from the previous phases. Hyperparameter tuning is performed to optimize model performance.

Implementation

Models Selection



— Basic Models

For the basic models, we opted for a diverse set of machine-learning regression models :

- **The Random Forest Regressor (RFR)** was chosen for its ability to handle complex relationships and non-linearity.
- **Linear Regression (LR)** was selected for its simplicity and interpretability, providing a baseline for comparison.
- **Support Vector Regression (SVR)** made the cut due to its effectiveness in capturing intricate patterns in the data.

— Advanced Models

Moving to advanced models tailored for time series analysis, we explored :

- **The (ARIMA)** model, which excels in capturing temporal dependencies and trends.
- **The (SARIMA)** was also considered, as it extends ARIMA to account for seasonality.
- **Convolutional Neural Networks (CNNs)**, which is ideal for detecting spatial patterns in the data
- **Long Short-Term Memory (LSTM)** networks, known for their ability to capture long-term dependencies in time series data.

Evaluation Metrics

— Mean Absolute Error (MAE)

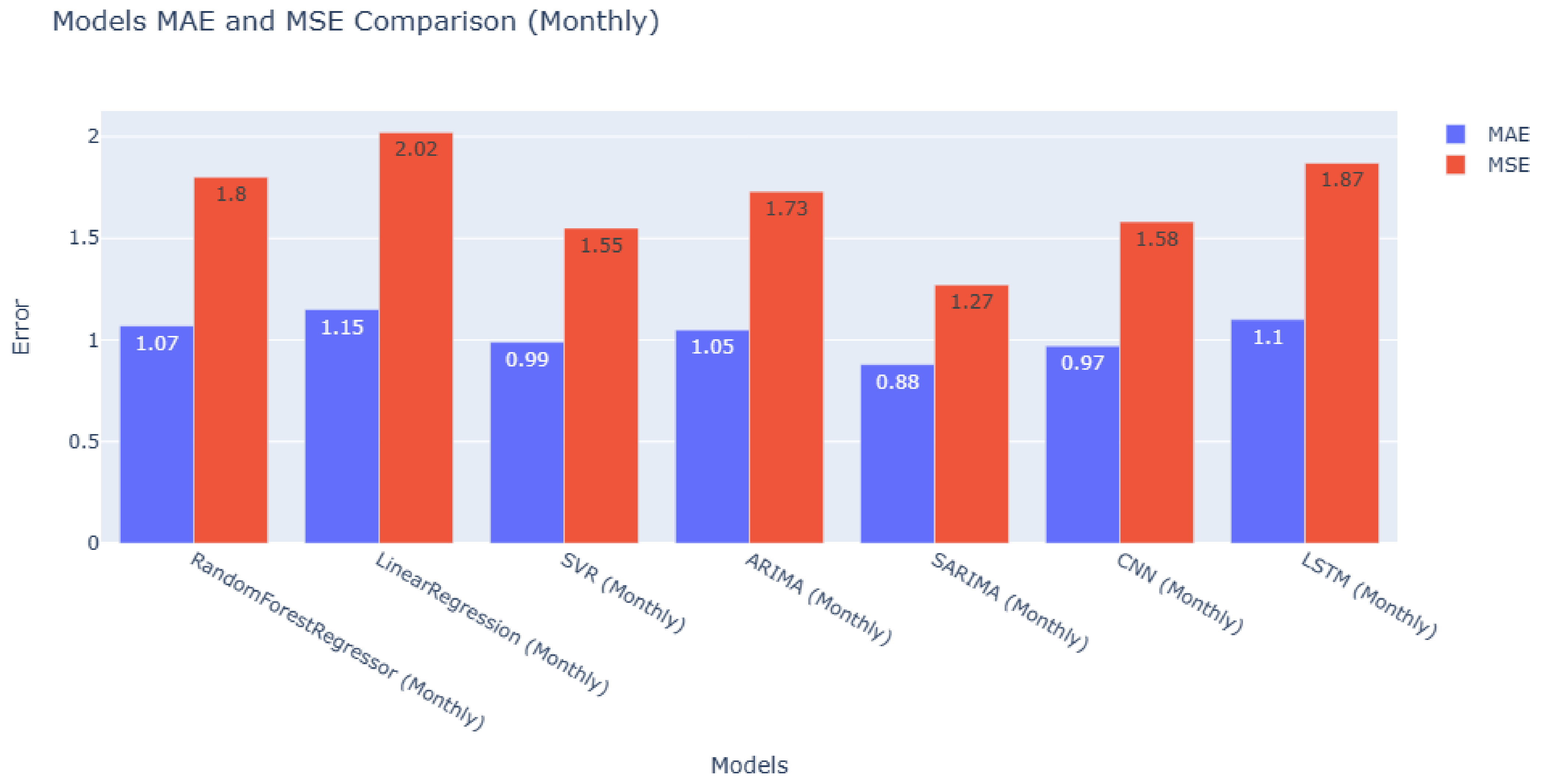
The MAE is the metric that measures the average absolute difference between the predicted values and the actual values. It gives equal weight to all errors, regardless of their direction. The MAE value represents the average magnitude of the errors, and a lower MAE indicates that the model has made, on average, smaller absolute errors in its predictions.

— Mean Squared Error (MSE)

The MSE is the metric that calculates the average squared difference between the predicted values and the actual values. Squaring the errors gives more weight to larger errors compared to MAE. A lower MSE value indicates that the model has made, on average, smaller squared errors in its predictions

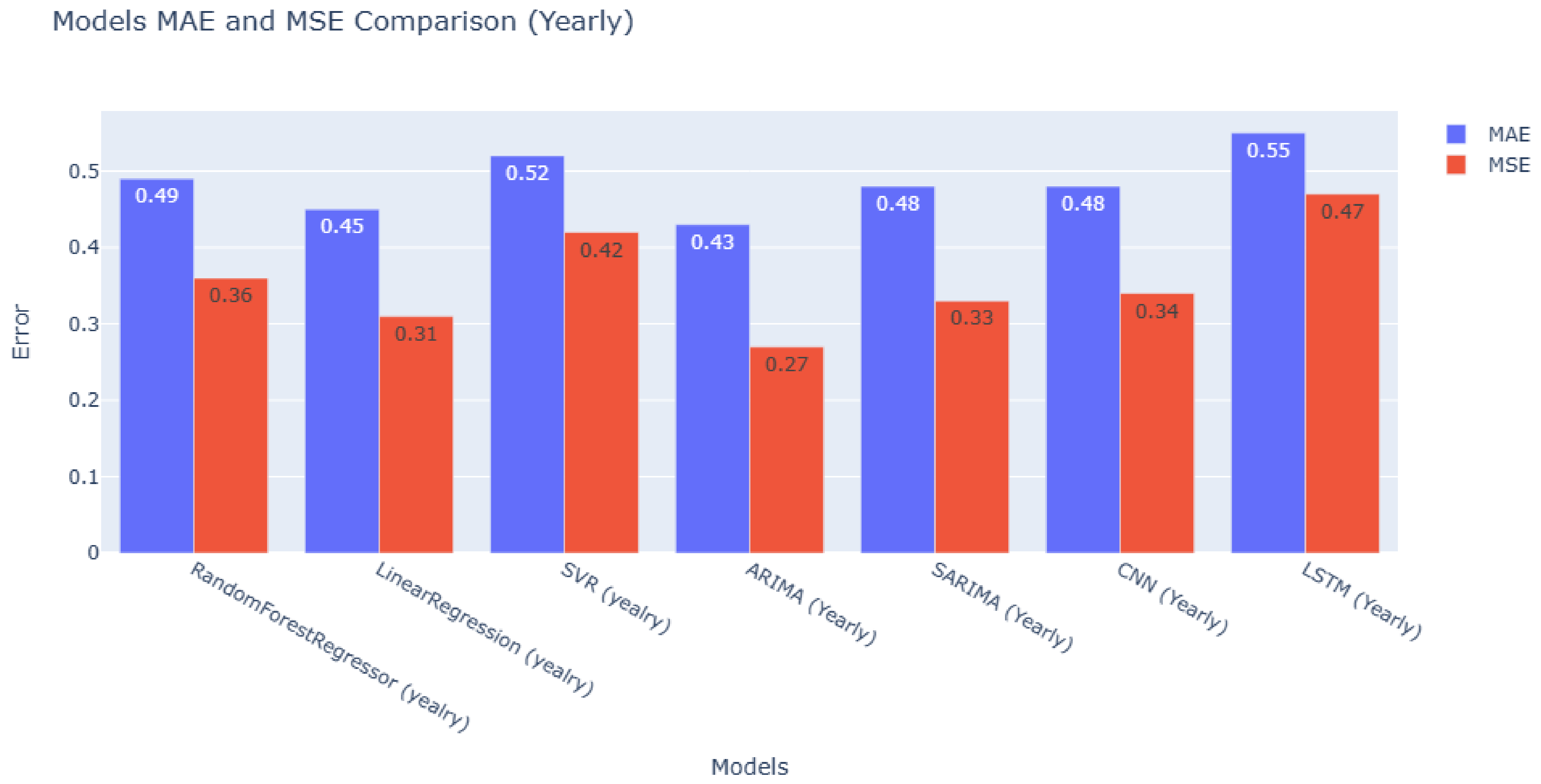
Results (Monthly Models)

Based on the results, **SARIMA** can be considered the top choice for predicting monthly average temperatures, followed closely by **CNN.** These models consistently outperformed others in terms of both MAE and MSE. However, it's worth noting that the **SVR** model has demonstrated competitive performance, indicating its potential as an effective forecasting tool for time series data.



Results (Yearly Models)

Based on the final results, it becomes evident that the **ARIMA** and **Linear Regression** models emerge as the most suitable options for forecasting yearly average temperatures. These models consistently outperform the others, demonstrating greater performance in both Mean Absolute Error (MAE) and Mean Squared Error (MSE) metrics.



Conclusion

In summary, this project provides a thorough assessment of various forecasting models for temperature prediction in the UK. Our findings highlight the effectiveness of models like SARIMA and CNN for monthly data, and ARIMA for yearly predictions. This research emphasizes the importance of aligning models with data characteristics. It also showcases the benefits of using a diverse set of models and sheds light on optimizing neural networks like CNN and LSTM for better predictions. This knowledge can serve as a valuable resource for climate analysis and decision-making in climate change strategies.

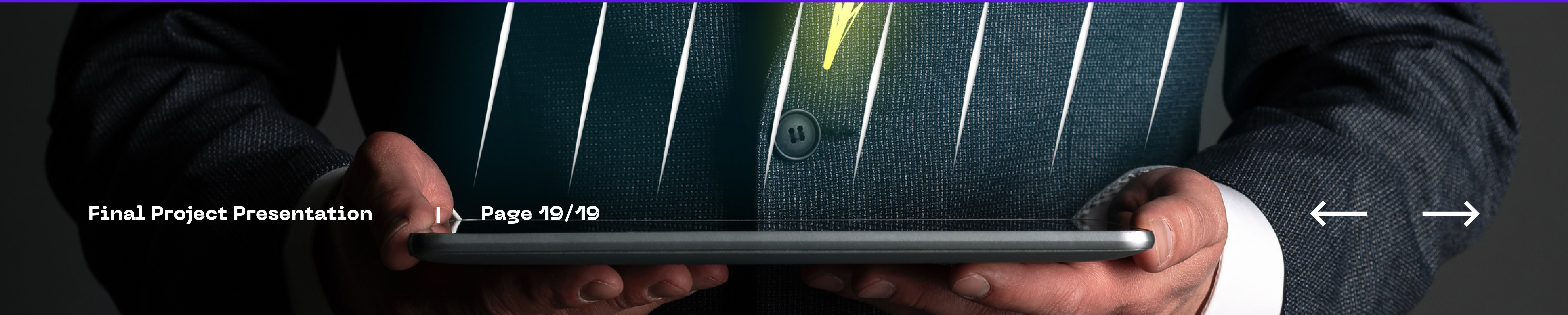
Key Takeaways

— Key 01

Effective Model Selection: We carefully compared various forecasting models, from basic regressions to advanced time series models. By evaluating their performance on monthly and yearly temperature data, we identified the top-performing models. This empowers us to choose the most suitable models for accurate temperature forecasting.

— Key 02

Data Quality Matters: Thorough data analysis and preprocessing were essential. We conducted in-depth exploratory analysis, addressed missing data, checked for stationarity, and uncovered seasonal and autocorrelation patterns. These steps enhanced data quality, reduced bias, and ultimately improved the precision of our forecasting models.



UNIVERSITY OF LONDON

Goldsmiths UNIVERSITY OF LONDON

2023

Thank You

Computer Science

Presented by Layal Ashi

Project Title: United Kingdom's Climate Change Analysis And Surface Temperature Forecasting