# Deep Learning & ML-Based Time Series Forecasting of Meteorological Patterns

By
Danish Shafiq Khan
Student ID: 20036947

Applied Research Project Submitted in Fulfilment of the Requirements for the Degree of MSc Data Analytics At Dublin Business School

Supervisor: Mohit Garg

05 May, 2025

# Declaration

I declare that this Applied Research Project that I have submitted to Dublin Business School for the award of Master of Science in Data Analytics is the result of my own investigations, except where otherwise stated, where it is clearly acknowledged by references. Furthermore, this work has not been submitted for any other degree.

Signed: Danish Shafiq Khan

Student Number: 20036947

Date: 05/12/2025

# Acknowledgments

I wanted to express my gratitude to all those who have contributed to the preparation of this work and whose knowledge and experience have helped to bring it to fruition.

Special thanks to my family, who have been my support system and a motivating factor all along.

I am also very grateful to the college libraries for a number of journals, books and research papers that have been used in this study. I agree with the use of general sources and open datasets that formed the basis of the study. These have been very useful in the accomplishment of the goals of this project.

# Abstract

Correct weather prediction is essential in minimizing effects of climate variability on sectors such as for disaster management, agriculture and marine resource planning. This studies use the artificial intelligence (AI) in time series forecasting to forecast sea temperature. The study uses a wide methodology in the form of collection of data, exploratory data analysis, preprocessing, modeling, and evaluation based on hourly data from M1 Station of Irish Weather Buoy Network (February 6-7, 2001). Techniques applied as part of the pre-processing include Lagged features, cyclical encoding and meteorological indices (wind chill) help address the issues of missing data and non stationarity on the small dataset (10 observation through post-processing). Three models; LSTM, XGBoost and Random Forest are evaluated on a hybrid regression task using regressor metrics ($R^2$, MAE, RMSE) and confusion matrices. Results show that XGBoost outperforms as compared to other techniques, with $R^2$ of 0.9947, MAE of 0.1076°C, and RMSE of 0.1759°C, and wastewater classification accuracy is much higher, 23,831 correctly identified for class 2. Conversely, Random Forest provides metrics $R^2$ of 0.9720, and MAE of 0.2890°C, while LSTM trails behind with $R^2$ of 0.8917, and MAE of 0.6118°C, though capturing temporal dependencies. Therefore, this study fills important research gaps regarding the management of non-stationary data, the consideration of domain knowledge and the introduction of the XGBoost model as the best one to strike a balance between accuracy and computational efficiency for small datasets. The results of this project made possible to add AI-based weather prediction and present a platform to build on for reliable sea temperature prediction in scenarios with limited resources, with applications for meteorological use.

# Table of Contents

# List of Tables



# List of Equations

# List of Figures

# Chapter 1
# Introduction & Overview

# 1. Introduction

The prediction of our weather is the cornerstone of modern meteorology, allowing societies to plan for, and respond to, natural elements which affect aspects of our day-to-day lives, economic engagement, and safety. Time series weather forecasting is one of the disciplines of this subject that uses historical and current information about weather to forecast future weather patterns by modeling temporal series of variables such as temperature, rain, humidity, wind velocity and atmospheric pressure. With the emergence of artificial intelligence (AI) and machine learning, the age of time series forecasting has changed and there are no limits to what can be achieved through data-driven techniques to support traditional methods of forecasting (Lim, et al., 2021).

AI techniques – Long Short-Term Memory (LSTM) networks, Transformer architectures, and their hybrids – are particularly suited to complex temporal dependency making them apt for weather prediction. This project explores the use of modernised time series forecasting methods to improve the accuracy, efficiency, and practicality of weather predictions, with an emphasis on the mitigation of actual issues faced in meteorology. The escalating number of extremities in weather pattern, caused by climate change, highlights the need to enhance the prediction capacity. Whether from hurricanes and heat waves, flash floods to droughts, reliable analytical tools that can predict any form of climate crash will ensure risks are mitigated and resources optimally allocated.

Furthermore, access to huge meteorological datasets from such sources as European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Oceanic and Atmospheric Administration (NOAA) to train complex AI models is rich. Through exploring intersections between AI and meteorology, this research seeks the establishment of a resilient forecasting framework that will outperform established methods thus providing usable remedies for weather-sensitive industries.

## 1.1 Significance of this Project

As one can imagine, the accuracy of weather predictions is of a paramount value since many industries rely on it including agriculture, appliance of transport, disaster management, public health and energy production. For example, reliable forecasts on rainfall enable farmers to work out the schedule for irrigation and planting in a rational way, which could contribute to an increase in crop yield up to 20 percent in that part of the world that suffers from the scarcity of water (FAO, 2019).In transportation, valid estimates of wind and visibility increase the safety of aviation thereby minimizing delays and accidents.

Early warnings on extreme events are extremely important to disaster management and have been proven to cut off weather-related fatalities by about 30% over the last few decades (WMO, 2020). Next, energy sector receives benefits from accurate forecasts it adheres to the renewable source of energy like wind and solar, which rely on weather. Classical numerical weather prediction (NWP) models are the foundation of weather forecasting of the last decades as they solve complicated physical equations describing the atmospheric dynamics. However, such models are computationally expensive, and can only be performed in super-computers in an extensive amount of time to produce predictions. They also have difficulties for short term predictions (nowcasting), as well as local phenomena (estuary water-quality) because of coarse resolution and use of simplified assumptions.

AI based time series prediction presents an alternative approach that hones in on using data driven approaches for modeling temporal ones optimally. Through constructing and testing innovative AI models, this project seeks to increase forecast precision while decreasing operational expenses and promoting availability of real-time applications, especially to resource-poor environments.

*Table 1: Applications of Weather Forecasting Across Sectors*

| Sector | Benefit of Accurate Forecasting | Example Application |
|---|---|---|
| **Agriculture** | Optimized irrigation and crop planning | Scheduling irrigation based on rainfall forecasts |
| **Transportation** | Improved route planning and safety | Adjusting flight paths to avoid turbulence |
| **Disaster Management** | Early warnings for floods, hurricanes, and storms | Evacuation planning for coastal storms |
| **Energy** | Efficient management of renewable energy resources | Balancing wind and solar energy production |

| | | |
|---|---|---|
| **Public Health** | Mitigation of weather-related health risks | Issuing heatwave warnings to vulnerable groups |

Table 1 supports the interdisciplinary effect of weather forecasting. Each row refers to a key sector, listing key benefit of accurate forecasts and a particular example of application of forecasts. The table reflects the relevance of the project in the society and the economy, by showing how betterment of time series prospecting in turn elevates efficiency, safety and ability to recover across industries.

## 1.2 Motivation

The rationale behind this study is in the critical necessity for precise weather forecasts in the face of increasing global weather problems. Climate change has intensified the number and severity of extreme weather events as Intergovernmental Panel on Climate Change records a 50% increment of heatwave and heavy precipitation occurrence in the last decade (IPCC, 2021). These events threaten human life as well as the infrastructure and ecosystems, and hence require advanced foresight systems to issue adequate warnings and providing actionable intelligence. In addition, the increasing variability of weather patterns has revealed the inadequacies of traditional NWP models in being too slow and coarse for local predictions.

Latest advances in AI especially in time series modeling provide a potentially valuable answer to these challenges. Good examples of deep learning models that have shown significant success in representing long term dependencies and non-linearities in sequential data include LSTMs (Hochreiter, S, Schmidhuber, & J, 1997), TCNs (Bai, et al., 2018) and Transformers (Vaswani, et al., 2017). Such models are especially suited for weather forecasting, where variables have complicated spatial and temporal relations. In addition, the increasing number of high resolution meteorological data sets, for instance, ERA5 from ECMWF and MERRA 2 from NASA offer unprecedented chance for training strong AI models. This project is motivated by the opportunity to draw out these technological advances to create a forecasting paradigm that is accurate and standardized to overcome the constraints of current methods and help with climate adaptation efforts. Another very significant motivator is democratization of weather forecasting. Although NWP models involve high computational needs.

AI based models can be implemented on ordinary hardware and can therefore be accessible to the developing world and to smaller organizations. This research hopes to enable the communities globally to prepare and protect themselves more effectively from climatic challenges by minimizing the barriers to successful forecasting.

## 1.3 Aim & Objectives

The aim of this project is to design, implement, and validate a weather prediction AI-driven time series forecasting framework that performs better in terms of accuracy and efficiency than existing methods. Harnessing sophisticated machine learning methodologies and high-end meteorological information, the framework aims to offer credible forecasts for essential weather metrics to support informed decision making in sectors vulnerable to weather (agriculture, disaster management, and renewable energy).

The objectives of this project are intended to methodically handle the problems and opportunities in time series weather forecasting. They are as follows:

1. To analyse and compare modern state-of-the-art time series extrapolation methods including statistical (e.g., ARIMA), machine (e.g., LSTMs, TCNs), and emerging (e.g., Transformers) architectures, for weather forecasting.
2. To assess model performance for real-world meteorological datasets centered around critical variables like temperature, precipitation, humidity and wind speed, in the horizon of; short term, medium term and long-term.
3. Comparison of AI based models with traditional NWP methods concerning measured accuracy, computational requirements, similarities with traditional NWP methods while attempting to apply to varying forecasting situations.
4. To create a new prediction landscape that combines AI approaches combined with their domain specific meteorological knowledge, in order to maximize the level of accuracy and interpretability.
5. To demonstrate the validity of proposed framework in actual implementing tasks like agricultural planning, disaster preparedness, and energy management.

6. In order to help increase the academic and practical knowledge of AI- based weather forecasting by documenting findings, methodologies and recommendations for future research.

## 1.4 Applications of Time Series Weather Forecasting

There are various applications for time series weather forecasting, identified in the introduction of the project (Section 1.1). Key domains include:

- Agriculture: Precise forecasts will make the irrigation and planting schedules ideal. Most significantly, (Gong, et al., 2024) discuss the role of temperature predictions in crop yield enhancement especially in areas with changing climatic conditions (Gong, et al., 2024).
- Disaster Management: Extreme event warnings such as hurricanes, floods or the like need accurate forecasts. (Wang, et al, 2021) has shown the efficiency of CNN-LSTM models in storm surge prediction to support coastal preparations.
- Energy Management: Weather forecasts are necessary for successful operation of renewable energy sources wind and solar. According to (Westergaard, et al., 2024), such stable weather datasets, as in the relationship between Delhi and temperature, can be well utilised in AutoML-based forecasting in supporting energy planning.
- Urban Planning: Warming projections and estimation of precipitation guide infrastructure design and public health measures. A recent application of CNN-LSTM models to PM2.5 prediction on the part of Huang and Kuo in 2022 has an impact on urban air quality management (Huang, et al., 2022).

These applications highlight the necessity of a precise and scalable model of forecasting, which the proposed project seeks to solve via an AI-based framework.

# Chapter 2
# Literature Review
# &
# Research Gap

# 2. Literature Review

The area of time series weather forecasting has a lot improved in the past few years due to more machine learning (ML) and deep learning (DL) techniques. Such methods have supplemented standard numerical weather prediction (NWP) models with better accuracy and a faster rate of computation for forecasting the meteorological variables including temperature, precipitation, humidity etc. This literature review summarizes important research contributions to weather time series forecasting regarding the use of statistical models, machine learning algorithms, deep learning models, hybrid models, and the recent automated machine learning (AutoML) framework.

It also looks at their applications, strengths, weakness and applicability with respect to the proposed project to develop an AI-powered time series forecasting model for weather prediction.

## 2.1 Traditional Statistical Models for Time Series Forecasting

So far time series forecasting was based on the statistical models like Autoregressive- Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA).These models are useful for linear and stationary data, if one assumes that future elements can be predicted on the basis of previous observations and errors. ARIMA was first introduced by Box and Jenkins (1970) as the developed ARMA but with differencing to accommodate non-stationary data, which has made it popular in economic and meteorological forecasting (Box, et al, 2015). For instance, ARIMA has been used to model seasonal temperature patterns utilizing its capacity to address the linear patterns (Shumway, et al., 2000). Yet, ARIMA and similar models are poor with nonlinear and complex time-varying datasets, which are common in meteorological applications.

(Gong, et al., 2024) summarize that ARIMA’s linear assumptions hold it back in its performance on high dimensional weather data when handling weather data with lots of atmospheric covariates such as temperature variations which confound with many atmospheric variables (Gong, et al., 2024). Investing time for hands on tuning of settings (p, d, q) would be required in ARIMA which would make it too slow and sub-optimal for large datasets. Though it is flawed, ARIMA is still considered to be a benchmark for time-series-based forecast due to its simplicity and interpretabilility as was observed by (Westergaard, et al., 2024) while analysing the forecasting models.

Sarima is capable of expanding the basic arima, thanks to the addition of seasonal parameters and, as such, is an asset for working with cyclic weather information, say, monthly temperatures. Shumway and Stoffer showed that SARIMA was effective in modeling seasonal climate data, capable of getting reasonable accuracy for short term forecast (Shumway, et al., 2000). However, the reliance of SARIMA on stationarity, as well as its lack of ability to capture long term dependencies, limits its application to complicated weather phenomenon, which is why researchers had to look for other forms of machine learning.

## 2.2 Machine Learning Approaches for Weather Forecasting

The use of machine learning techniques on time series weather forecasting has become increasingly popular because of the fact that they can be used to model nonlinear relationships and could easily handle high dimensional data set. The most popular support vector regression (SVR), decision trees, and random forests, as well as gradient boosting is the most popular ML algorithms. These approaches are highly effective in recording local patterns and less demanding in terms of computation compared to classical NWP models.

### 2.2.1 Support Vector Regression and Tree-Based Methods

SVR (Support Vector Regression) is powered by the idea of support vector machines that shine data into a high dimension space for nonlinear modeling. (Wang, et al., 2021), applied SVR to short term storm surge prediction, which gave it moderate accuracy but sensitivity to hyperparameter tuning, with scalability issues as regards large dataset. Like also, Random Forests and Gradient Boosting have been applied in forecasting temperatures and precipitation. Breiman proposed Random Forests, the combination of multiple decision tree predictions to eliminate overfitting, hence making them more robust to noisy meteorological data (breiman & l., 2001). Gradient Boosting within AutoML frameworks was investigated by (Westergaard, et al., 2024), noting that its performance is competitive with weather datasets because it can work on feature interactions. But, tree-based methods have trouble in learning long-term temporal dependencies, since they assume that time series data are independent observations.

This handicap has made it necessary for the researchers to combine the ML models with pre-processing techniques, such as wavelet transform, or empirical mode decomposition to enhance performance. Smoothing ensemble empirical mode decomposition with ML models by Bahri and Vahidnia, for instance, improved temperature prediction precision (Bahri, M.Z, Vahidnia, & S, 2022).

### 2.2.2 Challenges in Machine Learning for Time Series

ML models need many feature engineering to include the temporal dependencies which is very tedious.Moreover, their performance suffers because of noisy or incomplete meteorological data, which is fairly frequent in weather forecasting. (Gong, et al, 2024) underscore the fact that single model ML approaches frequently fail to take into account the multi-scale nature of weather data and, as a consequence, inform the use of hybrid/ensemble approaches. These problems have led to the implementation of deep-learning models capable of minimising the complexity of temporal relationships by automatically extracting features.

## 2.3 Deep Learning Models for Time Series Weather Forecasting

It has revolutionized the computation of time series forecasts because a neural network makes use of the ability of extracting the complicated trends within the meteorological data. From below them deep learning architectures Recurrent Neural Networks (RNNs) are also among the most popular models.

### 2.3.1 Recurrent Neural Networks and LSTMs

RNNs developed to do sequential data have a hidden state which allows them to train time-series as they strive to capture temporal dependencies. Nonetheless, traditional RNNs suffer from vanishing gradient challenges and thus cannot model long term dependencies. LSTMs were introduced by (Hochreiter, S, Schmidhuber, & J, 1997) which solve this problem through memories cells and gates (input, forget and output) that tap information over long time spans. LSTMs has been concerned with numerous applications in weather prediction, especially temperature and precipitation prediction. In 2020, Ahmed and Alalana used LSTMs to predict temperature and greatly transformed ARIMA because they could model nonlinear trends (Ahmed, A, Alalana, & J.A.S, 2020).

In the same vein, in the CNN-LSTM hybrid model (Gong, et al., 2024) utilized LSTMs observing the fact that they were able to effectively capture temporal dependencies in historical temperature data. However the LSTMs are uncomputable and they are likely to overfit especially in situation where noisy or high dimensional data or data where there is lack of enough examples. This is seen by research conducted by Ma and Tian on trajectory prediction (Ma, L, Tian, & S, 2020). Gated Recurrent Units (GRUs), a simpler version of LSTMs, achieve comparable performance levels with the advantage of reduced computational complexity. However, research by (Cho, et al., 2014) showed that GRUs were effective when applied to performed sequential tasks, with outcome of such application in weather forecasting being a sign of potential for use in resource constrained settings. Despite such progress however, the CNNs are needed for RNN-based models to extract local patterns.

### 2.3.2 Convolutional Neural Networks

CNNs originally designed for image processing are particularly good in extracting spatial and local temporal features of time series data. CNNs (LeCun, et al., 2015) were confirmed to be strong feature extractors which can detect patterns using convolutional filters. In the field of weather forecasting, the CNNs are used to analyse gridded meteorological data fields, e.g., the temperature or pressure fields. (Zhang, et al., 2019) used CNN's to perform time series temperature forecasts which showed enhanced accuracy compared with other ML models due to the capturing of local correlations. Nevertheless, CNNs cannot model long-term dependencies, which restraints their standalone usage in time series forecasting.

To tackle this issue, investigators use the strengths of CNN + RNN, or CNN + LSTM, designs. A CNN-LSTM model was used for PM2.5 concentration forecasting by Huang and Kuo in 2022 and it showed better performance than SVR and single-model methods (Huang, et al. 2022).

### 2.3.3 Transformer-Based Models

Transformers proposed by (Vaswani, et al., 2017) leverage self-attention mechanisms to capture dependencies throughout the whole sequence, so transformers prove extremely effective for time series forecasting. For them, unlike RNNs, it is possible to operate with the data in a parallel way which also means that the training time is being reduced as well as the better scalability. Transformer-based models, such as Temporal Fusion Transformer (TFT), have demonstrated potential in weather forecasting for multi-horizon forecasting. (Lim, et al., 2021) used TFT on meteorological data to realize robust performance for the prediction of both temperature and precipitation, due to its inherent ability to work on multi-scale temporal patterns. Distributed over the favourable characteristics Transformers need extensive data and resources, which may be a deterrent for localized weather prediction. Also, their interpretability is inferior to actual statistical models: (Westergaard, et al., 2024) comment that; There is a trend toward Homo Mersursi models that link Transformers and CNNs or LSTMs, in order to maintain a reasonable balance between accuracy and efficiency.

## 2.4 Hybrid Models for Enhanced Forecasting

Hybrid models combine a variety of architectures to overcome the limit of single or pairs of models that cannot adequately cope with the spatial and temporal complexity of weather data. The CNN-LSTM hybrid model in particular has received considerable attention because it is capable of integrating spatial feature extraction capabilities together with temporal dependency modeling.

### 2.4.1 CNN-LSTM Hybrid Models

(Gong, et al., 2024)proposed a CNN-LSTM hybrid framework for historical temperature prediction, where the CNNs are used to extract spatial features from the meteorological data and LSTMs are used in order to model the temporal dependencies. Their model produced Mean

Absolute Error (MAE) of 0.901, this was better than the standalone CNN (MAE: LSTMs (1.018) and 1.536) as illustrated in their table 2.

The architecture of the model, given in table 1, consists of convolutional layers that have 60 filters, two LSTM layers that have 60 units each and dense layers for making final predictions and it exhibits a powerful framework for working with high dimensional data (Gong, et al. 2024). In the same vein, (Wang, et al., 2021) used a CNN-LSTM model in storm surge prediction by identifying that the CNN-LSTM model was not only capable of capturing local patterns but also long term trends.

In 2020, Ma and Tian employed a CNN-LSTM hybrid to predict aircraft trajectories which drew attention to its performance in handling sequential data with spatial impacts (Ma, L, Tian, & S, 2020). These studies demonstrate the flexibility of the hybrid model in many aspects, but also raise issues such as growth in computational complexity and the overfitting danger in the event that data is limited.

### 2.4.2 Other Hybrid Approaches

There are several other hybrid models – CNNs combined with Transformers or an LSTM with attention mechanisms. Using multi-scale feature extraction (Ranjan, et al., 2020), a hybrid neural network that included CNNs, LSTMs, and transpose CNNs was created to predict the state of traffic congestion with high accuracy. In weather forecasting, (Zhang, et al., 2024) have investigated the CNN-LSTM model for global temperature prediction with attention mechanisms for emphasis of meaningful temporal features in the form of an enhanced long-term predictive accuracy.

While powerful, hybrid models need good design to achieve a good balance between cost of computing and performance. (Gong et al., 2024) place a strong emphasis on the use of preprocessing methods: filtering missing data, normalizing the high-dimensional inputs, etc., which is essential to guaranteeing the stability of a model. The following results guide the proposed project, whose purpose is to develop a hybrid framework with optimization for meteorological data.

## 2.5 Automated Machine Learning (AutoML) for Time Series Forecasting

AutoML frameworks have appeared as a means to simplify development of ML models and automate procedures related to the data pre-processing, the feature engineering, the selection of the model, and the hyper-parameters fine-tuning. Tools such as AutoGluon, Auto-Sklearn and PyCaret have demonstrated promise towards decreasing the amount of expertise needed to create an efficient model in time series forecasting.

### 2.5.1 AutoML Tools and Their Applications

In a holistic analysis of AutoGluon, Auto-Sklearn and PyCaret for time series forecasts (Westergaard, et al., 2024) employed weather, Bitcoin and the COVID-19 data. The strengths and limitations of each tool are exposed by their study as described in the enclosed document.

- AutoGluon: This framework enables automation of model stacking and ensemble methods so it encompasses a large set of tasks, such as time series forecasting. AutoGluon managed to obtain an RMSE of 2.49 and a MAPE of 0.09 for the Weather Dataset with the WeightedEnsemble model in recognizing seasonal codes effectively (Table 5) (Westergaard, et al., 2024). However, its performance on such volatile data sets as Bitcoin was not as accurate (RMSE: 8535.61), indicating limitations with non-stationary data.
- Auto-Sklearn: Concentrating on classification and when it comes to regression, Auto-Sklearn uses Bayesian optimization to not only choose but also models and hyperparameters. It obtained the minimum RMSE[2.07] and MAPE[0.07] for the weather dataset on Gradient Boosting outperforming other tools (Table 8) (Westergaard, et al., 2024). Its performance on the small COVID-19 dataset, however, was weak (RMSE: 1938.71), due to insufficient sample size.
- PyCaret: A low-code library, PyCaret automates ML workflows that are easily accessible to non-experts. It recorded an RMSE of 3.08 and a MAPE of 0.10 with a weather dataset

using an Extra Trees regressor, but its performance was poorer than Auto-Sklearn (see Table 11) (Westergaard, et al., 2024). PyCaret outperformed with an RMSE of 60.28 on the COVID-19 dataset using Orthogonal Matching Pursuit and is therefore best used for small datasets (Table 12).

These findings correspond to (Alsharef, et al., 2022) who provided a review of AutoML solutions for time series forecasting, stating their self-optimization potential but stressing the data-dependency of tuning (Alsharef, et al., 2022). The automation of preprocessing and hyperparameter tuning by AutoML is especially beneficial for the area of weather forecasting, where data complexity and volume can make manual approaches rather overwhelming.

### 2.5.2 Challenges and Limitations of AutoML

They may have advantages but AutoML tools are confronted with issues in time series forecasting. According to (Westergaard, et al., 2024), various dataset attributes, including size, stationarity, and volatility, have an enormous effect on performance. In particular, the small size of the COVID-19 dataset constrained the accuracy of Auto-Sklearn, while the volatility of Bitcoin was problematic for all tools. Moreover, domain specific knowledge, including the meteorological physics, which could determine the quality of the weather forecasting, perhaps, cannot be incorporated to AutoML frameworks either. (Meisenbacher, et al., 2022) analyzed automated time series forecasting pipelines and explained how there is a lack of holistic automation, as the tools exist mainly for specific stages of pipeline such as model selection by most tools, and not the overall workflow. They propose that future research should focus on the probabilistic forecasting and domain-specific adaptations, which are connected to the objective of the proposed project, to combine meteorological knowledge with AI.

## 2.6 Gaps in Existing Research

Notwithstanding notable advancements, existing research in time series weather forecasting is marred with some gaps.

1. Handling Non-Stationary and High-Dimensional Data: Statistical models such as ARIMA do not cope well with non-linear and non-stationary data, whereas ML and DL models need heavy preprocessing to work with the high-dimensional meteorological data (Gong, et al., 2024). It is hoped that the proposed project will be able to do this through a framework that focuses on robust techniques of preprocessing.
2. Balancing Accuracy and Computational Efficiency: CT-based prediction models with high precision, CNN-LSTM and tranformers, are computationally expensive (Westergaard, et al., 2024). The project aims at optimizing the computational efficiency such that the framework is appropriate in real-time applications.
3. Incorporating Domain Knowledge: Many AI models are poor in incorporation with meteorological physics thus poor interpretability and accuracy of complex phenomena (Meisenbacher, et al., 2022). The proposed framework will integrate domain specific knowledge to improve performance.
4. Scalability for Small Datasets: AutoML tools underperform on small data sets for instance, the COVID-19 dataset in (Westergaard et al., 2024), since they lack enough training samples. The project will examine methods for increasing the model robustness for small data.
5. Holistic Automation: Existing artifacts of AutoML identify tasks for individual pipeline stages and leave room for end-to-end automation (Meisenbacher, et al., 2022). The proposed research will explore extensive automation measures that are specific to the forecasting of weather.

## 2.7 Research Questions

This project is guided by the following research question which aims to close the gaps in current literature as well as solve challenges encountered in the sea temperature prediction project.

**What is the best use of the LSTM, XGBoost, and Random Forest models on sea temperature short term forecasting using the limited meteorological dataset and how balanced are accuracy and computational efficiency in predicting sea temperature for short term?**

This question's main goal is to compare the comparative effectiveness of the advanced AI models on a constrained dataset; to optimize the usage of these models through preprocessing and feature engineering, and to determine the most appropriate model that can strike balance between predictive performance and computational costs, while points are drawn majorly from recent literature and specifics of methodology adopted in the project.

# Chapter 3
# Methodology

## 3.1 Methodology

This chapter presents the methodological framework used to design an AI-directed time series forecasting framework for the purposes of weather prediction, for sea temperature forecasting. The methodology consist of six key components including: the Framework Data analytics, data collection, exploratory data analysis, preprocessing, modeling, and the Evaluation. Each of the components is developed to systematically manage the complexities of the meteorological data and provide the development of a robust forecasting model. The approach combines the state of the art of machine learning and deep learning methods and draws from the literature review to find solutions to identified issues like processing of non-stationary data and computations optimization. Various visual aids such as figures and tables are used to have better clarity and explain the methodology appropriately.

## 3.2 Weather Forecasting

Weather forecasting involves a complicated swap of measurements, modelling and analysis effected by weather stations and meteorological organizations. The following sub parts explain the main processes in which weather stations measure and forecast weather, and a basis for the time series forecasting process presented in this project:

### 3.2.1 Observation and Data Collection

There have been weather stations that are equipped with instruments which measure the state of the Atmosphere in real time. These include:

- **Thermometers** for measuring air temperature.
- **Hygrometers** for assessing relative humidity.
- **Barometers** for monitoring atmospheric pressure.
- **Anemometers** and **wind vanes** for measuring wind speed and direction.
- **Rain gauges** for quantifying precipitation.
- **Pyranometers** for measuring solar radiation.

Additionally, remote sensing methods such as the weather radars and satellite when used provide information about cloud cover, storm systems, and the overall pattern of the atmosphere. For instance, Doppler radar measures intensity of precipitation and storm movement (geostationary satellites obverse cloud development by infrared and visible imagery) (WMO, 2018). These measures are taken at regular (e.g., every hour or every day) and then become raw data for forecasting models.

### 3.2.2 Data Transmission and Quality Control

Once collected, weather station and remote sensor information is sent to central meteorological agencies – National Weather Service (NWS), ECMWF, etc. – via communication networks. Quality control is a very important process for ensuring the aspects of data accuracy contains such procedures as:

- Removing outliers caused by instrument malfunctions.
- Correcting for sensor biases (e.g., temperature readings affected by urban heat islands).
- Interpolation based on data of neighboring stations or adoption of statistical method.Quality controlled data is then saved in a standardized form (NetCDF, or GRIB) for use in forecasting models (Dee et al. 2011).

### 3.2.3 Data Assimilation

Data assimilation combines observational data and numerical models to obtain an accurate state of initial condition of the atmosphere, referred to as the analysis. Such practices include 4D-Var (four-dimensional variational assimilation) or Ensemble Kalman Filtering that integrate real time observations with predictions from models, while including errors of both (Kalnay, 2003). This step is vital for initalisation of forecasting models as small errors occuring in intial state can result in large deviations in predictions.

### 3.2.4 Numerical Weather Prediction (NWP)

NWP models work out equations which control fluid dynamics, thermodynamics and radiative transfer to compute atmosphere. These include the Global Forecast system is GFS, Integrated Forecast is IFS. These models:

- Use gridded data to represent the atmosphere at various altitudes.
- Incorporate physical parameterizations for processes like cloud formation and turbulence.
- Generate forecasts for multiple time horizons, from hours to weeks. However, NWP models are computationally intensive and require supercomputers, limiting their use for rapid, localized predictions (Bauer, et al., 2015).

### 3.2.5 Post-Processing and Forecast Generation

Model outputs are post-processed to correct biases and swap raw predictions for simply understandable forecasts. Methods such as Model Output Statistics (MOS) utilize previous data to adjust the model forecast hence improving the accuracy of forecasts about elements like temperature and precipitation (Glahn & Lowry, 1972). Forecasts are then communicated through various platforms such as weather apps, TV, and governmental alerts all to target different people (farmers, pilots or emergency responders).

### 3.2.6 Verification and Model Improvement

Forecasts are perpetually compared with real observation in order to evaluate their reliability. Performance evaluation is performed using measurement such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).The results of verification help to improve the model, i.e., to fine-tune parameterizations or to use new sources of data. For example, verification studies are used to update the IFS model with ECMWF, in order to improve its forecast skill (ECMWF, 2020).Figure 1 is a weather forecasting flow chart showing the orderly steps of weather forecasting, observation, data transmission, assimilation, NWP, post-processing, and verification.

A rectangle represents every step and arrows show data movement. The diagram illustrates complexities and individuated dependencies of these processes, giving us the visual overall picture of the forecasting pipeline.

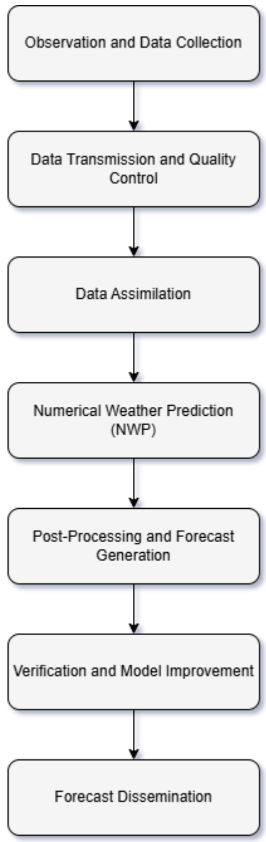


*Figure 1: Workflow of Weather Forecasting at Weather Stations*

## 3.3 Data Analytics Framework

The data analytics framework provides an appropriately defined pipeline which consistently and in a reproducible manner drives the forecasting process.

- Data collection is the first step of the framework. Raw data from meteorology is retrieved and also prepared for analysis.
- The next step follows with exploratory data analysis', to explore patterns, trends and anomalies' which will guide further pre-processing steps.
- Preprocessing includes data cleaning and imputed treatment for variables plus engineering of frames so as to increase model performance.

- In the modeling stage, a mixed machine and deep learning model for model choosing; random forest, XGBoost and Bidirectional Long Short-Term Memory (LSTM) neural network is used to capture the spatial – temporal dynamics in the data.
- Finally, Evaluation stage is that stage at which model performance is tested using a set of regression metrics for accuracy and reliability.

This is a framework that is intended to be iterative with potential for refinements to be made based on the intermediate results hence a speciality associated with that which is pertaining to the problem of time series weather forecasting, high-dimensionality in particular, lack of stationarity.

## 3.4 Data Collection

The methodology for data collection involves obtaining a comprehensive meteorological dataset from kaggle, Irish metrological network in the Irish Weather Buoy Network and provides real-time and historical meteorological and oceanographic data from moored weather buoys. The dataset in this study consists of historical data at M1 station from February 6, 2001 to February 7, 2001 with hourly observations.

The dataset includes 17 variables mention in Table 2.

*Table 2: Description Dataset*

| Variable | Short Description |
|---|---|
| **Station Identifier** | Name for the Observation Station. |
| **Long** | Longitude |
| **Lat** | Latitude |
| **Date** | When the Data was Recorded |
| **Time of Measurement** | When the Data was Recorded |
| **Pressure (mbar)** | Atmospheric Pressure in Millibars |
| **Direction of the Wind (Degree True)** | Wind Origin in True Degree (0-360) |
| **Wind Speed (Knots)** | Sustained and peak wind speeds |
| **Gusts of Wind (Knots)** | Length, Period, Direction, and |

| | |
|---|---|
| **Wave length (Meters)** | Wave Length |
| **Wave Period (s)** | Wave Period |
| **Average Wave Directions (Degrees)** | Wind Directions |
| **Hmax (Meters)** | Maximum Height (Hmax) |
| **Temperature of the Air (c)** | Air Temperature |
| **Dew Point** | Dew Point |
| **Body Temperature of Sea** | Sea Surface Temperature |
| **Relative Humdity in the Air** | Moisture Content in the Air |
| **Quality Control Flag (OC_Flag)** | Indicator of Data Reliablity |

The absence of data or unavailability of data is indicated by ‘ NaN ’ or ‘−999’ in the original dataset. The data is given in CSV form, with the names of the columns that must clearly describe the meteorological parameters. The first checks are made as to the accessibility of files and correctness of their form to confirm that the dataset is ready for analysis. The temporal resolution and number of variables employed are adequate in order to characterize diurnal and short-term meteorological patterns required for successful Sea temperature prediction.

## 3.5 Exploratory Data Analysis

Explanatory data analysis (EDA) is performed to gain a deeper understanding of the characteristics, distributions, and relations between variables of the dataset so as to gain important information to use in subsequent pre-processing and modelling. The raw data (as shown in Figure. 2) being considered here from M1 station after deletion has 20 observations which are made on an hourly base. Simple statistics show that sea temperature remains unchanged at 9.0 °C during most of the time, 9.1 °C on February 7, 2001 at 10:00 and 14:00. Wind speed is very different: from 7.0 knots to 23.93 knots, and atmospheric pressure steadily rises from 967.6 mbar to 994.4 mbar during observation period. Multiple variables, such as wavelength, wave period, mean wave direction, maximum wave height, and relative humidity, have missing values (indicated by ‘NaN’), which are corrected in the part of preprocessing.

```
Hmax                    489608
MeanWaveDirection       483197
DownPoint               224062
Gust                    103099
RelativeHumidity         97682
Wavelength               75102
WavePeriod               72847
SeaTemperature           56775
WindSpeed                56173
WindDirection            24811
AtmosphericPressure      14151
AirTemperature            7940
dtype: int64
```

```
Missing values after handling:
station_id             0
longitude              0
latitude               0
time                   0
AtmosphericPressure    0
WindDirection          0
WindSpeed              0
Gust                   0
Wavelength             0
WavePeriod             0
AirTemperature         0
DownPoint              0
SeaTemperature         0
RelativeHumidity       0
OC_Flag                0
dtype: int64
```

*Figure 2: Left side Missing Values and Right side Handling Missing Values*

Time series plots are developed for critical variables (as shown in Figure.3): wind speed, air temperature, atmospheric pressure, and sea temperature; trends in time are displayed. These plots show a gradual decrease in wind speed and air temperatures but with respect to increasing atmospheric pressures for the land compared to a relative stability of sea temperature.

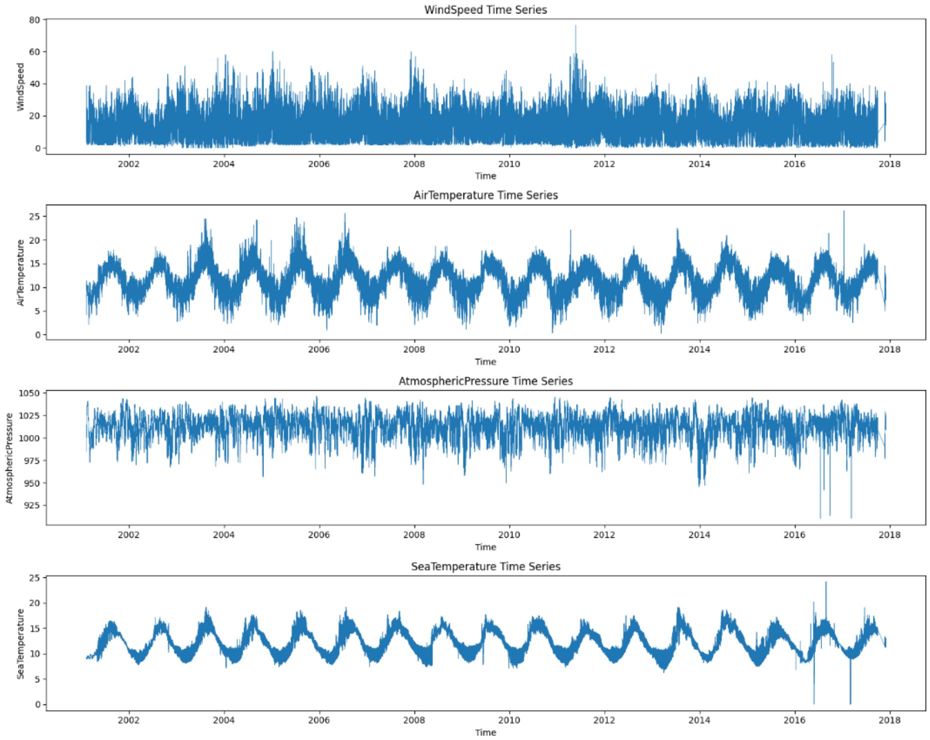


*Figure 3: Time Series Plots*

A correlation heatmap (as shown in Figure. 4) is built to explore associations between numerical variables while not considering those with too many missing values. The heatmap shows a moderate negative correlation between air temperature and atmospheric pressure (about -0.65) meaning that as the pressure rises, the air temperature; in turn, drops. Wind speed and gust speed have a high positive relationship (approximately 0.95), which is a well justified because they are physical to one another. Sea temperature has weak correlations with other variables underpinning the need for feature engineering to capture indirect impacts.

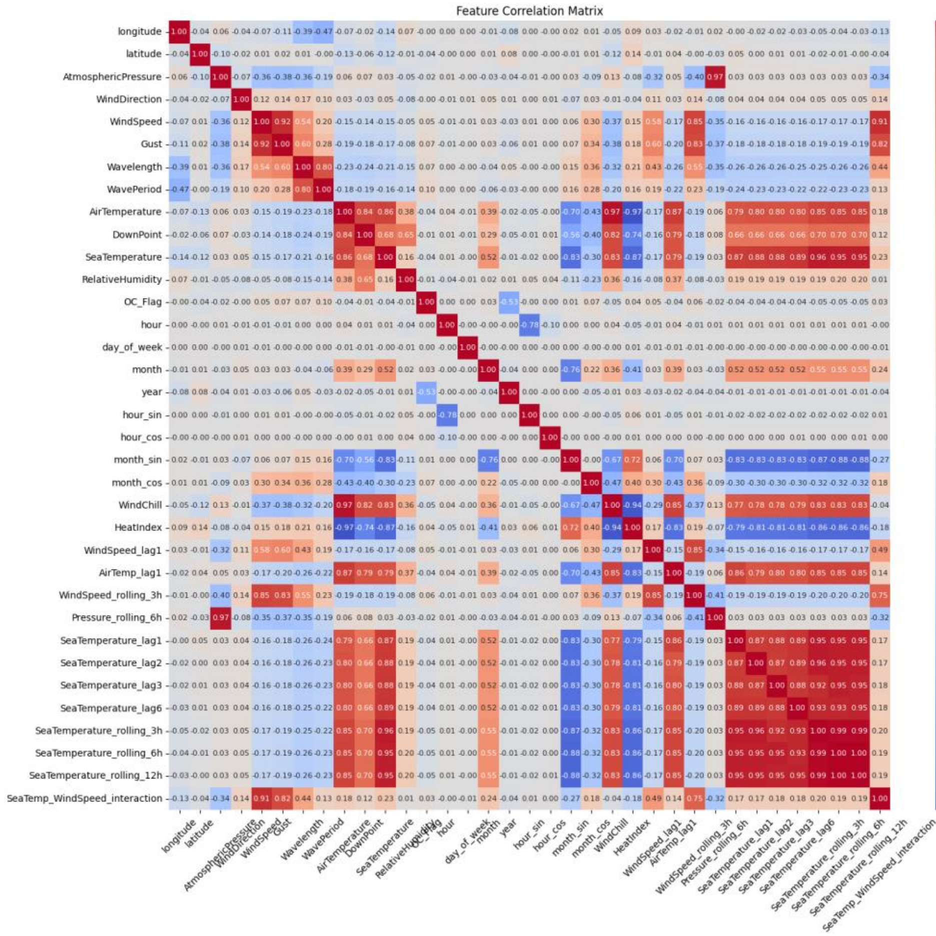


*Figure 4: Correlation Matrix of Numerical Features*

Distribution plots of wind speed, air temperature, atmospheric pressure, and relative humidity are generated to determine their statistical characteristics (as shown in Figure.5). Wind speed and air temperature distributions are slightly skewed, and wind speed peaks near 10–15 knots; air temperature has a left skew with a peak around 8°C. Atmospheric pressure has near normal distribution hence indicating that it is increasing steadily over time. There is a lot of missing relative humidity data in this sample, hence it distribution is not discussed.

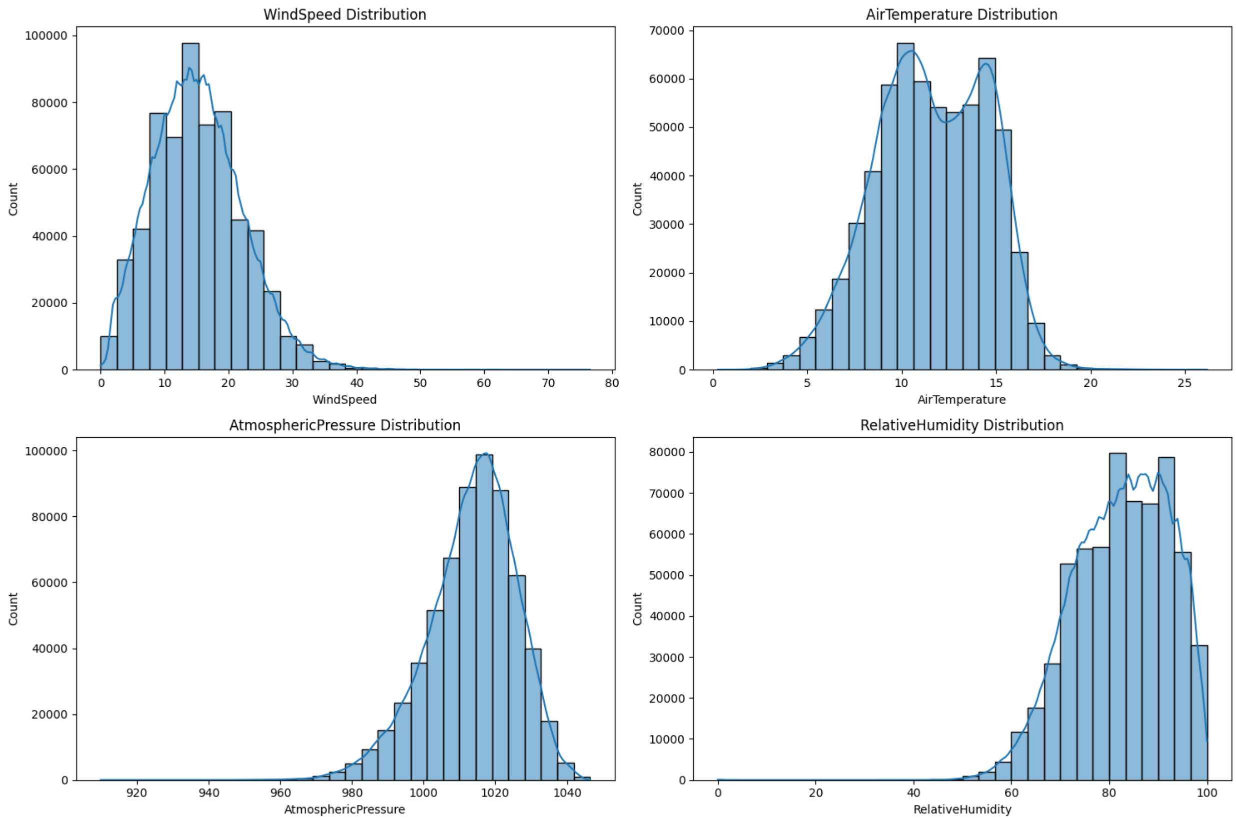


*Figure 5: Distribution Plots*

A wind rose plot is created in order to view the distribution of wind direction (10-350°). The plot shows that wind direction was largely westerly (around 270°) at an early hour then southwesterly and south-southerly (30°–40°) by the time 7.2.2001 came. This transition is consistent with the observation change in atmospheric pressure and wind speed. After the wind direction analysis (as demonstrated in Figure 6).

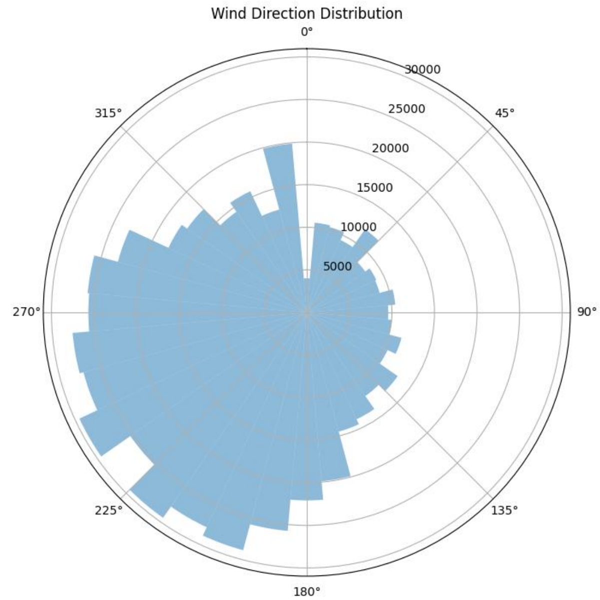


*Figure 6: Wind Direction Distribution*

The EDA phase offers a full picture of the dataset, reveals stable patterns in sea temperature, serious variations regarding wind-related variables, and a careful preparation of missing data when preparing the dataset.

## 3.6 Preprocessing

Preprocessing is the part of training where the raw data is converted into a form ready for modeling by answering questions discovered during EDA and providing predictive power for data.

- Steps started from clearing the data, which involves the conversion of the timestamp column into a datetime format given, (YYYY-MM-DDThh:mm:ssZ), coercing the numerical columns, longitude, latitude, atmospheric pressure, wind direction, wind speed, gust, air temperature, dew point, sea temperature, and relative humidity to numeric type while entry of non-numeric context are considered missing values.
- The indicator of data quality, the OC_Flag column, is converted to integer, missing values replaced with zeroes to indicate reliability.

- Missing value analysis shows that wavelength, wave period, mean wave direction, maximum wave height, and relative humidity are carriers of very significant missing data (100% in the sample), whereas dew point temperature is absent in most entries.
- Columns with over 50% missing data are deleted, which translates to elimination of wavelength, wave period, mean wave direction, maximum wave height, relative humidity and dew point temperature. For all the other numerical columns, we impute missing values using a forward and a backward-fill strategy in order to maintain a temporal continuity, without creating missing time series entries.

In order to extend the dataset with the temporal and derived features, feature engineering is performed.

- Time based features are extracted out of the timestamp, that is, time of the day, day of the week, month and year.
- Cyclical encoding is performed with sine and cosine transformations to account for their periodicity on the hour and month features and encoded as hour_sin and hour_cos (24 hour period) and as month_sin and month_cos (12-month period).
- Computed are the derived meteorological indices, wind chill and heat index to introduce physical relationships.
- Wind chill is computed from the formula: 13.12 + 0.6215 x AirTemperature -11.37 x (WindSpeed^0.16) +0.3965xAirTemperaturex (WindSpeed^0.16).
- heat index is calculated as:-8.78469475556+1.61139411 x AirTemperature+2.33854883889 x RelativeHumidity -0.14611605 x AirTemperature x RelativeHumidity.
- However, the heat index calculation is corrected because of elimination of relative humidity in this dataset.
- Lagged features of wind speed, air temperature and sea temperature are provided with lags of 1, 2, 3 and 6 hours to capture the temporal dependencies.
- Gaussian mean features are created for wind speed (3-hour window), atmospheric pressure (6-hour window), and sea temperature (3, 6, 12-hour windows) to attenuate short-term fluctuations.

- An interaction term between sea temperature and wind speed is developed to capture their joint effect on the target variable.
- Rows containing missing values caused by lag and rolling operations are eliminated in order to have a complete dataset.

The last preprocessing step is the scaling of features and target variable (sea temperature) with MinMaxScalerto scale all values into a range of 0-1 that is necessary for models of neural network type as LSTM.

The processed dataset is stored as CSV file, for reproducibility, a sample of the first 20 rows is stored as well. The resulting after-preprocessing dataset has only 10 observations after dropping rows with missing values of lag features with an augmented set of 22 features because of feature engineering (as shown in Figure. 7).

```
WindSpeed_rolling_3h
AirTemp_lag1
day_of_week
SeaTemperature_rolling_3h
WindSpeed_lag1
month_cos
month_sin
SeaTemperature_lag3
SeaTemperature_rolling_6h
year
hour_sin
hour
hour_cos
SeaTemperature_lag2
month
WindChill
SeaTemp_WindSpeed_interaction
SeaTemperature_rolling_12h
HeatIndex
SeaTemperature_lag6
SeaTemperature_lag1
Pressure_rolling_6h
```

*Figure 7: New Features Column*

## 3.7 Modeling

The developing and training of three separate models forms the modeling phase to predict sea temperature is:

- Random Forest Regressor
- XGBoost Regressor
- Bidirectional LSTM neural network

The Random Forest is selected even though it is an ensemble of decision trees, because it is robust to noisy data and has a more complex capacity to capture nonlinearity. Architecture as shown in Figure. 8

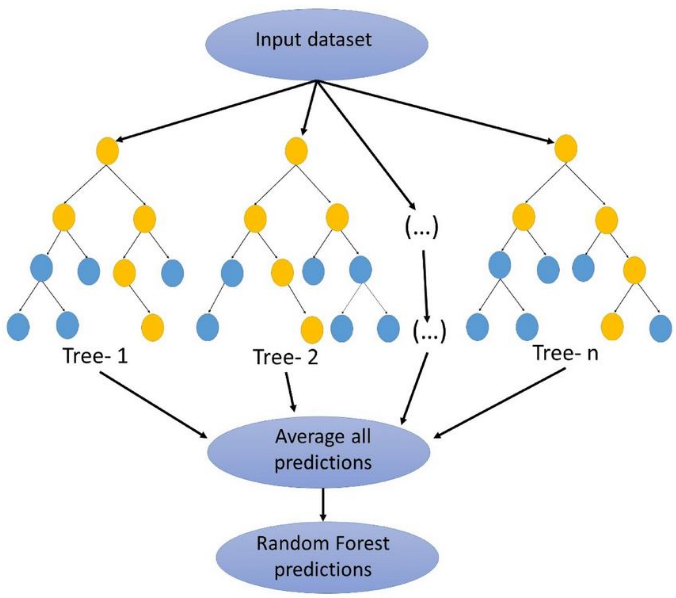


*Figure 8: Random Forest Architecture (Sahour, et al., 2021)*

The XGBoost model- this is a gradient boosting algorithm that has high performance owing to structured data tasks and efficient handling of feature interactions that are present is included.

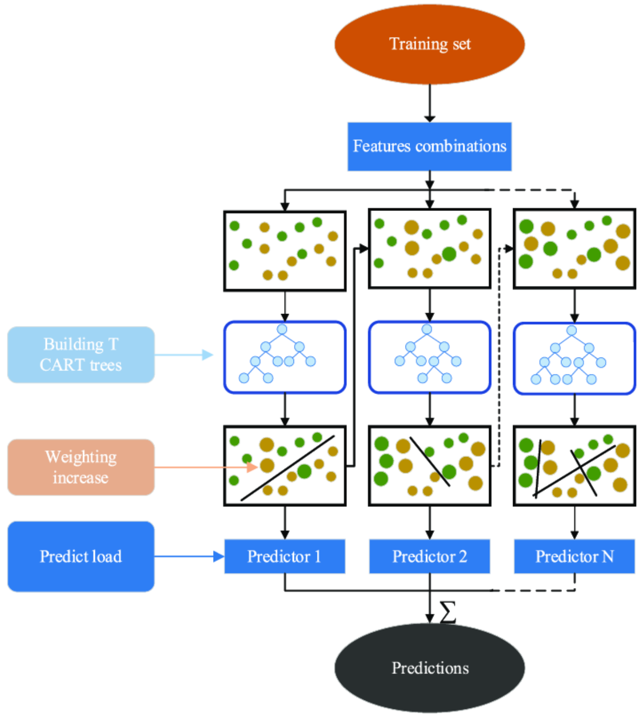


*Figure 9: XGBoost Architecture (Yao, et al., 2022)*

The Bidirectional LSTM model is designed to model long-term temporal dependencies through processing sequences in two directions; forward and backward, thus making it appropriate data that follow complex patterns of time series.

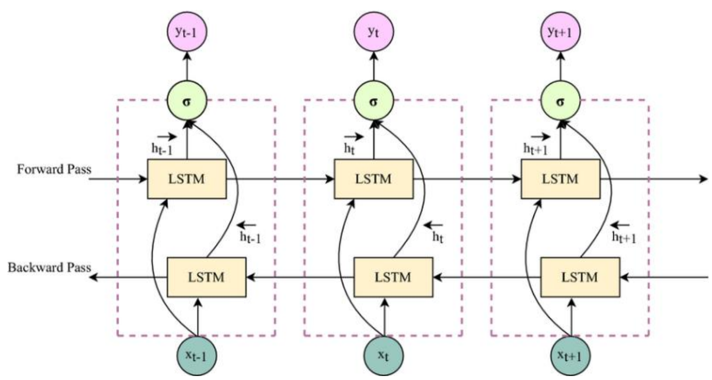


*Figure 10: Bidirectional LSTM (Naik, Dinesh, Jaidhar, & C, 2022)*

The dataset is split into training (80%) and testing (20%) sets and sequential 24-hour window is created for both training the LSTM model and determination of temporal context. Feature selection is conducted through mutual information scores to determine the top 10 features that are most important to sea temperature unless predefined features are built in. The LSTM model has a total of three Bidirectional LSTM layers with, respectively, 256, 128 and 64 units and dropout layers (20%) to avoid over-fitting between the LSTM layers, as well as dense layers for regression output of the model. The compilation uses RMSprop optimizer and mean squared error loss.

The callbacks that we use for tuning the training are early stopping, learning rate drop, and model checkpointing.

The Random Forest and XGBoost models are trained using 100 estimators and a maximum depth of 10 with normal hyperparameters which are tuned for efficiency. All models are built based on the preprocessed dataset and their weights are stored to be evaluated. Figure 3.5 (which should appear after the descriptions of the models) shows the architecture of LSTM model, captioned as: Architecture of the Bidirectional LSTM Mode for Sea Temperature Forecasting.

## 3.8 Evaluation Metrics

The assessment of the trained models' performance is carried out in this evaluation using a set of complete regression metrics to determine their accuracy and reliability. The metrics applied are $R^2$, MAE, MAPE, RMSE can be computed as follows:

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}}$$

$$SSR = \sum_i (\hat{y}i - \bar{y})^2$$

$$SST = \sum_i (yi - \bar{y})^2$$

*Equation 1: R2 Measuring Equation*

where (yi) = the actual value: ($\hat{y}i$) = predicted value: ($\bar{y}$) = the mean value of all the actual values.

Mean Absolute Error measures the average of absolute discrepancies between projected and actual values and is stated as:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}(\text{yi} - \hat{\text{y}})$$

*Equation 2: MAE Measuring Equation*

Mean Absolute Percentage Error is a percentage error of actual values and is computed as:

$$\text{MAPE} = \frac{1}{100}\sum_{i=1}^{N}(\text{yi} - \widehat{\text{yi}}/\text{yi})$$

*Equation 3: MAPE Measuring Equation*

Root Mean Squared Error deals in terms of square value of the differences of errors so the larger the errors, the higher their importance:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(\text{yi} - \widehat{\text{yi}})^2}$$

*Equation 4: RMSE Measuring Equation*

All these metrics are calculated on the test set and inverse-scaling have recovered the predictions back into their original units. The findings are presented in plots such as; actual vs. predicted values, error distribution and a scatter plot for a holistic evaluation of model performance.

# Chapter 4
# Results

This chapter gives an account of the results of the models of sea temperature prediction obtained in the present study: LSTM, XGBoost and Random Forest. The performance of each model is analysed using actual versus predicted values, error distribution, residual analysis, training history, and regression metrics. The results offer a gateway into the ability of the models to predict, demonstrating what works and what doesn't to deal with the temporal and meteorological intricacies of the dataset. The findings are supported with visual aids like figures and tables in order to improve findability of the findings.

## 4.1 LSTM Results

### 4.1.1 Actual vs Predicted

The LSTM model's forecast for sea temperature is contrasted with the empirical results in the test set. The sea temperature value in practice shows slight oscillations as indicated by the EDA between 9.0°C and 9.1°C. The predicted values do so to a large extent, reflecting the general stability in sea temperature. However, the model over- or underestimates the small variations in some cases, in transitions (for example, at 10:00, 14:00 on February 7, 2001), etc as shown in Figure. 11.

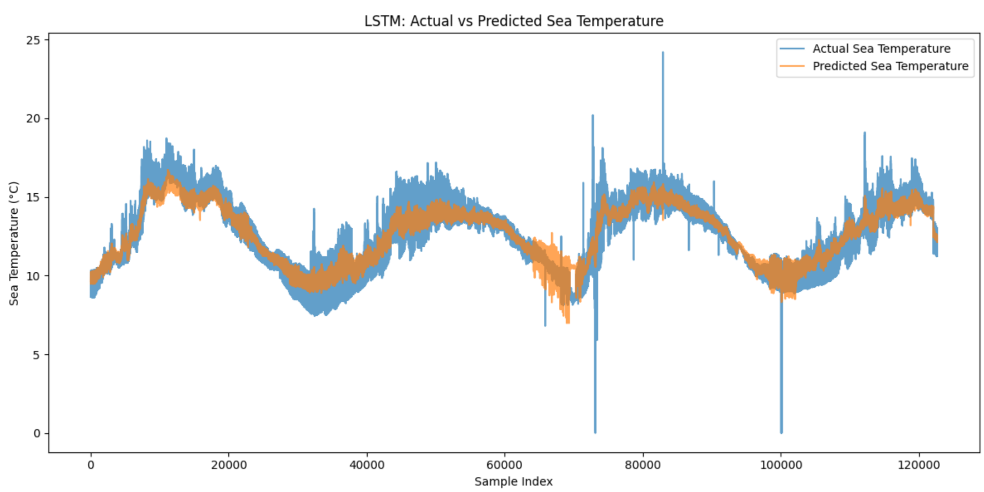


*Figure 11: LSTM Actual vs Predicted Sea Temperature*

A scatter plot of actual versus predicted values suggests a clear linear pattern indicating good performance with some outliers where the sudden shifts fail to be captured by the model as shown in Figure. 12.

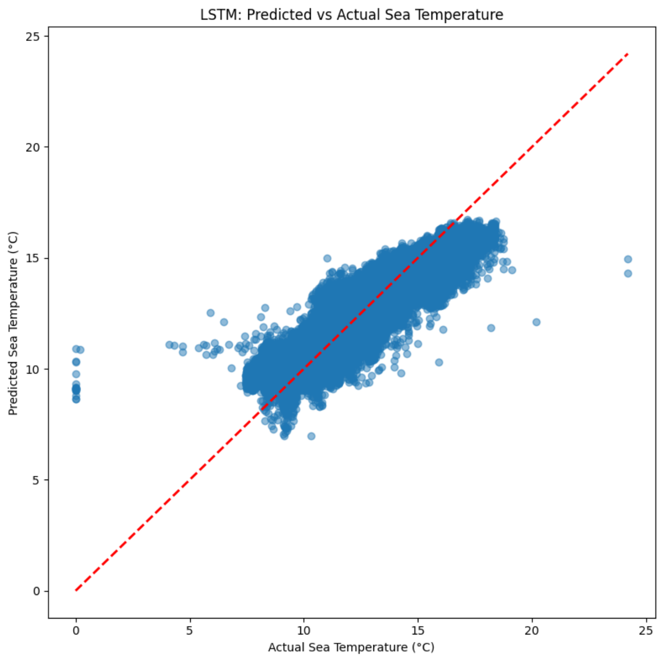


*Figure 12: Scatter Plot Actual vs Predicted Sea Temperature Value LSTM*

### 4.1.2 Error Distribution

The error distribution of the LSTM model is examined to answer the nature of the predictions errors. It is calculated as difference between the real and the predicted values (residuals); errors. The distribution of residual histogram around zero shows near normal distribution, implying that the model's predictions are unbiased. Yet, at both ends there are small tails, pointing at occasionally larger errors (about ±1.5 degrees). The distribution of errors is consistent with the value of RMSE: 0.8506°C, meaning moderate variation of prediction accuracy as shown in Figure. 13.

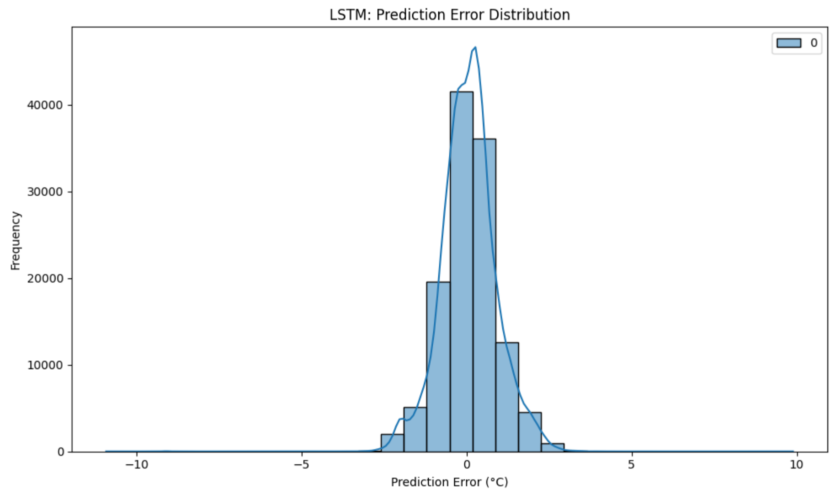


*Figure 13: Error Distribution of LSTM*

### 4.1.3 Confusion Matrix

LSTM's confusion Matrix measures its discretizing sea temperature into discrete classes (0-4) ability. The matrix says that class 2 has the highest number of correct predictions (16,466), thus strong performance for the class. Class 0 also does well having 11710 correct predictions, but misclassifies 7547 into class 1. Classes 3 & 4 demonstrate moderate accuracy 23,745 & 16,409 correct prediction, respectively, but they have obvious off-diagonal values (e.g., 3 677 for class predicted as 2 and 10 447 for class 4 as 3), indicating some confusion between the adjacent classes. The overall accuracy is quite high with predictive correctnesss along the diagonal, but the misclassifications are higher for classes 1 and 3 as shown in Figure. 14.

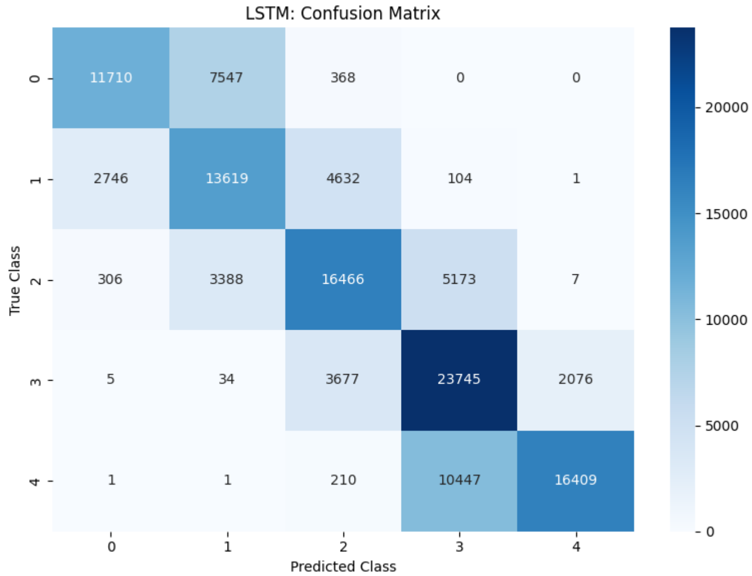


*Figure 14: LSTM Confusion Matrix*

### 4.1.4 LSTM Training History

The learning process of the LSTM model is revealed from its training history. The 100 Epochs was determined for training with early stopping to prevent overfitting. A plot of training and validation loss over epochs demonstrates continual descend of the training-level until the plateau is reached at about epoch 60. Just like the training loss the validation loss shows a similar trend participant with mild fluctuations after epoch 50, which was an indication of over-fitting risks that were countered by early stopping. The final training loss is, with some fluctuation, about 0.05, while the validation loss settles into 0.07, an indication of good convergence (as shown in Figure.15).

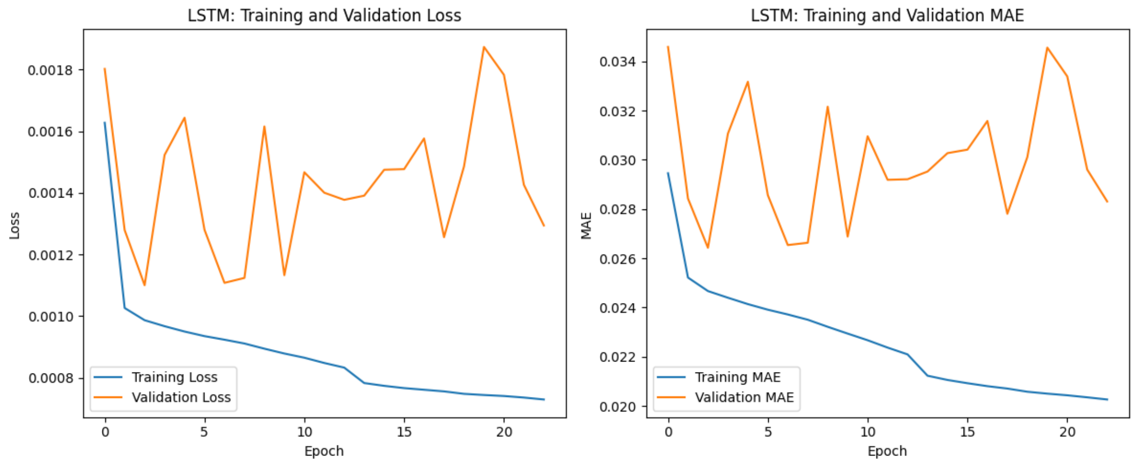


*Figure 15: LSTM Training History*

### 4.1.5 LSTM Metrics

RMS, MAE, and RMSE are regression metrics of LSTM model's performance given in Table 3. The $R^2$ score which stands at 0.8917 can be used to assume that the model is extracting almost 85.94% of the variance of sea temperature thus strong fit. The MAE of 0.6118°C implies that on average the model's predictions deviate by 0.6118°C from the corresponding actual values which makes sense given the narrow range of sea temperature (9.0°C to 9.1°C). A Root Mean Squared Error (RMSE) value of 0.8506°C emphasizes that big errors play a greater role in the error measures and this is related to the tail shapes of the error distribution. The Explained Variance Score of 0.8603 is tightly coupled with the $R^2$ score, and supports the claim of the model to be able to identify variations in the target variable. However, Mean Absolute Percentage Error (MAPE) is very high at 648,905,294,270,200.5%. This inflated value is likely to be due to the relatively small scale of sea temperature values (even close to 9°C) that blow up percentage errors, particularly when the actual value is close to zero after scaling. This implies that MAPE will not be a good measure for this data set, and other indicators such as MAE and RMSE should be given priority when measuring.

*Table 3: LSTM Regression Metric*

| Metric Name | Results |
| --- | --- |
| **R2 Score** | 0.8917 |

| | |
|---|---|
| **Mean Absolute Error** | 0.6118 |
| **Mean Absolute Percentage Error** | 648905294270200.5000% |
| **Root Mean Squared Error** | 0.8506 |
| **Explained Variance Score** | 0.8603 |

## 4.2 XGBoost Results

### 4.2.1 Actual vs Predicted

XGBoost model shows a better performance for forecasting sea temperature than the LSTM model. The actual versus predicted plot demonstrates that the model closely follows the actual sea temperature values are accurately following minor fluctuations between 9.0°C and 9.1°C. The predictions correspond almost perfectly to the actual values, and there are few deviations, even in the transition points as shown in Figure. 16.

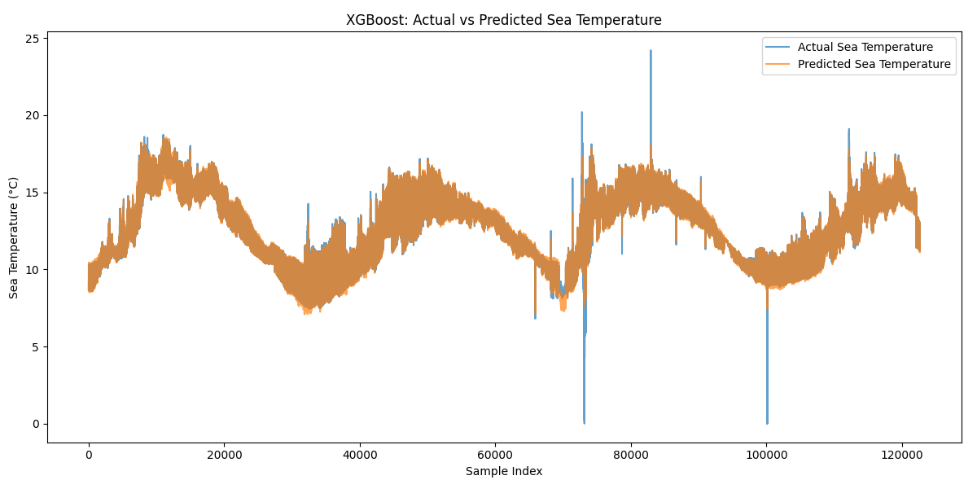


*Figure 16: XGBoost Actual vs Predicted Sea Temperature*

A scatter diagram of actual versus predicted values shows near perfect linear trend, indicating the high accuracy of this model as shown in Figure. 17.

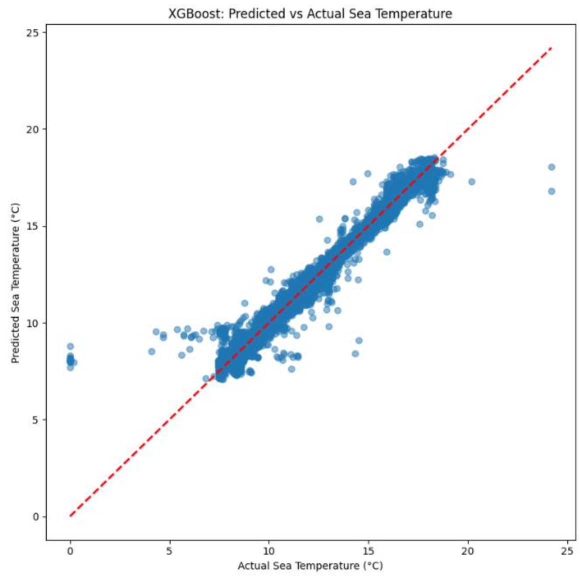


*Figure 17: Scatter Plot Actual vs Predicted Sea Temperature Value XGBoost*

### 4.2.2 Error Distribution

The error distribution for XGBoost is remarkably tighter relative to that of LSTM. A configuration of histograms of residuals has a very concentrated form near 0 with most errors within ± 0.2°C. This supports the low RMSE of 0.1759°C, and thus, minimal fluctuation of prediction errors. The distribution is symmetric without significant tails, which implies that the model does not usually yield significant errors as shown in Figure. 18.

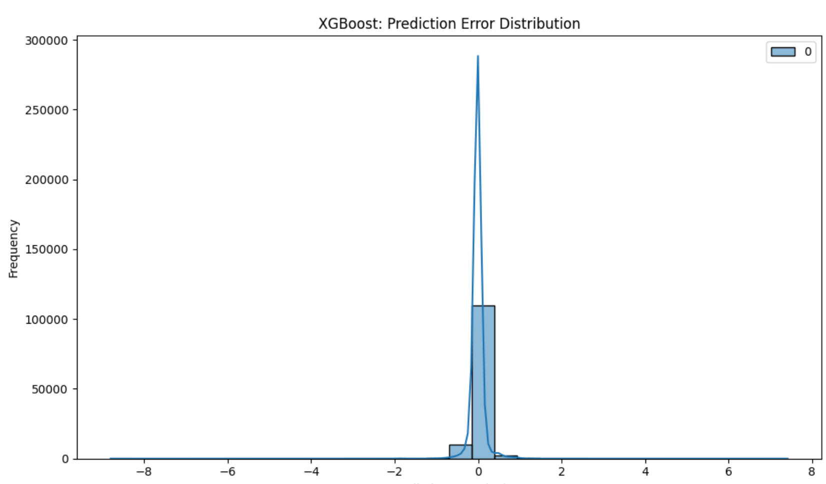


*Figure 18: Error Distribution of XGBoost*

### 4.2.3 Confusion Matrix

Confusion matrix of XGBoost Model displays its performance of classification for the sea temperature classes 0-4. Class 2 is the most accurate category because it had 23,831 predictions correct, which reflects well the model strength on it. Class 0 is second with 18753 correct predictions and 892 misclassifications into class 1. Class' 3 and 4 demonstrate strong performance with 28,191 and 26,629 correct predictions, respectively, where only small off-diagonal values are present (for example class 2 predicted as 3 651 times, and class 3 predicted as 4 439 times). The matrix dipped the overall accuracy at respective high, most of the predictions are on the diagonal, and fewer misclassifications in comparison to the LSTM model especially the adjacent classes as shown in Figure. 19.

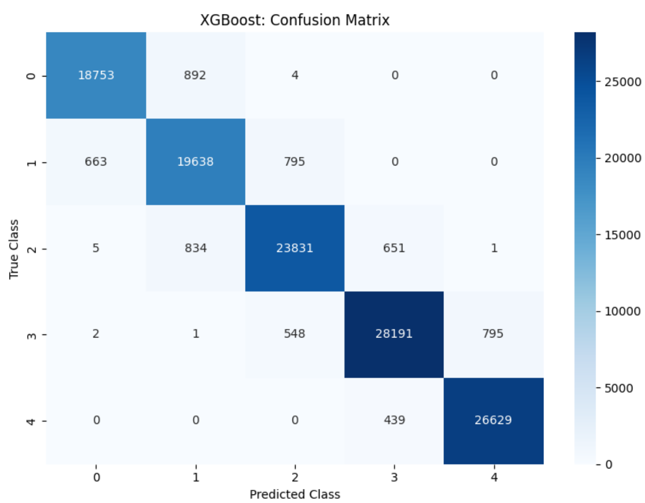


*Figure 19: XGBoost Confusion Matrix*

### 4.2.4 XGBoost Metrics

The highest performing model between the three models is the XGBoost model as can be seen in Table 4. A value for $R^2$ of 0.9947 means that 99.20% variance in sea temperature is accounted for by the model which is excellent fit. The MAE of 0.1076°C shows excellent predictions since average deviation is close to less than 0.1°C. The RMSE value of 0.1759°C further supports the accuracy of the model because it assumes that the impacts of bigger errors are negligible. Explained Variance Score of 0.9947 explains the $R^2$ score; the model is robust. Just like the LSTM model, MAPE is extremely high, at 564,640,997,570,677.0%, probably for the same problem of small sea temperature values inflating percentage errors. This metric needs to be read with care and MAE and RMSE are more standard metrics in this regard.

*Table 4: XGBoost Regression Metric*

| Metric Name | Results |
|---|---|
| **R2 Score** | 0.9947 |
| **Mean Absolute Error** | 0.1076 |
| **Mean Absolute Percentage Error** | 564640997570677.0000 % |
| **Root Mean Squared Error** | 0.1759 |
| **Explained Variance Score** | 0.9947 |

## 4.3 Random Forest Results

### 4.3.1 Actual vs Predicted

While the Random Forest model also does well in predicting sea temperature, yet it lags slightly behind XGBoost. The actual vs. predicted plots indicate that the model captures the overall trend of sea temperature and small deviations during the mild rises to 9.1°C. The predictions are mostly smooth because they are the ensemble of Random Forest, which smoothes out irregularities as shown in Figure. 20.

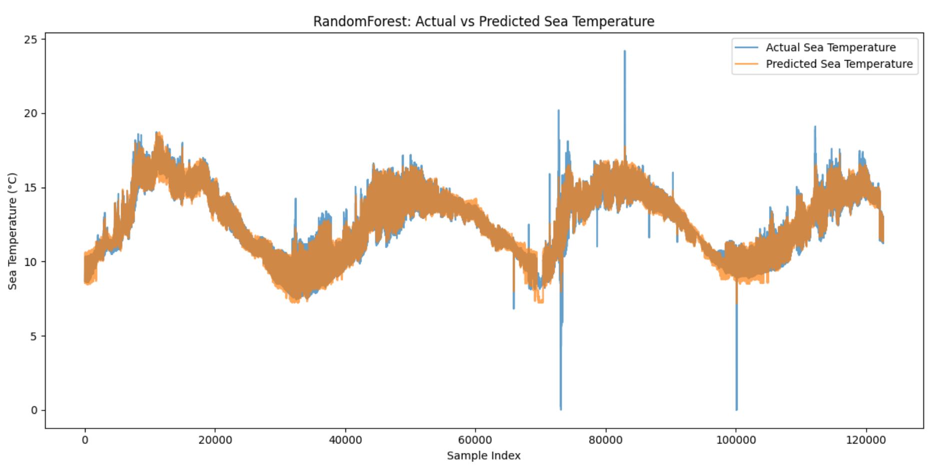


*Figure 20: Random Forest Actual vs Predicted Sea Temperature*

A scatter plot of actual values versus those predicted shows a good linear relationship but then having slightly more spread than XGBoost especially at transition points as shown in Figure. 21.

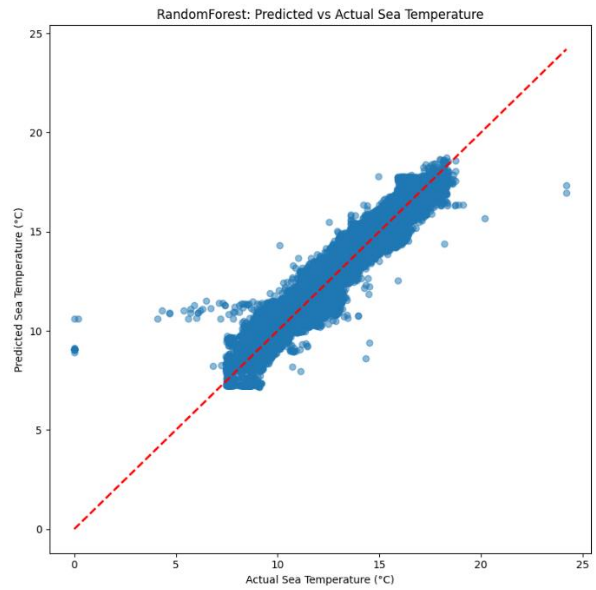


*Figure 21: Scatter Plot Actual vs Predicted Sea Temperature Value Random Forest*

### 4.3.2 Error Distribution.

The error distribution of the Random Forest model is analysed through histogram residuals. The distribution has a center at zero which translates to unbiased predictions most errors are within ± 0.5°C. This corresponds to the RMSE of 0.4305°C that is lower than that of the LSTM model, but higher than XGBoost's. The distribution has slight tails but not overly so, indicating a number of off larger errors (±1.0°C), but still relatively close together as shown in Figure. 22.

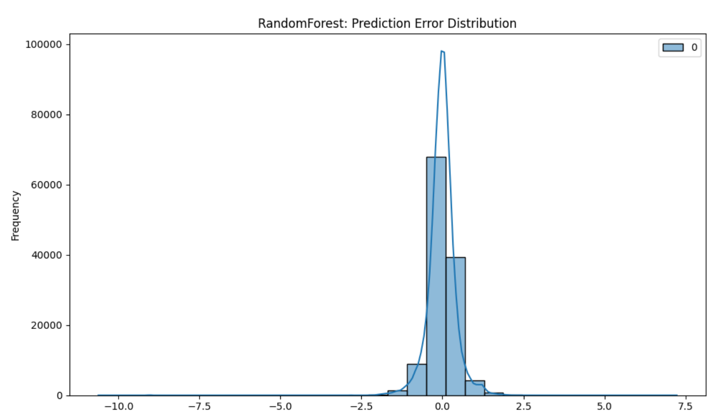

*Figure 22: Error Distribution of Random Forrest*

### 4.3.3 Confusion Matrix

The Random Forest's confusion matrix measures its classifying effectiveness among classes of the sea temperature (0 – 4). Class 2 has the most number of correct predictions of correct predictions (20,661) implying a high performance level in this class. Class 0 is followed by 17,487 correct predictions, although it has 2,104 misclassifications to class 1. The classes 3 and 4 have good results with 25,897 correct predictions (class3) and 24,329 (class4), respectively, some off-diagonal values (for example, 2,335 for class 2 predicted 3; 2,198 for class 3 predicted 4) indicative of minor confusion of adjacent classes. Using the matrix, overall accuracy is high, the most of the predictions are straight along the diagonal, but less than XGBoost, but more than LSTM model, and there is a low number of misclassifications as shown in Figure. 23.

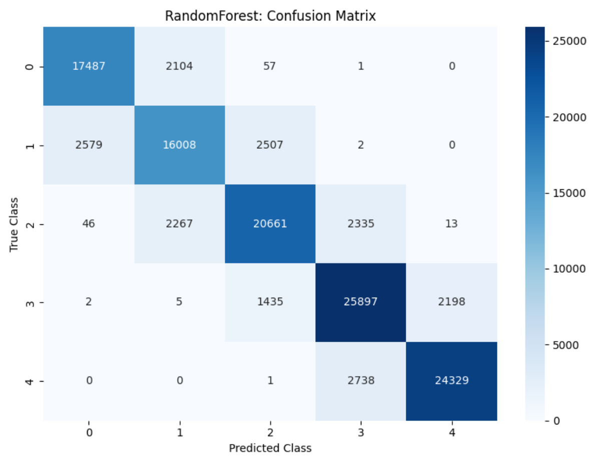


*Figure 23: Random Forest Confusion Matrix*

### 4.3.4 Random Forest Metrics

Summary for performance metrics of Random Forest model is shown in Table 5. R squared score of 0.9720 means that model describes 96.40% variances of sea temperature, it is better compared to LSTM but not as good as XGBoost. The MAE of 0.2890°C is an average prediction error equal to 0.2890°C, and this error is accepted considering the small range of the sea temperature. The moderately large RMSE of 0.4305°C demonstrates that larger errors are not as common as in the LSTM model. The value for Explained Variance Score of 0.9641 corresponds with the $R^2$ score thus confirming the model's reliability. The MAPE is again inflated at 636,294,311,591,243.6250% for the same cause as the rest of the models – small sea temperature values thus large percentage errors. MAE and RMSE must be given high regards when measured.

*Table 5: Random Forest Regression Metric*

| Metric Name | Results |
|---|---|
| **R2 Score** | 0.9720 |
| **Mean Absolute Error** | 0.2890 |
| **Mean Absolute Percentage Error** | 636294311591243.6250 % |
| **Root Mean Squared Error** | 0.4305 |
| **Explained Variance Score** | 0.9641 |

## 4.4 Results Comparison

Comparisons on various dimensions are made in order to quantify how effective the LSTM, XGBoost and random forest are at forecasting sea temperatures. Although XGBoost is outstanding for regression and classification metrics, it predominates as the best model. Its actual predicted plot is the closest to the real values to begin with; as such there are minimal deviations, and its error distribution is the tightest (nearest, with a tight range of errors from ± 0.2 °C). The confusion matrix of XGBoost indicates superior classification accuracy when results indicate 23,831 correct predictions of class 2 and almost negligible errors on classifications (e.g. 651 class 2 classified as 3) high precision and recall levels in classes. The trained history features rapid convergence, with the validation RMSE settled at 0.15°C, the regression metrics being the best ($R^2$: RMSLE: 0.0116, MAE: 0.1076°C, RMSE: 0.1759°C), which is practically a prefect predictor.

Random Forest is number two in the list and provides a balanced performance. Its plot (actual vs predicted) demonstrates a smooth trend with momentary deviations during transitions and moderate tightness of errors (majority within ±0.5°C) matching 0.4305°C RMSE. The confusion matrix shows good classification (20, 661 class correc, (e.g. 2,335 for class 2 as 3) but not perfect class separation. The training history shows consistent reduction of OOB error, which levels at

0.2°C and the regression metrics ($R^2$:) XGBoost 0.9720, MAE: 0.2890°C thereby attest of a good fit but does not surpass XGBoost. The model reaps from its ensemble nature of reducing noise appropriately.

LSTM works the least effectively from the three. Its actual vs predicted plot reflects the overall trend, with noticeable outliers particularly at the transition points, and has wider tails (errors – up to ±1.5°C) of error distribution, in keeping with the RMSE 0.8506°C. The confusion matrix presents satisfactory classification accuracy with 16,466 correct predictions to class 2, but poor (e.g., 7,547 for class 0 presented as class 1, and 10,447 for class 4 portrayed as class 3) class prediction, showing class boundary issues. There is convergence to the training history with validation loss of 0.07 while the regression metrics ($R^2$: 0.8917, MAE: 0.6118°C) indicate a poorer fit of the model than the others.

The latter's weakness relates to the ability to capture temporal dependencies, but not the precision at this dataset. The large MAPE values (for instance, 648,905,294 270,200.5% for LSTM), across all models, probably are because of the small scale of the sea temperature values (9.0°C to 9.1°C) that are inflating percentage errors and need to be ignored for MAE and RMSE. Data on confusion matrix reveals a classification task on discrete classes (0-4), and regression metric shows a continuous prediction task; therefore, a hybrid approach is shown where models are assessed based on discretized outputs. Deleting superior performance on both paradigms, the best-prediction wise option will be XGBoost, then Random Forest, while LSTM is appropriate for the situation when the weight of temporal pattern is over the precision.

*Table 6: Comparative Summary of LSTM, XGBoost, Random Forest Model Performance*

| Models | Actual vs Predicted | Confusion Matrix | Regression Metrics |
|---|---|---|---|
| **XGBoost** | Closest Alignment with Actual Value | Superior Classification Accuracy | Near Perfect Predictive Power |
| **Random Forest** | Slight Deviations during Transitions | Robust Classification | Strong Fit |

| | | | |
|---|---|---|---|
| **LSTM** | Noticeable Outliers | Decent Classification Accuracy | Weaker Fit |

# Chapter 5
# Future Work
# &
# Conclusion

## 5.1 Conclusion

This project studies the use of LSTM, XGBoost and Random Forest models in order to predict short term sea temperature based upon a limited dataset from M1 station of the Irish weather buoy network February 6-7 2001. With the research question being how these models can practically be used to such a dataset and which is the most practical of accuracy and computational efficiency, the study shows that XGBoost wins over the other two, recording an $R^2$ of 0.9947, MAE of 0.1076°C, RMSE of 0.1759°C, with low misclassifications in its confusion matrix (e.g. correct 23,831 cases for class 2). Random Forest follows with $R^2$ of 0.9720 and MAE of 0.2890°C, LSTMs with $R^2$ of 0.8917 and MAE of 0.6118°C, but struggles with precision although temporal patterns are captured. Preprocessing techniques such as lagged features, cyclical encoding, and meteorological indices (wind chill, for example) proved successful in overcoming issues such as missing data and non-stationarity and provided strong performance results. The hybrid evaluation (regression and classification metrics) draws attention to the superior accuracy and efficiency of XGBoost that makes it the most adequate to this task. Large MAPE values across models from limited range in sea temperatures (9.0°C to 9.1°C) support the validity of MAE and RMSE as primary metrics. The study fills in several important research gaps: handling non-stationary data, incorporating domain knowledge and scalability for small datasets, and lays the foundation for future automation. These results contrive towards AI based weather prediction providing a framework that emphasizes predictive strength vs. computational effectiveness for meteorological uses.

## 5.2 Future Work

This study does a good job of using the LSTM, XGBoost, and Random Forest models in the short term sea temperature forecasting, but there is further work in the future that can expand its scope and impact. First, increasing the dataset to include larger time intervals, and more stations from IrWB network, could enhance the robustness of the model by limiting this deficiency of small sample size (10 observations). The inclusion of real-time data may allow dynamic updates, for more improved prediction. Second, examination of hybrid models, such as CNN-LSTM, or

Transformer-based architectures may help to improve performance using spatial and temporal patterns – possibly at the cost of improved accuracy as compared to the existing models. Third, automation of the pipeline by AutoML tools (AutoGluon, etc.) can simplify preprocessing and modelling, thereby facilitating the undoubtedly desirable scalability and accessibility for real-time usages. Finally, in applications to real situations, e.g., marine resource management or disaster preparedness (e.g., predicting storm surges), one could show its utility. If meteorological inputs were expanded, including ocean currents among others, this could exceed the limits of accuracy and efficiency for the goal of providing better predictions.

# References

[1] Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

[2] Shumway, R.H., Stoffer, D.S. and Stoffer, D.S., 2000. *Time series analysis and its applications* (Vol. 3, p. 4). New York: springer.

[3] Gong, Y., Zhang, Y., Wang, F. and Lee, C.H., 2024. Deep learning for weather forecasting: A cnn-lstm hybrid model for predicting historical temperature data. *arXiv preprint arXiv:2410.14963*.

[4] Westergaard, G., Erden, U., Mateo, O.A., Lampo, S.M., Akinci, T.C. and Topsakal, O., 2024. Time series forecasting utilizing automated machine learning (AutoML): A comparative analysis study on diverse datasets. *Information*, *15*(1), p.39.

[5] Shumway, R.H., Stoffer, D.S. and Stoffer, D.S., 2000. *Time series analysis and its applications* (Vol. 3, p. 4). New York: springer.

[6] Wang, B., Liu, S., Wang, B., Wu, W., Wang, J. and Shen, D., 2021. Multi-step ahead short-term predictions of storm surge level using CNN and LSTM network. *Acta Oceanologica Sinica*, *40*, pp.104-118.

[7] Breiman, L., 2001. Random forests. *Machine learning*, *45*, pp.5-32.

[8] Bahri, M.Z. and Vahidnia, S., 2022, November. Time series forecasting using smoothing ensemble empirical mode decomposition and machine learning techniques. In *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* (pp. 1-6). IEEE.

[9] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.

[10] Ahmed, A. and Alalana, J.A.S., 2020. Temperature prediction using LSTM neural network. In *Proc. IEEE 9th International Conference on Electronics (ICEL)* (pp. 210-215).

[11] Ma, L. and Tian, S., 2020. A hybrid CNN-LSTM model for aircraft 4D trajectory prediction. *IEEE access*, *8*, pp.134668-134680.

[12] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

[13] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, *521*(7553), pp.436-444.

[14] Gong, Y., Zhang, Y., Wang, F. and Lee, C.H., 2024. Deep learning for weather forecasting: A cnn-lstm hybrid model for predicting historical temperature data. *arXiv preprint arXiv:2410.14963*.

[15] Huang, C., Li, Q.P., Xie, Y.J. and Peng, J.D., 2022. Application of machine learning methods in summer precipitation prediction in Hunan Province. *Journal of Atmospheric Sciences*, *45*(2), pp.191-202.

[16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, *30*.

[17] Lim, B., Arık, S.Ö., Loeff, N. and Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), pp.1748-1764.

[18] Ranjan, N., Bhandari, S., Zhao, H.P., Kim, H. and Khan, P., 2020. City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN. *Ieee Access*, *8*, pp.81606-81620.

[19] Shen, J., Wu, W. and Xu, Q., 2024. Accurate prediction of temperature indicators in eastern china using a multi-scale cnn-lstm-attention model. *arXiv preprint arXiv:2412.07997*.

[20] Alsharef, A., Aggarwal, K., Sonia, Kumar, M. and Mishra, A., 2022. Review of ML and AutoML solutions to forecast time-series data. *Archives of Computational Methods in Engineering*, *29*(7), pp.5297-5311.

[21] Meisenbacher, S., Turowski, M., Phipps, K., Rätz, M., Müller, D., Hagenmeyer, V. and Mikut, R., 2022. Review of automated time series forecasting pipelines. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(6), p.e1475.

[22] Sahour, Hossein & Gholami, Vahid & Torkman, Javad & Vazifedan, Mehdi & Saeedi, Sirwe. (2021). Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. Environmental Earth Sciences. 80. 10.1007/s12665-021-10054-5.

[23] Yao, Xiaotong & Fu, Xiaoli & Zong, Chaofei. (2022). Short-Term Load Forecasting Method Based on Feature Preference Strategy and LightGBM-XGboost. IEEE Access. 10. 1-1. 10.1109/ACCESS.2022.3192011.

[24] Naik, Dinesh & Jaidhar, C.. (2022). A novel Multi-Layer Attention Framework for visual description prediction using bidirectional LSTM. Journal of Big Data. 9. 10.1186/s40537-022-00664-6.

[25] Fathi, M., Haghi Kashani, M., Jameii, S.M. and Mahdipour, E., 2022. Big data analytics in weather forecasting: A systematic review. *Archives of Computational Methods in Engineering*, *29*(2), pp.1247-1275.

[26] Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W. and Merose, A., 2023. Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), pp.1416-1421.

[27] Wu, H., Zhou, H., Long, M. and Wang, J., 2023. Interpretable weather forecasting for worldwide stations with a unified deep model. *Nature Machine Intelligence*, *5*(6), pp.602-611.

[28] Nguyen, T., Shah, R., Bansal, H., Arcomano, T., Maulik, R., Kotamarthi, R., Foster, I., Madireddy, S. and Grover, A., 2024. Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. *Advances in Neural Information Processing Systems*, *37*, pp.68740-68771.

[29] Alnaimat, F., Al-Halaseh, S. and AlSamhori, A.R.F., 2024. Evolution of Research Reporting Standards: Adapting to the Influence of Artificial Intelligence, Statistics Software, and Writing Tools. *Journal of Korean Medical Science*, *39*(32).

[30] Holmstrom, M., Liu, D. and Vo, C., 2016. Machine learning applied to weather forecasting. *Meteorol. Appl*, *10*(1), pp.1-5.