Vaasan yliopisto
UNIVERSITY OF VAASA

Wubshet Solomon

A Comparative Analysis of the Use of Deep Learning and Machine Learning in Weather Forecasting: Using Meteorological Dataset on Vaasa

School of Technology and Innovations
Master's thesis in
Industrial Systems Analytics

Vaasa 2023

UNIVERSITY OF VAASA

School of Technology and Innovation

Author: Wubshet Solomon

Title of the thesis: A Comparative Analysis of the Use of Deep Learning and Machine Learning in Weather Forecasting: Using Meteorological Dataset on Vaasa

Degree: Master of Science in Technology

Programme: Industrial System Analytics

Supervisor: Assistant Professor Emmanuel Ndzibah

Year: 2023 **Pages:** 50

ABSTRACT:

This study presents a comparative analysis of two prominent technologies, namely deep learning, and machine learning, in the context of weather forecasting. The main research question is "How can machine learning and deep learning algorithm be implemented to obtain near-accurate weather forecasting"?

The objectives of this research are identifying the fundamental differences between deep learning and machine learning algorithms handling weather-related dataset and to ascertain the accuracy of using deep learning as compared to machine learning in weather forecasting. The study begins by providing a detailed overview of deep learning and machine learning techniques, explaining their fundamental principles, and highlighting their respective implementation in weather dataset.

In addition, the focus of the research is on the application of technologies such as polynomial regression, gradient boosting, neural prophet, and recurrent neural network models to the process of weather forecasting. The study applied quantitative methodology and used an open-source dataset from Finnish Meteorological Institute which is a weather record collected from the city of Vaasa. The comparative analysis involves employing those techniques to capture nonlinear relationships between weather variables and the pattern within the dataset. Moreover, the study investigates the performance of each technology and evaluates its effectiveness in forecasting weather conditions over different interval of time using performance evaluation matrices.

The outcomes of the comparative analysis provide valuable insights into the application of recent machine learning and deep learning methods with regard to the quality and the amount of data applied for the process. This includes proper implementation of data pre-processing techniques, that significantly impact the accuracy of models.

KEYWORDS: Deep Learning, Machine Learning, Weather Forecasting,

Contents

List of Figures

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Networks
AR	Auto Regression
DL	Deep Learning
ML	Machine Learning
NWP	Numerical Weather Predictions
ANN	Artificial Neural Network
NN	Neural Network
WF	Weather Forecasting
RNN	Recurrent Neural Network
LSTM	Long short-term memory
MAE	Mean Absolute Error
MSE	Mean squared Error.
RMSE	Root Means squared Error

1 Introduction

1.1 Background of the study

Weather forecasting is an intensive process that involves collections and analysis of atmospheric observations based on location and time (Chen et.al. ,2022). As a crucial element of daily human activities, traditional forecasting techniques have rapidly transformed into data-driven technologies. One of the pioneer mathematical models in this field is Numerical Weather Prediction (NWP), which aims to translate hydrodynamic activities in the atmosphere using collection of equations (Rozas ,2019). These equations iteratively process current weather observations to forecast future weather conditions. Despite the success of NWP, the output retains uncertainty due to the equations used in the method (Cho et al,2022). Moreover, the NWP method has encountered challenges in understanding the patterns of observation data. Additionally, high-performance computing resources are needed to process the massive amount of data required for accurate predictions (Ren et al.,2021).

ML and DL techniques are increasingly being applied in weather forecasting and significant progress has been made in addressing challenges such as handling large datasets, improving computational capabilities, and increasing prediction accuracy (Schultz et al.,2020).

ML algorithms are designed to train a dataset and predict the future considering the behavior trained from the input data. There are several ML techniques used for weather prediction, such as regression and random forest which are popular choices.
DL algorithms, on the other hand, involve training huge datasets using neural networks that mimic the structure of the human brain.

DL is particularly suitable for capturing complex and non-linear relationships in weather data, which makes it powerful technique for improving weather forecasting accuracy.

This research paper aims to review distinct ML and DL technologies applied in weather forecasting, explain the theoretical background of these technologies ,how to handle the process .Moreover , it will evaluate the performance of both technologies in terms of accuracy ,precision and different scores.

1.2 Research Gap, Questions and Objectives

Based on the main key words used in this research, different scientific papers reviewed to find out the research gap. Different academic publications databases used for searching relevant resources and some articles and journals which published within the last five years from IEEE database listed in the Table 1

Table 1. Research gaps

Keywords	Timeline 2017 - 1022	Database	Hits	Description
Machine Learning & Deep Learning, Weather Forecasting	2021	IEEE	3	Air Temperature Forecasting using Traditional and Deep Learning Algorithms (Li et.al,2020)
	2021	IEEE	69	Air Temperature Forecasting using Traditional and Deep Learning Algorithms (Chengsi et.al,2021)
	2016	IEEE	176	Weather forecasting using deep learning techniques (Ayman et.al

				2021)
	2022	IEEE	47	Rainfall Prediction using Different Machine Learning and Deep Learning Algorithms (Mahadware et al.,2022)
	2021	IEEE	55	Forecasting of Temperature by using LSTM and Bidirectional LSTM approach: Case Study in Semarang, Indonesia (Nizar et al.,2021)

According to Chengsi et.al. (2021), weather forecasting requires enormous amounts of data as input. In addition, the dynamic nature of the collected data resulting more complex behavior during interpretation of this data to thoughtful conclusions, this significantly affects the accuracy of the prediction. The emerging of ML technologies play magnificent role in discovering the hidden patterns in massive data processing and produce near-accurate prediction

DL strategies have an extraordinary capacity to investigate and grasp subtle patterns contained within enormous and complex datasets, which ultimately results in the production of results that are dependable and correct. Nowadays, DL has shown tremendous breakthroughs in the area weather forecasting, which has allowed them to better serve their customers (Ayman et.al 2021).

The research will apply selected ML and DL technologies on the sample data and evaluate the accuracy of each method using different performance matrices. Based on the results obtained during the process, the study targets to answer the following research questions.

- RC1: How can different deep learning and machine learning algorithms be measured, analyzed, and compared using performance matrices based on the dataset of the Finnish Meteorological Institute on Vaasa?
- RC2: How can such an algorithm be implemented to obtain near-accurate weather forecasting.

The main objectives of this study are:

1. To identify the fundamental differences between deep learning and machine learning algorithms handling time series datasets, specifically weather-related dataset.
2. To identify the pros and cons of deep learning and machine learning for weather forecasting.
3. To ascertain the accuracy of using deep learning as compared to machine learning in weather forecasting.

1.3 Definitions and Limitations

A weather dataset is an accumulation of meteorological data that comprises different atmospheric characteristics collected at places over a period. These parameters are recorded at specific locations over some interval of time. The science of meteorology makes use of these datasets for a wide variety of purposes, including different scientific studies, forecasting weather, analyzing environmental conditions, and other similar endeavors.

Weather datasets applied for future prediction based on numerical weather prediction models, which contain innate errors. The uncertainty nature derives from the complicated structure of weather systems, model assumptions, and limits in input data assimilation. However, the study limited to models that can learn the pattern of dataset instead of using a set of mathematical rules, that is numerical weather predictions.

Weather forecasting is the process of predicting atmospheric conditions of the future for a specific location and time using scientific techniques (Singh and Chaturvedi ,2019). The raw data used for weather prediction have time series presentation format, that obtain repeated value over time (Afteniy,2021). In recent years, there are enormous amounts of technologies which have developed to handle such prediction effectively.

There is still a degree of uncertainty connected with weather forecasting, even though contemporary systems for making forecasts have substantially increased their accuracy. It may be challenging to provide an accurate forecast of meteorological events that occur on a smaller scale, such as thunderstorms or enormous amount of rainfall. In addition, short term and long-term forecasting are uncertain and challenging due to the complex relations of numerous atmospheric processes and the limitations models.

According to Soori et.al. (2022), machine learning defined as "a technology that represents important evolution in computer science and data processing systems which can be used in order to enhance almost every technology enabled service, product and industrial applications ". In addition, it learns from the data and draws a pattern that is used for prediction or classification. In the case of weather forecasting, ML be trained from enormous amounts of meteorological data to improve their ability to predict the weather. The algorithms behind these technologies are intelligent to anticipate the weather, by analyzing this dataset and discovering patterns and associations.

To provide reliable forecasts, machine learning algorithms need vast volumes of high-quality data. In addition, there is a possibility that the data coming from various sources contains errors or inconsistencies. These limitations have an impact on the accuracy of ML models.

Deep Learning is one of subspecialized machine learning technology that consist of neural network, resembles human brain structure, with multiple layers (LeCun et al.,2015). The layers are between neural network and pass the processed data to the next network until the data capture the required feature. The number of layers and neural networks on the structure depend on the volume data. From the perspective of weather predictions, DL describes the process of using neural network models that have numerous layers to evaluate and predict weather trends. These models strive to identify complicated links and patterns in the data, which gives them the ability to create forecasts based on previous observations.

The training procedure for DL models is time-consuming and expensive, which is one of the limitations of these models. Because of these models' intricate design and extensive list of parameters, significant computing resources, such as high-powered graphics processing units or specialized hardware, are necessary. It takes a substantial amount of time to train DL models on big meteorological datasets, which may restrict their ability to provide accurate real-time predictions.

1.4 Research process

This study focuses on analyzing the performance of selected ML and DL technologies that implement data driven techniques. The theoretical background of these technologies has been reviewed from different scientific papers. The study applies quantitative methodology, and the research process is divided into different phases as listed below.

- Fetching Input Data

  The dataset used for this research was retrieved from Finnish Meteorological institute and describes hourly record of minimum and maximum temperature for the city of Vaasa.
- Pre-processing Data

  The first step includes converting the raw input data into a proper format and data type according to specification of ML and DL algorithms. In addition, null values, duplicate values, and error values are handled and dropped. For this purpose, one of the python data manipulation and analysis libraries, Panda, will be applied throughout this task.

- Splitting Data

  The pre-processed data is divided into two independent datasets, namely the training, and testing sets. The training dataset covers 70 -80 % of the whole dataset which is used to train the algorithm to learn the behavior of data and build a new model to make prediction for a given scenario. The testing dataset that contains the rest part of the dataset is used to evaluate the performance of the newly created model.

- Building Model

  After suitable cleaning and preparation of the dataset, the training dataset feed into selected ML and DL algorithms. It results a new model which forecast the temperature for given time.

- Testing Model

  The testing dataset is used here to check the performance of the models. The accuracy level of the model is also calculated using this dataset.

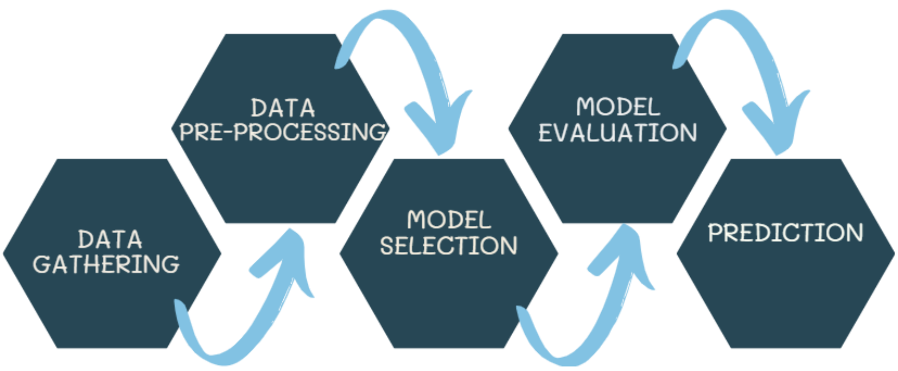

Figure 1. Research Process

1.5 Structure of the study

This research paper contains five chapters, each of which sequentially explains the topic, starting from its background to the conclusion. The detailed description of each chapter is presented as shown in the following lists.

- Chapter 1: Introduction

  This chapter starts with a brief description of the study background and explains the research gap, research questions and main research objectives respectively. It also includes the definition of main key words, limitation of the study and the research process.
- Chapter 2: Literature review

  This part includes detailed literature review of ML and DL technologies related to weather forecasting.
- Chapter 3: Methodology

  This chapter deals with four different methods representing ML and DL algorithms according to steps described in the research process.
- Chapter 4: Result and Discussion

  This part is dedicated to answering the research questions and explain the result obtained using the methods described in previous chapter.
- Chapter 5: Conclusion

  This is the last part of the study and present summarization of the result based on research objectives defined in the first chapter.

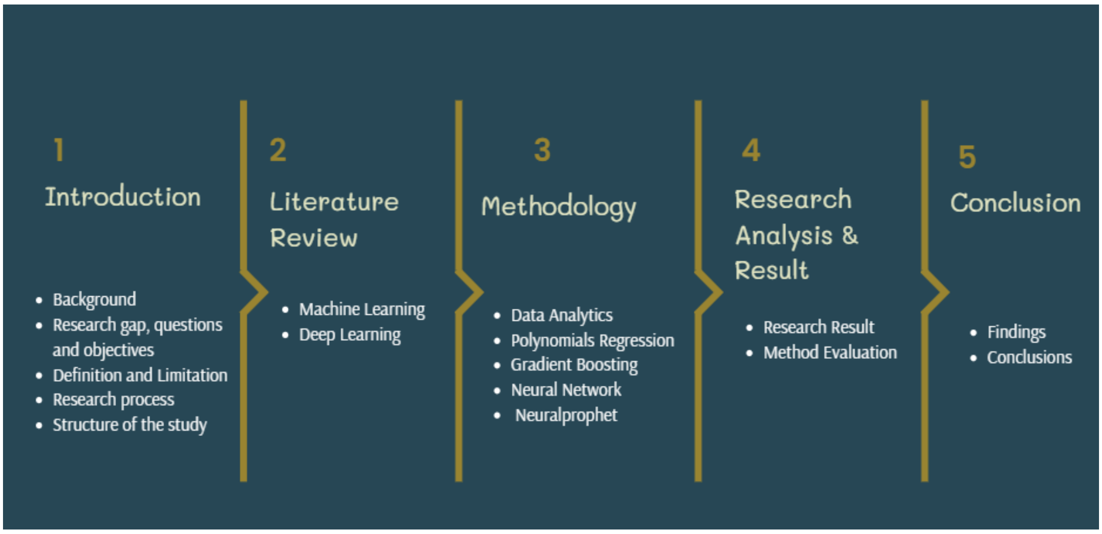

Figure 2 Study Structure

2 Review Literatures

ML and DL have made tremendous advancement in time series forecasting over the past few years. These technologies enable scientific forecast based on historical time-stamped observational data. This chapter is dedicated to discuss relevant academic literature which focuses on how machine learning and deep learning technologies approach and implemented for weather forecasting.

In ML, input data is processed to extract relevant collection of features that are used to train model. The model then learns how to map features with the desired output, using statistical techniques such as classification, regression, or clustering. The performance of the model tested against a test dataset using evaluation method. On the other hand, DL, involves feeding the input data directly into a network of nodes that is made up of several layers. The NN extracts features from input data in each and uses it to make predictions or classification. The network is then trained by using an algorithm that optimizes the weights of the node in order to minimize the difference between the predicted and actual output (Hinton ,2015).

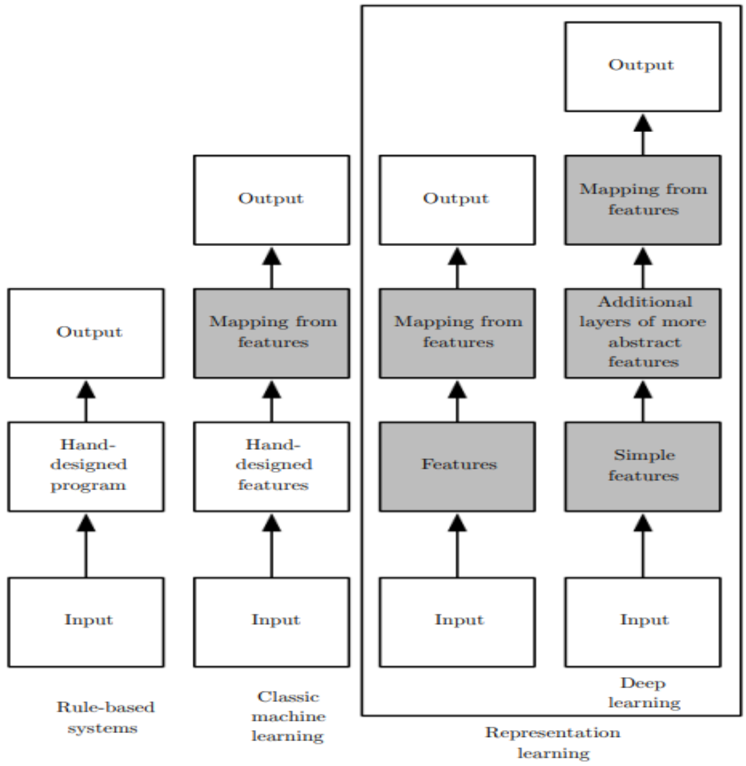

Figure 3 Flow chart

“Flow chart shows how the different parts of an AI system relate to each other within different AI disciplines. Shaded boxes indicate components that are able to learn from data” (Goodfellow et al ,2016)

2.1 Machine Learning for Weather forecasting

Machine learning has shown tremendous promise in terms of increasing the accuracy of weather forecasting. Machine learning algorithms able to generate forecasts and give useful insights into future weather conditions because uses past observations on weather conditions and training models to understand complicated relationships and correlations (Holmstrom et al., 2016).

In recent years, research on weather forecasting applying ML has been broadly increasing in all sectors of science.ML technology combine mathematical techniques with prior knowledge to enhance its performance to generate precise forecasts. The experience refers to the past information available to the learner, that typically takes the

form of digital data collected and made available for analysis. This data could be in the form of digitized human-labeled training sets, or other types of information obtained via interaction with the environment. The quality and size of this data are critical to the success of the predictions made by the ML model (Mohri et al. ,2016).
Machine learning is classified into three categories, namely supervised, unsupervised, and reinforcement learning, based on the learning algorithm, input data type, and problem type to be addressed (Sah,2020).

2.1.1 Supervised Learning for weather for forecasting

Supervised learning has been extensively utilized for weather forecasting, with the goal of using past meteorological data to educate prediction models, has seen widespread adoption. This method makes it possible to create precise models that are able to provide predictions depending on the characteristics that are fed into them. The use of supervised learning methods in weather prediction has been the subject of several research, which has shown the efficiency of these methods in capturing historical trends and boosting prediction accuracy (Brown et al., 2019).

In supervised ML, the given weather data is a combination of labels $\{(X,Y)\}_{i=1}^{N}$. A feature vector is a collection of all the element X_i among N, in which each of the elements, i = 1, ..., N, has a value that in some way characterizes the sample. This value is known as a feature and is represented by the symbol X (i). The label Y_i can be element of any fixed set of classes, that used to categorize the element belong to a feature vector X_i.
The main objective of supervised learning approach is to generate a model from the dataset that accepts a feature vector as an input and outputs a model that can be used to infer a label for the feature vectors it takes as input (Burkov, 2019).

Supervised learning algorithms can be further grouped into classification and regression problems.

1 Classification: It is the process of identifying the class to which a new data point belongs, based on a dataset that already contains observations with known class membership. Classes are commonly known as targets or labels and serve as categories for grouping items. For instance, the process of detecting spam in email service providers entails binary classification, which involves solely two classes (Campesato, 2020).

The field of machine learning covers various classification algorithms, which enumerated as follows (Campesato, 2020).

- Decision trees: It is one of classification algorithm that utilizes a structure resembling a tree. In addition, the positioning of a data point is established through uncomplicated conditional reasoning.
- Random Forests: Considered as an extension of decision trees, wherein the classification process requires the use of multiple trees, the quantity of which is predetermined by the user.
- kNN (k Nearest Neighbour): It is a classification technique, that classification of data points into the same class is determined by their proximity to one another. Upon the introduction of a novel point, it is assigned to the same class as most of its closest neighbours.
- Logistic regression: It is a statistical method that serves as both a classifier and a linear model, producing a binary output. Its' method deals with multiple independent variables and utilizes a sigmoid function to compute probabilities.
- Naïve Bayes: It is a type of probabilistic classifier that draws inspiration from the Bayes theorem. The Naive Bayes classifier operates under the assumption of conditional independence among attributes and has demonstrated effective performance even in cases where this assumption is not strictly upheld. This claim significantly diminishes computational expenses and constitutes a straightforward algorithmic implementation that solely necessitates linear time.
- SVM (Support Vector Machines): Apply to a supervised machine learning algorithm that is capable of addressing classification or regression problems. Sup-

port Vector Machines (SVM) have the capability to operate with data that is not only linearly separable but also nonlinearly separable.

2 Regression: The linear regression algorithm is widely used in regression analysis to learn a model that is a linear combination of input features (Burkov,2019). The objective of linear regression is to determine the optimal line of best fit that accurately reflects a given dataset. It is imperative to bear in mind two fundamental aspects. The optimal regression line may not necessarily intersect with the majority, or all, of the data points within the dataset. The objective of determining a best fitting line is to reduce the vertical deviation between said line and the data points within the dataset. It should be noted that linear regression is not capable of determining the optimal polynomial fit. This task requires the identification of a polynomial of higher degree that intersects with a significant number of data points within a given dataset (Campesato, 2020).

Moreover, it is possible for a dataset within a two-dimensional plane to comprise of two or more points that are situated on a common vertical line. This implies that these points share an identical x value. It is important to note that a function is incapable of passing across a pair of points if two points, namely (x_1,y_1) and (x_2,y_2), share the same x value. In such cases, it is imperative that the y value of both points be identical (i.e., y_2=y_2). Conversely, it is possible for a function to exhibit multiple points that are situated on a common horizontal axis.

2.1.2 Unsupervised Learning for weather forecasting

Unsupervised learning is widely use in weather forecasting to find hidden patterns within meteorological data without considering the presence of complex patterns and lacking target variables that have been identified explicitly. However unsupervised learning does not directly provide predictions, it can nevertheless give insightful information and help with feature mining, irregularity identification, and clustering for weather analysis (Lin et al., 2019).

Unsupervised learning involves utilization of techniques for the purpose of detecting trends within data sets that do not possess any classification or labeling of data points. The algorithms possess the capability to classify, label, and group data points within datasets autonomously, despite any external direction (Dridi, 2021).

Unsupervised machine learning methods are utilized when a target feature is not present, and instead, the model fundamental structure inherent in the descriptive features of a given dataset. The previously framework is commonly represented through newly created characteristics that can be added to the initial dataset, thereby enhancing, or supplementing it (Kelleher et al. ,2020).

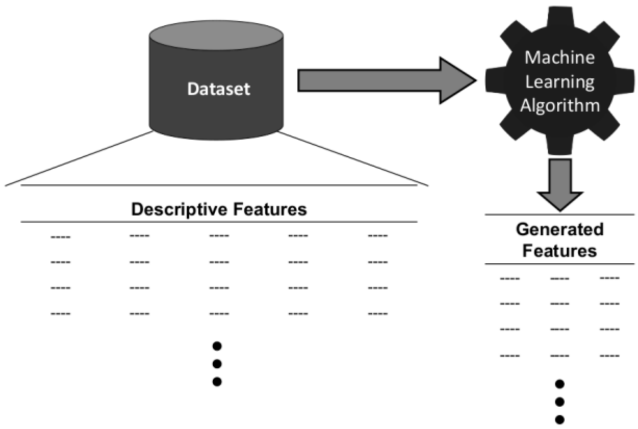

Figure 4 Unsupervised ML as a single-step process

(Kelleher et al. ,2020)

Clustering is one of the unsupervised learning algorithms that entails the utilization of a distance metric and the iterative relocation of comparable entities in closer proximity. Upon completion of the process, the items that exhibit the highest density clustering around n centroids are deemed to be categorized within that particular group. K-means clustering is a well-known variant of clustering within the field of machine learning (Patterson & Gibson, 2017).

K-Means clustering, and hierarchical clustering are two widely recognized unsupervised clustering algorithms. The K-means clustering technique is a well-established method for clustering and is considered a prominent example of unsupervised learning. Due to its straightforward concept, superior efficiency, and uncomplicated execution, this approach has garnered extensive utilization across various domains. (Chong,2021).

2.2 Deep Learning for weather forecasting

Due to its capability of automatically capturing complicated patterns and obtaining sequential correlations in dataset, deep learning has gathered a substantial amount of interest in weather forecasting in recent years. The continuous growth of meteorological data in volume, contribues to the envolement of intelligent technologie, starts to play significant role in the weather forecasting (Chen et al.,2019).

The technique of deep learning for image analysis and recognition is utilized extensively in the identification of atmospheric radar and satellite cloud images, as well as in the prediction of inversions that will occur later. This results in obtaining automatic observation of metrological phenomena (Chen et al.,2022).

DL has obtained huge popularity in recent years due to its capacity processing enormous amount of data and produce near-to-accurate prediction output. According to Ekman (2021)," DL is a class of machine learning algorithms that use multiple layers of computational units where each layer learns its own representation of the input data". The fundamental building block of DL is ANN, that simulate biological neurons present in human brain. These networks consist of billions of interconnected neurons through synapses, that exchange electrical signals by adding values to the input received. The activation function determines the activation status of each neuron by computing the total weight plus the constant called bias, that turn the activation function to the positive or negative part (Géron 2022).

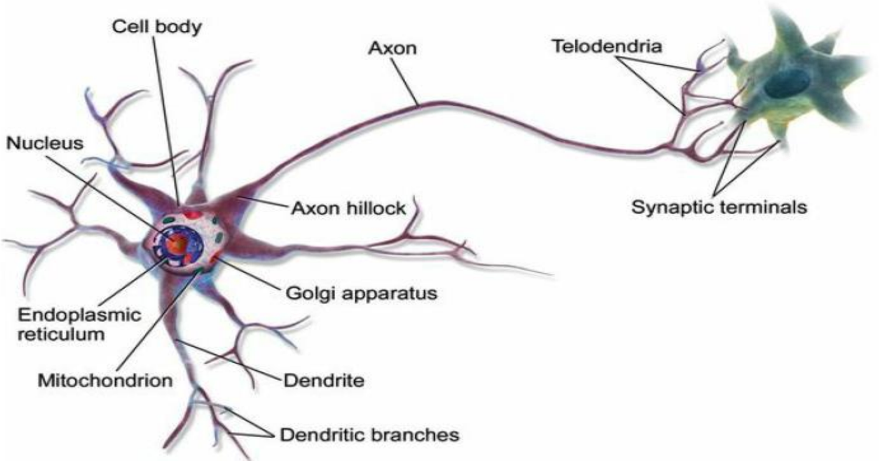

Figure 5 Biological Neuron (Géron 2022)

DL transform conventional ML to more efficient technology by introducing more complex behaviour into the model. This result obtained by adding extra layer to NN design. Moreover, DL entails modifying data with different functions that permit sequential description in several layers of abstraction. This enables DL models, resulting in higher accuracy in variety of applications including weather forecasting (Kamilaris et al ,2018).

2.2.1 Neural Networks applications in weather forecasting

The field of meteorology utilizes neural networks for a wide variety of applications, including weather forecasting. The process of a neural network is dictated by the network topology, the connection strength, and the processing that is carried out at computation components, also known as nodes. A neural network is a system that is built of many basic handling parts that operate at the same instance. The adaptable nature of neural networks is one of the most fundamental aspects of these systems. Because of this property, the ANN approaches are especially attractive in application areas of weather forecasting for resolving highly nonlinear events (Baboo et al.,2010).

A neural network can be considered a mathematical function, similar to other machine learning model.

$$y = f_{NN}(x) \tag{1}$$

The function f_{NN} exhibits a specific structure, that it is an interconnected function.

$$y = f_{NN}(x) = f_3(f_2(f_1))) \tag{2}$$

f_1 and f_2 can be expressed as:

$$f_1(z) \stackrel{\text{def}}{=} g_l\ (w_l z + b_i) \tag{3}$$

The variable "l" is commonly referred to as the layer index, and its range of values extends from 1 to an arbitrary number of layers. The activation function known as "gl" is classified as a mathematical function utilized in neural networks. The data analyst typically selects a non-linear function prior to commencing the learning process. The matrix w_l and vector bl for each layer are acquired through gradient descent optimization, with the specific cost function being dependent on the task at hand (Burkov, 2019).

Currently, there exist three prevalent categories of deep neural networks that are widely employed.

1. Multilayer Feed-Forward Networks

   The multilayer feed-forward network is a type of neural network that comprises an input layer, one or more hidden layers, and an output layer. Each stratum comprises of one or multiple synthetic neurons. The artificial neurons exhibit resemblance to their perceptron predecessor, but their activation function varies based on the layer's distinct purpose within the network (Patterson & Gibson, 2017).

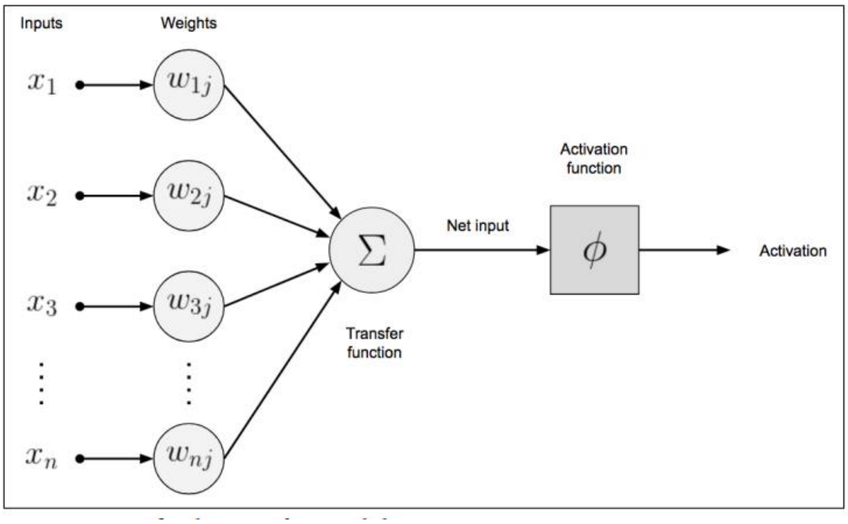

Figure 6 Multilayer perceptron ANN (Patterson & Gibson, 2017)

2. Convolutional neural network (CNN)

The Convolutional Neural Network (CNN) is a distinct type of Feedforward Neural Network (FFNN) that effectively minimizes the parameters in a complex neural network with multiple units, while maintaining a satisfactory level of model accuracy. Convolutional Neural Networks (CNNs) have been utilized in various domains such as image and text processing, exhibiting superior performance compared to earlier recognized targets (Burkov, 2019). The effectiveness of Convolutional Neural Networks in the field of image recognition is a significant factor in the widespread acknowledgement of the capabilities of DL (Gibson, 2017).

3 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) are a highly expressive model category that is commonly used for tasks involving sequences (Sutskever et al.,2019). It possesses the ability to handle input of varying lengths, similar to RNN Neural Networks. Recurrent Neural Networks possess the capability to represent the hierarchical structures present in the training dataset, which sets them apart from other types of neural networks (Gibson, 2017).

Traditional NN have inputs and outputs that are not reliant on one another in any way. However, in situations in which it is necessary to anticipate the next word in a phrase, it is necessary to remember the prior words. As an outcome, it is necessary to remember the earlier words. As a result, RNN brought an innovative idea to resolve this problem with the assistance of a hidden Layer. RNN's hidden state, which remembers certain information about a sequence, is the property that is considered to be its primary and most significant characteristic. Memory State is another name for this condition because it stores information about the most recent input that was made to the network. It implements the same job on all the inputs or hidden layers in order to generate the result, and so employs similar weight for each input it receives. In contrast to other NN, this simplifies the relationship between the constraints.

3 Methodology

This chapter focuses on the research methods and technologies used to collect, analyze, and interpret the dataset for the study. Moreover, it provides a clear and detailed description of how the study was conducted in terms of research design process, data selection, model selection and data analysis tools.

3.1 Research Design

The research design of this research involves a comparison of the performance of ML and DL models for weather forecasting. Based on the nature of the research topic, the research methodology is quantitative, focusing on retrieving and analyzing the dataset obtained from open source. Mathematical, statistical, and computational tools are utilized to analyze the data and obtain results.

The dataset was obtained from Finnish Meteorological institute and contains hourly historical weather observations from the automatic observation station of Vaasa, wester Finland. The dataset is in CSV format due to its fast-processing times when importing and exporting data. It includes various variables, such as temperature, atmospheric pressure, humidity, wind, and solar radiation. However, this study focuses specifically on forecasting maximum and minimum temperature.

A	B	C	D	E	F	G	H	I
Year,m,d,Time,Time zone,Maximum temperature (degC),Minimum temperature (degC)								
2010,5,28,00:00,UTC,11.3,7.4								
2010,5,29,00:00,UTC,12.6,7								
2010,5,30,00:00,UTC,12.8,5.6								

Figure 7 Dataset Header

Based on the selected dataset, the research proposes models that combine both ML and DL technologies to perform weather forecasting. The model design involves the use of a variety of algorithms including polynomial regression, gradient boosting, and recurrent neural networks. The main purpose of the models is to accurately forecast the temperature based on the dataset, by using the techniques. The process contains the following two main steps.

- Train the model on 75% of the entire data, applying the algorithms stated above. These technologies involve feeding the dataset into the newly created model for the purpose of learning the pattern and trend of input data.
- Evaluate the model applying the remaining 25% of the dataset to verify its performance. This process is used to discover the accuracy and reliability level of the model.

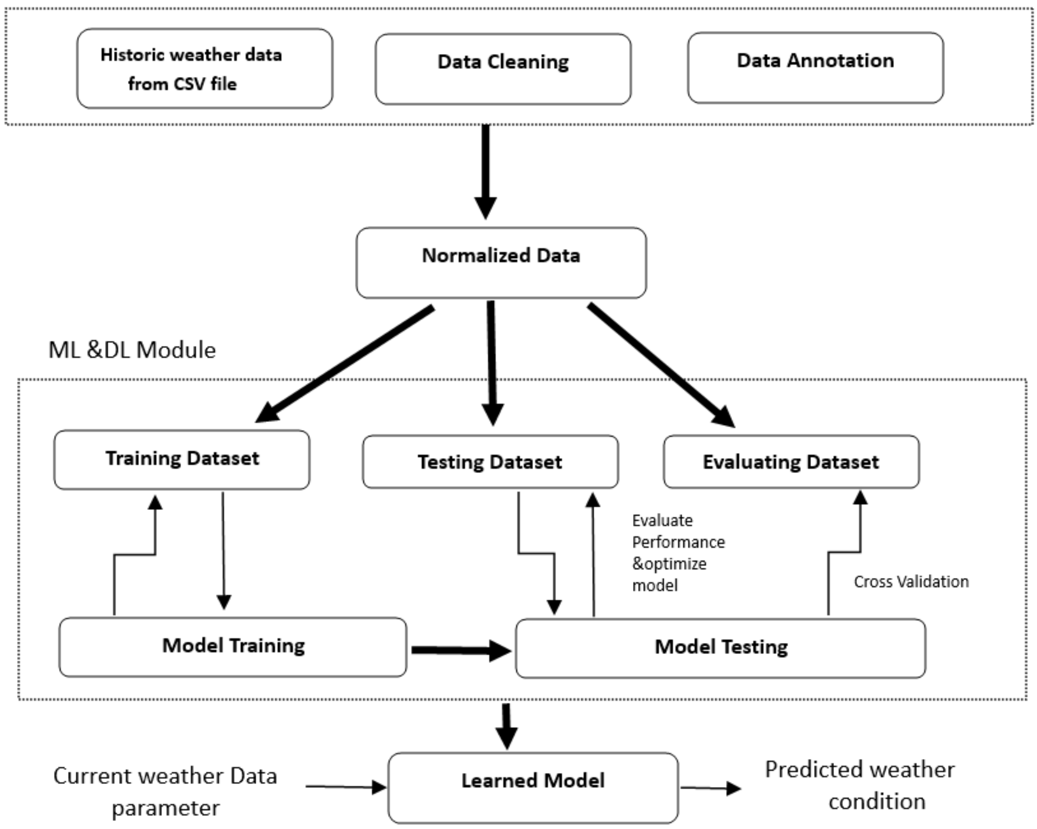

Figure 8 Proposed Model

3.2 Data Preprocessing

The study undertakes several pre-processing steps to guarantee the weather dataset is suitable for applying ML and DL algorithms.

- Data Cleaning: The historical weather data checked for incomplete, errors, missing values and outliers removed using appropriate data analysis tools.
- Feature Engineering: New variable created from the existing variable to provide more valuable information.
- Data Normalization: The dataset passed through the process of cleaning and standardizing to ensure that all variables have the same format, scale, and range.
- Dimension Reduction: The dimension of the data reduced into a low dimension space to maintain meaningful properties of the raw data.

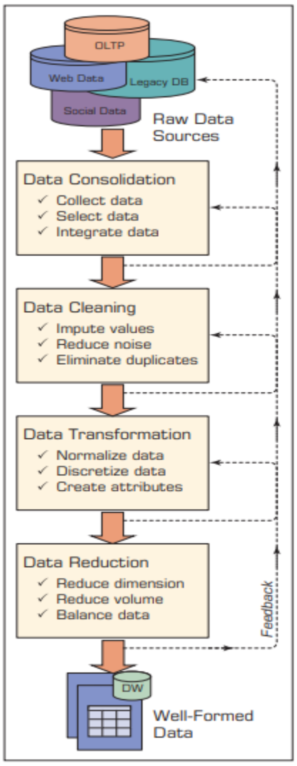

Figure 9 Data Pre-processing steps (Sharda et al.2021)

3.3 Model Selection

Four different models selected from both machine learning (ML) and deep learning(DL), that are suitable to handle weather forecasting.

3.3.1 Polynomial Regression for weather forecasting

The process of weather forecasting often requires the analysis of time-series data, in which the variables fluctuate over the course of a period. Polynomial regression assumes that the connection between the independent and dependent variables is constant, which means that it may not be able to capture the time-based patterns and dynamics of meteorological dataset.

Any weather observations data will typically exhibit a nonlinear pattern of activity in its overall behavior. As a direct consequence of this, the linear regression model will be very challenging to visualize and will not accurately forecast any of the data. Because of this, it will be quite challenging to construct the optimal line that accounts for the majority of the meteorological data. As a result, the prediction of the weather forecast will be too uncertain, and polynomial regression become preferred option since it allows to match the data curve while maintaining a minimal error value.

According to Peck et al. (2012), polynomial regression is a sort of regression analysis in which the relationship between an independent variable (X) and a dependent variable (Y) is modeled as an nth-degree polynomial. Polynomial regression is also known as polynomial modeling. Additionally, it is one of the ML models that fits a non-linear regression curve to obtain a non-linear relation between the two variables.

Polynomial regression represented by the equation:

$$Y = \beta + \beta_1 X + \beta_2 X^2 + \ldots + \beta_n X^n + e \tag{1}$$

Where Y is dependent variable, X is independent variable, β, β_1, $\beta_{2,\dots}$ β_n are coefficients of the equation, n is the degree of polynomial equation and e is error value.

The ability of polynomial regression to capture non-linear trends in the data makes it an ideal choice for use in the forecasting of time series data. The relationship between the dependent variables and the independent variables in a time series dataset will typically result in a polynomial regression, which is able to capture non-linearity.

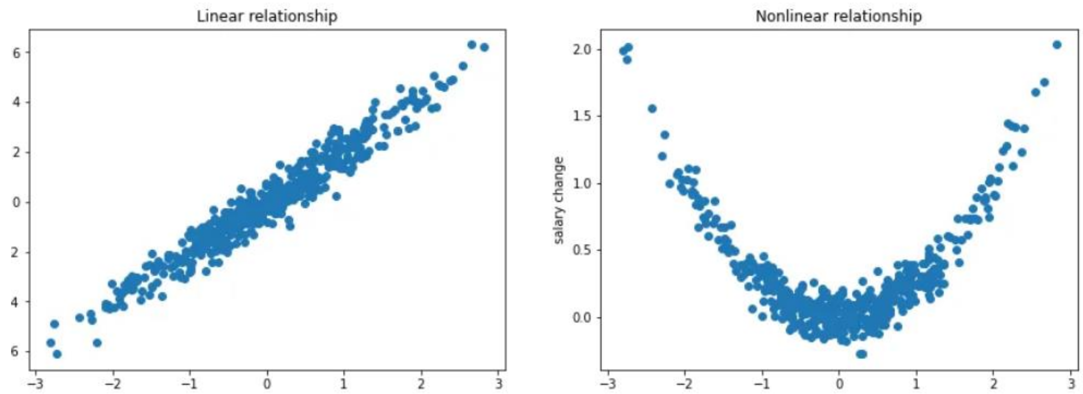

Figure 10 Linear vs Non-Linear relationship (Cukrowski, 2022)

3.3.2 Gradient Boosting for weather forecasting

Gradient boosting is a modern machine learning approach that is able to apply to predict weather. The process of weather forecasting includes making predictions about future weather conditions based on observation from the past, and gradient boosting algorithm is one tool that enhances the accuracy of these forecasts.

According to Friedman (2001), "Gradient boosting of regression trees produces competitive, highly robust, interpretable procedures for both regression and classification, especially appropriate for mining less than clean data". It is a common ML technology that has recently gained popularity and functions by integrating a group of simple or ineffective learners into a single and more successful model. This method has been demonstrated to be extremely effective in a variety of applicates including weather forecasting.

The concept of boosting serves as the foundation for another efficient ensemble learning approach known as gradient boosting. First, investigate the gradient boosting technique for regression. In order to construct a robust regressor, start with a model in which f is equal to f0.

$$f = f_0(x) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} y_i \tag{2}$$

Subsequently, the labels of each example i, where i ranges from 1 to N in the training set, are modified in the following manner:

$$\hat{y} \longleftarrow y_i -- f(x_i) \tag{3}$$

Where $\hat{y}$ is residual and x_i is the new label.

The revised training set, which utilizes residuals in place of primary labels, is employed to construct a novel decision tree model, denoted as f1. The current definition of the boosting model is represented by f, which is defined as $f \stackrel{\text{def}}{=} f_0 + \alpha f_1$, where α denotes the learning rate, a hyperparameter.

3.3.3 Recurrent NN (RNN) for weather forecasting

Recurrent neural networks are a prominent kind of deep learning model that is applied for time-series modeling applications such as weather forecasting. RNN is especially useful for applying with sequential data because they are able to describe sequential relationships and generate predictions based on the circumstance of previous records. This makes the algorithm an ideal tool for implementing sequential datasets. However, in some cases coping with long-term dependence or sudden shifts the sequence of the weather data, cause difficulty. To solve this problem, more sophisticated architectural designs, such as transformer-based models, applied by combining peripheral inputs into the model, such as geographical or satellite data, added to boost the capability of the output to make accurate forecasts.

A Recurrent Neural Network (RNN) is an architectural design that originated in the 1980s. RNNs are a fitting choice for datasets that feature sequential data (Campesato, 2020). Additionally, weather forecasting, stock prices forecasting, predicting energy demand are time series prediction problems. In those examples, events happen in the time-ordered sequence, where the previous event affects the current and future events. RNNs are meant to learn from data sequences in order to tackle time series issues by transmitting the hidden state from one step in the sequence to the next and mixing it with the input. However, the memory in RNN is generally short-term memory, in particular, RNN works by storing and merging the right before short-term memory in the current event. From that, RNN attempts to handle time-based or sequence-based data (Peter, 2021).

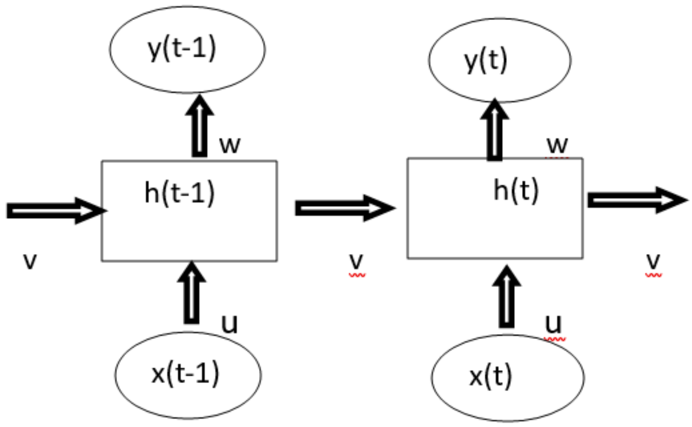

Figure 11 RNN Structure

Assume that the input sequence is denoted as x_1, x_2, x_3, ..., x(t), ... Additionally, assume that the hidden state sequence is denoted as h_1, h_2, h_3, ..., h(t). It should be noted that both the input sequence and hidden state are represented as a vector of size 1xn, where n corresponds to the number of features.

During time t, the input is determined by the amalgamation of h(t-1) and x(t). Subsequently, an activation function is employed on this combination, which may also encompass the inclusion of a bias vector. An additional distinction pertains to the feedback mechanism inherent in recurrent neural networks, which operates between successive temporal intervals. The recent inner state is computed by integrating the previous output with the present input, as per the operational procedure. The sequence {h(0), h(1), h(2), ..., h(t-1), h(t)} is utilized to denote the internal states of a Recurrent Neural Network (RNN) over a period of time {0, 1, 2, ..., t-1, t}. Additionally, it is assumed that the sequence {x(0), x(1), x(2), ..., x(t-1), x(t)} represents the inputs throughout the same time frames (Campesato, 2020).

The equation below represents primary correlation for a recurrent neural network (RNN) at a given time t:

$$h(t) = f(w * x(t) + u * h(t-1)) \tag{4}$$

where w and u are weight matrices, and f is tanh activation method.

3.3.3 NeuralProphet application in weather forecasting

NeuralProphet is a time series forecasting algorithms created based on Facebook's Prophet algorithm (Catherine,2022) which is currently applicable for weather forecasting. Initially, the Prophet algorithm provided a simple, practical, customizable, and reasonable tool to forecast time series. However, a persistent issue remained, poor performance. In order to address this matter, NeuralProphet was developed by combining neural networks with Prophet.

NeuralProphet emphasis on configurability and interpretability, which means end user permitted to customize the model's hyperparameters to best fit own approaches and provides analytical tools allow user to evaluate the model's performance an identify part for enhancement. Furthermore, the modular architecture of NeuralProphet includes a feature that enables the addition of new components as required (Yu et al., 2022).

The NeuralProphet model consists of six different modules, with each module contributing an additional element to the time series prediction. According to Triebe et al.(2021), " a core concept of the NeuralProphet model is its model it modular composability ". The full model is summation of each module as shown in the equation 2, where h is the number of steps predicted in the future and $\hat{y}$ is predicted value.

$$
\begin{aligned}
\hat{y}_{t+h-1} = \; & T(t+h-1) && \text{trend} \\
& + S(t+h-1) && \text{seasonal effects} \\
& + E(t+h-1) && \text{event \&holiday} \\
& + F(t+h-1) && \text{regression effect for future} \\
& + A(t+h-1) && \text{auto-regression effect} \\
& + L(t+h-1) && \text{regression effect for lagged observation of variable}
\end{aligned} \tag{5}
$$

It is possible to configure each individual module of the model components and merge to form the complete model.

- **Trend**: One of the most common ways to model trends is to use a combination of offset value, represented by m, and a growth rate, denoted by k. The trend impact at each given time t is assumed to be driven by multiplying the growth rate by the time difference between the beginning point t_i and the current time t_c , plus the offset m (Triebe et al. ,2021).

$$T(t_c) = T(t_i) + k\,(t_c - t_i) \tag{6}$$

- **Seasonality**: The seasonality of a model refers to the extent to which a given dataset exhibits a periodic pattern. This characteristic is typically represented using the following Fourier term equation.

$$S(t) = \sum_{i=0}^{n=k} \left(a_i \cos\left(\frac{2\pi i t}{p}\right) + b_i .\sin\left(\frac{2\pi i t}{p}\right)\right) \tag{7}$$

- **Auto-Regression:** This module is a commonly employed time series model for capturing temporal dependence among the stochastic variables within a series.
- **Lagged Regressors:** The utilization of lagged regressors is a common practice in order to establish correlation between the target time series and other observed variables. The variables in question are commonly denoted as covariates. In contrast to future regressors, the trajectory of lagged regressors remains uncertain (Triebe et al. ,2021).
- **Future Regressors:** Refer to variables that are anticipated to be recognized in the future. The value of these variables is identified at every time
-

3.4 Performance Evaluation

The performance of all selected models will be evaluated applying the following set of performance metrices.

1. Mean Squared Error (MSE): It is a widely used measure of the average squared difference between actual and predicted values in regression problem. The value is computed using the following formula.

$$MSE = (1/n) * \sum(Yi - \hat{Y}i)^2 \quad (8)$$

Where: n is the observations in the data

Yi the actual value of the corresponding dependent variable in the observation.

$\hat{Y}i$ the predicted value of the corresponding dependent variable in the observation.

Squaring the difference results in a non-negative value and guarantees that the MSE always return positive number or zero. An MSE zero is returned only by perfect model with no errors, but in actual case this does not occur. The closer the MSE value to zero, the model considered more accurate.

2. Mean Absolute Error (MAE): This metrices is used to evaluate the performance of a regressions model and defined as the measurement of average absolute difference between the actual observation and the predicted observation.

   The MAE value can be computed as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| Yi - \hat{Y}i \right|^2 \quad (9$$

   Where: Yi the actual

   $\hat{Y}i$ the predicted value

   One of the advantages of MAE is that it measures the average size of the model's errors in its predictions. In addition, it is used for evaluating the performance of a model when the errors are uniformly spread within the data.

3. R-squared (R^2): This metric measures the proportion of variance in the target variable that is explained by the regression model. It also measures how much the data is closer to the fitted line.

   The formula for R^2 is as follows:

$$R^2 = 1 - \frac{RSS}{TSS} \quad (10)$$

   Where: Rss is sum of residual square, which measure difference the pre predicted value and actual values.

   TSS is total sum of squares, that measure the difference between actual values and the mean of the actual values.

$$\text{RSS} = \sum(\hat{Y}i - Yi)^2$$

$$\text{TSS} = \sum(Yi - \bar{Y})^2$$

Where: Yi the actual

$\hat{Y}i$ the predicted value

$\bar{Y}$ mean of actual value

The higher R^2 implies a better performance in terms of fitting the model to the data and its value varies between 0 and 1.

The models trained and tested using k fold cross validation technique to verify that the results are robust and not impacted by the selecting training and testing dataset. The cross-validation aids in reducing over-fitting and provide a more accurate prediction. In addition to the performance metrices different qualitative analysis is used to evaluate the model's ability to capture complex patterns between variables, such as visualizing the result of model prediction and comparing it with the real observation data.

4 Research result and Analysis

This part is dedicated to answering the research questions and explaining the result obtained using the methods described in the previous chapter.

4.1 Analysis of the dataset

Analyzing the properties and relationships present in the dataset is essential prior to constructing machine learning and deep learning models for weather forecasting. This study facilitates the process of making learned decisions regarding feature engineering, data preprocessing, and model selection. Several crucial procedures have been conducted before developing a model in this study.

- Main preprocessing methodology conducted to assure the quality of the dataset. The dataset used in this research is considered as time serious data type, the data patten checked in the first steps underlying pattern and characteristics of the data was important for effective analysis and modeling. In addition, time serious data demonstrates seasonal dependencies, that means that the values at different points in time are likely to be related.

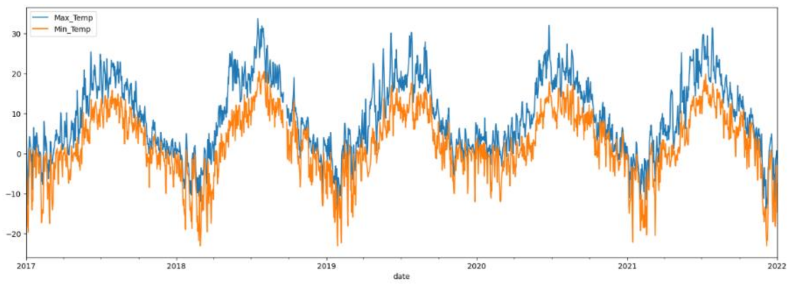

Figure 12 Yearly distribution of the feature, temperature

Figure 5 shows the year distribution of the maximum and minimum températures by placing the daily count each year on top of each other. The graph reveals some interesting patterns in the early phase and gives some hypotheses about the trend of maximum tempéra-ture:

- Trend and seasonality detection performed in the second steps of the process, which refers to recurring occurrence at regular intervals, such as weekly, monthly, and yearly cycles, and represents organized changes in the dataset over time. This process played a significant role in selecting suitable models that capture the overall direction and magnitude of the dataset evaluation.

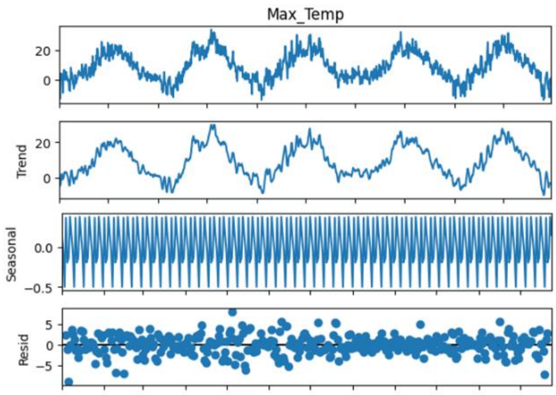

Figure 13 Trend and Seasonality simulations

- The dataset was carefully examined in order to detect any instances of null values or values that were missing. Upon closer inspection, it became clear that the dataset did not include any null values, and there were no entries that were missing. The result of this process is shown in the following code snippet.

```
data.isnull().sum()
```

```
precipitation    0
Max_Temp         0
Min_Temp         0
date             0
dtype: int64
```

Figure 14 Number of Null values in the dataset

- In this process, an accurate analysis was conducted to identify any instances of duplicated values within the dataset. It was observed that a small portion of the dataset showed duplicated values, indicating the presence of replication.

```
data1.duplicated().value_counts()
```

```
False    1809
True       18
dtype: int64
```

Figure 15 Duplicated value analysis

- Most time series data have a high probability of including outliers, which are values that dramatically depart from the norm of the rest of the data points. Outlier detection methods were used to the dataset, specifically focusing on two different features. Following the implementation of the detection functions, outliers were detected in one of the features. The analysis showed the presence of outliers, suitable actions were made to manage and fix these issues in the succeeding phases.

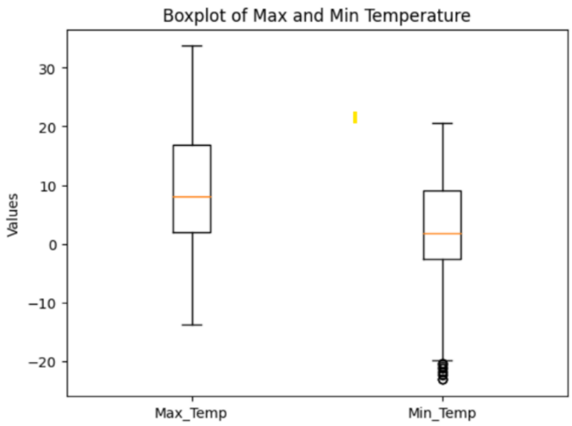

Figure 16 Boxplot for Maximum temperature

- Feature extraction was conducted using heatmap to find out which feature is highly correlated to the other feature. The extracted feature selected based on the correlation value and provides a more brief and understandable presentation of the basic dataset patterns and easily applied in selected models.

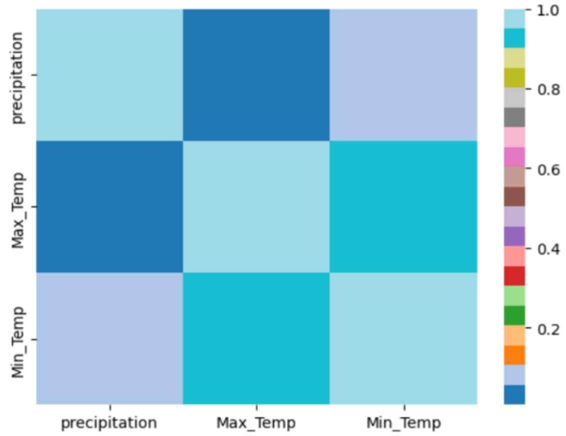

Figure 17 Heatmap for feature selection

4.2 Analysis of Model results

The first model approach is polynomial regression. The minimum and maximum temperature shows a seasonal pattern (yearly, monthly, daily) that needs to be identified and modelled separately before studying other factors. To achieve this, the dataset is divided into separate train and test sets, and the model is adjusted with various degrees. The results obtained after training the model shown in the following figure.

```
Mean Squared Error: 8.187729485572392
Mean Absolute Error: 2.1403545727125484
r2_score: 0.908561231305821
```

Figure 18 Polynomial Regression Evaluation matrices

The evaluation metrics indicate the accuracy and performance of the weather forecasting model. RMSE value of 8.18 indicates the average difference between the predicted temperatures and the actual temperatures in the initial unit of the target variable. A lower MSE value implies better accuracy. In this case, it means that, on average, the predictions deviate by approximately 8.18 units from the actual temperatures. The MAE value is 2.14, it indicates that on average the absolute difference between the predicted values and actual value. The R^2 score is 0.908, which indicates that approximately 90.08% of the variance in the dependent variable can be described by the independent variables in the model. This implies a strong correlation and a good fit between the predicted values and the actual values.

In general, the results indicate that the regression model performed well. The low MSE and MAE values suggest that the predictions are generally close to the actual values. Additionally, the high R2 score indicates a high level of explanation and prediction accuracy in the model.

date	actual	predictions
2020-08-01	19.5	20.125251
2020-08-02	20.6	19.051569
2020-08-03	18.4	20.088855
2020-08-04	21.4	17.981976
2020-08-05	20.2	20.976877
...	...	...
2021-12-27	-1.9	-11.160451
2021-12-28	-0.7	-1.199091
2021-12-29	0.7	-0.370503
2021-12-30	0.4	1.117576
2021-12-31	-2.8	0.831008

518 rows × 2 columns

Figure 19 Actual and prediction value comparison

The second model applied on the give dataset was gradient boosting. The model scored 0.91 for R^2 ,which indicated the fraction of the total variation in the dependent variable that can be attributed to the predictability of the independent variables. A higher R^2 score in gradient boosting suggests that the ensemble of decision trees is able to explain a bigger amount of the variation in the data. This reveals, 91.03% of the variation that was found in the dependent variable was able to be explained by the independent variables that were used in the gradient boosting model. This suggests that there is a substantial connection between the projected values and the actual values, as well as a good match between the two sets of data.

```
Mean Squared Error: 8.032383613564566
Mean Absolute Error: 2.1415816563773205
r2_score: 0.9102960999630171
```

Figure 20 Gradient Boosting model evaluation metrices

The dataset format has the same features as time series dataset, so the RNN model, in particular LSTM, is applied with kerras to learn and predict the sequence of maximum and minimum temperatures from year to year. Moreover, it handles complicated models with multivariate input variables and promotes the creation of a time series-based forecasting system.

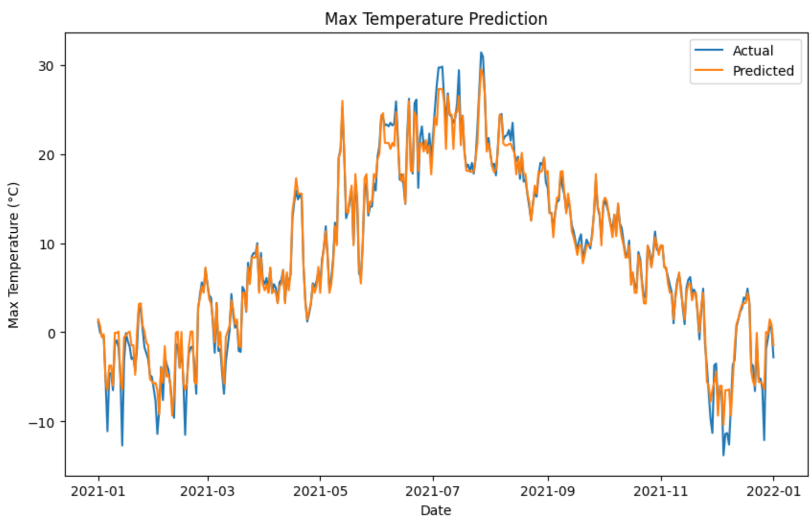

Figure 21 LSTM model prediction

According to the MAE value of 0.8449, the LSTM model's predictions deviated by around 0.8449 units from the actual values, on average. MAE scores that are lower are indicative of improved prediction performance. A MSE score of 1.6527 indicates that the LSTM model's predictions, on average, drifted from the actual values by around 1.6527 units squared. Along the same lines as MAE, lower MSE values suggest improved predicting ability. R^2 score is 0.91, which indicates that approximately 91.03% of the variance in the dependent variable can be explained by the independent variables used in the gradient boosting model. A higher R2 score generally indicates better model performance.

```
Mean Absolute Error: 0.8449819194324616
Mean Squared Error: 1.6527214410373152
R-squared: 0.9841206007350898
```

Figure 22 LSTM model evaluation metrices

5.Conclusion

5.1 Key findings

Following an analysis of the result from those models described in the previous chapter, a number of important key points were observed, including the following:

- The LSTM model had the least amount of mean absolute error (MAE) compared to the other models, indicating that it had the best accuracy in predicting the target variable. It is possible that its capacity to grasp temporal relationships in sequential data was a contributing factor in this outstanding performance.

- The MAE and MSE were not significantly different between the results obtained by the Polynomial Regression and Gradient Boosting models. Even though Polynomial Regression had a somewhat higher MSE, Gradient Boosting did marginally better in terms of its R^2 score, which indicates that it is able to explain a bigger proportion of the variation in the target variable.

- R^2 ratings that were more than 0.9 were achieved by each of the three models, indicating that they each displayed a satisfactory level of predictive ability. This suggests that a significant percentage of the variation in the target variable may be traced back to the predictors that are included in each model.

- The LSTM model had the best R2 value, which indicated that it provided a better overall fit to the data. It is important to note, however, that LSTM models may be computationally costly and may need more extensive data preparation than other models.

- Having quality data significantly plays essential roles in achieving consistent and accurate predictions across different models and reduces the accuracy gap between machine learning and deep learning models.

5.2 Conclusions

Comparing the performance of different forecasting models in weather forecasting can be a complex task as it depends on various factors such as data quality, model configuration, hyperparameter tuning, and the specific weather patterns being predicted. Both deep learning and machine learning algorithms have their strengths and weaknesses when applied to weather-related time series datasets.

Deep learning models, such as recurrent neural networks with alternatives like LSTM, shine in capturing complex sequential dependencies and patterns in weather data. They can automatically learn feature representations from raw data, eliminating the need for manual feature engineering. Deep learning models can handle large-scale datasets effectively, making them suitable for weather forecasting tasks. However, this technology often requires substantial amounts of labeled training data and computational resources for training. Interpretability can also be a challenge with deep learning models.

In summary, selecting the appropriate forecasting model for weather-related time series datasets involves considering decisions between accuracy, interpretability, computational requirements, and data availability. It is recommended to experiment with different models, evaluate their performance using appropriate metrics, and choose the model that best suits the specific requirements of the weather forecasting task at hand. In general, having data of a high quality is very necessary in order to make accurate and trustworthy predictions using machine learning and deep learning models. While DL models have the benefit of being able to learn from raw data, ensuring the quality of the data via preprocessing, ensuring accurate labeling, and reducing bias are still critical issues for both techniques.

Reference

A. Burkov (2019). The Hundr-page Machine Learning BOOK, Page (1-5).
Academic Journal of Computing & Information Science, 2021, 4(5)

A. Géron (2022). Hands-On Machine Learning withScikit-Learn, Keras, and TensorFlow.

A. Goodfellow,Y.Bengio,A.Courville (2016,chapter 3).Deep Learning https://www.deeplearningbook.org/contents/mlp.html

A.Kamilaris , F.Prenafeta-Boldú(2018). "Deep learning in agriculture: A survey." computers and Electronics in Agriculture

A.Salman ,B.Kanigoro ,Y.Heryadi.(2015). A Deep Learning-Based Weather Forecast System for Data Volume and Recency Analysis"

A.Peck,G.Vining ,D.Montgomery(2012)."Introduction to Linear Regresssion Analysis"

A.Kamilaris , F.Prenafeta-Boldú(2018). "Deep learning in agriculture: A survey." computers and Electronics in Agriculture

B.Chong(2021). K-means clustering algorithm: a brief review.

C.Li ,M.Zhao,Y.Liu,F.Xu (2020) .Air Temperature Forecasting Using Traditional and Deep Learning Algorithms.

C.Lin , Y.Yu , L.Wu · J.Cao(2019). Unsupervised Learning on U.S. Weather Forecast Performance.

C.Brown, J. Johnson,J.Mith(2019).Supervised Learning approaches for weather fore casting".Journal of Meteorological applications

D.Mishra , P.Joshi (2021).A Comprehensive Study on Weather Forecasting using Ma chine Learning .

D.Cho , C.Yoo ,B.Son ,J.Im ,D.Yoon ,D.Cha (2022). A strategy for operational implementation of 4D-Var, using an incremental approach. In Quarterly Journal of the Royal Meteorological Society 120.519, pp. 1367–1387.

P.Rozas (2019). "Application of machine learning techniques to weather forecasting"
I.Sutskever,J.Martens,G.Hinton(2019). Generating Text with Recurrent Neural

Networks.

M. Abdalla, H. Ghaith, A. Tamimi (2021). Deep Learning Weather Forecasting Tech niques: Literature Survey.

J.Booz ,W.Yu , G.Xu ,D.Griffith ,N.Golme (2019)."A Deep Learning-Based Weather Fore cast System for Data Volume and Recency Analysis .

J.Kelleher ,B.Namee,A.D'arch (2020). Fundamentals of Machine Learning for Predic-tive Data Analytics (2nd ed.)

J. Segovia , J. Toaquiza , J. Llanos ,David R. Rivas(2013), Meteorological Variables Fore casting System Using Machine Learning and Open-Source Software.

N.Singh ,S.Chaturvedi ,S.Akhter (2019).Weather Forecasting Using Machine Learning Algorithm. Published in Toward Data Science.

L.Chen, B.Chen, J.Tingting(2019). Application analysis of meteorological Big data Ser vice in cloud computing Environment. *China Sci. Technol. Inf.* 2019, *11*, 88–90.

M.Afteniy (2021). Predicting time series with Transformer"

M.Mohri ,A.Rostamizadeh ,A.Talwalkar (2016)."Foundations of Machine Learning, pages 1-8

M.Cukrowski(2022)."Polynomial Autoregression: Improve your Forecastsbin 2 minutes"

M.Homstrom , D.Liu,C.Vo (2016)." Machine Learning Applied to Weather Forecasting. Stanford University

M.Soori ,B.Arezoo ,R.Dastres (2022)." Machine learning and artificial intelligence in CNC machine tools, A review ".

N.Triebe,H.Hewamalage ,P.Pilyugina , N.Laptev,C.Bergmeir ,R.Rajagopal(2021). Neural Prophet: Explainable Forecasting at scale.

O.Campesato (2020) .Artificial Intelligence Machine Learning and Deep Learning, Page 128 – 150.

P.Flach(2012)." MACHINE LEARNING The Art and Science of Algorithms that Make Sense of Data"

S.Sah(2020). Machine Learning: A Review of Learning Types. https://www.pre prints.org/manuscript/202007.0230/v1

Y.Lecun,Y.Bengio ,G.Hinton. (2015). Deep Learning . *Nature* volume 521, pages436–444 (2015)

Y.BENGIO, Y. LECUN,, G.HINTON.(2021).Deep Learning for AI.

S.Dridi (2021). Unsupervised Learning – A systematic Literature Review. https://osf.io/kpqr6

R.Sharda,D.Delen,E.Turban(2021) . Analytics, Data Science and Artificial Intelligence

X.Ren ,X.Li, K.Ren ,J.Song ,Z.Xu ,K.Deng X.Wang (2021) “Deep Learning -Based Weather Prediction: A Survey”

V.Catherine (2022).In-Depth Understanding of NeuralProphet through a Complete Example. Published in Toward Data Science.

Z.Yu, K.Niu, X.Chen, Z.Guo, D.LI (2022). A Hybrid Model Based on NeuralProphet and Long short-Term Memory for Time Series Forecasting.