Final Project Presentation

2024

Shawn Nagar

Task Descriptions:

Tasks included:

Task 1 is time-series prediction with neural networks
Task 2 is the decomposition based anomaly detection
Task 3 is the prediction-based anomaly detection
Task 4 is the clustering-based anomaly detection

Task 1.1 Prediction with synthetic series using MLP, RNN, and LSTM

1. An equal-difference series starting from 0, ending to 1 (excluding 1), with a length of 200 points (step = 0.005). Design an MLP for one-step prediction. The output vector has a size of 1. Let the input vector be a size of 4

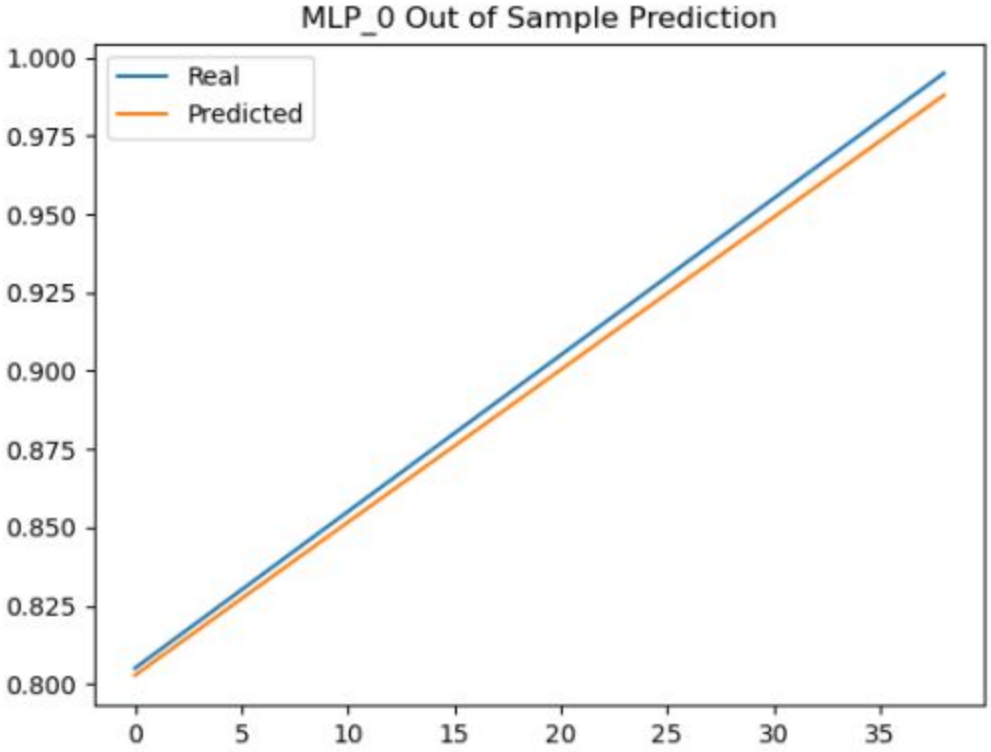
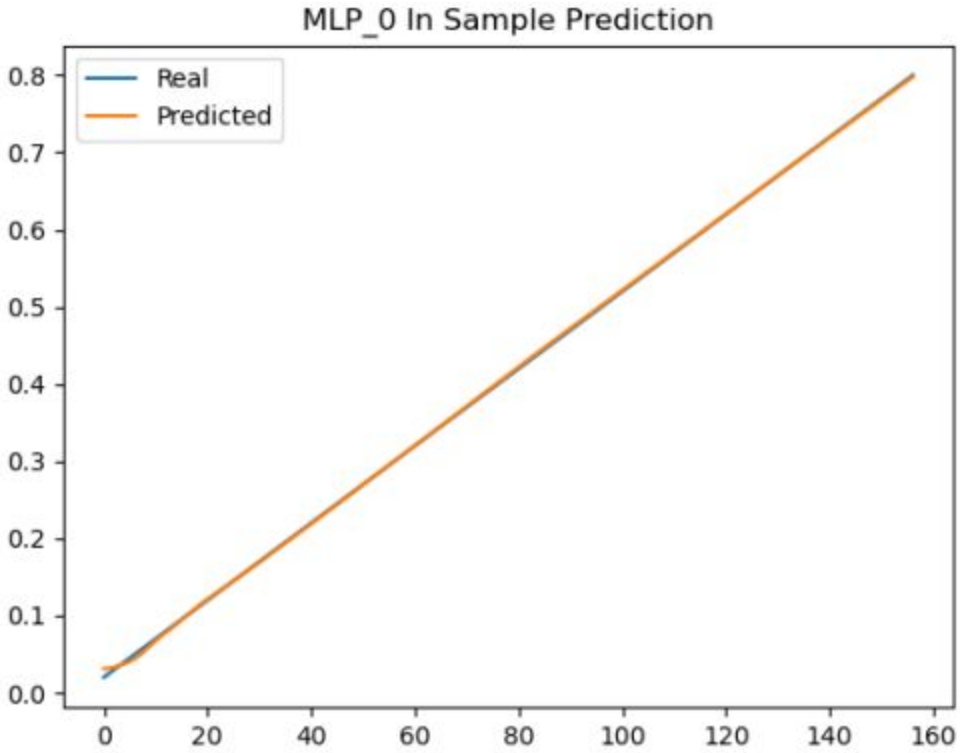


Figure 1: MLP Out of Sample Prediction


MLP_0 In Sample Prediction
Real
Predicted
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
120
140
160

2. An equal-difference series starting from 0, ending to 1, with a length of 200 points (step = 0.005), plus white noise i.e., random variable with zero mean and 1 variance. You may need to control the amplitude of the noise series in order to control the signal-noise ratio. Design an MLP for onestep prediction. The output vector has a size of 1. Let the input vector be a size of 4

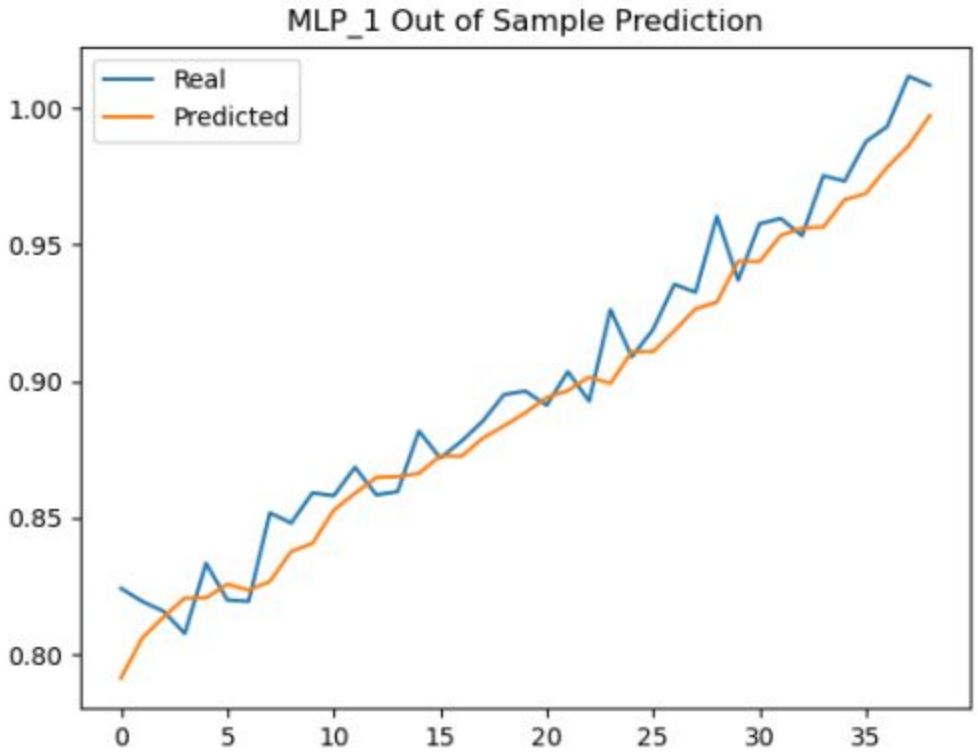


Figure 3: MLP Out of Sample Prediction

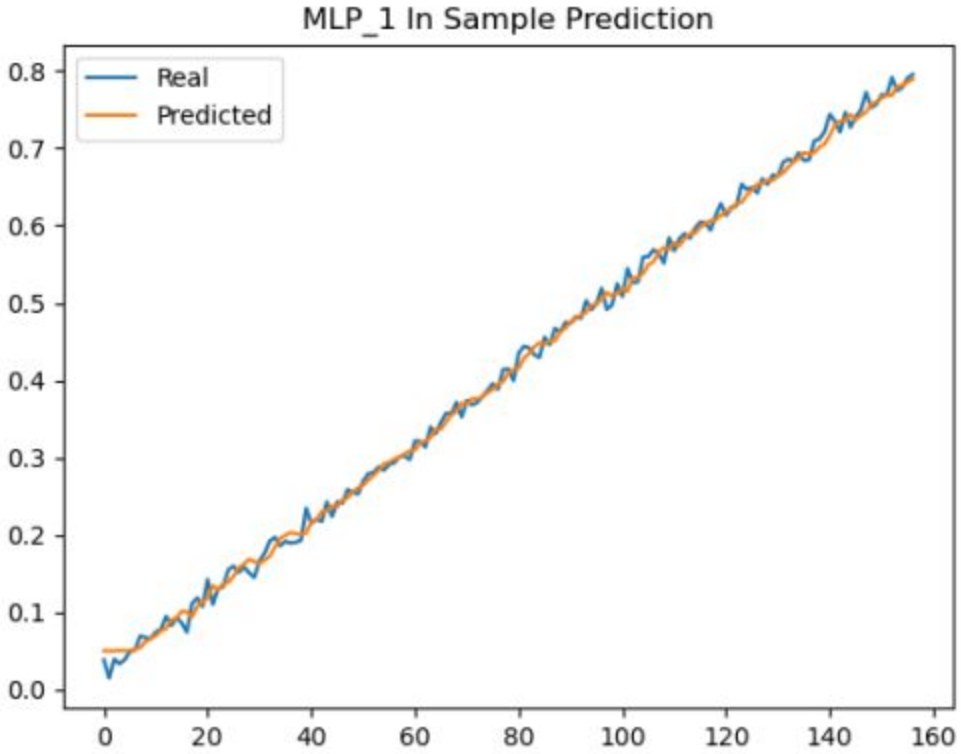


Figure 4: MLP In-Sample Prediction

3. A deterministic series sampled from a sinusoidal wave with period 20 seconds, with a sample rate of 100 Hz. Generate sufficient samples (at least 3 periods of data) as needed to achieve good performance, e.g. MSE (mean squared error) below 0.5. Design an RNN and a LSTM for two-step prediction. The output vector has a size of 2. Set the input vector size by yourself.

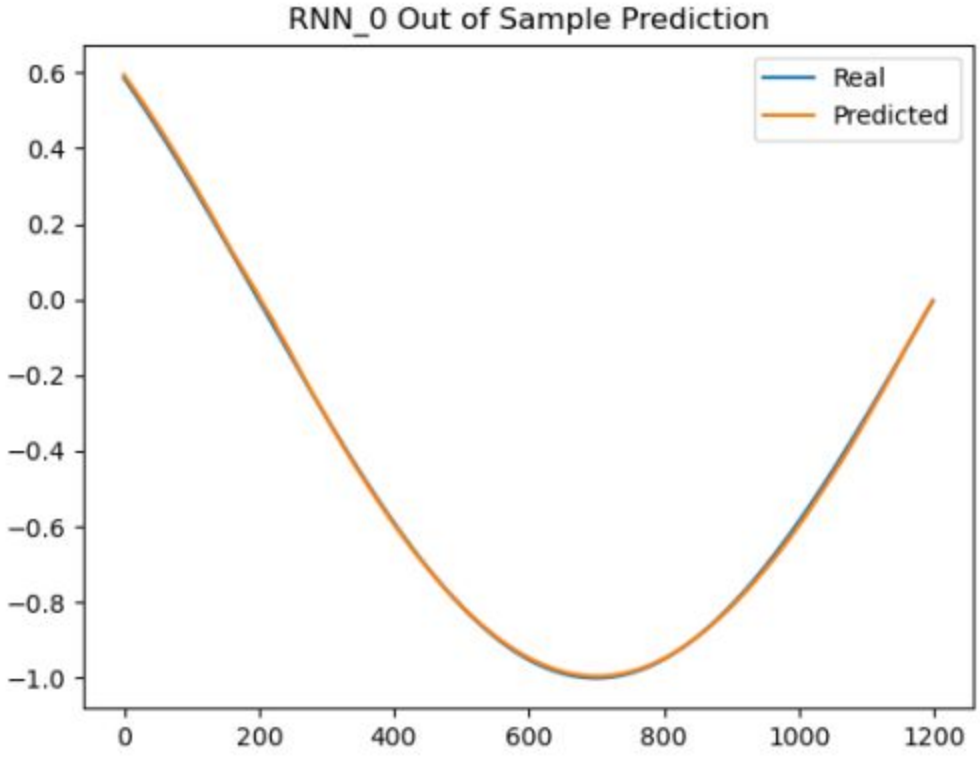


Figure 5: RNN Out of Sample Prediction

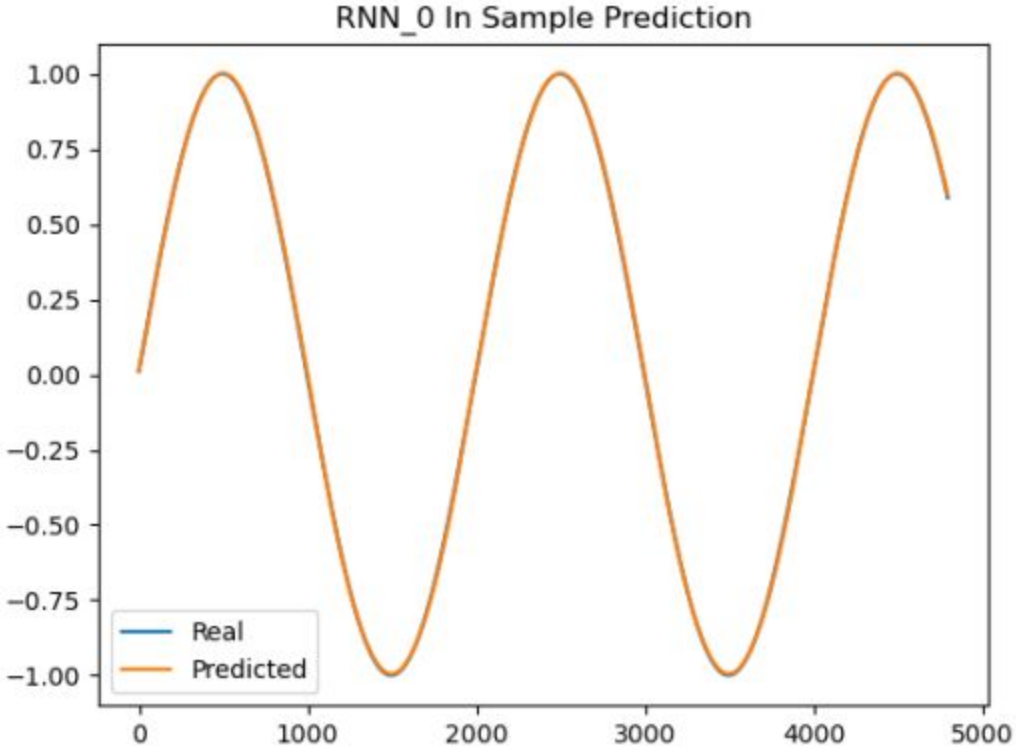


Figure 6: RNN In Sample Prediction

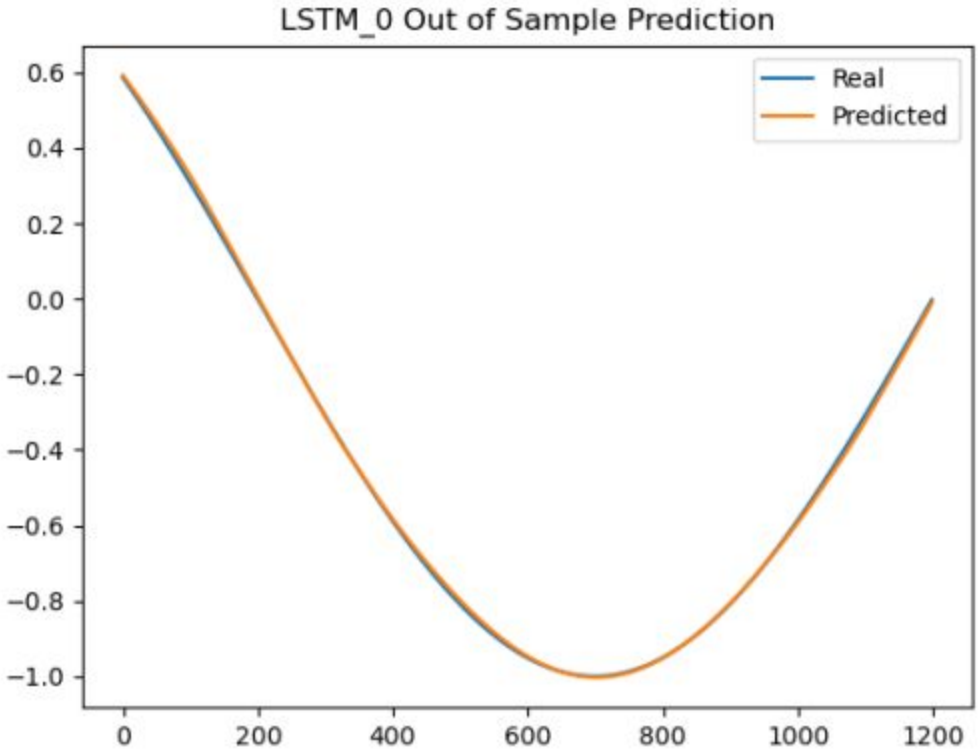


Figure 7: LSTM Out of Sample Prediction

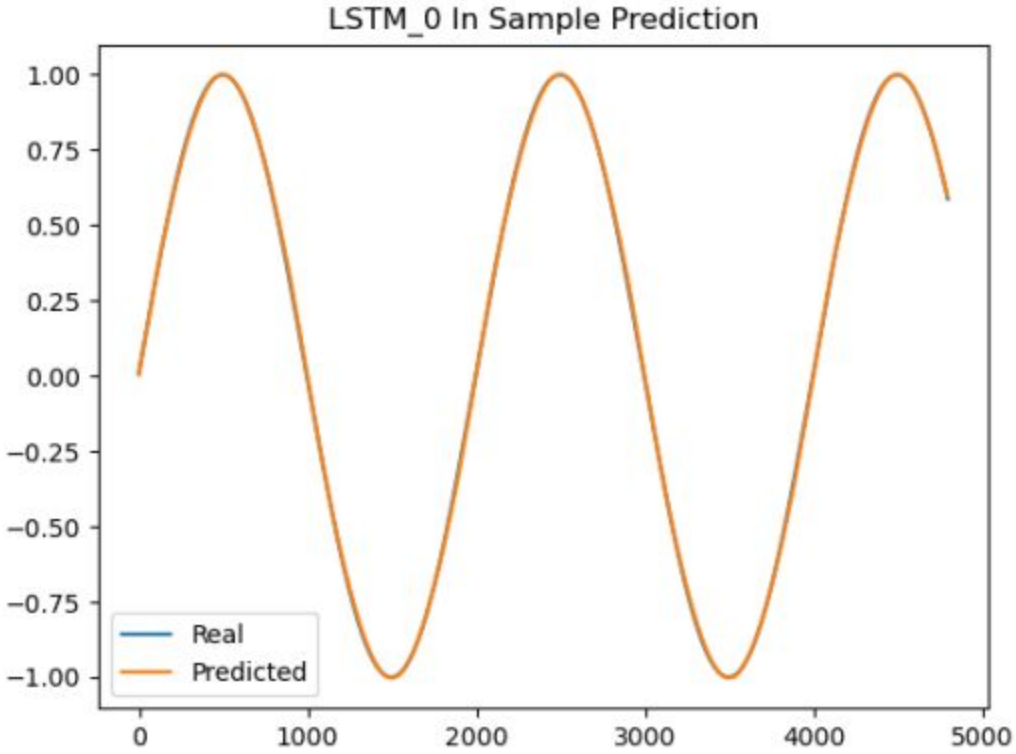


Figure 8: LSTM In Sample Prediction

4. A stochastic series sampled from a sinusoidal wave with period 20 seconds, with a sample rate of 100 Hz, plus random white noise i.e., random variable with zero mean and 1 variance. Control the amplitude of the noise with a fractional number, e.g. 0.1. Design an RNN and a LSTM for two-step prediction. The output vector has a size of 2. Set the input vector size by yourself.

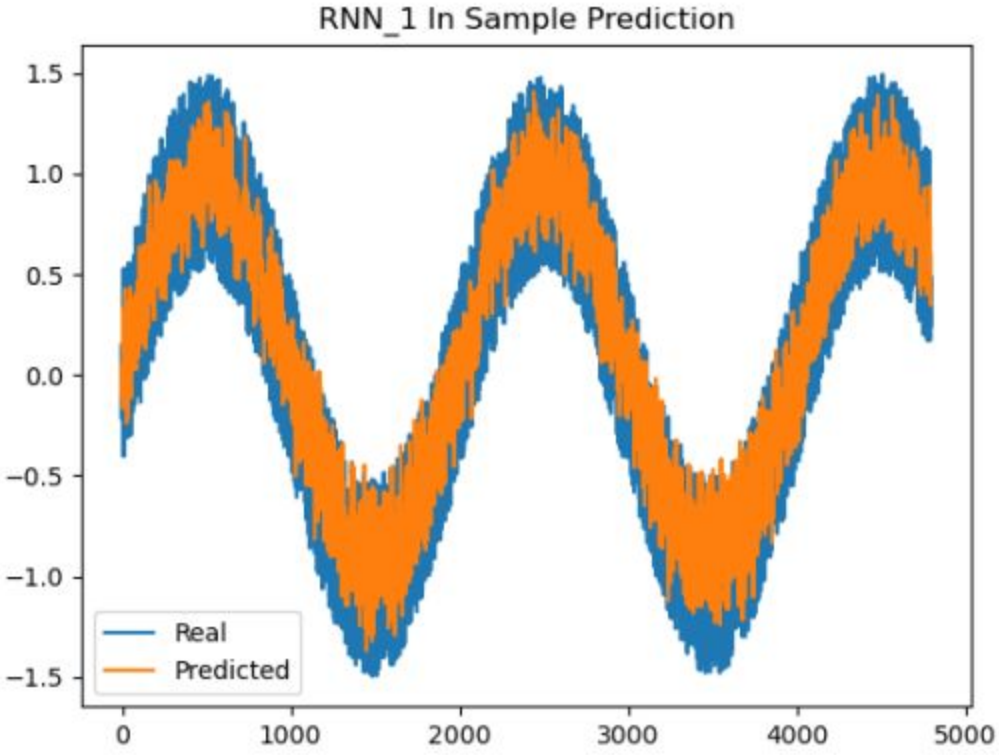


Figure 10: RNN In Sample Prediction

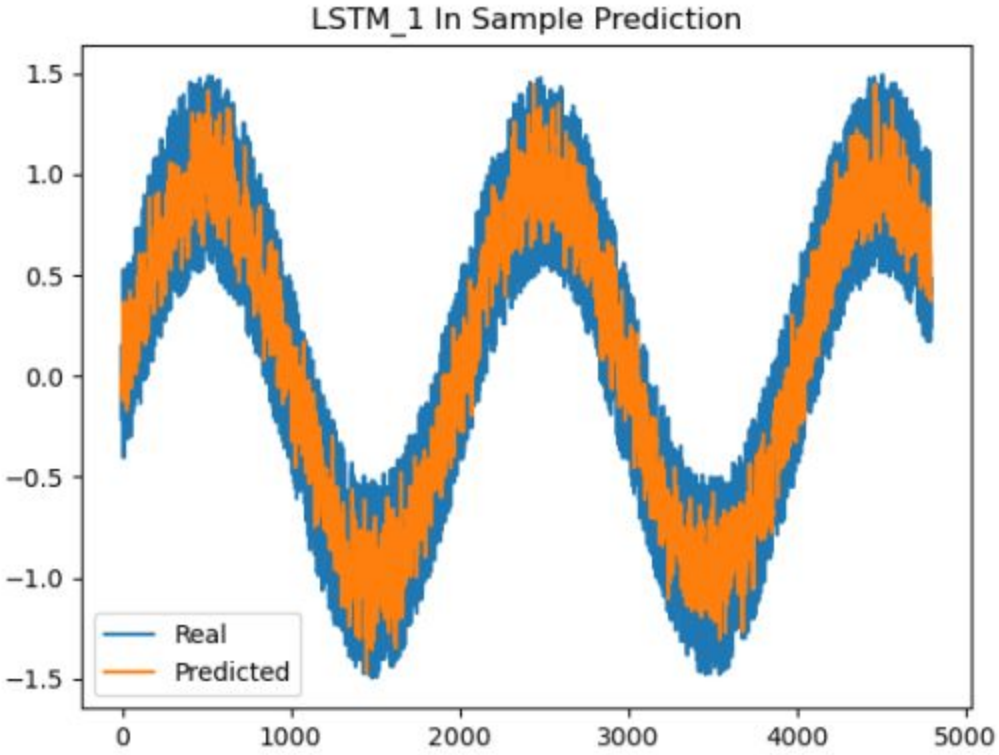


Figure 12: LSTM In Sample Prediction

Task 1.2 Predict white noise, random walk, an ARMA process using neural networks

Each series of modelled by an RNN,LSTM,MLP and ARIMA model. A comparison was made for each dataset.

Data sets:

1. A pure white-noise signal.
2. A random-walk series.
3. A stationary series generated by an ARMA(2, 2) process

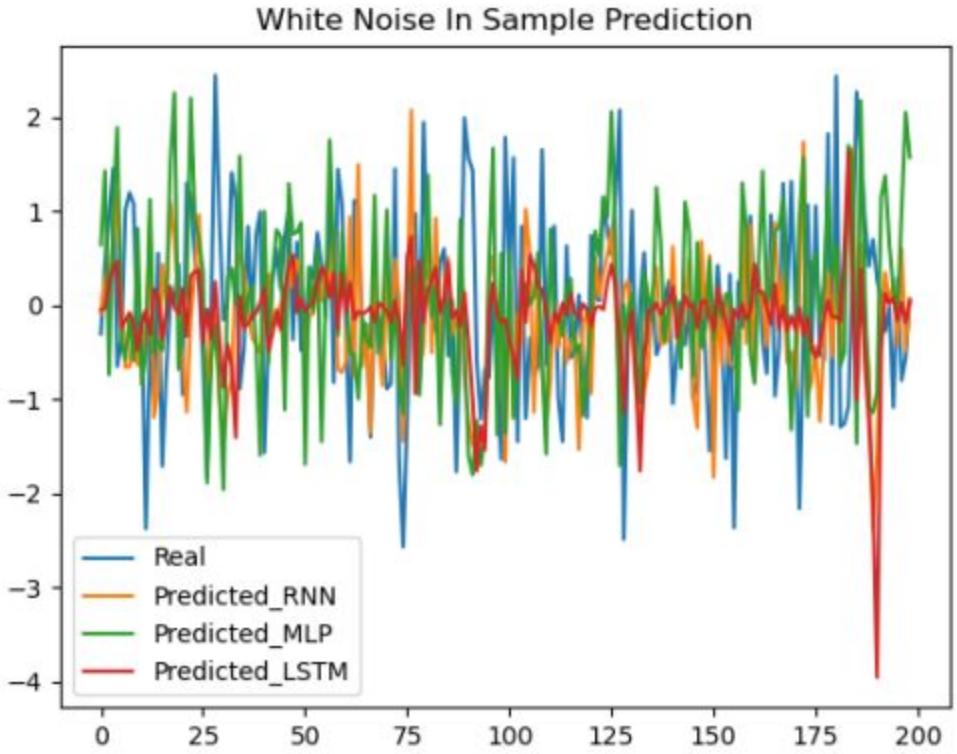


Figure 13: White Noise In Sample Prediction

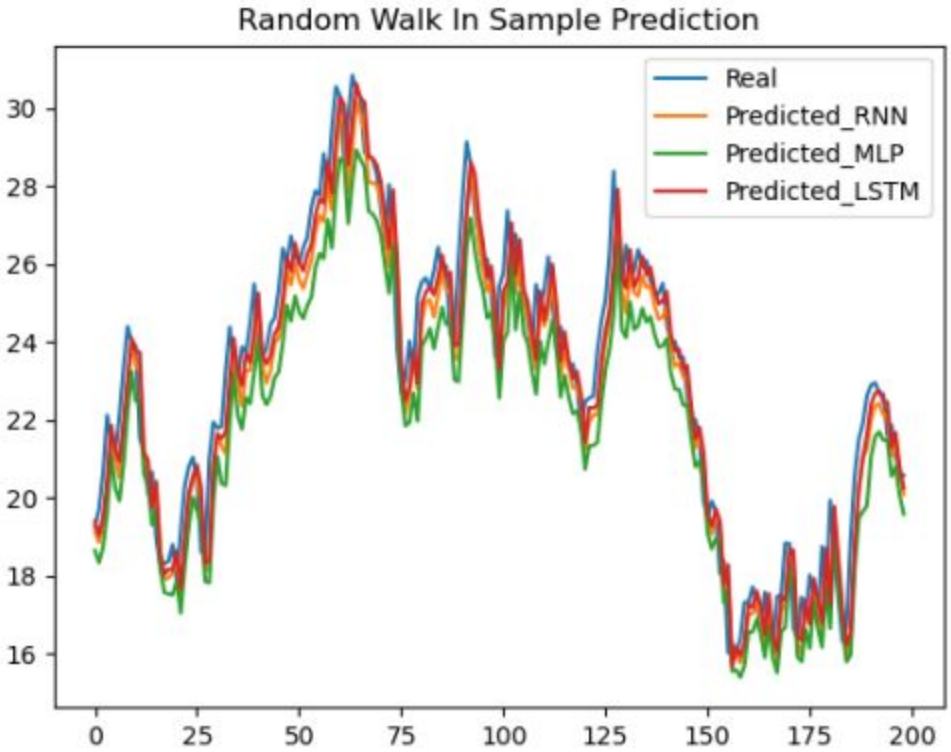


Figure 14: Random Walker In Sample Prediction

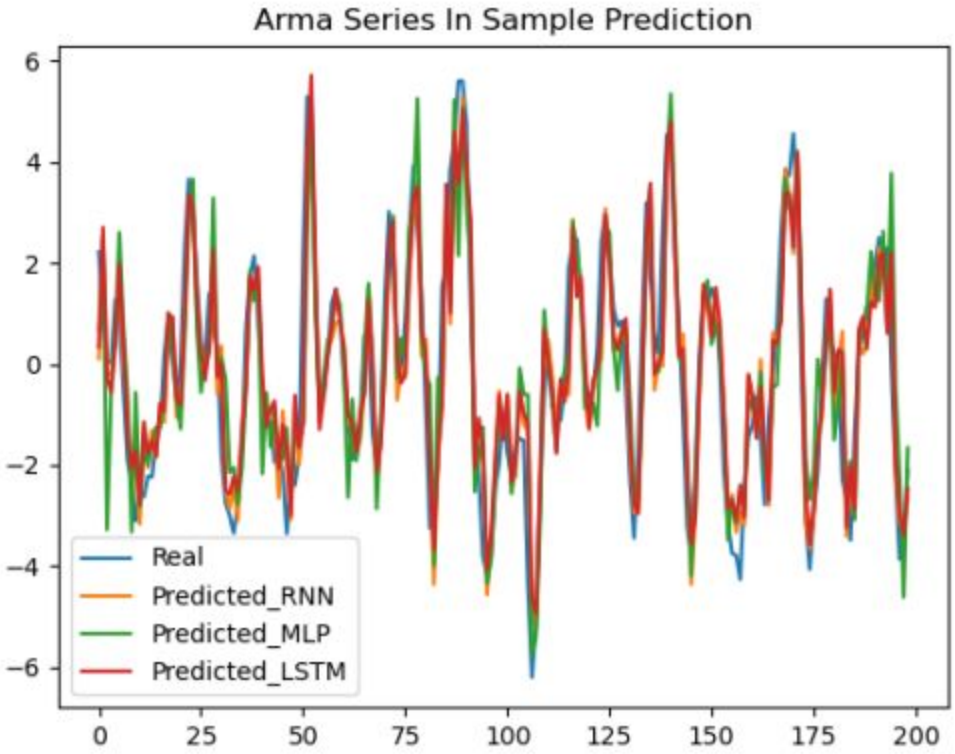


Figure 15: ARMA In Sample Prediction

Using MSE as a measurement of accuracy, LSTM seems to model all series the best.

Task 1.3 Comparison with ARIMA-based modeling and prediction

Generated a Fibonacci series and added standard Gaussian noise.

4 models (MLP,RNN,LSTM and ARIMA) built to model and predict future values

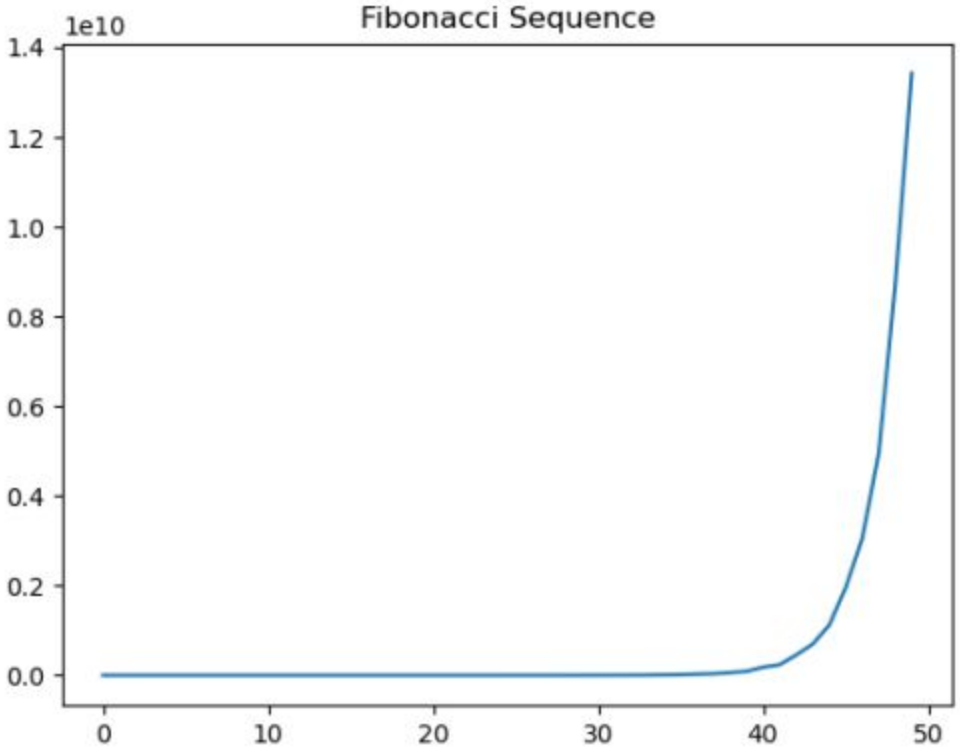


Figure 16: Fibonacci Sequence

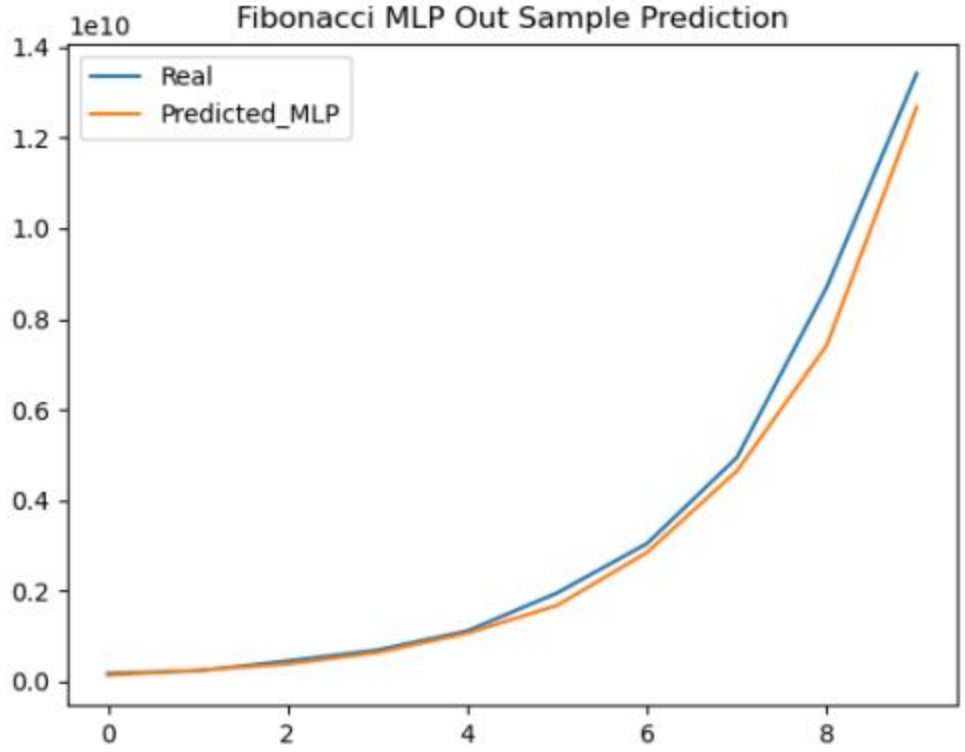


Figure 17: Fibonacci Sequence MLP Out of Sample Prediction

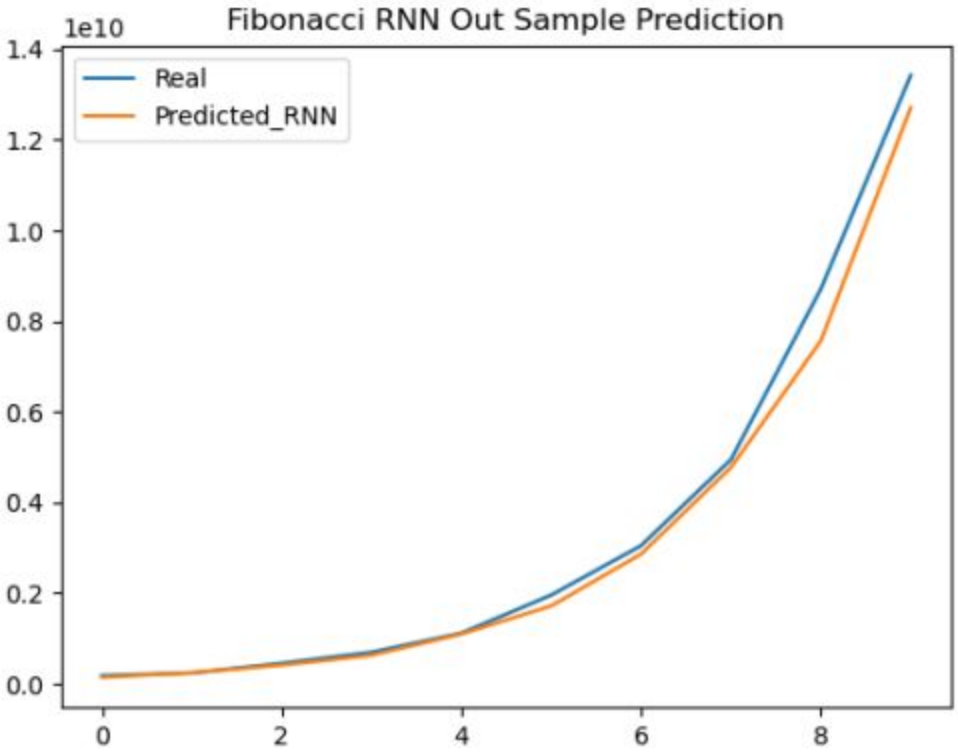


Figure 18: Fibonacci Sequence RNN Out of Sample Prediction

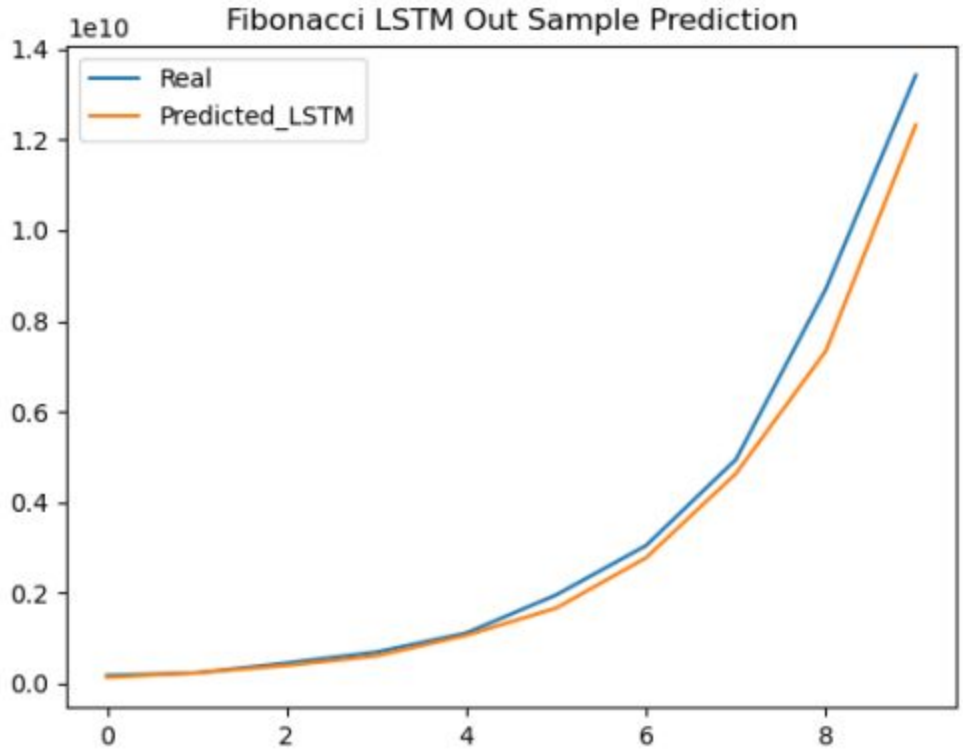


Figure 19: Fibonacci Sequence LSTM Out of Sample Prediction

For ARIMA:

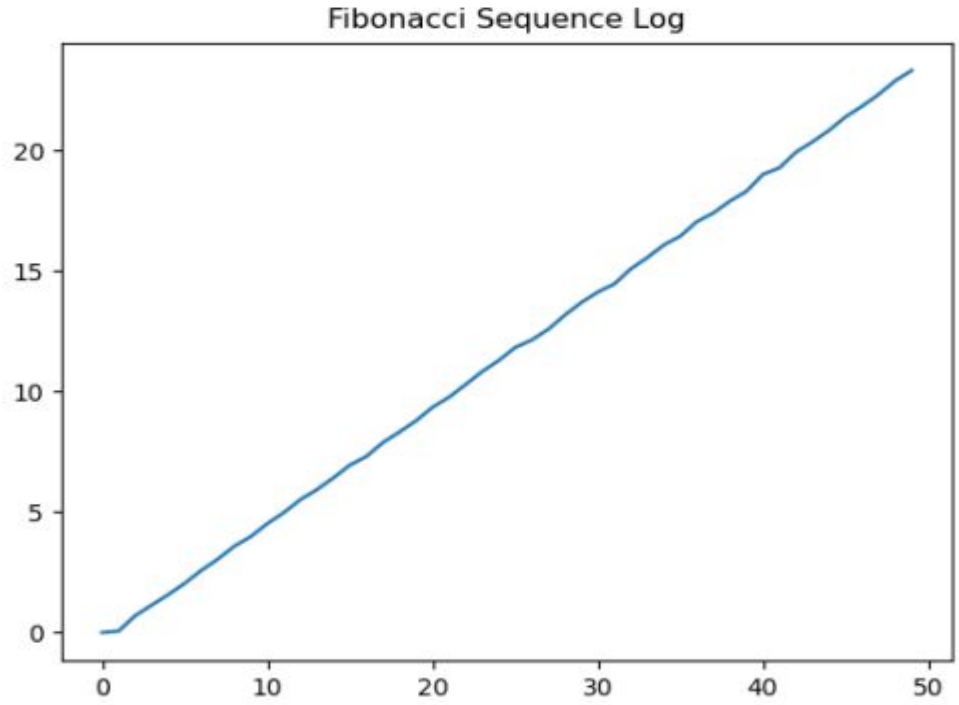


Figure 20: Fibonacci Sequence Log Transformed

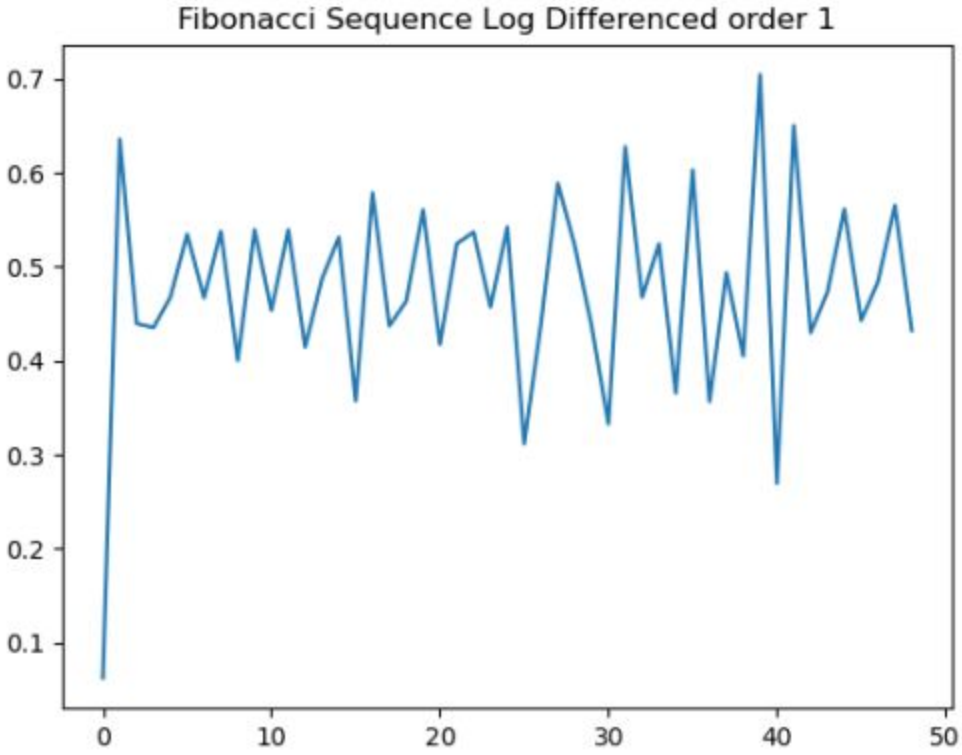


Figure 21: Fibonacci Sequence Log Transformed - Difference Order 1

ARIMA: (2,1,2)

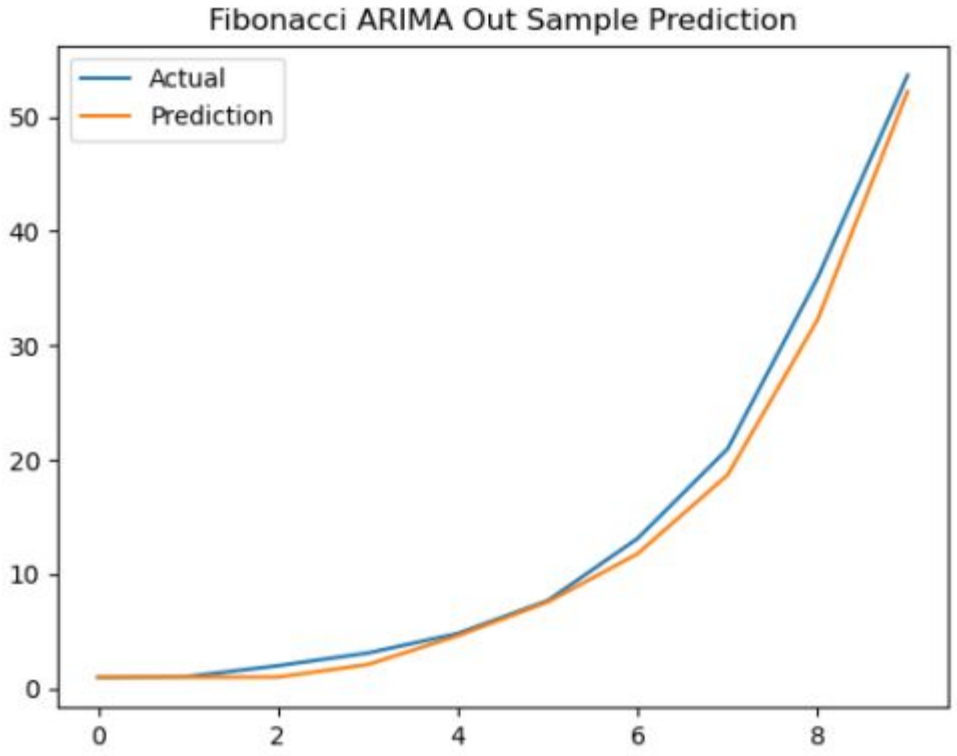


Figure 24: Fibonacci Sequence ARIMA Prediction

Clear from Accuracy metrics that ARIMA models the data best

	R2	MAE	MSE	MAPE
MLP	0.992	1.91e10	1.28e15	0.102
RNN	0.984	2.88e10	2.28	0.102
LSTM	0.986	2.72e10	2.04	0.105
ARIMA	0.966	0.154	0.06	0.17

Table 14: Fibonacci Sequence Accuracy Metrics

Task 2. Anomaly identification in global land temperature changes

One anomaly point inserted in input dataset (’LandAverageTemperature’).
Anomaly point was identified using seasonal decomposition.

An anomaly was placed 1995-01-01.

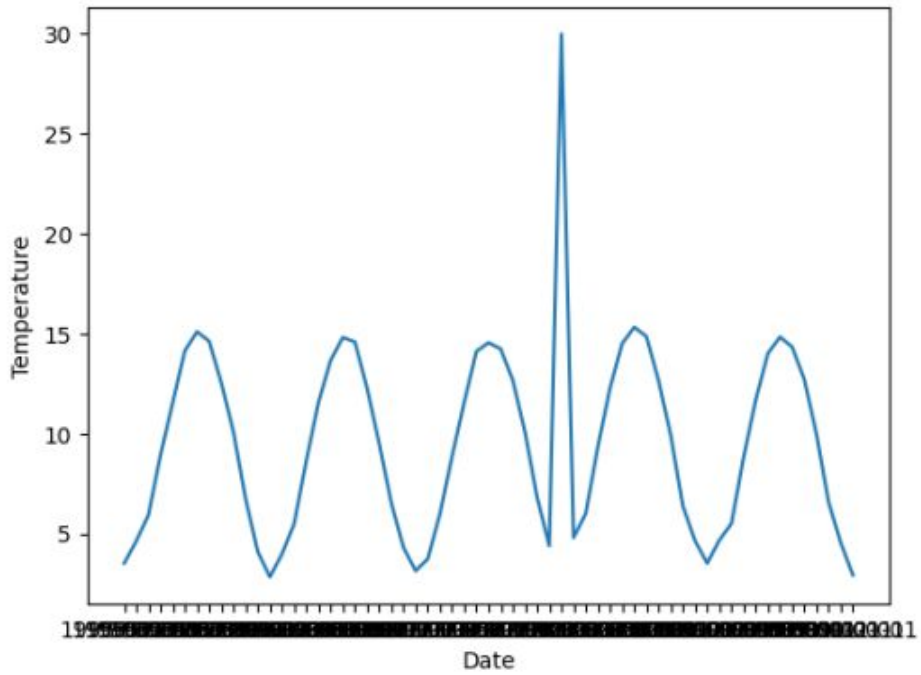


Figure 25: Global Temperature with Anomaly

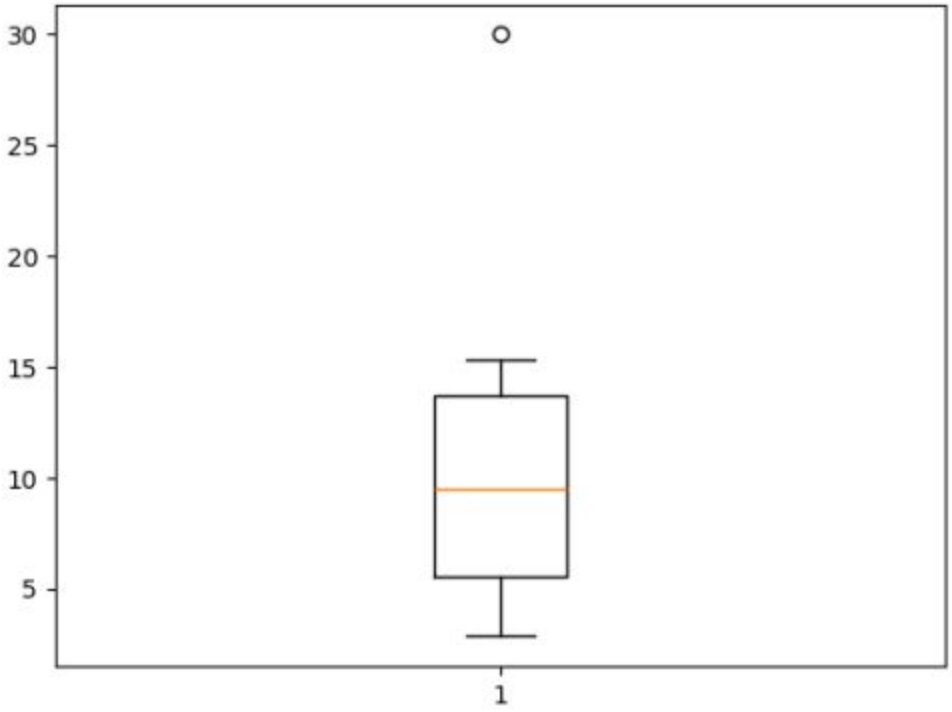


Figure 26: Global Temperature Anomaly Box Plot

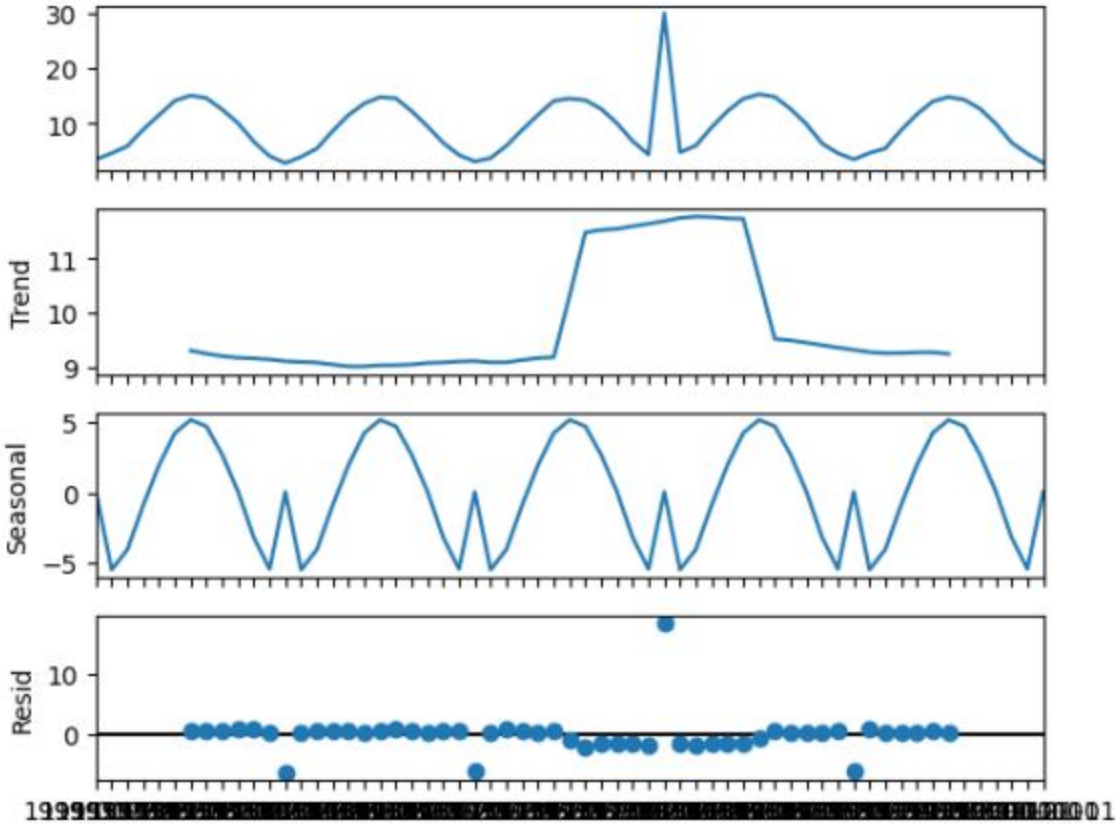


Figure 28: Global Temperature Seasonal Decomposition

Task 3.1 Anomaly detection for uni-variate series with ARIMA

To identify anomaly points from the Global Land Temperature Anomaly data set using the prediction-based anomaly detection with ARIMA

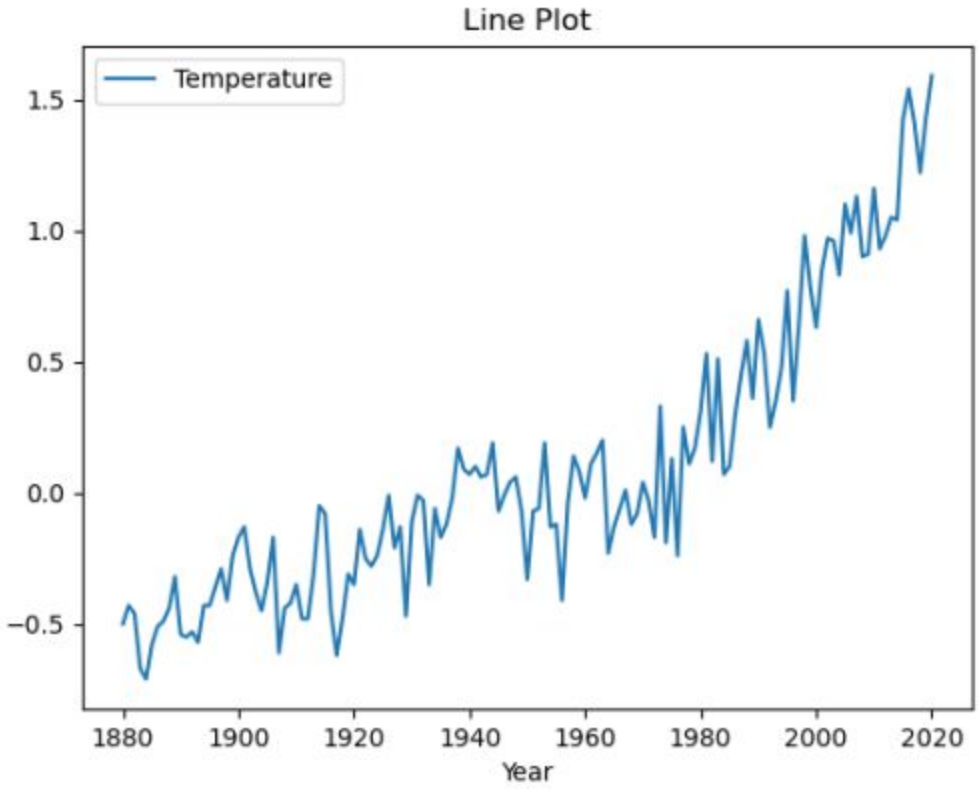


Figure 29: Global Temperature Line Plot

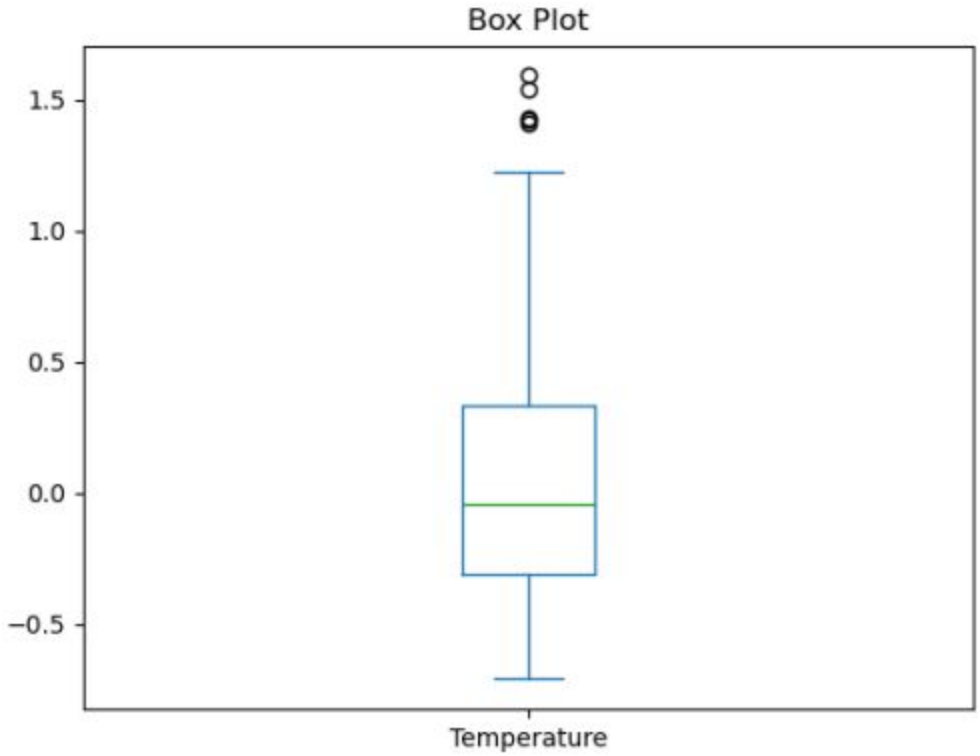


Figure 33: Global Temperature Box Plot

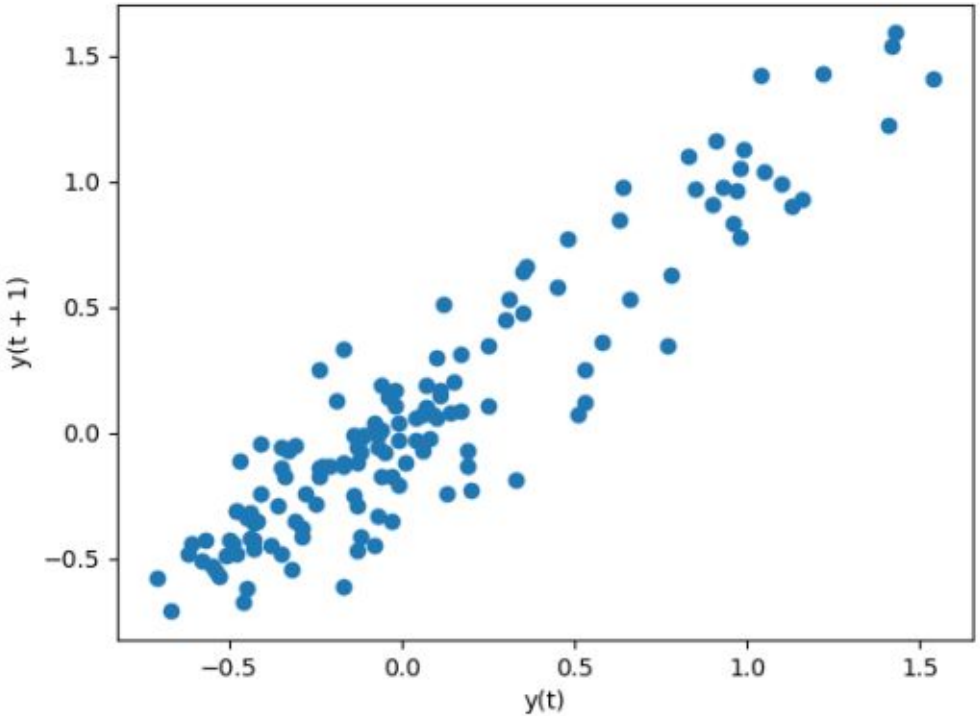


Figure 34: Global Temperature Lag-1 Plot

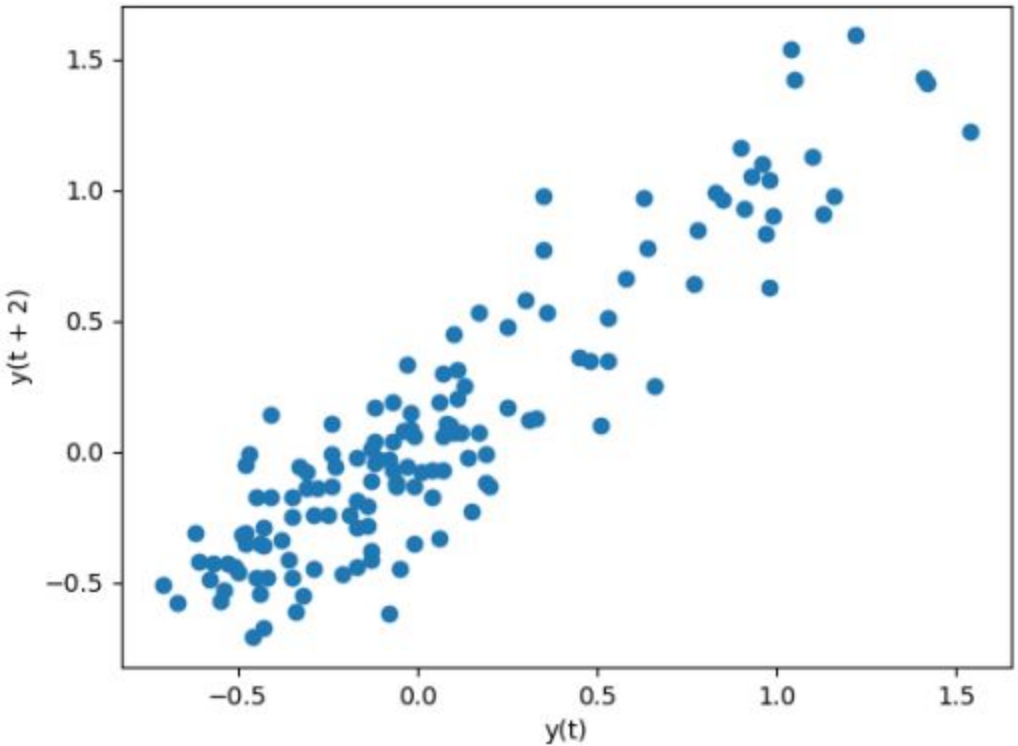


Figure 35: Global Temperature Lag-2 Plot

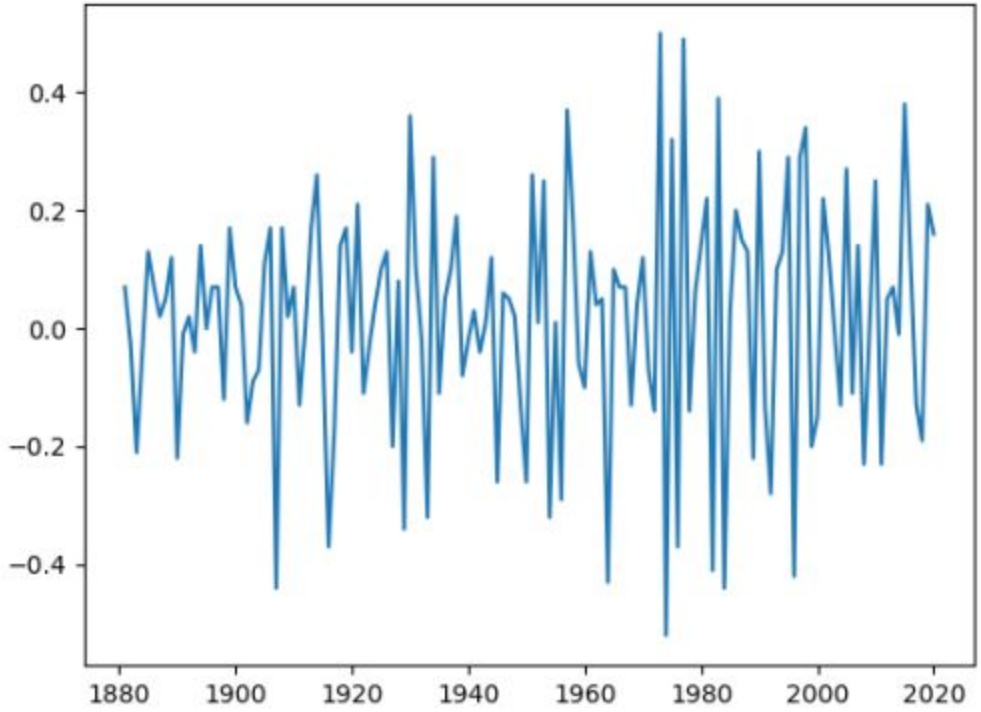


Figure 36: Global Temperature Differenced 1 Plot

ARIMA (2,1,4)

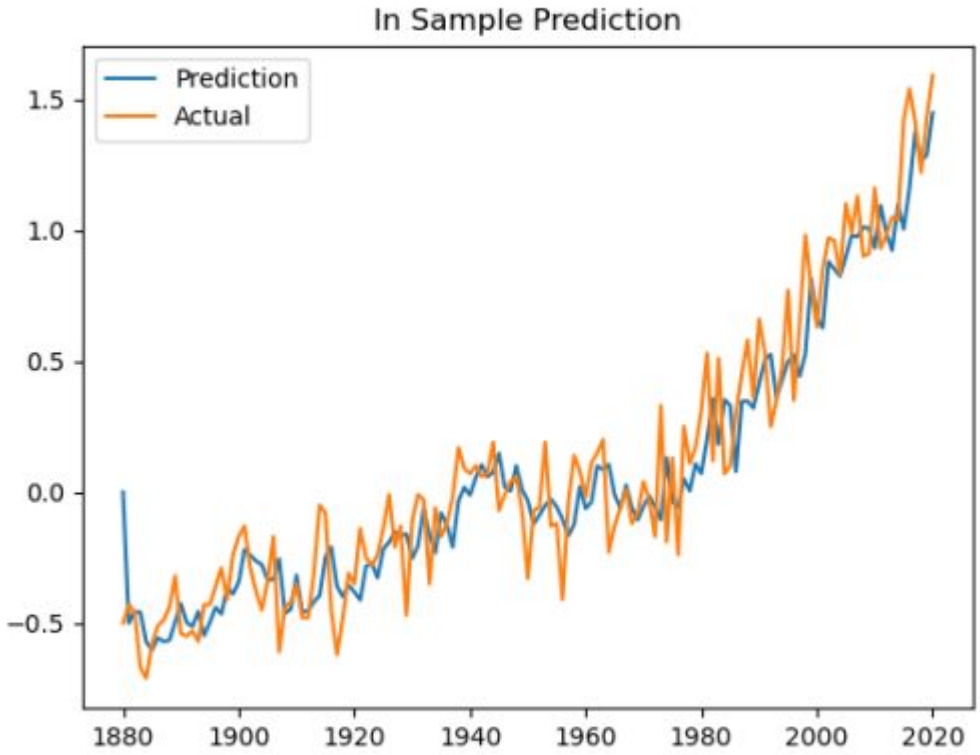


Figure 39: Global Temperature ARIMA Model Prediction

Ljung-box test: p value of 0.25 indicates the residual data is random

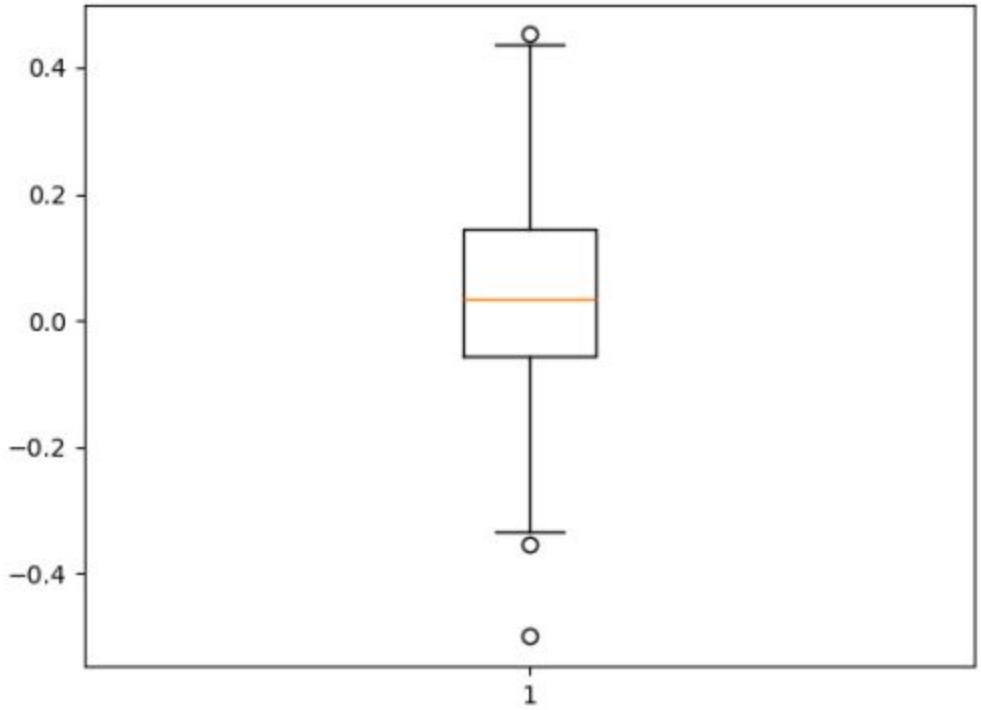


Figure 40: Global Temperature Series Residuals Boxplot

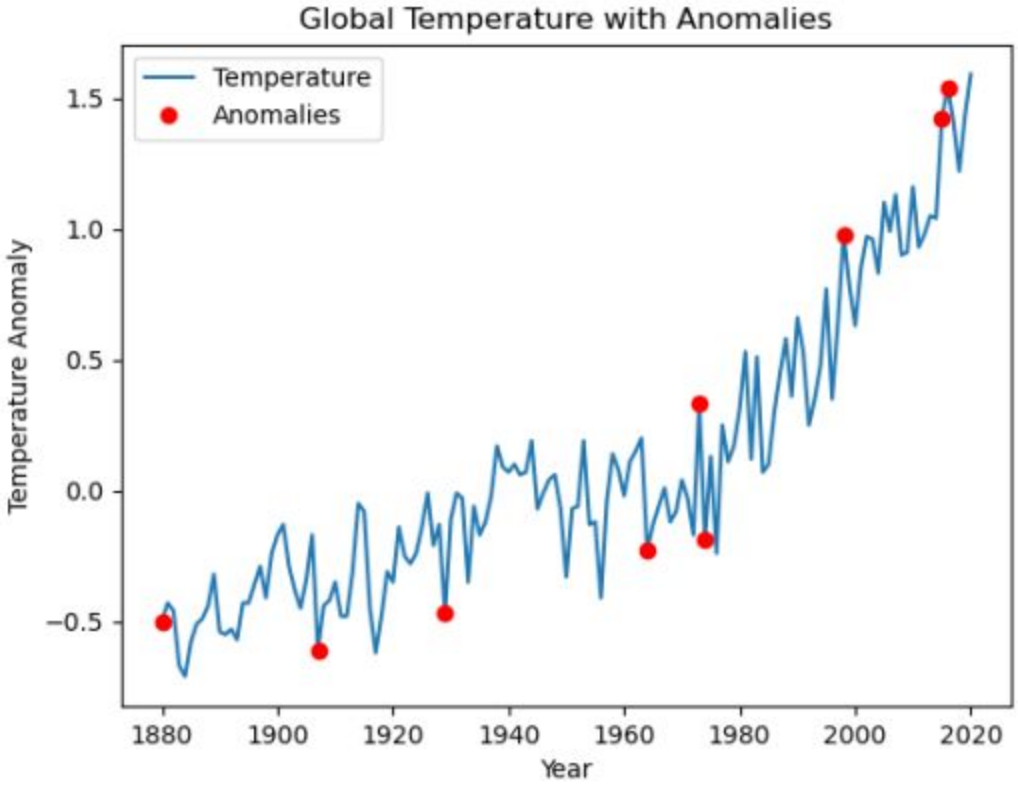


Figure 42: Global Temperature Series with Anomalies

Task 3.2 Anomaly detection in ECG signals with LSTM

ECG signals from the MIT-BIH Arrhythmia database were used to build a prediction model for the ECG series using LSTM.

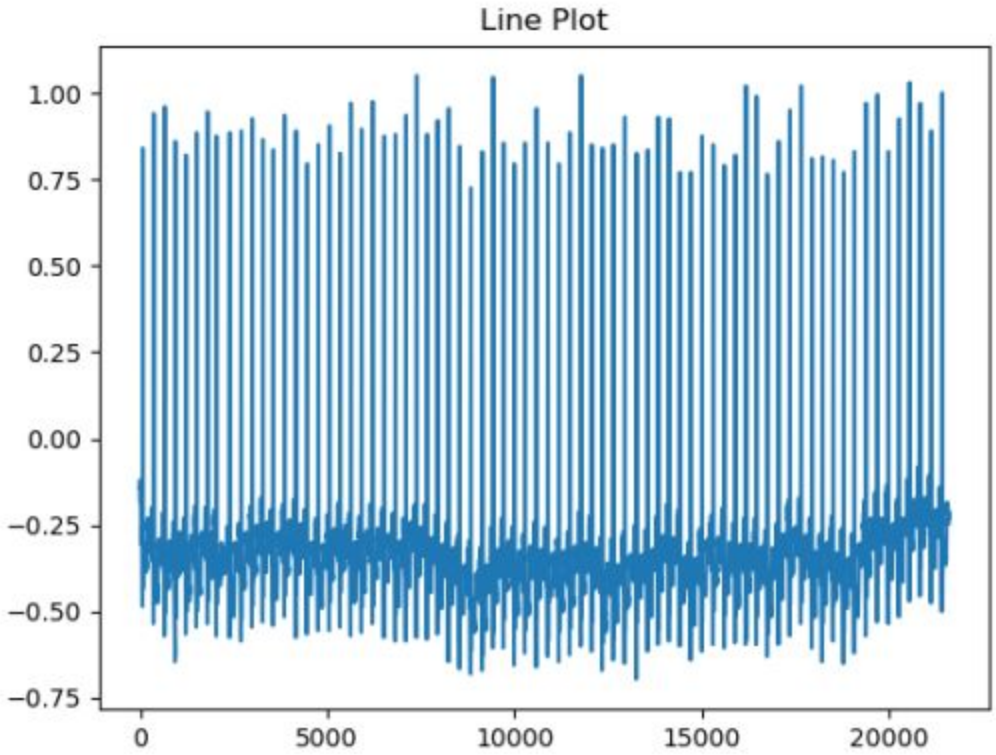


Figure 43: ECG MLII Line PLot

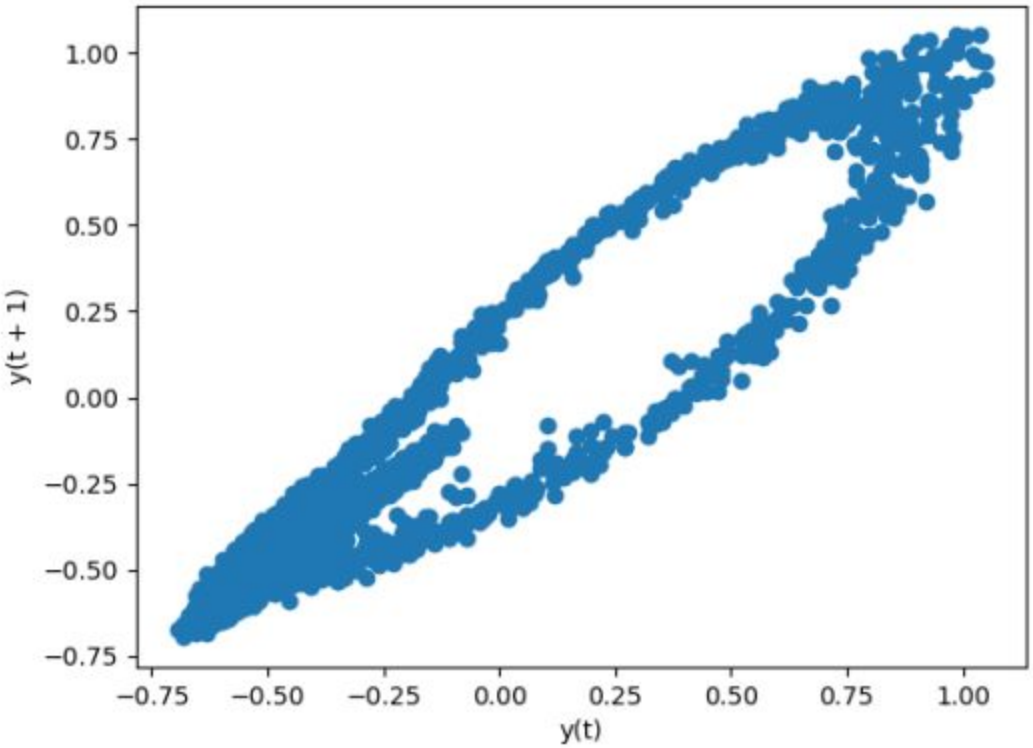


Figure 44: ECG MLII Lag-1 PLot

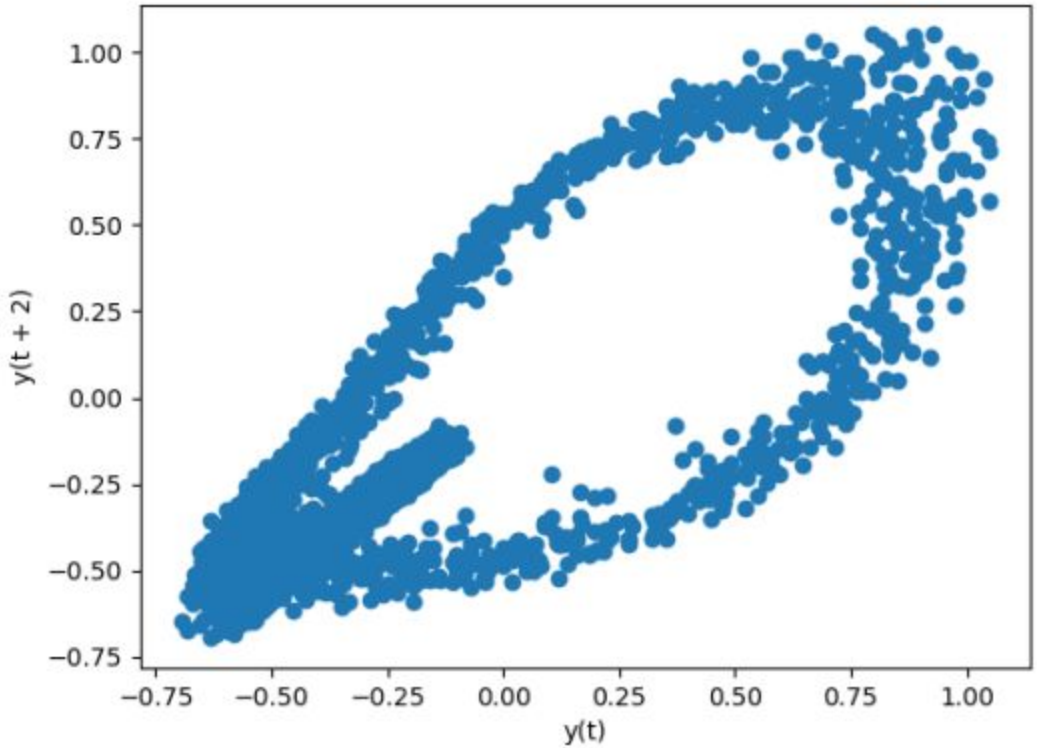
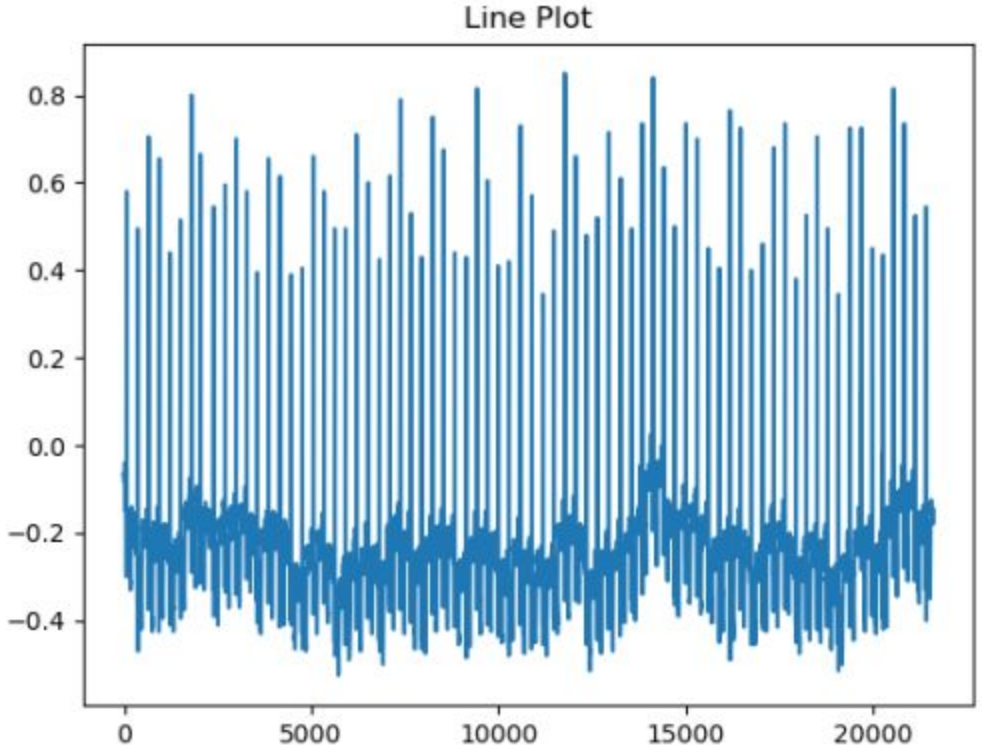


Figure 45: ECG MLII Lag-2 PLot



ECG V5 Line PLot

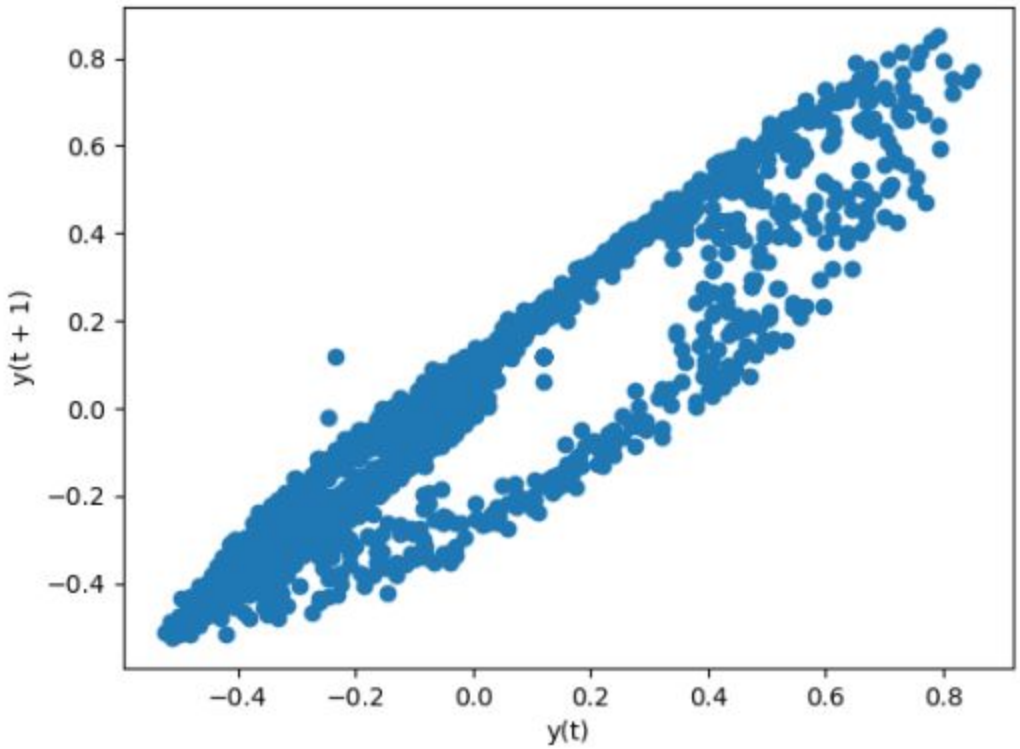


Figure 47: ECG V5 Lag-1 PLot

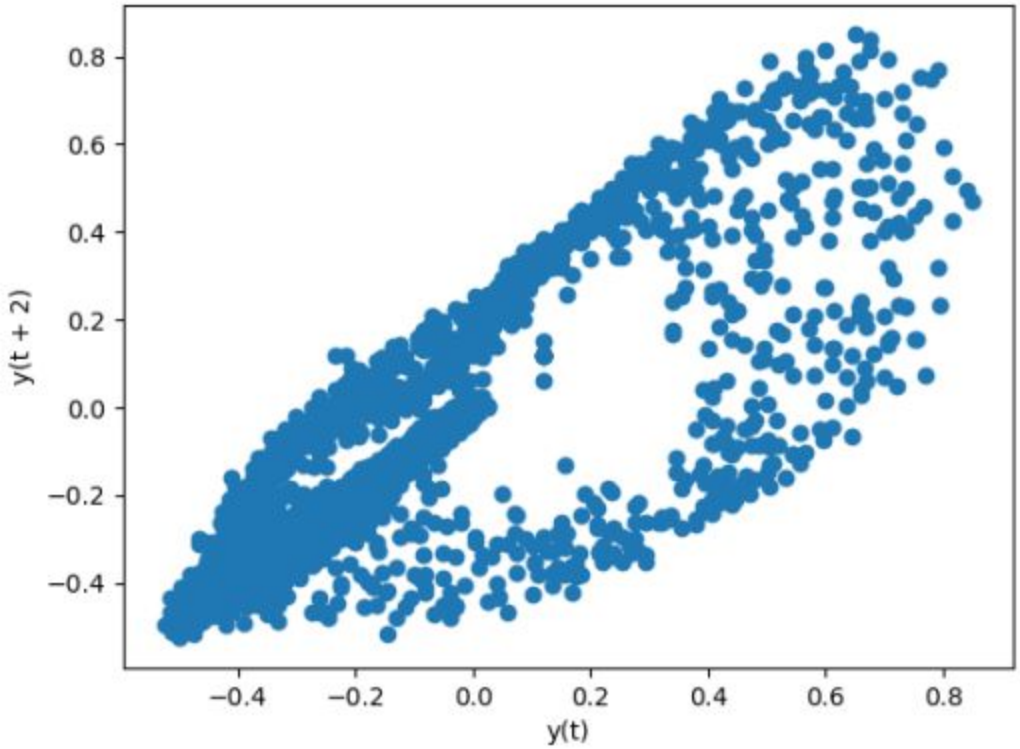


Figure 48: ECG V5 Lag-2 PLot

Models were constructed with input vectors of different dimensions (n = 4,8,16). Another consideration was to treat the data sets as univariate (V5 and MLII) or bi-variate(V5 & MLII)

No LSTM model significantly stands out in terms of accuracy

	n=4	n=8	n=16
MLII	0.00019	0.00015	0.00015
V5	0.00030	0.00021	0.00021
MLII - V5	0.00024	0.00018	0.00017

Table 16: Univariate / Bivariate Model MSE

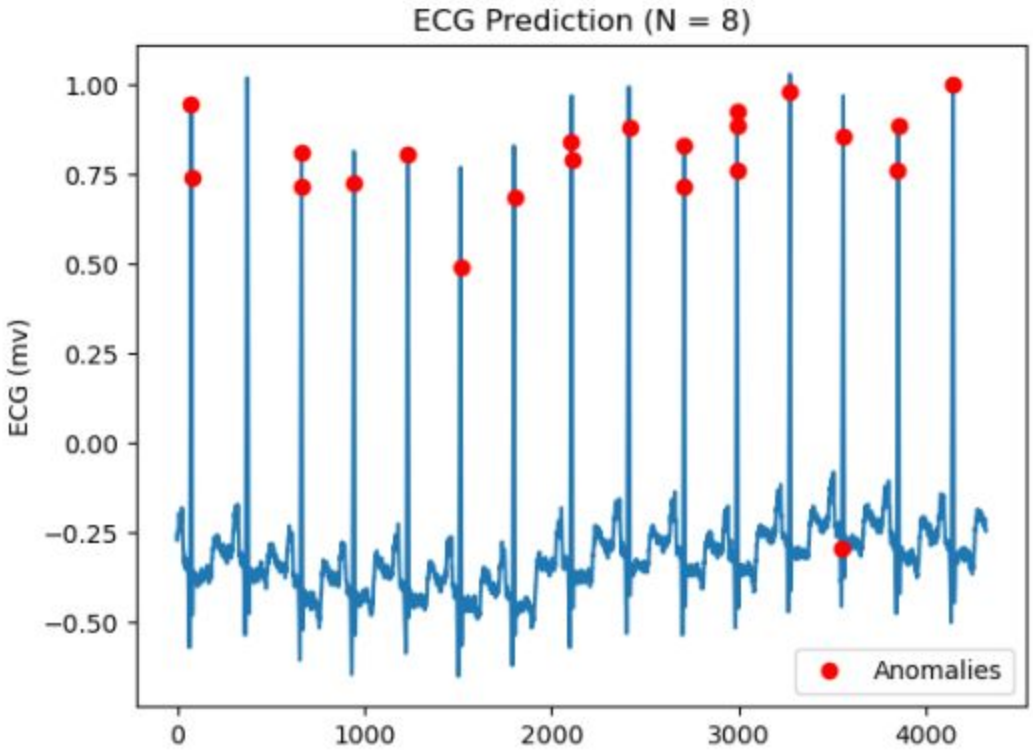


Figure 50: LSTM MLII Prediction W/ Anomaly (N=8)

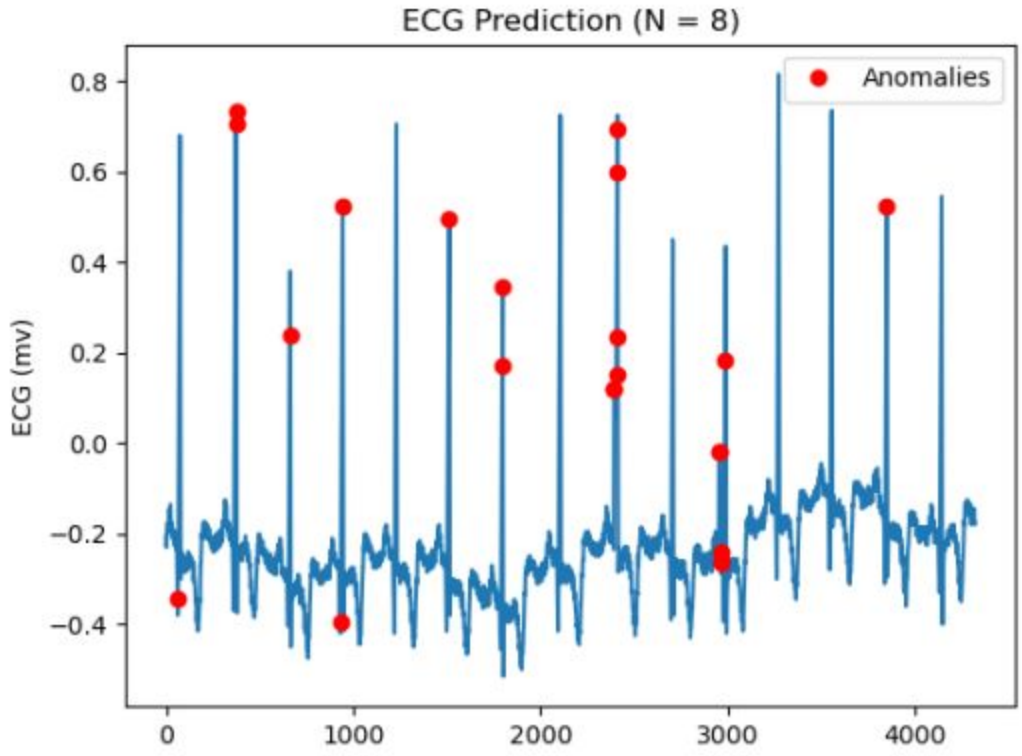


Figure 53: LSTM V5 Prediction W/ Anomaly (N=8)

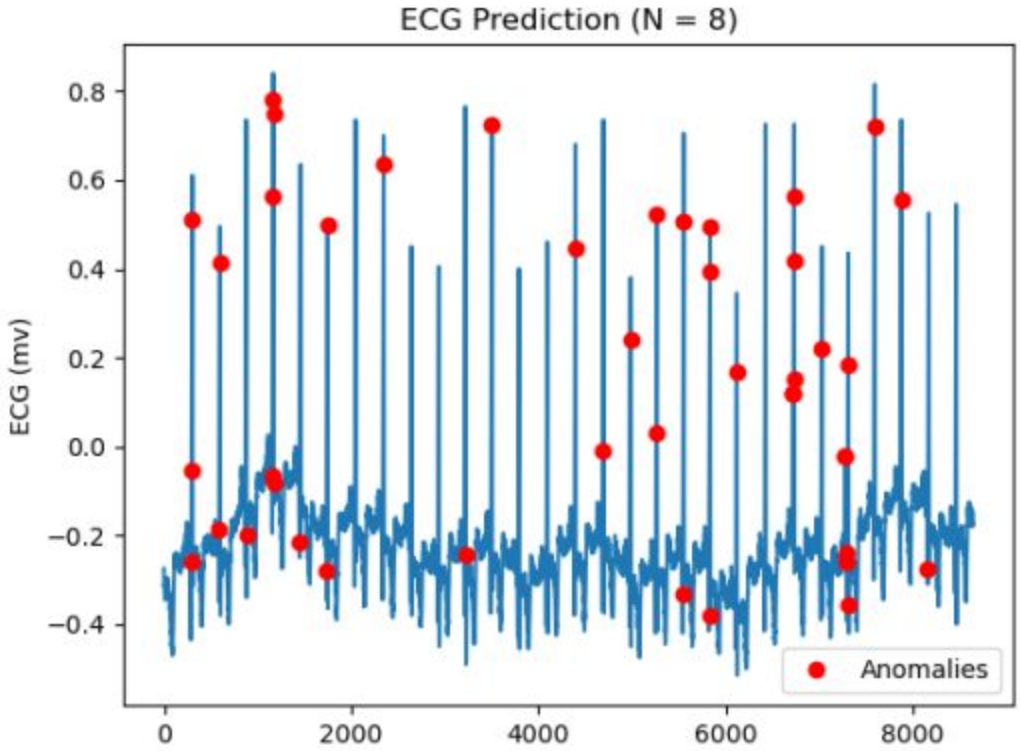


Figure 56: LSTM MLII-V5 Prediction W/ Anomaly (N=8)

Task 4. Anomaly detection in a bivariate series

Two-variable (X1, X2) stochastic time series, each variable with 200 data points were generated.

Clustering using K-means and SOM was implemented.

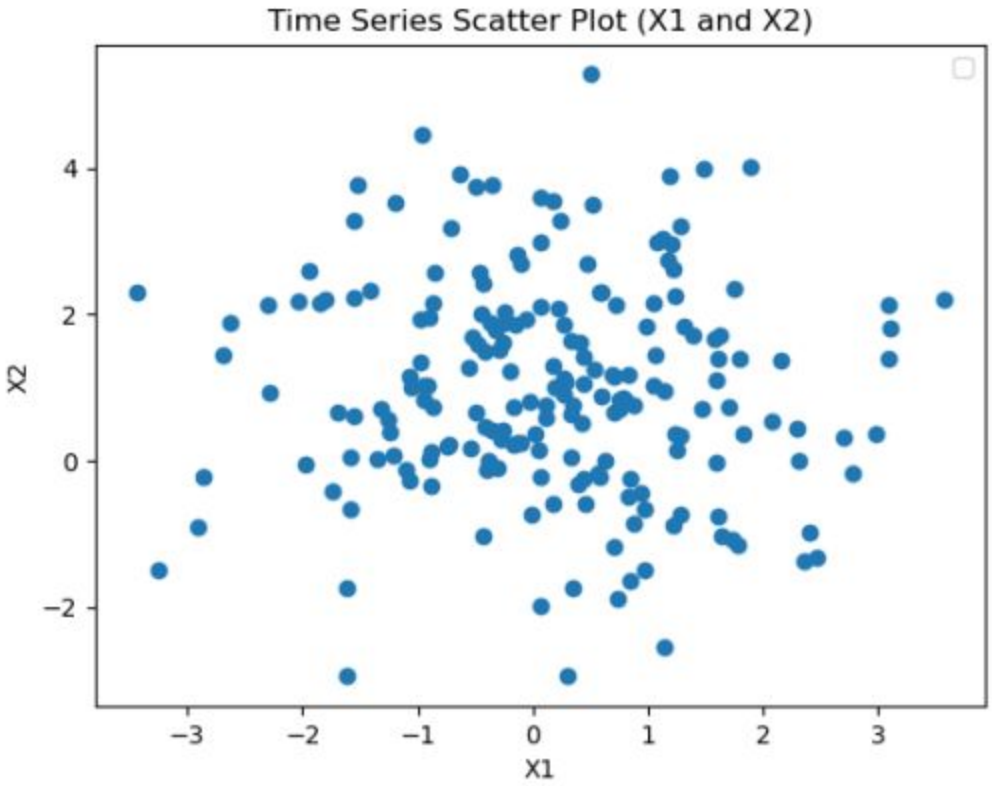


Figure 58: Cluster Time Series Scatter Plot

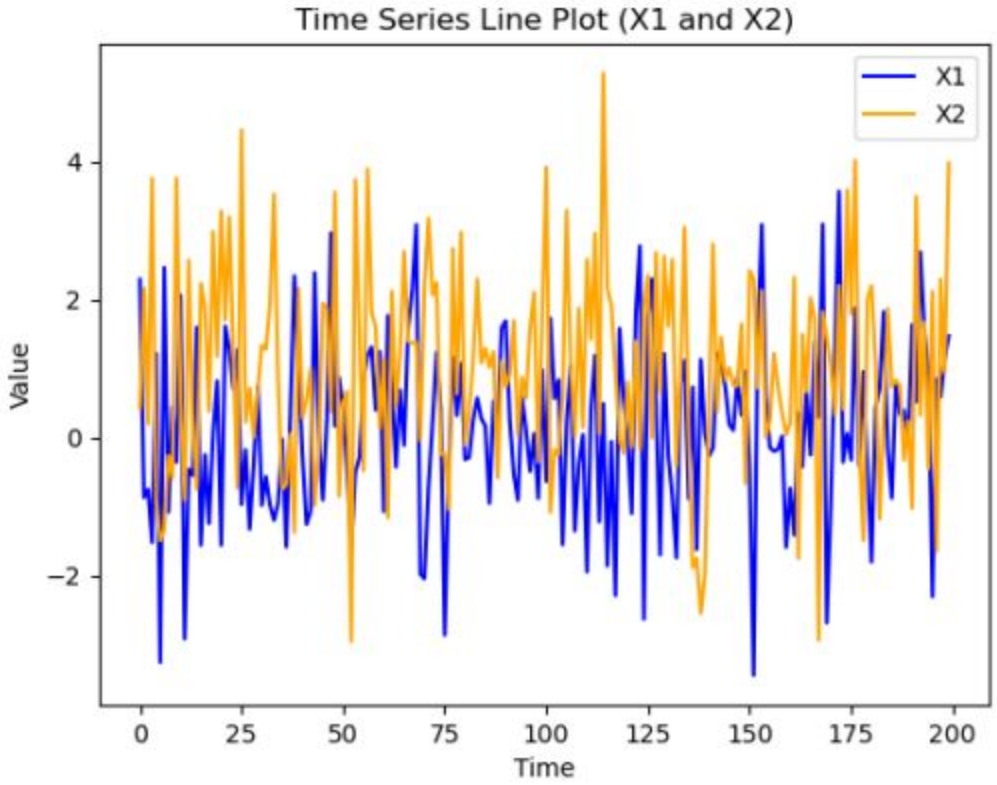


Figure 59: Cluster Time Series Line Plot

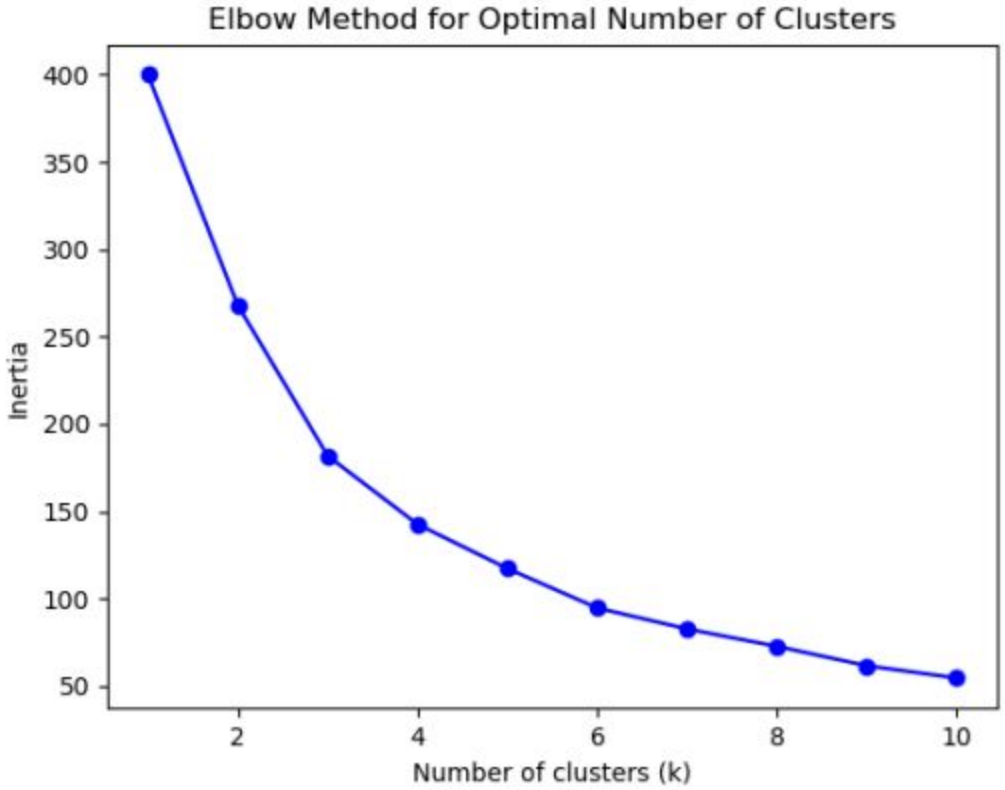


Figure 60: Cluster Elbow

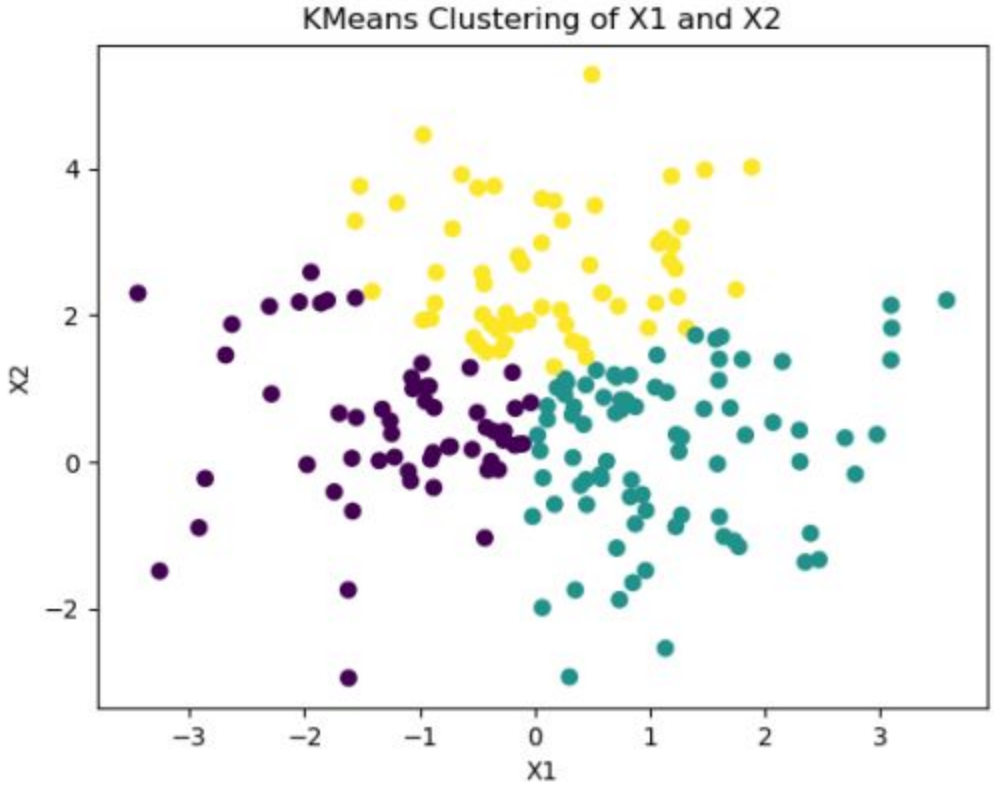


Figure 61: Kmeans Clusters

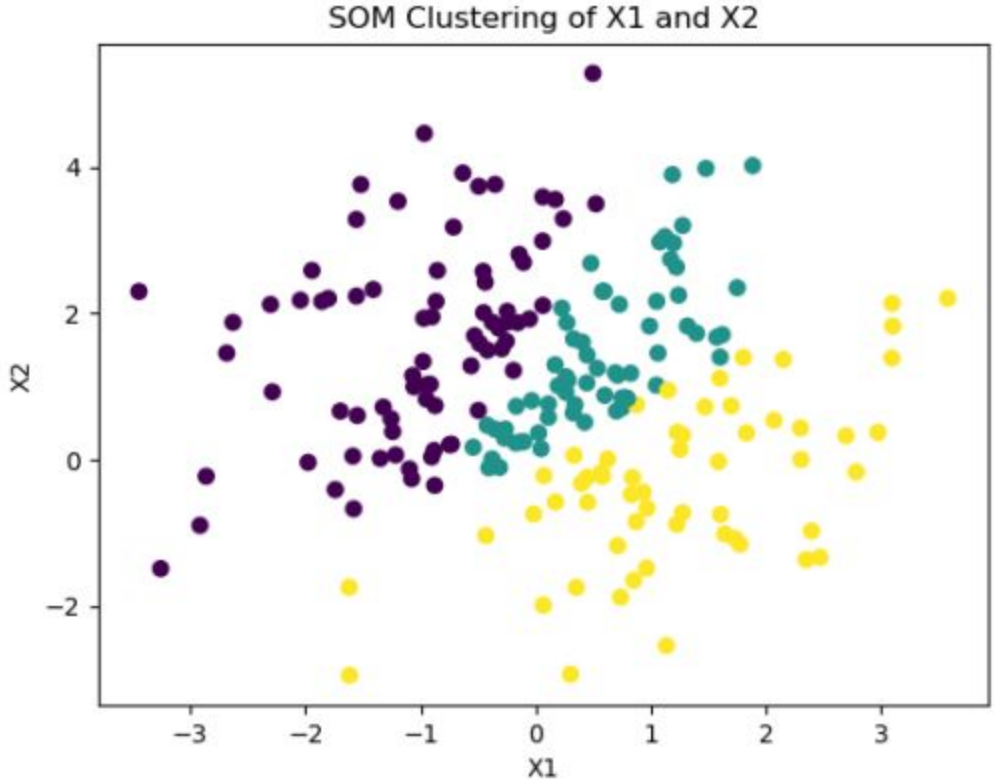


Figure 62: SOM Clusters

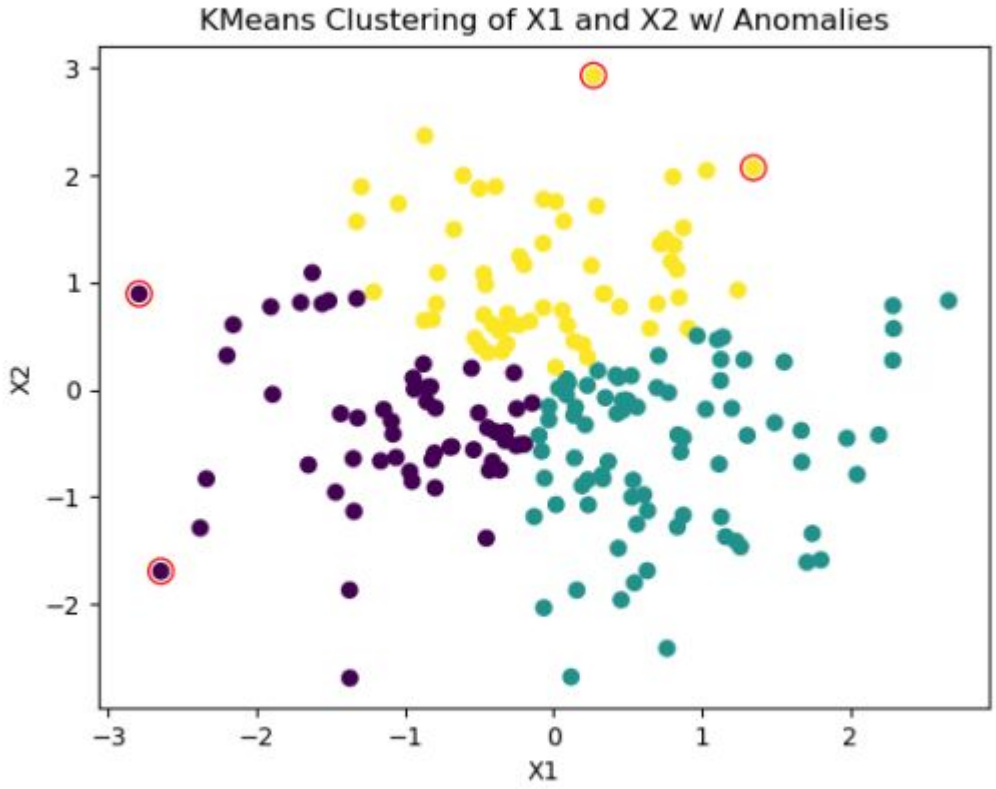


Figure 63: K Means Clusters w/ Anomalies

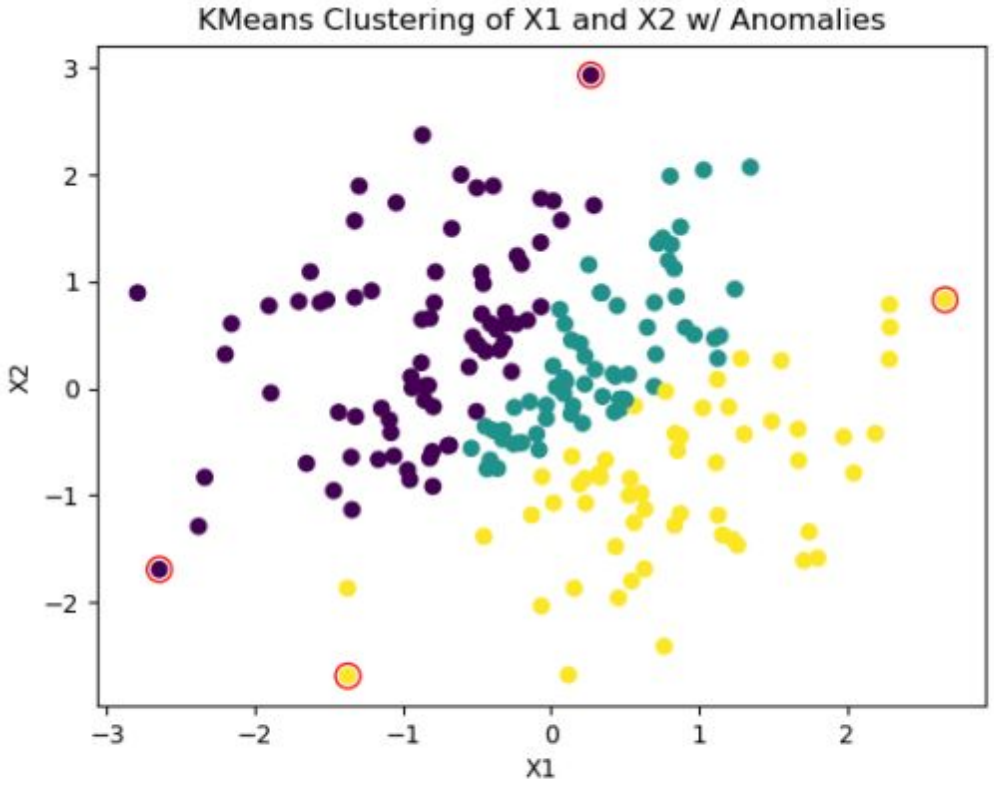


Figure 65: SOM Clusters w/ Anomalies

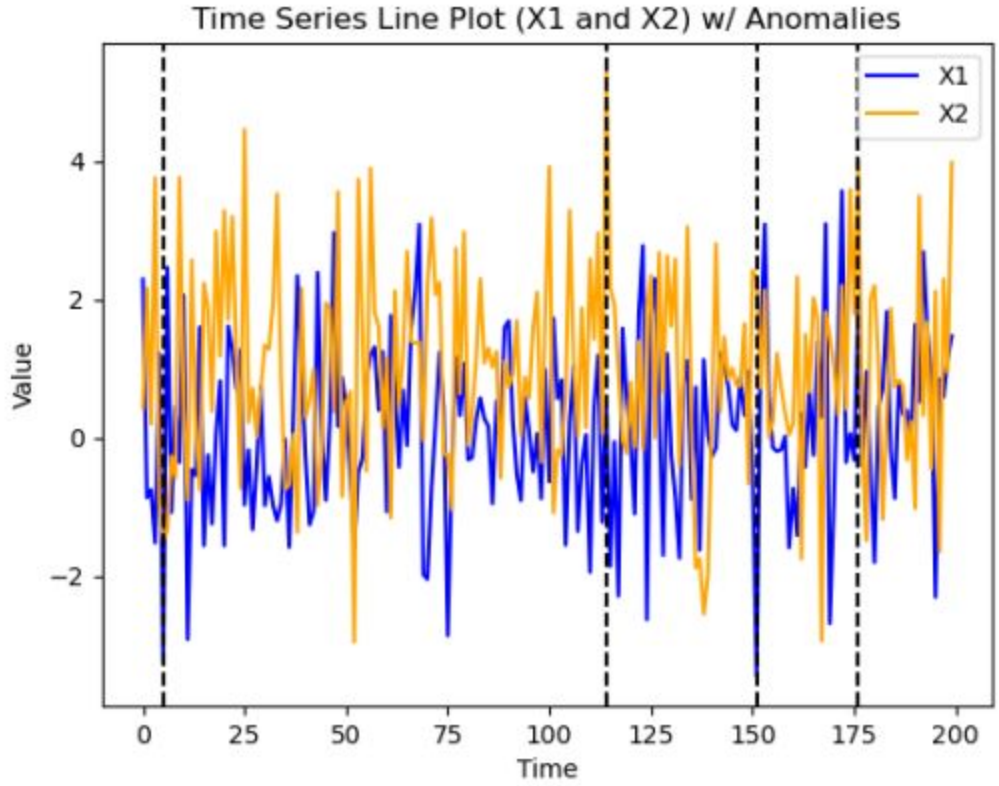


Figure 64: K Means Time Series w/ Anomalies

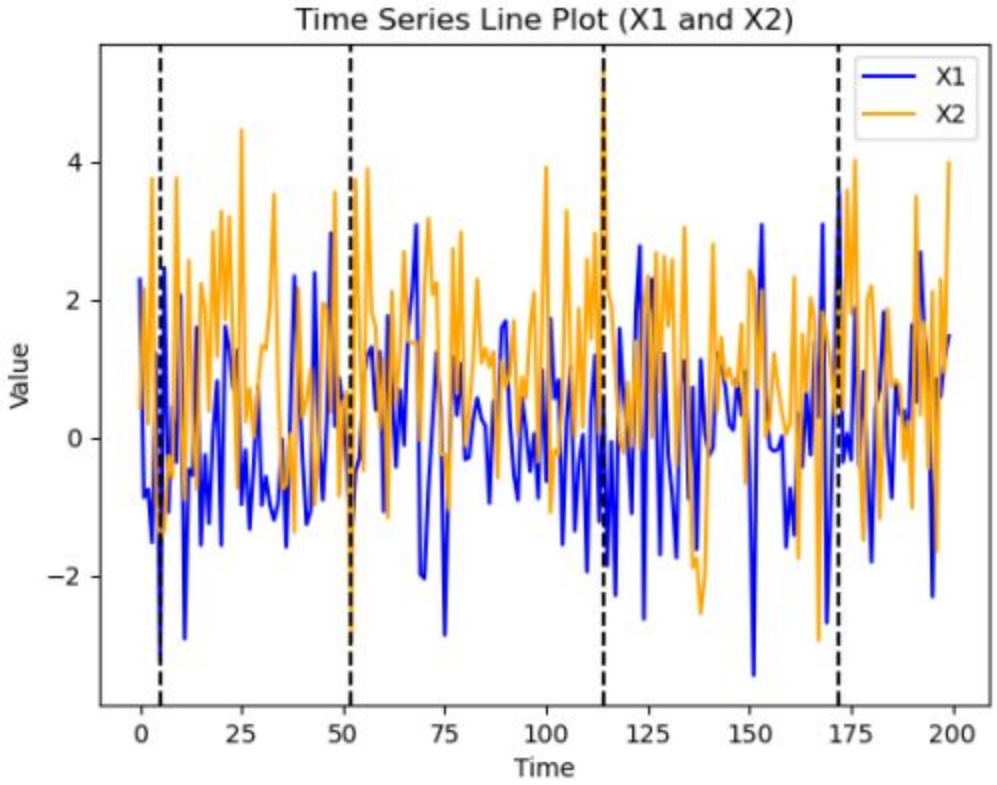


Figure 66: SOM Time Series w/ Anomalies

Conclusions

Tasks included:

Task 1 is time-series prediction with neural networks
    Datasets: Linear, sinusoidal, white noise, walker, arma and fibonacci series

Task 2 is the decomposition based anomaly detection
    Datasets: Land temperature series

Task 3 is the prediction-based anomaly detection
    Datasets: Land temperature and ECG series

Task 4 is the clustering-based anomaly detection
    Datasets: Stochastic series