WOA7015 Advanced Machine Learning

Alternative Assessment – January 2023

Group 3

|  |  |  |
| --- | --- | --- |
| No | Team Members | Matric Number |
| 1 | Chong Chia Hsing | S2159070/1 |
| 2 | Darkhan Baibulat | S2172558/1 |
| 3 | Xue Ying Kung | S2149135/1 |

Table of Contents

[2 Introduction – Chong CH 3](#_Toc125636888)

[2.1 Aim and Objectives 3](#_Toc125636889)

[2.2 Background – Chong CH 4](#_Toc125636890)

[2.3 Team Overview – Chong CH 4](#_Toc125636891)

[3 Datasets 6](#_Toc125636892)

[3.1 Characteristics – Darkhan 6](#_Toc125636893)

[3.2 Preprocessing – Chong CH 6](#_Toc125636894)

[4 Baseline Model or Benchmark – Xue Ying 8](#_Toc125636895)

[5 Machine Learning Model – Xue Ying 8](#_Toc125636896)

[5.1 Regression – Xue Ying 8](#_Toc125636897)

[5.2 Support Vector Machine for Regression – Xue Ying 9](#_Toc125636898)

[6 Performance Metrics – Chong CH 11](#_Toc125636899)

[7 Experiment Setup and Implementation – Chong CH 12](#_Toc125636900)

[8 Results and Discussion – Chong CH 21](#_Toc125636901)

[9 Conclusion – Chong CH 27](#_Toc125636902)

[10 References 28](#_Toc125636903)

# Introduction – Chong CH

This document is a report on the machine learning project where the specific plan and construct of the project are listed. The purpose of the report is to provide context on the data sets that will be used in the project and subsequently the data handling process will be investigated and justified for its usage.

Furthermore, the machine learning algorithms that will be used for the machine learning model training will be stated with their pros and cons to justify their application in solving the problem at hand. Then, the algorithms will be chosen with some of them being used as the baseline model and benchmark while the other will be the model being investigated for solving the problem.

Finally, the project will be concluded with an analysis of the performance of the trained model with relevant evaluation metrics. The metrics will be applied to both the baseline model and the model trained with the chosen algorithm for comparison and further analysis.

## Aim and Objectives

The aim of the study is to obtain a prediction model, , for heart rate, , where the dependent variable is the respiratory signal, .

The stated function will be a regression model given that this is a prediction of a time series data from another time series data.

To achieve the stated aim, some of the objectives need to be fulfilled. These objectives are shown below.

1. To read the electrocardiogram (ECG) and respiratory signal (RESP) from the files.
2. To perform data preprocessing and cleaning.
3. To calculate the heart rate from ECG signals.
4. To train the regression model with heart rate and RESP data.
5. To assess the performance metrics obtained for the models.

## Background – Chong CH

The project uses data sets from a research project conducted by Gee et al. (2017) where it is used to investigate and predict the onset of bradycardia for pre-term infants. In the paper, the author stated the use of instantaneous mean and variance as features for the detection of the onset of bradycardia. (Gee & Barbieri, 2017)

The method used is a stochastics method where the data within a 3-minute window is taken on 2 instances where the first instance would be marked as the control window (CW) while the second instance would be time window before the occurrence of bradycardia which will be labeled as pre-bradycardia window (PBW). Then, the prediction model is trained on 2 features: (1) mean and (2) variance and generate a cumulative density curve over distance to centroid instantaneous variance against instantaneous mean plot. Then, on each new instance of evaluation window (EW), the generated cumulative density curve is compared to the curve for PBW and CW using the Euclidean distance to identify the possible onset of bradycardia. (Gee A. H., Barbieri, Paydarfar, & Indic, 2017)

Other than the main theme of bradycardia, the author also indicated that the respiratory data could improve the prediction of the ECG heartbeat although the respiratory does not affect the prediction of the bradycardia. (Gee et al, 2017, “Analysis of Prediction Algorithm” section, para. 1). Furthermore, it is stated that the future direction of the study is to investigate the inclusion of other physiological signals to improve the prediction with emphasis on respiratory data where the author also cited 2 other journal articles by Moody et al. (1985) and Helfenbein et al. (2014) for the study on derivation of respiratory signals from ECGs.

## Team Overview – Chong CH

Team members information:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Name | Matric # | Email | Phone |
| 1 | Chia Hsing Chong | S2159070/1 | [chiahsing.chong@outlook.com](mailto:chiahsing.chong@outlook.com) | 60135836818 |
| 2 | Xue Ying Kung | S2149135 | [kungxueying@gmail.com](mailto:kungxueying@gmail.com) | 60182093219 |
| 3 | Darkhan Baibulat | s2172558 | [xxlabbesxx@gmail.com](mailto:xxlabbesxx@gmail.com) | 77089402662 |

The GitHub repository links:

|  |  |  |  |
| --- | --- | --- | --- |
| No | Name | GitHub | Repo |
| 1 | Chia Hsing Chong | cchsing | <https://github.com/cchsing/AML_Project_G3> |
| 2 | Xue Ying Kung | kungxueying | <https://github.com/kungxueying/AML_Project_G3> |
| 3 | Darkhan Baibulat | darkhan-b | <https://github.com/darkhan-b/AML_Project_G3> |

# Datasets

In this section, the dataset that will be used in the study will be discussed with its characteristics and background information.

## Characteristics – Darkhan

There are 2 sets of data namely electrocardiogram (ECG) and the respiratory signal (RESP) obtained from the website, “https://physionet.org/content/picsdb/1.0.0/”. The data are obtained from ten infants aged 29 to 34 weeks and weighing from 843 to 2100 grams with an average 1468 grams.

Single-channel ECG and respiratory recordings are provided in the standard WFDB format. The start time of the recording is synchronized within each baby. ECGs are recorded at a frequency of 500 Hz, except for infants 1 and 5 years old, which are recorded as a composite ECG signal with a frequency of 250 Hz.

On the other hand, the respiratory (RESP) data is recorded using the inductance plethysmography method. The sampling frequency for the respiratory data is 50 Hz except for infant 1 which has a sampling frequency of 500Hz.

## Preprocessing – Chong CH

There are 4 steps involved in data preprocessing for the ECG and RESP data. The steps are data smoothing with interquartile range, data normalization, heartrate and respiration rate calculations, and data fixup and resampling.

The first step of data smooth is to reassign the outliers to a maximum and minimum range which is obtained from the 90th and 10th percentile values. Subsequently, the data is normalized to a -1 to 1 values. Then, the RR interval calculation is used to calculate the instantaneous value of the heart rate with a unit of beat per minutes (BPM) and the respiration rate of unit breath per minutes (BPM). Lastly, the heartrate and respiration rate calculated would contain null or NaN value which need to be fixed or replaced with a zero value. The heart rate and respiration rate would also need to be resampled to have the same cardinality due to different sampling frequency. The steps are listed below.

1. Data smoothing with interquartile range.
2. Data normalization.
3. Heart rate and respiration rate calculations.
4. Data fixup and resample.

# Baseline Model or Benchmark – Xue Ying

A baseline model is a simple model which acts as a reference in a machine learning project. It serves as a benchmark for a trained model and helps us to improve the understanding of our data (Nair, 2022). There are a few requirements for a good baseline model:

1. A good baseline model should be simple, as simple models are easy to implement and less likely to overfit.
2. A good baseline model should be interpretable, so that it helps us to understand our data better (Sennikova, 2020).

In our project, basic linear regression and polynomial regression will be used as the baseline models. Both are simple and help us in inspecting the correlation between the 2 variables: ECG signal and respiratory signal.

|  |  |
| --- | --- |
| **Sklearn modules** | Description |
| [**linear\_model.LinearRegression**](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html#sklearn.linear_model.LinearRegression)(\*[, ...]) | Ordinary least squares Linear Regression. |
| [**preprocessing.PolynomialFeatures**](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html#sklearn.preprocessing.PolynomialFeatures)([degree, ...]) | Generate polynomial and interaction features. |

# Machine Learning Model – Xue Ying

## Regression – Xue Ying

Regression analysis is a statistical method that shows the relationship between two or more variables. Usually expressed in a graph, the method tests the relationship between a dependent variable against independent variables. Typically, the independent variable(s) changes with the dependent variable(s) and the regression analysis attempts to answer which factors matter most to that change (What is Regression Analysis?, 2020).

In this project, the heartbeat rate of infants will be predicted by using their respiratory rate. Hence, regression analysis is used.

There are many types of regression algorithms, which includes linear regression, polynomial regression, neural network regression, support vector machines (SVM), random forest and more.

## Support Vector Machine for Regression – Xue Ying

Support Vector Machine (SVM) is a supervised learning model with associated learning algorithms that analyze data for classification and regression. (Support vector machine - Wikipedia, 2022). The idea of SVM is to find a hyperplane that best divides a dataset into two classes. The distance between the hyperplane and the nearest data point from either set is known as the margin. The goal is to choose a hyperplane with the greatest possible margin, so that new data coming in has a higher possibility for being classified correctly (Bambrick, 2016).

Although SVM is more popular and commonly used for classification problems, the method of Support Vector Classification (SVC) can be extended to solve regression problems. This method is called Support Vector Regression (SVR) (1.4. Support Vector Machines — scikit-learn 1.2.0 documentation, n.d.). The basic idea behind SVR is to find the best fit line. In SVR, the best fit line is the hyperplane that has the maximum number of points (Raj, 2020).

The strength of using SVM in regression are as follows:

1. SVM has L2 Regularization feature. Hence, it has good generalization capabilities which prevent it from overfitting (Kumar, 2019).
   * Preventing overfitting issues in the healthcare domain research is important because an inaccurate prediction may lead to serious consequences.
2. Linear SVM handles outliers better, as it derives maximum margin solution (Varghese, Comparative Study on Classic Machine learning Algorithms | by Danny Varghese | Towards Data Science, 2018).

* Even if there is any infant in the training data who is having health problems and has a significantly different relationship between his/her heartbeat rate and respiratory rate, his/her training data will not have too much negative impact on the model.

1. SVM can perform better than a few other algorithms when there is limited training data. Other algorithms, for example Neural Network needs large training data for sufficient accuracy (Varghese, Comparative Study on Classic Machine learning Algorithms , Part-2 | by Danny Varghese | Medium, 2018).
   * This is an important point because in our project, there is only data for 10 infants.

# Performance Metrics – Chong CH

There are 2 types of performance metrics being used in the study. The first type is correlation metrics where it is used to inspect the possible correlation between the 2 sets of data. The second type is the error metrics where it is used to assess the quality of the model trained.

The correlation metrics are the covariance matrix, Pearson’s coefficient, and Spearman’s coefficient. (Brownlee, 2020) The first metric is a covariance matrix where the diagonal of the matrix shows the covariance within the variable individual while the other 2 values (assuming only 2 variables) will indicate the correlation between the 2 variables. A sample of covariance matrix is given below.

The value of 12 indicates the 2 variables have positive correlation and the higher the magnitude the better the correlation. Same goes for the negative value which indicates a negative correlation. A value of zero would indicate no correlation for the variables.

The other 2 correlation metrics are Pearson’s and Spearman’s coefficients. These metrics indicate positive and negative correlation with the use of 1 and -1 respectively. A value of zero would mean no correlation between the variables. The difference between these 2 coefficients is that Spearman’s coefficients can detect non-linear correlation for the variables.

As for the metrics used to assess the model, there are 3 namely, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The 3 metrics are used to calculate the error or difference between the predicted and sampling value. The libraries that will be used for the metrics are shown below.

|  |  |
| --- | --- |
| **Sklearn modules** | **Description** |
| [**metrics.mean\_absolute\_error**](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html)**(y\_true, y\_pred, \*)** | Mean absolute error regression loss. |
| [**metrics.mean\_squared\_error**](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html)**(y\_true, y\_pred, \*)** | Mean squared error regression loss. |

# Experiment Setup and Implementation – Chong CH

The setup for the model training experiment is shown as a data flow diagram below. The experiment is separated into 3 main sections.

Diagram

Description automatically generated

The first section of the experiment will be the reading of the dataset. The datasets are in the standard WFDB format. Thus, the wfdb.rdsamp() function is used to read the data into the local environment. The picture below shows a sample of ECG data in a tuple format with the first component containing the signal data in array while the second component of the tuple is the metadata for the signal.

Text

Description automatically generated

Then, the data is separately processed by using a for loop to iterate through each portion of the data. As shown in the figure below, the starting hour (startHrs) of the data is 10, ending hour (endHrs) is 12, and the step or bin size is 0.25 which is a quarter of an hour or 15 minutes. Thus, the data from 10th until 12th hour are used for the training with a data bin of 15 minutes.

Text

Description automatically generated

Due to the computation, the full use of the dataset (more than 40 hours) is impractical. A 2 hours dataset should suffice given the sampling rate of 50 Hz and 500 Hz for the RESP and ECG data respectively.

In the second section, the data preprocessing consists of 4 steps as mentioned previously. There are 2 functions written for this step as shown in the figure below. One of the functions is to calculate the interquartile range with the use of numpy.percentile() while the other is a custom function written to handle the outliers.

Text

Description automatically generated

The data normalization step will involve the use of numpy.preprocessing.MinMaxScaler() function from the Sci-Kit learn library. The data are normalized to a minimum of -1 and maximum of 1 value as shown in the figure below.

Text

Description automatically generated

The next step would be the calculation of heart rate and the respiration rate. Both of the calculations are similar where the time interval between the local peaks are taken to calculate the beat per minute or breath per minute as it is assumed that the peaks represent the heart beat for the ECG while an inhale or exhale for the respiratory signal. The figure shown below are the functions used for the calculation.

A screenshot of a computer

Description automatically generated with medium confidence

Within the functions, the processing.qrs.gqrs\_detect() and processing.peaks.find\_local\_peaks() modules are taken from the wfdb library for the calculation of peaks in both the ECG and respiratory data respectively. Given that the features (heart rate and respiration rate) calculation will be the same, the processing.hr.compute\_hr() is used for both ECG and RESP signals.

The last step in the preprocessing is the data fixup and resampling. The heart rate and respiration rate calculated would contain some null or “NaN” value which is not allow for subsequent models training. Thus, a simple fixup would be required by filling in zero value in place of null. Furthermore, the data sizes for heart rate and respiration rate are different due to the different sampling frequencies. Thus, a scipy.signal.resample() function is used to synchronize the data size between the heart rate and respiration rate. The code for this processing is shown below.

Text

Description automatically generated

The last section of this data pipeline is the model training section. There are 4 steps in this pipeline. The first step is the splitting of the data into training and testing sets. A 75% to 25% is used for training and testing set respectively as shown below.

Text

Description automatically generated

The training data is then used to train the linear regression model, polynomial regression model, and the support vector regression model. The code for linear regression model training is as shown below.

Text

Description automatically generated

The code for training polynomial regression model is shown below.

Text

Description automatically generated

The code for training SVR model is shown below.

Text

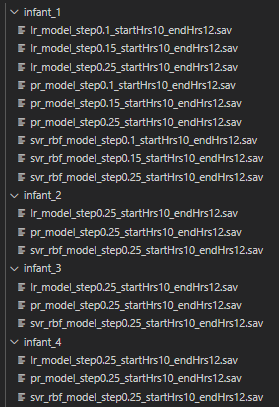
Description automatically generated

It should be noted that the model training is done repeatedly on the data bin (15 minutes range) within the stated 10th to 12th hours interval due to limitation of the computation power. The model is fitted repeatedly on each bin to preserve the information trained on the previous bins. This is visualized as diagram shown below.

Diagram

Description automatically generated with medium confidence

The whole training procedure will be performed on infant 1 until infant 8 while infant 9 and infant 10 datasets will be used as testing set. The data leakage can be prevented by excluding the 1 or 2 infants dataset to inspect the out of sample predictions for the models trained to simulate real-world application of the models. The figures below show the models trained for infant 1 until infant 8.

 A screenshot of a computer

Description automatically generated with medium confidence

These trained models will in turn be tested using the dataset from infant 9 and 10. Infant 9 or infant 10 ECG and RESP data will be fed to the data pipeline first to convert into heart rate and respiration rate before the prediction will be performed. Once the heart rate and respiration rate data are obtained, the data are fed to the 8 models without splitting as shown below.

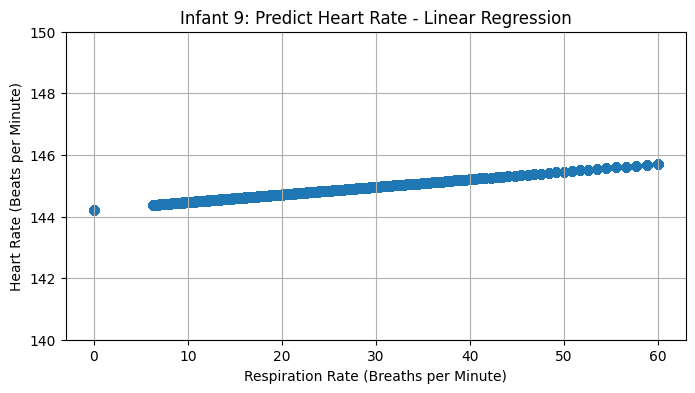
**Graphical user interface, text, application, chat or text message

Description automatically generated**

Then, there will be 8 sets of predicted heart rate generated from the models. These predicted will be averaged and a single mean value of predicted heart rate will be obtained. The mean predicted heart rate will then be compared with the ground truth values to obtain the 3 performance metrics i.e., MAE, MSE, and RMSE.

# Results and Discussion – Chong CH

The results for testing on infant 9 and infant 10 are shown below with the graph plotted for the predicted heart rate against the respiration rate.



Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Infant 9: Linear Regression | | | |  | Infant 10: Linear Regression | | | |
| Model | MAE | MSE | RMSE |  | Model | MAE | MSE | RMSE |
| 1 | 10.71 | 512.33 | 22.63 |  | 1 | 13.91 | 501.52 | 22.39 |
| 2 | 21.35 | 897.26 | 29.95 |  | 2 | 28.69 | 1152.01 | 33.94 |
| 3 | 20.59 | 862.98 | 29.38 |  | 3 | 27.68 | 1091.23 | 33.03 |
| 4 | 20.71 | 848.51 | 29.13 |  | 4 | 17.69 | 668.22 | 25.85 |
| 5 | 14.32 | 604.16 | 24.58 |  | 5 | 14.36 | 519.22 | 22.79 |
| 6 | 17.42 | 743.76 | 27.27 |  | 6 | 24.80 | 943.45 | 30.72 |
| 7 | 14.50 | 610.13 | 24.70 |  | 7 | 14.51 | 526.26 | 22.94 |
| 8 | 17.57 | 747.98 | 27.35 |  | 8 | 24.80 | 942.13 | 30.69 |
| Average | 9.85 | 513.59 | 22.66 |  | Average | 15.97 | 576.64 | 24.01 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Infant 9: Polynomial Regression | | | |  | Infant 10: Polynomial Regression | | | |
| Model | MAE | MSE | RMSE |  | Model | MAE | MSE | RMSE |
| 1 | 10.75 | 512.93 | 22.65 |  | 1 | 13.85 | 499.59 | 22.35 |
| 2 | 21.44 | 898.72 | 29.98 |  | 2 | 28.37 | 1128.55 | 33.59 |
| 3 | 20.67 | 865.38 | 29.42 |  | 3 | 27.44 | 1072.41 | 32.75 |
| 4 | 20.69 | 847.56 | 29.11 |  | 4 | 17.64 | 666.82 | 25.82 |
| 5 | 14.15 | 600.03 | 24.50 |  | 5 | 14.34 | 517.75 | 22.75 |
| 6 | 17.83 | 756.45 | 27.50 |  | 6 | 24.95 | 946.56 | 30.77 |
| 7 | 14.58 | 612.49 | 24.75 |  | 7 | 14.70 | 532.37 | 23.07 |
| 8 | 17.85 | 756.79 | 27.51 |  | 8 | 24.68 | 929.23 | 30.48 |
| Average | 9.86 | 514.04 | 22.67 |  | Average | 15.91 | 574.21 | 23.96 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Infant 9: Support Vector Regression | | | |  | Infant 10: Support Vector Regression | | | |
| Model | MAE | MSE | RMSE |  | Model | MAE | MSE | RMSE |
| 1 | 10.75 | 512.93 | 22.65 |  | 1 | 13.85 | 499.59 | 22.35 |
| 2 | 21.44 | 898.72 | 29.98 |  | 2 | 28.37 | 1128.55 | 33.59 |
| 3 | 20.67 | 865.38 | 29.42 |  | 3 | 27.44 | 1072.41 | 32.75 |
| 4 | 20.69 | 847.56 | 29.11 |  | 4 | 17.64 | 666.82 | 25.82 |
| 5 | 14.15 | 600.03 | 24.50 |  | 5 | 14.34 | 517.75 | 22.75 |
| 6 | 17.83 | 756.45 | 27.50 |  | 6 | 24.95 | 946.56 | 30.77 |
| 7 | 14.58 | 612.49 | 24.75 |  | 7 | 14.70 | 532.37 | 23.07 |
| 8 | 17.85 | 756.79 | 27.51 |  | 8 | 24.68 | 929.23 | 30.48 |
| Average | 9.86 | 514.04 | 22.67 |  | Average | 15.91 | 574.21 | 23.96 |

There are 24 models trained which can be separated into 3 models: (1) linear regression model, (2) polynomial regression model, and (3) support vector regression (SVR) model. Thus, there are a total of 8 models for each regression type trained on 8 different datasets corresponding to the 8 infants ECG and RESP data respectively.

From the graph plotted, the predicted heart rate ranges from 144 bpm to 146 bpm which corresponds to 10 to 60 bpm (breath per minute) for the respiratory data. Furthermore, the curve for polynomial regression is similar to SVR curve. This similarity is also reflected in the performance metrics given in the tables where value predicted from polynomial regression has the same MAE, MSE, and RMSE with the value predicted from SVR model.

The failure to gain an accurate model might be attributed to the noise in the data as the heart rate and respiration rate calculated also have some outliers that are not filtered. Furthermore, the techniques of IQR smoothing used in the data preprocessing section might not be suitable and a frequency-based filtering approach should be taken which are the band-pass filter, low-pass filter, and high-pass filter. This technique filters out the noise frequencies by processing the frequency domain of the signal through the Fast Fourier Transform methods.

In addition, hyperparameters tuning should be used for the selection of the optimum parameters of the SVR model training which might help improve the models. The quality for the model can also be improved with the use of cross validation method.

There is also another challenge facing the current study which is computational power required for the model training. It is encountered that if the datapoints goes more than 20 minutes. The script would take more than 2 to 3 hours to train an instance of the models.

# Conclusion – Chong CH

In conclusion, the data pipeline consists of reading the data, data preprocessing, and model training is built and it sets the foundation for future exploration and research. There are 24 models trained using the first 8 infants’ datasets with each dataset trained on 3 different models i.e., linear regression, polynomial regression, and support vector regression. The model is then used to obtain the predicted heart rate for infant 9 and infant 10. However, the root mean squared error (RMSE) shows a range of 22 to 24 for the predicted heart rate data on infant 9 and infant 10.

As mentioned previously, there are gaps to be explored in the processing of the noise and outlier within the ECG and RESP signals along with the heart rate and respiration rate calculation. Furthermore, improvement can be made on the model with parameters tuning and cross-validation methods which is yet to be explored.

# References

*1.4. Support Vector Machines — scikit-learn 1.2.0 documentation*. (n.d.). Retrieved from scikit-learn: https://scikit-learn.org/stable/modules/svm.html#:~:text=Support%20vector%20machines%20(SVMs)%20are,than%20the%20number%20of%20samples.

Bambrick, N. (July, 2016). *Support Vector Machines: A Simple Explanation - KDnuggets*. Retrieved from KDnuggets: AI, Analytics, Data Science, Machine Learning Feed: https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html

Gee, A. H., Barbieri, R., Paydarfar, D., & Indic, P. (2017, September). Predicting Bradycardia in Preterm Infants Using Point Process Analysis of Heart Rate. *IEEE Transactions on Biomedical Engineering, 64*(9), 2300-2308. doi:10.1109/TBME.2016.2632746

Gee, A., & Barbieri, R. (2017, February 9). *Preterm Infant Cardio-Respiratory Signals Database v1.0.0*. doi:10.13026/C2QQ2M

Helfenbein, E., Firoozabadi, R., Chien, S., Carlson, E., & Babaeizadeh, S. (2014). Development of three methods for extracting respiration from the surface ECG: A review. *Journal of Electrocardiology, 47*(6), 819-825.

Kumar, N. (1 March, 2019). *The Professionals Point: Advantages and Disadvantages of SVM (Support Vector Machine) in Machine Learning*. Retrieved from The Professionals Point - Atom: http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of-svm.html

Moody, G. B., Mark, R. G., Zoccola, A., & Mantero, S. (1985). Derivation of Respiratory Signals from Multi-lead ECGs. *Computers in Cardiology, 12*, 113-116. Retrieved from https://archive.physionet.org/physiotools/edr/cic85/

Nair, A. (5 April, 2022). *Baseline Models: Your Guide For Model Building | by Aashish Nair | Towards Data Science*. Retrieved from Towards Data Science: https://towardsdatascience.com/baseline-models-your-guide-for-model-building-1ec3aa244b8d#:~:text=A%20baseline%20model%20is%20essentially,a%20necessity%20for%20many%20reasons.

Raj, A. (3 October, 2020). *Unlocking the True Power of Support Vector Regression | by Ashwin Raj | Towards Data Science*. Retrieved from Towards Data Science: https://towardsdatascience.com/unlocking-the-true-power-of-support-vector-regression-847fd123a4a0

Sennikova, T. (2020, June 6). *How to Build a Baseline Model | Towards Data Science*. Retrieved from Towards Data Science: https://towardsdatascience.com/how-to-build-a-baseline-model-be6ce42389fc

*Support vector machine - Wikipedia*. (24 December, 2022). Retrieved from From Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Support\_vector\_machine

Varghese, D. (11 December, 2018). *Comparative Study on Classic Machine learning Algorithms , Part-2 | by Danny Varghese | Medium*. Retrieved from Towards Data Science: https://medium.com/@dannymvarghese/comparative-study-on-classic-machine-learning-algorithms-part-2-5ab58b683ec0

Varghese, D. (7 December, 2018). *Comparative Study on Classic Machine learning Algorithms | by Danny Varghese | Towards Data Science*. Retrieved from Towards Data Science: https://towardsdatascience.com/comparative-study-on-classic-machine-learning-algorithms-24f9ff6ab222

*What is Regression Analysis?* (21 September, 2020). Retrieved from TIBCO Software: https://www.tibco.com/reference-center/what-is-regression-analysis