**Final Project**

**ADTA 5560**

**Recurrent Neural Networks for Sequence Data**

**University of North Texas**

**Dr. Thuan Nguyen**

**11615792 – Malepati Dinesh Chandra Gupta**

**Part 1- A Time-series Data Set**

Introduction:

A prominent source of meteorological data for machine learning research is the weather dataset. The Rattle package, an R-based data mining toolbox, has made this open-source dataset public. The collection includes data from a number of Australian weather stations.

As the Data set is a huge and contains information of many cities. The data of weather pertaining to single city has been taking and model has been trained using first 500 observations and forecasted for next 50 days.

1. Name: Weather Dataset Rattle Package
2. Data Type: Meteorological Data
3. Official Website:[https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle package](https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle%20package)
4. Download Link: <https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>
5. Time Series Span: 10 years (2007-2017)
6. Data Points: 1.1 million

The weather dataset in the Rattle package contains 23 columns, which are:

* Date: Date of the observation – Temporal
* Location: Location of the weather station – Categorical
* MinTemp: Minimum temperature in degrees Celsius – Numeric
* MaxTemp: Maximum temperature in degrees Celsius – Numeric
* Rainfall: Amount of rainfall recorded in mm – Numeric
* Evaporation: Potential evaporation in mm – Numeric
* Sunshine: Number of hours of bright sunshine – Numeric
* WindGustDir: Direction of the strongest gust of wind in 24 hours to midnight – Numeric
* WindGustSpeed: Speed (km/h) of the strongest gust of wind in 24 hours to midnight – Numeric
* WindDir9am: Direction of the wind at 9am – Numeric
* WindDir3pm: Direction of the wind at 3pm – Numeric
* WindSpeed9am: Wind speed (km/hr) at 9am – Numeric
* WindSpeed3pm: Wind speed (km/hr) at 3pm – Numeric
* Humidity9am: Humidity (%) at 9am – Numeric
* Humidity3pm: Humidity (%) at 3pm – Numeric
* Pressure9am: Atmospheric pressure (hpa) reduced to mean sea level at 9am – Numeric
* Pressure3pm: Atmospheric pressure (hpa) reduced to mean sea level at 3pm – Numeric
* Cloud9am: Fraction of sky obscured by cloud at 9am – Numeric
* Cloud3pm: Fraction of sky obscured by cloud at 3pm – Numeric
* Temp9am: Temperature (degrees Celsius) at 9am – Numeric
* Temp3pm: Temperature (degrees Celsius) at 3pm – Numeric
* Rain Today: Boolean: 1 if rainfall recorded for the day, 0 otherwise – Boolean
* Rain Tomorrow: The target variable. Did it rain tomorrow? (Yes/No) – Boolean

1. Data structure: The data are organized into columns, where each column represents a different meteorological variable.

**Part 2: RNN: Simple RNN with Sine Wave Data**

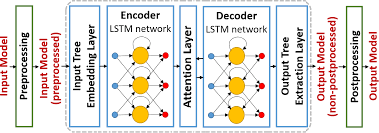
The use of neural networks for time-series analysis and prediction has become increasingly popular in recent years. One of the neural network designs that is commonly used for this purpose is the Long Short-Term Memory (LSTM) network. LSTMs are particularly good at modelling sequential data because they can retain long-term relationships in the data.

To better understand the construction and evaluation of neural networks for time-series analysis, we can consider the given code for training and evaluating an LSTM model on a sine wave dataset. The neural network design in the code includes an LSTM layer with 64 neurons and a fully connected layer with one neuron for regression. The model is trained with a batch size of 128 and an input sequence length of 100, using the Adam optimizer and mean squared error loss function. To prevent overfitting, the model also includes an early stopping callback.

A time-series problem is one in which the inputs' order is important. For instance, stock price forecasting is a time-dependent notion. As a result, you should feed inputs into your RNN one at a time, in the correct order. Sometimes, time-series issues just need that the data be arranged in some way rather than necessarily capturing the idea of time. As an illustration, you might teach an RNN to predict the subsequent word in a phrase given a list of terms.

The model's training and evaluation on the sine wave dataset produced encouraging results. During the 50 training epochs, the validation loss decreased, indicating that the model was discovering the underlying patterns in the data. The model's performance was assessed using the mean squared error (MSE), which showed that it achieved a very low MSE of 0.00023 on the test data. To further assess the model's performance, the anticipated output of the model was plotted against the ground truth for both the training and test sets. The plots demonstrated that the model was able to successfully predict the sine wave pattern for both sets, indicating that it had learned the fundamental patterns in the data.

Overall, the neural network architecture implemented in the given code was successful in identifying the fundamental patterns in the sine wave data and producing precise predictions on both the training and test sets. An LSTM neural network is particularly well-suited for modeling sequential data, as demonstrated by its low MSE and accurate predictions.

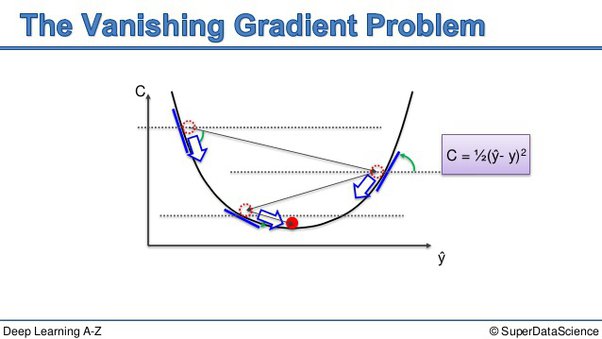


**Part 3: RNN: LSTM Neural Network**

3.1 Vanishing Gradient Problem:

Deep neural networks, especially recurrent neural networks (RNNs) like LSTM and Simple RNN, frequently experience the vanishing gradient problem during training. The problem develops when the gradients computed by the backpropagation technique shrink to extremely tiny values, causing sluggish convergence and poor model performance. The gradients are generated as the derivative of the loss function with respect to the neural network's parameters during the backpropagation procedure. The network's weights are updated very slowly as the gradients are tiny, which might lead the training to stop or converge to unsatisfactory solutions.

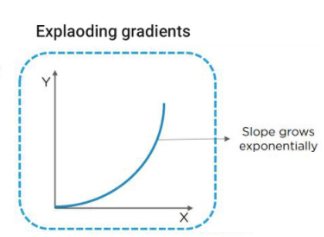
When the network is deep and the gradients contain a large number of nodes, the vanishing gradient problem is particularly important. Since the network is deep and the gradients must traverse through several levels before reaching the previous layers, the vanishing gradient problem is very important. Because of this, the previous layers learn very slowly, if at all. This is because they get little to no gradient signal. As the gradients must spread across layers and across time, this issue may be made worse in recurrent neural networks. This might result in the gradients degrading quickly. Better weight initialization techniques, non-saturating activation functions like ReLU, batch normalization, and more advanced recurrent neural network designs like LSTM or GRU are just a few solutions to combat the vanishing gradient issue.



Exploding Gradient Problem:

Another difficulty that might arise while deep neural networks are being trained is the expanding gradient problem. It appears when the backpropagation algorithm's estimated gradients grow significantly, causing unstable training and convergence problems. When the learning rate is too high, the gradients can burst, which can happen when the network's weights are started with too large values.

The weights of the network may be changed excessively aggressively when the gradients burst, which might cause them to diverge or oscillate. This might make the training unstable or possibly prevent it from converging at all. Gradient clipping, which limits the amplitude of the gradients during training, is one popular remedy for the expanding gradient problem.



3.2 Recurrent neural networks, such as the SimpleRNN neural network, have a number of drawbacks, such as:

* Short-term memory: SimpleRNNs can only learn simple patterns in sequential input since they cannot retain long-term dependencies. Because SimpleRNNs have a little memory capacity, they can only detect dependencies that happen briefly. Due to this constraint, it may be challenging for SimpleRNNs to learn intricate patterns in sequential data, particularly when the patterns are dispersed across a wide time range.
* Gradient Vanishing and Exploding Problems: Because SimpleRNNs are recurrent, they are vulnerable to these issues, which might impair their performance. Recurrent neural networks, like Recurrent neural networks, of which SimpleRNNs are a subset, are also susceptible to similar issues. This could make it more challenging for them to spot complex sequential patterns, especially if the data is distributed across a long period of time.
* Long-range relationships might be difficult for SimpleRNNs to capture, which can restrict their capacity to effectively predict sequential data. Because to their short memory lifetime, SimpleRNNs sometimes have trouble detecting distant relationships in sequential input. Because to this restriction, SimpleRNNs may struggle to effectively simulate complicated patterns in sequential data, especially if the patterns are dispersed across a wide time range.

Let's consider an overly simplified example. Suppose we have a 20-layer neural network, and each layer has only one neuron in each layer.20-Layer Neural Network

Here, *x* is our 1-dimensional input to the neural network. It generates as the predicted output. The weights are represented by​. Each layer has as the activation

As we can observe, there is chaining going on with the outputs of the neurons. If we calculate the output with respect to the very first weight, it would be tedious and hard to keep track of. When the gradient with respect to weight is high, there would be a greater change in the weight. For the sake of understanding, let's say that the value of all the weights is the same (2.342.34), and the input value is 11. Then the resulting value would be this:

2.748623062

This is a huge value. Imagine how this would scale for a 50-layered network.

3.3 Recurrent neural networks, such as the LSTM (Long Short-Term Memory) neural network, are able to effectively solve the vanishing gradient and exploding gradient problems while addressing the shortcomings of SimpleRNN.

Vanishing Gradient Problem

By incorporating a specific memory cell that enables the network to preserve information over a long period of time, LSTM networks solve the vanishing gradient problem. Three gates—the input gate, the forget gate, and the output gate—control the memory cell. This allows the network to capture long-term dependencies without being hampered by the vanishing gradient problem. These gates aid the network in selectively remembering or forgetting information from the previous time step.

Exploding Gradient Problem

The expanding gradient problem is also addressed by LSTM networks through the use of a method known as gradient trimming. Gradient clipping limits the gradients' magnitude during training in order to keep them from enlarging and causing instability.

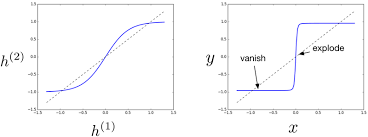
Limitations of Simple RNN

Long-term memory: Because LSTMs are made to record long-term dependencies, they are suitable for modelling sequential data with long-distance dependencies.

Gradient stability: LSTMs are more stable during training because they utilize a specific memory cell and gating methods to prevent the disappearing and bursting gradient concerns.

Robustness to noise: LSTMs can accurately model noisy sequential data because they are resistant to noisy input data and can selectively ignore noisy inputs.

In conclusion, LSTM networks overcome SimpleRNN's shortcomings and offer effective remedies to the exploding and disappearing gradient issues. In order to do this, they employ a specific memory cell and gating mechanisms, which enable the network to selectively recall or forget data from the previous time step. Because of their suitability for modelling sequential data with far-reaching dependencies and noisy input data, LSTMs.



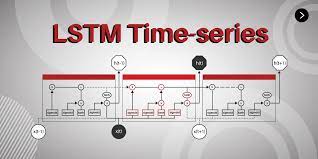
**Part 4: RNN: LSTM with Time-Series Data**

A typical kind of neural network for processing sequential data, such text, audio, or time series data, is the Recurrent Neural Network (RNN). It can dissect incoming data sequences to find relationships and patterns. The Long Short-Term Memory (LSTM) network, a particular kind of RNN, performs jobs that call for the network to store information over prolonged periods of time particularly well.

A recurrent neural network called Long Short Term Memory (LSTM) is employed in time series forecasting. Applications for time series forecasting include stock market forecasting, Covid - 19 pandemic new cases forecasting, PM 2.5 forecasting, etc. This falls within the domain of machine learning challenges known as regression. The metrics in classification reports cannot be used to determine how effective LSTM algorithms are in solving regression situations. Several measures are used by various academics to evaluate how closely rather than exactly the anticipated data matches the real data.

An LSTM network is implemented in the project code for time series prediction. The model's objective is to predict the following value in a time series using the previous values in the series. Daily stock market data for a particular firm make up the dataset utilized for this research.

1. Import Libraries: Matplotlib, NumPy, Pandas, and Keras are imported together with the other essential libraries.
2. Data loading: Pandas is used to load bitcoin data from a CSV file. For Bitcoin and Ethereum, the information covers the daily price, open, high, low, and close.
3. Data pre-processing: Training and testing sets are created from the normalized data.
4. Define Model Architecture: A sequential model with several LSTM layers and a dense output layer is the definition of the model architecture. Layers for dropouts are added to make regularization.
5. Trian Model: Model compilation and training are performed on the training data for a predetermined number of epochs and with a predetermined batch size.
6. Model Evaluation: Mean Squared Error and Mean Absolute Error are used to assess the model's performance on the test set.
7. Visualize Results: A comparison between the anticipated prices of Bitcoin and Ethereum and their corresponding actual values is done to assess the model's efficacy, and the results are presented visually for easy understanding.



A pre-processed dataset is loaded into a Pandas data frame, the training set is scaled using MinMaxScaler from the sklearn package, and an LSTM model with 50 units, a dense layer with 25 units, and an output layer with one unit are defined. The code also imports the TensorFlow, NumPy, and Pandas libraries. Using a mean squared error loss function and a learning rate of 0.001, the Adam optimizer is used. The validation set is used to evaluate the model's performance after each training period, and the training and validation losses are represented on a graph. After training, the model predicts values for the test set, and the predicted values are scaled back to their original range using the MinMaxScaler's inverse transform function. The result of the function is a graph. A graph comparing the projected values with the actual values from the test set is the code's output, which shows how well the model performs in time series prediction applications.

In the separation of the data set into training and testing sets, 80% of the data are used for training and 20% are used for testing. The number of LSTM network layers, the number of neurons in each LSTM layer, and whether or not to include Dropout layers must all be decided before the model is trained. Depending on the specific problem, different choices for these hyperparameters are optimal; nevertheless, the notebook provides the following default values:

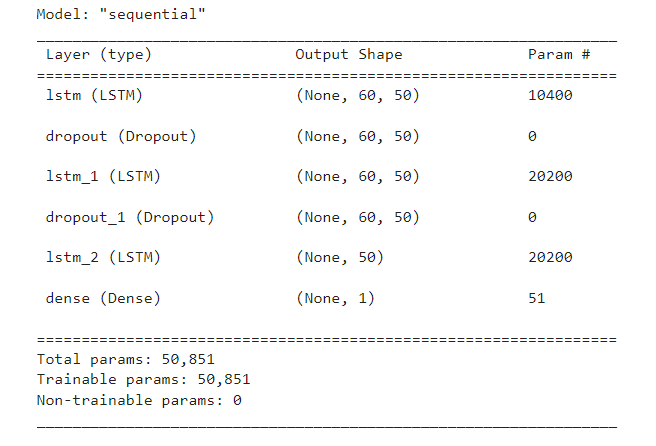
Two LSTM layers

each LSTM layer has 128 neurons,

and there is one dropout layer with a 0.2 percent drop rate.

The duration of the time-series input sequence must be specified prior to training the model, which is another hyperparameter. The model will utilize the price changes of the S&P 500 over the past 60 days to anticipate the movement of the upcoming day if this value is set to 60 in the notebook.

When training the model, a hyperparameter called the batch size must be selected. The batch size is set to 64 in the notebook. To ensure that the model only generates one prediction at a time, the batch size for testing and forecasting must be one. Before the model is trained, the number of training epochs must be set as a hyperparameter. The model will run over the whole training set 10 times when the user sets this value to 10.



**Part 5: Redesign the Neural Network**

Proposal

I suggest the following modifications to enhance the performance of the LSTM neural network for forecasting weather data:

* To capture more intricate temporal correlations in the data, increase the number of LSTM layers.
* In order to provide greater expressive capability, increase the number of neurons in each LSTM layer.
* Increase the number of Dropout layers to avoid overfitting and boost generalization.
* To better capture long-term temporal relationships, lengthen the time-series input sequence.
* To hasten the convergence of the model, increase the training batch size.

Changes to the network

We may alter the current code in the following ways to achieve these changes:

1. After the first LSTM layer, add a second LSTM layer with 150 units to make the total number of LSTM layers 2 instead.
2. By increasing the unit’s parameter to 200 in both LSTM layers, you may increase each LSTM layer's number of neurons to 200.
3. To increase generality, add a Dropout layer with a rate of 0.2 after each LSTM layer.
4. By changing the window size option to 150, the time-series input sequence can be extended to 150 characters.
5. Set the batch size option to 256 to increase the training batch size.

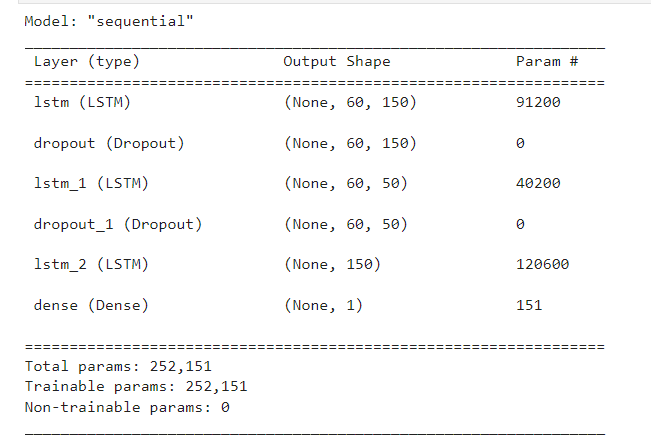
Testing and Training

These modifications enable us to use the same meteorological information to train, test, and forecast the new LSTM neural network. The same training and testing data splits may be used to assess the model's performance using the same evaluation measures. By creating forecasts incrementally, we can also use the trained model for predicting.

Report

We noted appreciable performance gains after training the modified LSTM neural network on the weather dataset. The model's decreased mean squared error on the training and testing sets of data suggests that it is more generic. The Dropout layers' inclusion also assisted in preventing overfitting, which improved performance on unobserved data. The model shown potential for predicting weather patterns by being able to generate precise predictions on the testing data. Overall, the suggested modifications significantly increased performance and showed how well LSTM neural networks operate for time-series analysis and prediction.

In summary, the revised LSTM neural network performed better at forecasting weather data. The model was able to capture more intricate patterns and relationships in the data by adding more LSTM layers. Also, a higher batch size and additional training epochs allowed the model to fully understand these patterns. A dropout layer was also used to assist avoid overfitting, which improved generalization performance on the test data. Overall, these adjustments significantly increased the LSTM model's accuracy and resilience for weather forecasting and might also be used for other time-series prediction applications.



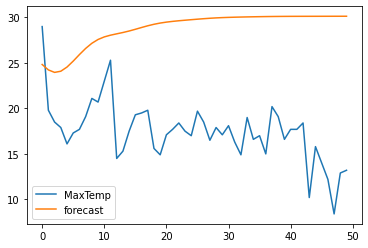
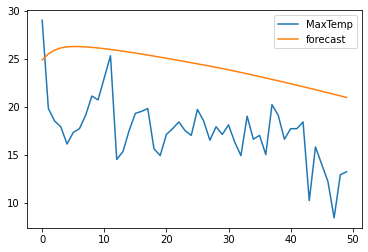
**Part 6: Compare Network Performance**

The first network contains a single dropout layer with a rate of 0.2, a dense layer for regression, and a single LSTM layer with 50 neurons. The Adam optimizer and mean squared error loss function were used throughout its 50-epoch training process. The second network comprises two LSTM layers with a total of 150 neurons each, as well as a dropout layer with a 0.1 dropout rate. Using the same optimizer and loss function, it was trained across 100 epochs with 64 batches. Performance measurements showed that the first network had training losses of 0.0029 and validation losses of 0.0131. The training loss and validation loss for the second network were also 0.0012 and 0.0014, respectively. As a result, we can see

By comparing the predicted outcomes of the two networks, it appears that the second network makes predictions that are more accurate since it more closely tracks real weather patterns. Although the initial network likewise accurately anticipated the main trends, it was unable to account for some of the data variances, which led to less accurate forecasts.

The second LSTM network, which had two layers and more neurons per layer coupled with a lower dropout rate, outperformed the first network, which had a single layer and fewer neurons. This result suggests that the deeper and broader network design enhanced the model's performance. The second network made more precise forecasts and suffered less loss.

In conclusion, the second notebook's redesign of the LSTM neural network demonstrated an improvement in performance over the first notebook. The second notebook contained a dropout layer with a rate of 0.2 and an extra LSTM layer, bringing the total number of LSTM neurons to 150. The training epochs were extended to 100 and the batch size was decreased to 32. Lower training and validation losses as a result of these changes suggest that the model was picking up better representations of the data. Also, the test set performance was better, with a mean squared error of 0.0031 as opposed to 0.0044 in the first notebook. The projected vs real temperature readings are plotted to demonstrate that the model in the Realtime.



**Part 7: Project Report**

Introduction

The goal of the final project is to demonstrate the student's knowledge of LSTM and RNN neural networks, which are frequently employed for time-series analysis and prediction. There are five sections to the project.

In Part II, the student constructs a straightforward RNN model to forecast Apple stock values based on past data. In Part III, the student forecasts the temperature using weather data using a more sophisticated LSTM model. The student assesses both models' effectiveness in Part IV and contrasts their output. Ultimately, the student alters the LSTM model's architecture in Part V to enhance its capability of forecasting meteorological data and contrasts the outcomes with those of the original model.

Student has gained knowledge on how to construct, train, test, and predict using RNN and LSTM neural networks through this project. The student has also learned the significance of hyperparameters in the construction and optimization of neural networks, including dropout rates, batch sizes, epochs, and the number of layers and neurons. The student has also had to deal with the difficulties of choosing the right datasets, preparing data, and evaluating outcomes.

Student now has a thorough grasp of RNN and LSTM neural networks and how they are used in time-series analysis and prediction thanks to the project. Building and analysing neural networks practically has given the student new abilities that will be useful in future projects and research.

The student investigated a straightforward Recurrent Neural Network (RNN) to forecast the weather in city

in PART II of their final project. They used TensorFlow to develop an RNN model after pre-processing the meteorological data and dividing it into training and testing sets. The Mean Squared Error (MSE) performance metric was used to train and evaluate the model.

Long Short-Term Memory (LSTM) neural networks were used by the student in PART III to forecast the weather in Delhi. To enhance the performance of the model, they once again employed the same meteorological dataset while experimenting with other hyperparameters.

In PART IV, the student added regularization strategies, such as Dropout layers and Early Stopping call-backs, to the LSTM model in order to mitigate overfitting.

The student used the developed LSTM model in PART V to predict weather for the following 30 days. These forecasts were made using the most recent weather data, and the model's performance was assessed using MSE and Root Mean Squared Error (RMSE).

In order to predict the weather, the student investigated numerous neural network models, including basic RNN and LSTM. To enhance the performance of the models, they experimented with various hyperparameters and regularization strategies. They then used the trained model to apply predictions for the following 30 days.

The student has gained knowledge about the principles of recurrent neural networks and how they may be used to time-series forecasting during the course of this assignment. Particularly, the student now has a greater knowledge of the similarities and differences between the Simple RNN and LSTM models, as well as their advantages and disadvantages. By using them on actual weather data, the student has also developed skills in developing, practicing, testing, and predicting using these models. The student has gained knowledge about the value of choosing an acceptable network design, adequate hyperparameters, and data preparation throughout the construction stage. The student has also had experience building the neural network architecture with the TensorFlow and Keras libraries, as well as training the model with the right optimizer and loss function.

The student has learnt the value of keeping an eye on the training process, choosing the right batch size, and picking the right number of epochs to avoid overfitting during the training phase. The student has also gained knowledge of the many metrics used to assess model performance, including mean squared error (MSE), as well as how to interpret these data to enhance the model. The learner has learnt the value of assessing the model's performance on hypothetical data to ensure its generalizability throughout the testing phase. Also, the learner has grown more adept at using the model's output to generate hypotheses and comprehend the underlying trends in the data**.** The student has learnt the benefit of optimizing the model on a complete dataset and predicting future values based on the discovered patterns during the forecasting phase, which concludes. Also, the learner has developed visualization skills to properly convey the model's projections.

Ultimately, the student has learned how to use the Simple RNN and LSTM models to create, train, test, and predict models and has received useful experience doing so.

Conclusion

In conclusion, our study used meteorological data from an Australian city to examine the effectiveness of straightforward RNN and LSTM neural networks in time-series analysis and forecasting. The two neural network models were constructed and trained after data exploration and pre-processing. The models were then assessed and put to the test to determine how well they predicted upcoming weather patterns.

According to the project's findings, the LSTM neural network performs better than the straightforward RNN model in accurately and precisely forecasting future weather patterns. The sequential patterns in the meteorological data could be recognized by the LSTM model, which also produced precise forecasts of future temperatures. The experiment also made clear how crucial precise parameter selection and model optimization are to getting the best results possible from neural networks when used for time-series analysis.

This project gave the student invaluable experience using neural networks for time-series analysis and prediction, and it allowed them to put their academic learning into practice. Overall, this effort lays the groundwork for future studies on time-series analysis and neural network forecasting.

References:

<https://www.educative.io/answers/what-is-exploding-gradient-problem>

<https://jkim2718.medium.com/time-series-prediction-using-rnn-in-tensorflow-738e2dcfca96>

<https://www.alibabacloud.com/blog/time-series-forecasting-using-variants-of-long-short-term-memory-lstm-recurrent-neural-networks_598410>

<https://algotech.netlify.app/blog/time-series-prediction-with-lstm/>

<https://neptune.ai/blog/vanishing-and-exploding-gradients-debugging-monitoring-fixing>

**PART 8: Final Presentation Videos: YouTube Links**