CMP7229 Timeseries

Marios Kyriacou - St. ID: 21136828

Time series is a cluster of measurements taken at regular intervals [Hayes, 2021]. The dataset consists of five columns (Timestamp, Lifetouch-Heart-Rate, Lifetouch-Respiration-Rate, Oximeter, Oximeter-Pulse) and 227 samples. The project aims to forecast the Lifetouch-Heart-Rate (LHR) over the next 20 minutes and provide uncertain information. The data consists of the LHR in minutes for the date 17/05/15, starting at 15: 09 and ending at 18: 54. This project employs ARMA, ARIMA, and a combination of Exponential smoothing and ARIMA to find the best model for forecasting out of the sample.

Procedure of replacing the High Values	*Why*
Replace the high values.	This project does not eliminate the high values since it will influence the model, and the dataset includes only 227 samples.
Replace the high value with left and right nearest neighborhood of each high value.	If the high values are replaced with mean, median, or mode indicates a constant mean representing a zero variance.
No seasonality.	Heart Rate is unpredictable.
ADF and KPPS test for stationarity.	Use the AF and KPPS test to check when time series are stationary.

Table 1: Summary of the high values and the way have been replaced.

This report uses a technique known as Nearest Neighbor for Time-series imputation to replace the high values of the dataset [Moritz and Bartz, 2020].

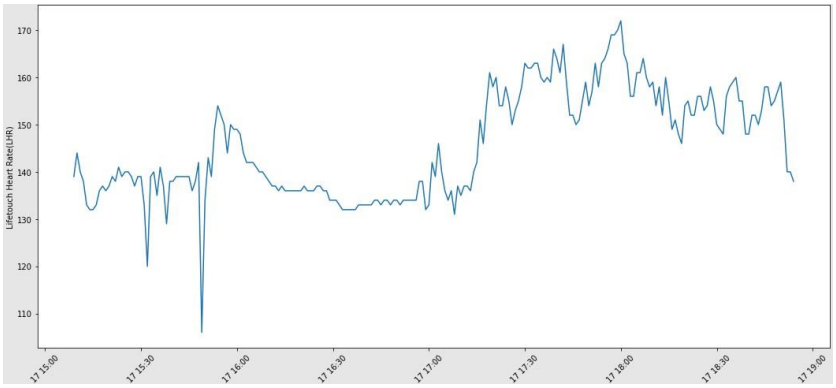


Figure 1: LHR Timeseries

The null hypothesis of the KPSS test implements stationary time series, which is dissimilar from the ADF test. If the p-value is more significant than the significand threshold (0.05 in our case), the series is non-stationary. However, the ADF test implements that the series is stationary [Prabhakaran, 2019]. The PACF plot illustrates the number of lags in an AR model and indicates

the connection between a lag and the series. The ACF refers to the number of MA terms that are necessary [Prabhakaran,2021].

	ARMA
POSSIBLE P-VALUES	1, 2, 10, 16, 18, 19
D-VALUE	0
POSSIBLE Q-VALUES	1, 3, 16, 17
OPTIMAL VALUES P,D,Q VALUES	18, 0, 16

Table 2: The ARMA model.

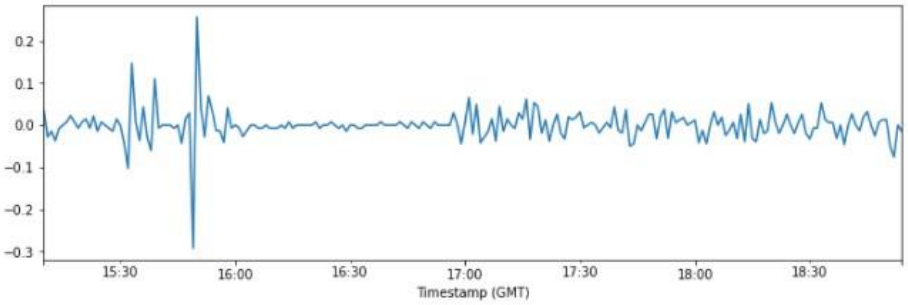


Figure 2: Transformed Time series.

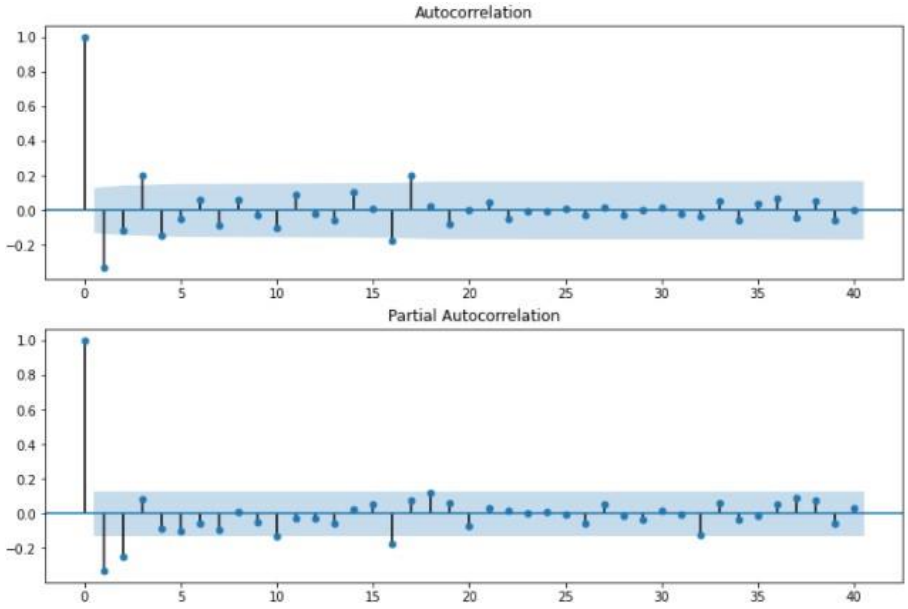


Figure 3: ACF and PACF plots.

We tested the ARMA model by dividing the time series into 70% training and 30% testing. The AIC and root mean squared error equals -550.665 and 0.028868, respectively. The AIC is a mathematical tool for determining how well a model fits the data from which it was created [Bevans, 2021]. Note the model is away one standard division (0.00141) from the mean (-3.2089e-05), which indicates a good model. Lastly, a negative AIC suggests less information loss than a positive AIC.

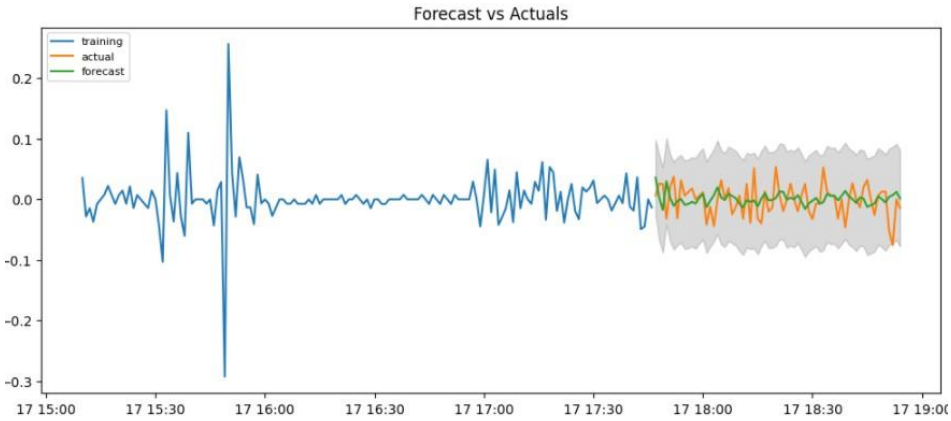


Figure 4: ARMA (18,0,16).

After the analysis of the process and the model selection, it is needed to calculate the forecasts with 95% confidence level for the next 20 minutes. The last step is to reverse the forecasted values to the actual.

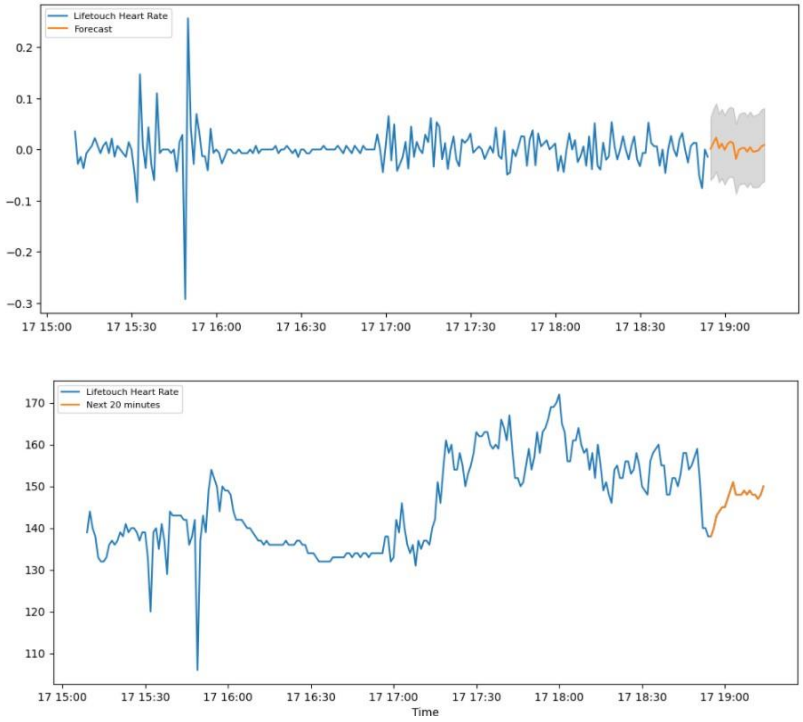
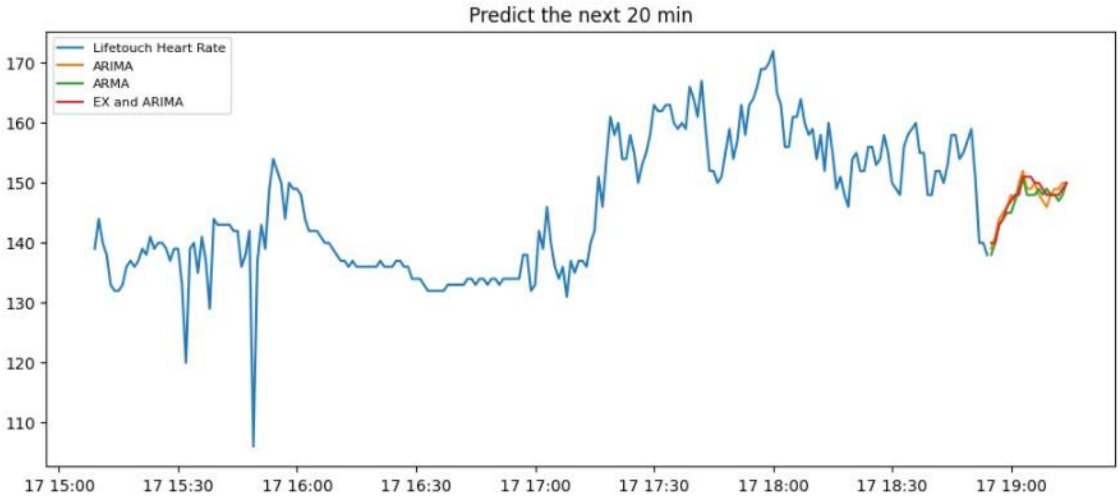


Figure 5: The next 20 minutes.

314 words

The ARIMA, ARMA and Exponential Smoothing and ARIMA forecasts:



	ARMA	**ARIMA**	**Exponential smoothing and ARIMA**
Best values	(18,0,6)	(19,1,10)	Exponential (α=0.8, β = 0.1) &ARIMA (19,1,10)
AIC	-550.665	1058.577	1312.670
RMSE	0.028868	7.45475575	High RMSE value
Mean	-3.2089e-05	145.73	High Mean value
Variance	0.00141	124.4823	High Variance value
Loss of Information (based on AIC)	No	Yes	Yes

Please find attached my code link for more models:

https://colab.research.google.com/drive/1XUNU0UsRmQxNyVG2LhrCMpTsA3sbBwNd?usp=sharing

References:

Hayes. A (2021) What is Time Series? Available at: https://www.investopedia.com/terms/t/timeseries.asp

[Accessed 15 Feb 2022]

Moritz.S and Bartz.T (2020) imputeTS: Time Series Missing Value Imputation in R. Available at: https://cran.r-project.org/web/packages/imputeTS/vignettes/imputeTS-Time-Series-Missing-Value-Imputation-in-R.pdf [Accessed 15 Feb 2022]

Prabhakaran. S (2019) KPSS Test for Stationarity. Available at: https://www.machinelearningplus.com/time-series/kpss-test-for-stationarity/#:~:text=A%20key%20difference%20from%20ADF,that%20the%20series%20is%20stationary.&text=That%20is%2C%20if%20p%2Dvalue,the%20tested%20series%20is%20stationary. [Accessed 12 Feb 2022]

Prabhakaran. S (2021) ARIMA Model – Complete Guide to Time Series Forecasting in Python. Available at: https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/ [Accessed 12 Feb 2022]

Bevans.R (2021). Akaike Information Criterion | When & How to Use It. Available at: https://www.scribbr.com/statistics/akaike-information-criterion/#:~:text=The%20Akaike%20information%20criterion%20(AIC,best%20fit%20for%20the%20data. [Accessed 12 Feb 2022]