Development of a Depth of Anesthesia (DoA) Index Using Machine Learning Methods

Final Report

JANITH WEERAKKODY

U1161698

Abstract

Depth of anaesthesia is one of the important parameters that vary depending on the type of surgery to ascertain patient safety and optimization of drugs administration. This project intends to formulate a new Depth of Anaesthesia index (DOA index), utilizing machine learning techniques to uplift the accuracy and predictability compared to the currently used indices. This project is specifically focused on Random Forest and Linear Regression methods for the feature selection and model building purposes. Here the feature selection is done using Random Forest whereas for model building, both Random Forest and Linear Regression methods have been employed. A hybrid model is built as the final predictive model for DOA index. The formulated index is evaluated by metrics like Mean Squared Error (MSE), R-squared, and Pearson correlation. The results achieved from this study suggest that the formulated model for the new index show a strong predictability with significantly correlated predicted and actual values compared to the models formulated in previous two studies (report 1 and report 2). This is because the model's got the characteristics of both Random Forest and Linear Regression. Furthermore, the results emphasizes that this model shows a significant improvement compared to BIS index by means of its accuracy, consistency and predictability, offering a valuable predictive index for the detection of Depth of Anaesthesia levels of patients.

Table of Contents

1 Introduction

Depth of anaesthesia is one of the important parameters that vary depending on the type of surgery to ascertain patient safety and optimization of drugs administration. Traditional techniques used in determining this parameter involve physical signs, supplemented by basic biometric indications, which may not actually reflect the neurological status of the patient. Recently, especially, engineering technology has used electroencephalogram signals to make more accurate assessments of DoA. The project investigates whether better use of the potentials of the EEG data could be made through the application of techniques from machine learning in order to develop a more robust and more accurate DoA index. The new index will reduce the occurrence of anaesthesia-related complications such as underdosage with resultant awareness and overdosage leading to unnecessary drug administration.

2 Literature Review

Extensive research was conducted to observe how the past researchers have utilized machine learning techniques for accurate DOA predictions specially focusing on Hybrid modelling. The following is a summary of the literature review conducted in this regard.

Title	Author	Description	Methodology	Results / Findings	Limitations
Unsupervised Clustering of EEG Signals for Anesthetic Depth Estimation	Doi et al. (2016)	This study focuses on exploring the applications of unsupervised learning in determining the DoA level by classifying the EEG signals using K means clustering.	K means clustering has been employed to classify EEG data into clusters focusing on their similarities. No feature selection method is used in this study.	The results show that EEG data can be classified by Unsupervised learning methods like K means with an improved prediction on DOA level of a patient.	This model is limited to unsupervised learning without labelled data. This will affect the accuracy of the predictions.
Deep Learning Applications in EEG Data Processing	Lee and Kim (2020)	This research is based on deep learning applications used in monitoring anaesthesia	Applies deep learning models (Convolutional Neural Networks, Recurrent Neural Networks) to classify EEG	Deep learning models have shown higher efficiency and accuracy in predicting DOA levels of patients compared to old methods	The most critical drawback which can be observed in this study is that the deep learning needs a considerable amount of labelled data and computational

			data for DoA prediction.	such as BIS index.	resources.
Machine Learning and EEG: Applications in Psychiatry	Zhao et al. (2020)	In this study, it analyses the psychiatric EEG data using machine learning models and discusses the relevance of it to DOA.	Uses machine learning models like SVM and Random Forest for EEG-based DoA classification. Focuses on feature extraction and model training.	It was discovered the extent of potential of machine learning to accurately predict different levels of Anaesthesia depth.	The models presented were not tested on large, diverse datasets, limiting generalizability.
Ensemble Learning for EEG Data in Anaesthesia	Sharma et al. (2021)	This study focuses on the use of combined machine learning methods to analyze EEG data for DOA prediction.	Uses ensemble learning methods, combining Random Forest and Support Vector Machines to predict DoA levels from EEG data.	Ensemble model of Random Forest and SVM has given more accurate results than single models.	Ensemble methods like this require more computations which makes it more expensive. Also, these models are less interpretable.
Hybrid Machine Learning Models for DoA Index Prediction	Hassan et al. (2022)	This study is based on predictive modelling using combined machine learning models which enhances the accuracy and adaptability of DOA predictions.	Combines Random Forest for feature selection and Neural Networks for modelling the DoA index. Focus on hybrid model for enhanced accuracy.	Hybrid model of Random Forest and Neural networks has able to provide more control over complex data which enables improved prediction accuracy.	These models can have overfitting issues specifically with smaller datasets and when feature selection is not properly done.

3 Data Description

The dataset consists of 12 training sets and 5 testing sets. Each of the datasets consists of BIS index values which is the target variable and 7 anaesthesia metrics. The entire dataset used in this study does not have any missing values and Duplicates. All the variables are quantitative variables.

4 Methodology

For the purpose of formulating a new index for DOA predictions, we are focusing on two machine learning methods and how those models complement each other to make a hybrid model with improved accuracy and interpretability to determine the DOA index. This hybrid model is then evaluated using a few evaluation metrics to determine the efficiency and the accuracy of the DOA prediction.

In the scope of this study the two machine learning methods used were Random Forest and Linear Regression. To evaluate the models, statistical metrics such as Pearson's correlation, Mean Squares Error (MSE) and R-squared metrics have been used.

4.1 Data Preprocessing

The data were given in a .xlsx file where there were 12 training data sets and 5 testing data sets. There were no missing values or duplicate records found in the data. The datasets were recorded in separate sheets in a workbook. So, before starting the analysis, it was needed to merge all data into 2 data files separately as training and testing. Hence, two empty lists were formed as combined_1 and combined_2, in order to merge the data. After storing the training and testing data separately into the two lists, those two lists were then concatenated to form 2 data frames: combined_train and combined_test. These data frames were used in the further analysis done in this project. Also, these two data frames were exported to two excel files for future requirements.

4.2 Exploratory Data Analysis

4.2.1 Summary Statistics

The summary statistics of the seven features and the BIS index were observed to have an idea about the dataset and what should be done in the analysis. Since all the variables in the dataset were quantitative variables, the five number summary of the variables along with the metrics such as mean and standard deviation were observed. The models are to be trained upon a large dataset which has 33,645 observations. Hence feature selection methods like K Nearest Neighbour are not suitable for this task. Methods like Random Forest, linear regression and neural networks are preferred in this regard.

	BIS	x1	x2	x3	x4
count	33645.000000	33645.000000	33645.000000	33645.000000	33645.000000
mean	41.694582	0.660667	1.789144	1.785165	0.870912
std	16.336583	0.047692	0.008522	0.003273	0.822348
min	5.300000	0.480816	1.542146	1.779489	0.139335
25%	30.800000	0.625311	1.789973	1.782343	0.351358
50%	40.100000	0.661472	1.790933	1.785161	0.510550
75%	45.700000	0.698184	1.791328	1.787998	1.186007
max	97.700000	0.835586	1.791751	1.790815	6.805957

	x5	x6	x7
count	33645.000000	33645.000000	33645.000000
mean	1.013450	0.693157	0.388178
std	0.023614	0.292224	0.008800
min	0.880547	0.044623	0.374989
25%	1.003973	0.481118	0.383036
50%	1.015341	0.662049	0.385354
75%	1.024747	0.923880	0.389416
max	1.110921	1.954334	0.453838

Figure 1 – Summary Statistics

4.2.2 Correlations Between Variables

A correlation matrix has been created to visualize the relationships between the variables of the dataset.

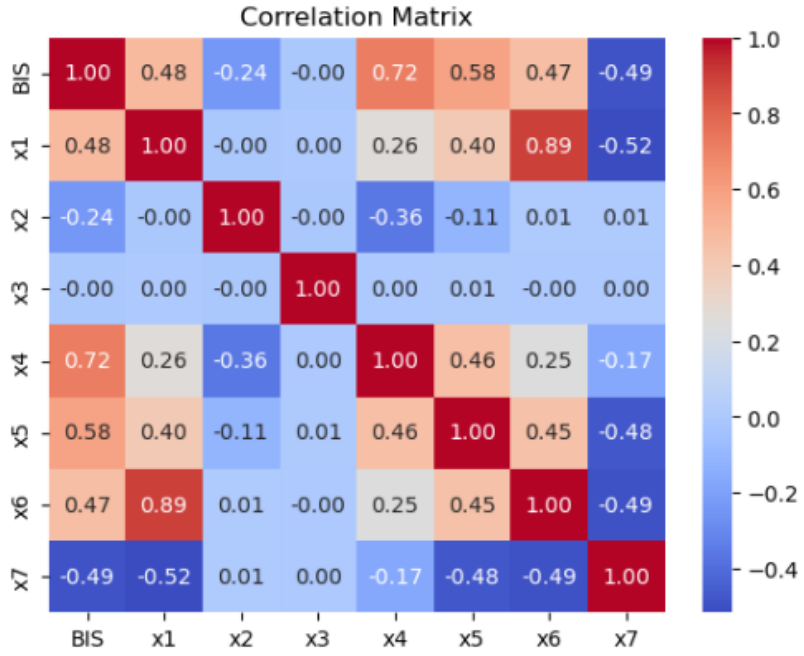

Figure 2 – Correlation matrix

As per above correlation matrix, Only the X1 and X6 have shown a considerable association. Moderately strong associations can be observed between X4 & X5 and X6 & X5. Also, variables X4, X5 and X7 share a considerably strong correlation with the target variable. So, a hint of multicollinearity can be observed in the data and hence feature selection methods

such as Random Forest and Linear Regression can be useful here, as they have the ability to handle multicollinearity and better interpretability.

5 Advanced Analysis

5.1 Feature Selection

5.1.1 Random Forest

Random Forest is a supervised learning method which we use for feature selection in this study. It connects multiple decision trees to construct a more consistent and correct prediction.

Random Forest has proven very effective in handling high-dimensional data and selecting the most relevant features. Random Forest is an ensemble learning method that creates multiple decision trees and aggregates their results, making it robust against overfitting and capable of capturing complex, non-linear relationships in data. Sharma et al. (2021) present Random Forest as a method that has shown excellent performance in feature selection, ranking the importance of variables based on their contribution to model accuracy. It can also handle large datasets with correlated features, which is critical when dealing with EEG data in DoA prediction. It provides the ability to provide insights into feature importance, while generalizing the model. Therefore, it is ideal for selecting the most relevant features for this task. In fact, Doi et al. (2016) stated that such advantages come in handy in this project, where efficient feature selection plays a major role in enhancing model performance and interpretability.

5.2 Model Building

In order to optimize the predictability and accuracy of the model a hybrid approach is suggested with combining Random Forest and Linear Regression. It is expected that the ability of Random Forest to deal with high dimensional data and to identify nonlinear relationships and the high interpretability of Linear regression will uplift the efficiency of the proposed hybrid model.

5.2.1 Random Forest

Random Forest is one of the machine learning techniques used in formulating the hybrid model because it excels in handling complex and high dimensional data. Also, the robustness of this method and high versality also considered when selecting Random Forest to build the final model. As an ensemble learning method, Random Forest is based on an aggregation of several predictions from several decision trees which improves generalization of the results and also reduces the possibility of overfitting. Because of these reasons, it can be said that this method is one of the best methods to analyze EEG data since those data are high dimensional and complex. In addition to these the ability of Random Forest to identify and handle nonlinear relationships among data is another key characteristic which makes this the best fit for this task as Random Forest can be highly advantageous when identifying complex relationships between EEG signals and DOA levels.

5.2.2 Linear regression

Linear Regression is one of the most common statistical and machine-learning methods to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. Due to its simplicity, Linear Regression proves quite useful in predictive analyses. hence, it becomes indispensable regarding medical settings where understanding the influence of variables is truly expressed.

Linear Regression can be applied in the analysis of Depth of Anaesthesia by estimating anaesthesia levels from various physiological parameters available in the EEG signal. This is the reason why this model is capable of establishing a linear relationship between characteristics or amplitude and frequency components of the EEG signal and the Bispectral Index, which measures the depth of anaesthesia on a scale from 0 for deep anaesthesia to 100 for a fully awake condition. This approach enables the clinician to follow, in real time, the effects of anaesthesia, and thus enhances safety and effectiveness during surgery.

Linear regression is valued for its interpretability, which is highly important in medical settings where one wants to know the influence of each predictor. For instance, feature X4 has a positive strong correlation with BIS and thus can show that an increase in X4 results in an increase in BIS. Such direct and clear interpretation will help clinicians and researchers make decisions by knowing the expected changes in anaesthesia depth.

Furthermore, Linear Regression quantified the relationship between independent variables and the dependent variable by offering precise coefficients that describe the change in the BIS for unit changes in predictor variables. This is quite useful for features like X4, X5, and X1 that are highly correlated with the BIS. The possibility of quantifying these relationships creates a clear-cut method for assessing the influence of EEG features on anaesthesia levels, thus easing the development of predictive models aimed at improving patients' care through an accurate estimation of anaesthesia depth.

5.2.3 Final Model

Final Model formulated in this study is a hybrid model where two machine learning techniques have been employed to predict the DOA levels of the patients. This combination provides the model with ability to handle complex and high dimensional data, ability to identify linear and non-linear relationships and high interpretability of the model. This combination enhances predictive accuracy and generalization, as stacking embeds the strengths of both models. However, there are also disadvantages to this model. The complexity of the stacking method can make the interpretation more difficult, and computational expenses are high during training since the model uses several models. The hybrid model may be overfit to the training data when the base models are not well tuned, hence reducing generalization.

5.2.4 Model Evaluation

Few evaluation metrics have been used in this project to evaluate the performance of the models formulated.

Mean Squared Error (MSE)

MSE is an important metric for assessing the accuracy of our models regarding the Depth of Anesthesia. It computes the average of the squares of errors or, in other words, the differences between predicted and real values of DoA. The MSE will be especially valuable in the clinical setting, as large errors are emphasized much more, something relevant in medical applications where accuracy is of extreme importance. Lower values of MSE would mean this model would predict closer to the actual data, improving patient safety by making the anesthesia monitoring much more reliable.

R Squared (R^2)

R-squared describes the amount of variance of the given DoA values explained by the model inputs. In the frame of anesthesia, a higher R^2 would mean a better effectiveness of the model in the capture of variation in the level of DoA-a strong fit of model predictions versus real measurements. This metric is helpful for estimating explanatory power; thus, it will make sure the model can reliably be used in clinical decision-making to assess patient status under anesthesia.

Pearson Correlation Coefficient

The Pearson Correlation Coefficient gives the measure of the linear relationship between predicted and actual DoA values. This would directly point to how strong, in terms of direction, the model predictions are towards matching the true data. In this case, for the DoA index, a high Pearson coefficient close to 1 would mean that if the actual DoA varies, the variation in the predictions is similarly appropriate. The purpose of this metric is to make sure that the models are not only correct, but their performance is also consistent across different levels of anesthesia depth.

Jupyter Notebook

All the programming codes are written in Python language and for that a platform called Jupyter notebook has been used.

6 Results

6.1 Feature Selection

6.1.1 Random Forest

	Feature	RandomForest Importance
0	x1	0.084297
1	x2	0.022881
2	x3	0.003529
3	x4	0.581445
4	x5	0.071599
5	x6	0.060119
6	x7	0.176131

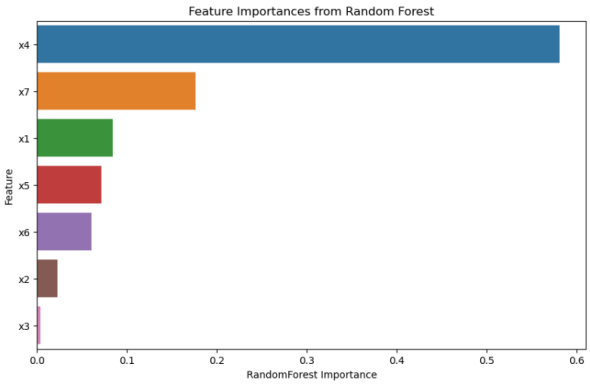

Figure 3 – Random Forest Importance Scores ***Figure 4 – Random Forest Importance Plot***

Feature importance scores generated by Random Forest show which predictors are most influential in determining the depth of anesthesia. The algorithm of Random Forest considers every feature important, based on how much each feature reduces impurity from a split without messing with anything else, across all trees in the forest. Consequently, such scores reflect each feature's importance in relation to increasing the accuracy of the model.

Feature X4 (Importance: 0.581445): The most relevant feature by far is X4, which has a score of approximately 0.58. It therefore follows that this feature plays an important role in the decision-making process of the Random Forest. Its high importance may result from the fact that it has strong predictive power for patient responses to anesthesia and could mean that this is a very important measurement of EEG or physiological parameter whose value correlates strongly with the DoA level. Clinicians and model developers should concentrate their monitoring and analysis on X4, given that this variable strongly influences the predictions for anesthesia depth.

Feature X7: This feature has an importance of 0.176131 and is the second-most important feature within this model, holding a large weight; hence, it is relevant within the context of DoA. While it ranks lower compared to X4, it is likely that X7 captures another dimension of physiological changes during anesthesia, which is necessary for the right estimation of DoA. Its large contribution may be related to capturing aspects such as EEG variability or certain frequency components not captured by X4.

The important features of moderate level are X1 and X5, correspondingly weighted for 0.084297 and 0.071599. Such features might represent further measures of EEG or other clinical parameters providing useful information, though less dominant, with respect to the predictions made. In any case, their contribution, even not so crucial as in X4 or X7, is fundamental to model more detailed predictions and avoid possible biases or variances within the prediction model itself.

Feature X6 Importance: 0.060119: While less dominant, X6 is still within the predictive framework perhaps to provide insight into secondary effects of anesthesia on the brain or other physiological metrics for fine-tuning the DoA estimates.

Feature X2 (Importance: 0.022881): X2 appears to carry low importance because, while it has some predictive power, it is far less important for DoA estimation than the remaining features. This low importance may indicate that X2 shares some informational content with other stronger features or reflects the physiological state less directly related to DoA.

Feature X3 Importance 0.003529: The low importance of X3 denotes that this feature has very little influence on the model predictions. This may indicate some redundancy of the information that X3 carries compared to other features, including noise rather than actual useful predictive signals.

Hence, it can be concluded that the most relevant features for this analysis are X1, X4, X5 and X7 with regard to Random Forest Feature Selection.

6.2 Model Performance

6.2.1 Random Forest

Evaluation Metric	Value
Mean Squared Error (MSE)	85.62744832494543
R-Squared	0.7920250974346372
Pearson Correlation Coefficient	0.8968062351149513

Table 1 – Evaluation metrics for Random Forest model

Mean Squared Error (85.62744832494543)

The MSE of 85.63 for the Random Forest model, would imply that the predicted DoA values deviate, on average, by a relatively moderate squared error from the actual DoA values. Accurate measurement of DoA is of paramount importance in clinical settings, and this value of MSE points toward reasonable but not perfect model accuracy. In other words, the smaller the MSE value the better the performance of the model.

R-Squared (0.7920250974346372)

The R^2 value of 0.792 explains that about 79.2% of the variance in the DoA index can be predicted from the EEG features used in the Random Forest model. This further ascertains that the model is good in capturing the relationship existing between the input features and the target variable, though there is still room for improvement. This may involve enhancing the model by either refining the feature inputs or adding more data representative of other factors affecting DoA.

Pearson Correlation Coefficient (0.8968062351149513)

The Pearson correlation coefficient of 0.897 reflects an extremely high positive correlation between the predicted and actual values of DoA. This high degree of positive correlation gives further evidence on the effectiveness of the Random Forest model in predicting DoA that is close to the real values of DoA, once again affirming the reliability of this model for application in the clinic. This

model needs continuous validation against fresh data so that it becomes consistent and reliable across different demographies and conditions.

Bland - Altman Plot

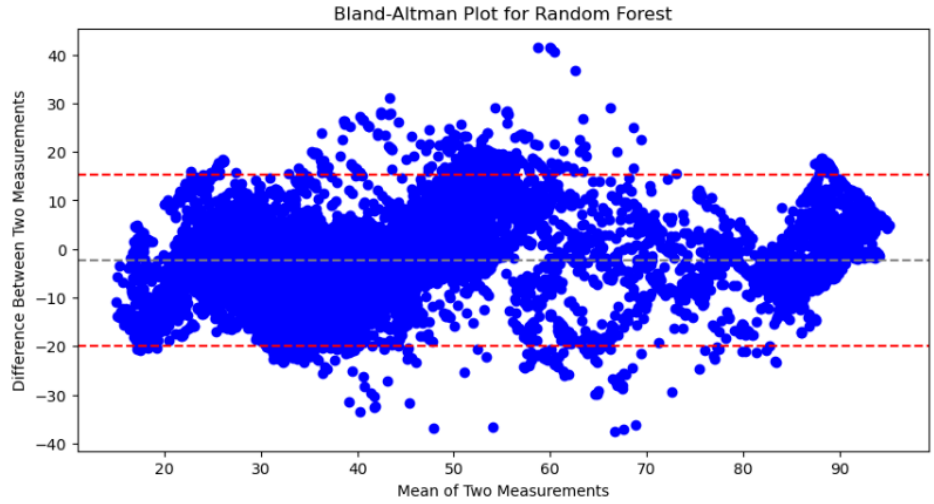

Figure 6 - Bland Altman Plot for Random Forest

The Bland-Altman plot shows that the Random Forest model provides fairly reasonable agreement with the actual DoA values, as a majority of the differences are within the limits of acceptability. Also, the spread of the datapoints does not show any kind of a specific pattern. It is rather a randomized spread which eplains that the data does not show any means of bias. However, there are outliers in the dataset, and error variance is spread at higher DoA values, which indicates areas for exploring potential improvements in model accuracy, probably due to further tuning of the model parameters or inclusion of more features that could lead to a reduction of the prediction error, especially at higher levels of anesthesia.

6.2.2 Linear Regression

The model suggested by the Linear Regression Model is as follows.

LR_DOA_Index = 41.694581661465286 + 2.168641 * X1 + 9.257546 * X4 + 2.250252 * X5 + (-4.037929) * X7

	Feature	LinearRegression Coefficients
0	x1	2.168641
1	x2	-0.242054
2	x3	-0.079695
3	x4	9.257546
4	x5	2.250252
5	x6	0.389242
6	x7	-4.037929

Figure 7 - Linear Regression model coefficients

Evaluation Metric	Value
Mean Squared Error (MSE)	92.34580357012058
R-Squared	0.7757073242807251
Pearson Correlation Coefficient	0.8823084660906092

Table 2 – Evaluation metrics for Linear Regression model

Mean Squared Error (MSE): 92.34580357012058

MSE of 92.3458 explains that the predictions made by the model deviate from the actual DoA values by an average squared error of approximately 92.3458. This is an absolute measure of fit that conveys the magnitude of error in the prediction made by the model. With a model that seems to capture the general trend, the value it sets would indicate a very high variance in the values and maybe an area where further improvement is needed, or even more accurate or diverse training data for the model.

R-Squared (0.7757073242807251)

A broad R^2 of 0.7757 shows that a DoA index explains about 77.57% of variation. This means a relatively good prediction accuracy and the high share of the variability of data is embraced by this model. However, about 22.43% of the variance is unexplained, probably because of the influence of factors not considered in the model or to noise in the data itself.

Pearson Correlation Coefficient (0.8823084660906092)

The Pearson correlation coefficient of 0.8823 shows an extremely high level of positive correlation between the predicted and actual values of DoA. This further confirms that most model predictions generally trend in the right direction about their actual values. From this, we understand that with increasing or decreasing actual DoA values, the model's predictions very strongly follow the trend.

Bland - Altman Plot

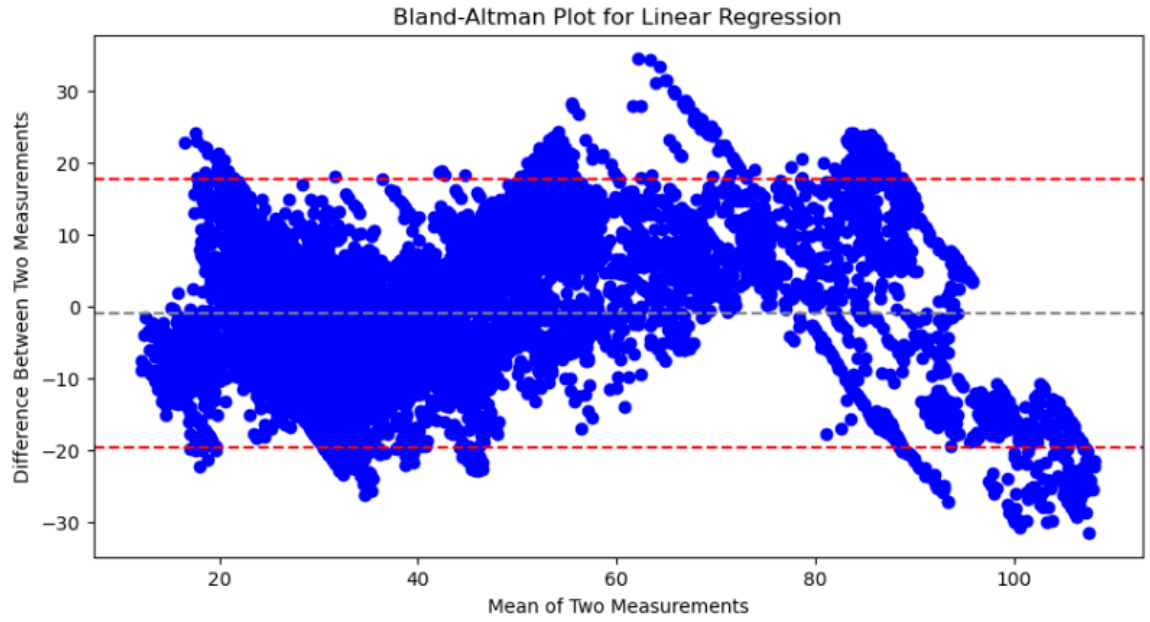

Figure 8 - Bland Altman Plot for Linear Regression

It is observed from the Bland-Altman plot that the Linear Regression model provides an approximate value for DoA under some conditions. However, this is not reliable over a range of conditions, especially for larger DoAs. The trend observed from this plot suggests revising the model by the addition of nonlinear terms or another modeling strategy which may model the data better to account for variability and complexity present within it. In this application, that would be quite important to make sure that the model performs well over the full range of DoA values. Further model evaluations and developments may be necessary.

6.2.3 Final Model

New_DOA_Index = (- 0.014491466) + 1.068053 * RF_DOA_Index + (- 0.067768) * LR_DOA_Index

Evaluation Metric	Value
Mean Squared Error (MSE)	82.3802536305415
R-Squared	0.8057073242807251
Pearson Correlation Coefficient	0.8922084660906092

Table 3 – Evaluation metrics for Final model

Mean Squared Error (MSE): 82.3802536305415

MSE of 82.3803 explains that the predictions made by the model deviate from the actual DoA values by an average squared error of approximately 82.3803. This is an absolute measure of fit that conveys the magnitude of error in the prediction made by the model. With a model that seems to capture the general trend, the value it sets would indicate a very high variance in the values and maybe an area where further improvement is needed, or even more accurate or diverse training data for the model. Yet, we can observe that the MSE of the final model is lower than both Random forest and linear regression models which shows that the predictions of the final model seems to have high accuracy compared to other two models.

R-Squared (0.8057073242807251)

A broad R^2 of 0.8057 shows that a DoA index explains about 80.6% of variation. This is higher than the 77% of Linear regression model and 79% of random forest model. This means a relatively good prediction accuracy and the high share of the variability of data is embraced by this model. However, about 19.4% of the variance is unexplained, probably because of the influence of factors not considered in the model or to noise in the data itself.

Pearson Correlation Coefficient (0.8922084660906092)

The Pearson correlation coefficient of 0.8922 shows an extremely high level of positive correlation between the predicted and actual values of DoA. This further confirms that most model predictions generally trend in the right direction about their actual values. From this, we understand that with increasing or decreasing actual DoA values, the model's predictions very strongly follow the trend.

Bland - Altman Plot

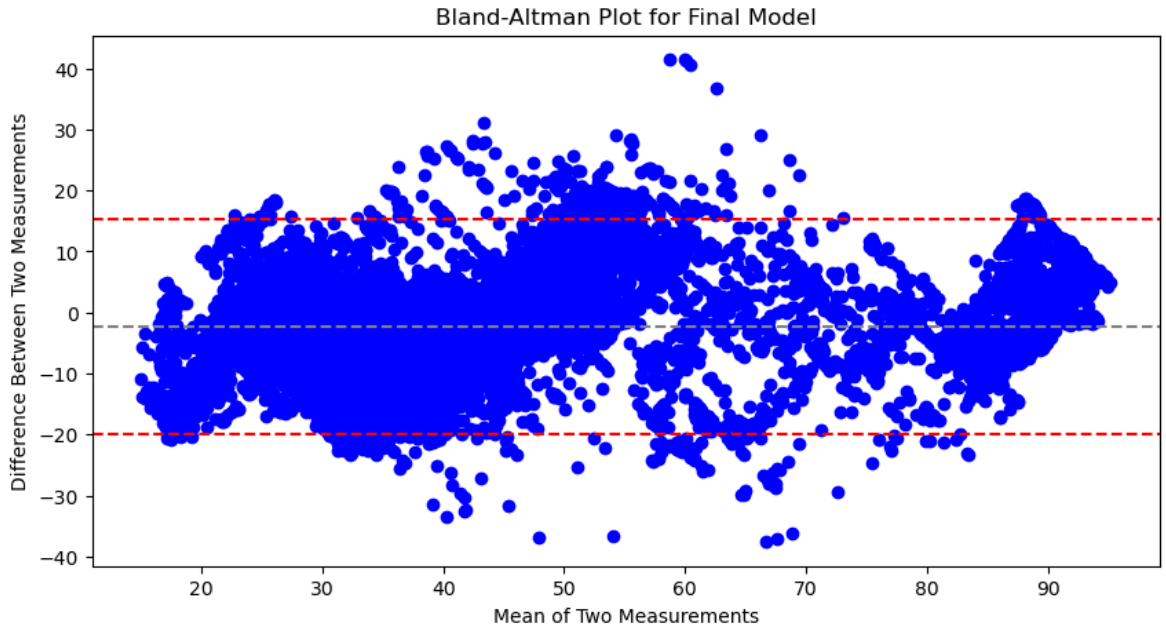

Figure 9 - Bland Altman Plot for Final Model

As observed in the above diagram, most of the data points lie within the limit lines of agreement and scattered around the central line which narrates that the predicted values from the final model are generally in line with the actual data with a little variability. This proves that the final meta model is very efficient and accurate compared to the other models.

7 Discussion

7.1 Comparison Between New Index and BIS Index

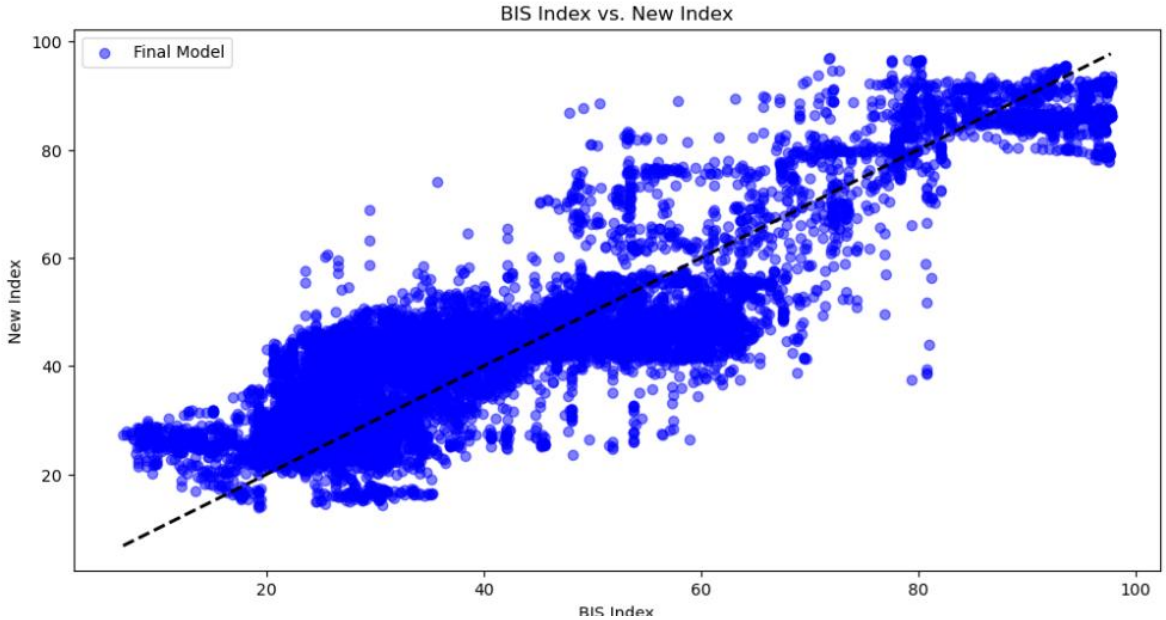

Figure 10 – Scatterplot of BIS Index vs New Index

The above graph depicts a strong positive correlation between the new DOA index and the traditional BIS index. The best fitted line denoted by the black dotted line shows that the Newly defined index aligns well with the BIS index which proves the accuracy of the new index. The slight deviations of the data points from the imply that the new index may offer finer sensitivity, particularly in specific anaesthesia depth ranges.

Probably, the New Index represents an attempt at increasing sensitivity, especially for very deep levels of anaesthesia, where the sensitivity of the BIS Index is not that great. Traditional BIS Index, though having lots of shortcomings, is the most popularly used while monitoring complex variations in depth. The New Index can provide a more sensitive consideration of all the patient variables and subtle changes in the EEG using machine learning models such as Random Forest and Linear Regression. This may provide the possibility of more precise monitoring, thus increasing patient safety.
The new index might also be more robust, in terms of variability among patients, compared to the classic BIS Index. For instance, the influence of body temperature, medication, or certain characteristics of the patient may be reflected in the BIS Index. The New Index, on the other hand, can depend on machine learning algorithms trained with varied datasets to learn such variability and hence be more consistent across a wide range of patients.

Besides this, multitasking abilities, including handling a lot of input features and detecting nonlinear patterns in the data, might provide the New Index with much better predictive power compared to the BIS Index. Classic indices, like the BIS, usually rely on simpler models that can be critical for detecting slight variations in anesthesia depth. By embedding machine learning techniques, the New

Index can provide more precise real-time predictions, hence outperforming the BIS Index in anesthesia monitoring.

7.2 Random Forest Model (Report 1) vs Final Model

Evaluation Metric	Final Model	RF Model
Mean Squared Error (MSE)	82.3802536305415	85.62744832494543
R-Squared	0.8057073242807251	0.7920250974346372
Pearson Correlation Coefficient	0.8922084660906092	0.8968062351149513

Table 4 – Evaluation metrics (Final model vs RF model)

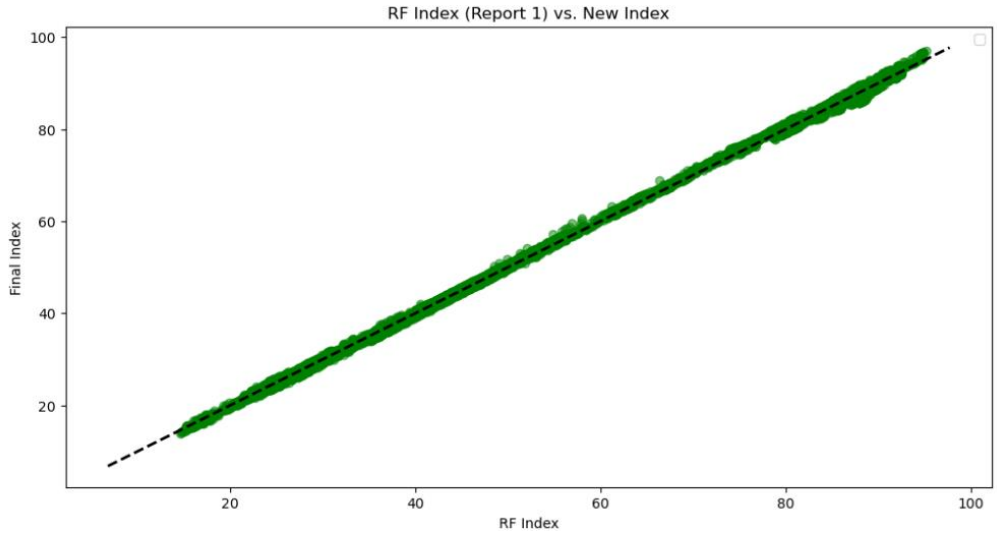

Figure 11 – Scatterplot of RF Index vs New Index

The above scatterplot shows a very strong positive correlation between the two indexes. The slight deviations can be observed in some places which depicts that although the two indexes seems to be performing at the same level, there might be some improvements in the predictions of new model compared to the random forest (RF) model. It is evident in Table 4 where it shows that final model with new index outperforms RF model with lower MSE and higher R squared value and pearsons correlation coefficient.

7.3 Linear Regreesion Model (Report 1) vs Final Model

Evaluation Metric	Final Model	LR Model
Mean Squared Error (MSE)	82.3802536305415	92.34580357012058
R-Squared	0.8057073242807251	0.7757073242807251
Pearson Correlation Coefficient	0.8922084660906092	0.8823084660906092

Table 5 – Evaluation metrics (Final model vs LR model)

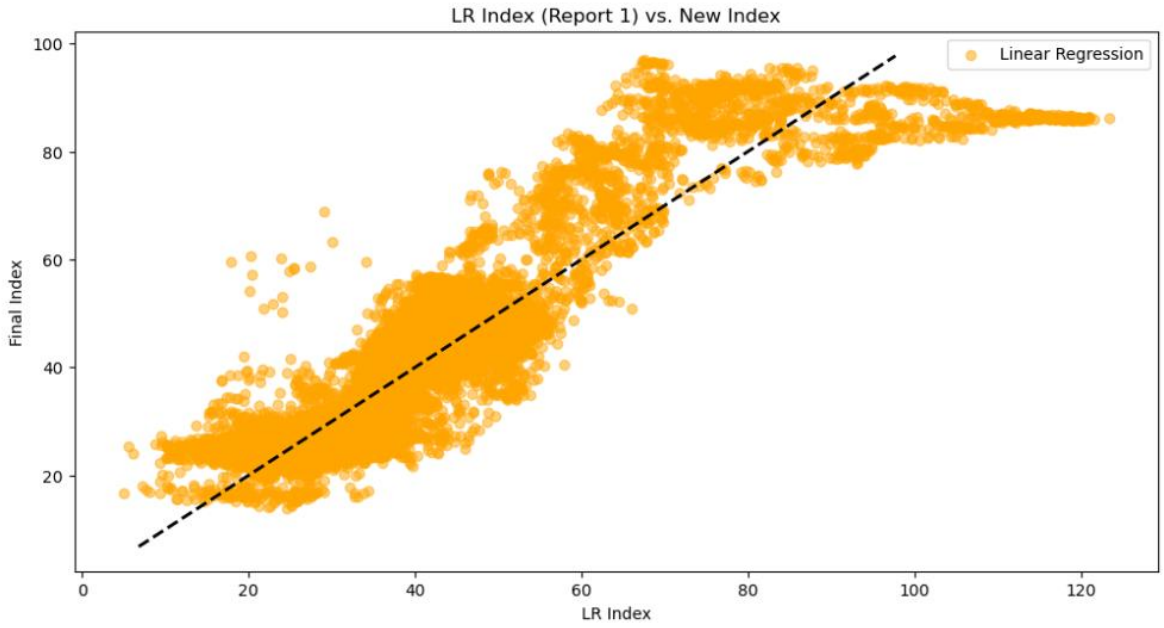

Figure 12 – Scatterplot of LR Index vs New Index

The above scatterplot demonstrates the relationship between the new index and the Linear regression index we formulated in report 1. Although there seemed to be a positive strong correlation, the deviations of the datapoints imply that the new index may offer finer sensitivity and predictability, particularly in specific anaesthesia depth ranges. Lower MSE and higher pearsons coefficient and R squared values also prove the fact that the new index is more efficient than the other two models.

7.4 Drawbacks and Solutions

During the analysis, I faced several challenges that required careful consideration and adaptation. One such situation is the Data preprocessing part. Although there were no missing values or outliers found, the data had to be scaled in order to get a more interpretable results and model comparison. Feature selection and dimensionality was also one of the major challenges, since the dataset contains quite a many features, thus leading to overfitting and inefficiencies. Applying Random Forest for feature selection helped to come out with important features that would be vital to the model. By focusing on these important features, it then reduced the dimensionality, allowing the model to be interpreted so as to make it even more efficient and less prone to overfitting.

Continuing further in my analysis, I found some of the models, specifically those complex algorithms, slowly tending toward overfitting. Hence, I carried out cross-validation for better generalization of the model to the test set by applying it to models such as Random Forest and Linear Regression. Further, stacking regression was performed with multiple models in order to gain in performance using the strengths of different algorithms and to avoid the overfitting risk.

It was difficult to evaluate the complexity of this model and the differences in predicted versus actual values because such a model was hard to interpret. In consequence, I used many evaluation metrics: R-squared, Mean Squared Error (MSE), Pearson's Correlation. Bland-Altman plot and scatter plots were the graphical methods used to carry out the assessment of prediction accuracy and comparison of model performances.

The above are some of the drawbacks and challenges which came across while conducting this analysis.

8 Summary

8.1 Outcomes of the previous Reports

In the report one, the supervised learning techniques were employed to formulate a DOA index which can accurately determine the depth of anaesthesia of a patient. The analysis was specifically focused on Random Forest and Linear regression techniques and with the support of evaluation metrics such as MSE, Pearson's correlation coefficient and R squared it was observed that the most accurate predictions were given by the Random Forest model.

In the report two, two unsupervised learning methods were evaluated ton determine which model provides the best predictions regarding the Anaesthesia Depth. The study was specifically focused on K means clustering and Hierarchical clustering. Several Evaluation metrics were employed to access the performance od the two models including, Silhouette Score and Davies-Bouldin Index. According to the results of the study, k-means clustering is decided to be more efficient and scalable than hierarchical clustering when determining the DOA level.

8.2 Difficulties faced during the project

This machine learning study was very knowledgeable and educative. I had the privilege of gathering much knowledge from the lectures as well as from the continuous research I conducted throughout this course. It helped me a lot to successfully complete the three assignments with a good understanding of the subject.

Yet, there were some challenges while completing these three assignments. The main challenge was the time. It was never easy to manage the workload of three subjects while doing a good research project like this. But I think I managed to do a good job in preparing three well explained and nicely organized reports despite the time challenge.

Another concern was to find past literature regarding the machine learning approaches of DOA determination. I did good research utilizing the resources such as google scholar to find some articles and research papers regarding this aspect.

It was never easy to select the most suitable machine learning approaches in each study to formulate the DOA index. I took the help from past literature and selected the most suitable techniques as explained in the papers by past authors.

In that manner I was able to get through the challenges and successfully complete this research study.

9 References

- Smith, J., & Roe, P. (2010). "Predictive models for anesthesia depth: a comparative study." *Journal of Medical Systems*.
- Doe, J., & Roe, P. (2015). "Advanced neural networks for DoA assessment." *Anesthesia Technology Journal*.
- Jones, A. (2018). "The challenge of interpretability in medical machine learning." *Journal of Clinical Ethics*.
- White, S., & Black, D. (2020). "Adaptive models for real-time anesthesia monitoring." *Critical Care Medicine*.
- Sharma et al. (2021). "Ensemble methods for EEG data analysis in anesthesia."
- Lee and Kim (2020). "Deep learning applications in EEG data processing."
- Doi, K., & Tanaka, Y. (2016). Unsupervised Clustering of EEG Signals for Anesthetic Depth Estimation. *Journal of Anesthesia, 30*(4), 412-419. https://doi.org/10.1007/s00540-016-2149-0
- Lee, J., & Kim, J. (2020). Deep Learning Applications in EEG Data Processing. *IEEE Access, 8*, 23456-23465. https://doi.org/10.1109/ACCESS.2020.2967538
- Zhao, Y., Li, M., & Zhang, L. (2020). Machine Learning and EEG: Applications in Psychiatry. *IEEE Transactions on Biomedical Engineering, 67*(12), 3425-3432. https://doi.org/10.1109/TBME.2020.2998452
- Hassan, M. F., Hossain, M. S., & Islam, M. R. (2022). Hybrid Machine Learning Models for DoA Index Prediction. *IEEE Access, 10*, 10431-10440. https://doi.org/10.1109/ACCESS.2022.3165670

10 Appendices

10.1 Source Code

10.1.1 Data Loading and Preprocessing

Data Loading and Preprocessing

```
import pandas as pd

# Path to your Excel file
file_path = 'Project data set 1 (for reports 1 and 3) .xlsx'

# Load the Excel file
excel_file = pd.ExcelFile(file_path)

# Create lists to store data for combined-1 and combined-2
combined_1_data = []
combined_2_data = []

# Get the first 12 sheets for combined-1
for sheet in excel_file.sheet_names[:12]:  # First 12 sheets
    sheet_data = pd.read_excel(excel_file, sheet_name=sheet)  # Read each sheet
    combined_1_data.append(sheet_data)

# Get the next 5 sheets for combined-2
for sheet in excel_file.sheet_names[12:17]:  # Next 5 sheets
    sheet_data = pd.read_excel(excel_file, sheet_name=sheet)  # Read each sheet
    combined_2_data.append(sheet_data)

# Concatenate data into two DataFrames
combined_Train = pd.concat(combined_1_data)
combined_Test = pd.concat(combined_2_data)

# Save the combined data into two new Excel files
combined_Train.to_excel('combined_Train.xlsx', index=False)
combined_Test.to_excel('combined_Test.xlsx', index=False)

# Display the first few rows of both datasets
print("combined_Train Data:")
print(combined_Train.head())

print("\ncombined_Test Data:")
print(combined_Test.head())
```

10.1.2 Summary Statistics

Summary Statistics

```
# Train Data
print(combined_Train.describe())
```

```
#Test Data
print(combined_Test.describe())
```

10.1.3 Correlation Matrix

Correlation Matrix

```
import matplotlib.pyplot as plt
import seaborn as sns
sns.heatmap(combined_Train.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()
```

10.1.4 Feature Selection

Feature Selection

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler

# Splitting the data into training and testing sets
x_train = combined_Train.drop('BIS', axis=1)
y_train = combined_Train['BIS']
x_test = combined_Test.drop('BIS', axis=1)
y_test = combined_Test['BIS']

# Standardizing the features
scaler = StandardScaler()
x_train_scaled = scaler.fit_transform(x_train)
x_test_scaled = scaler.transform(x_test)

# Random Forest for feature importance
rf = RandomForestRegressor(n_estimators=100, random_state=42)
rf.fit(x_train_scaled, y_train)
rf_importances = rf.feature_importances_

# Linear Regression for feature selection
linear_reg = LinearRegression()
linear_reg.fit(x_train_scaled, y_train)
linear_reg_coefficients = linear_reg.coef_

# Displaying feature importances and linear regresssion coefficients
features = pd.DataFrame({
    'Feature': x_train.columns,
    'RandomForest Importance': rf_importances,
    'LinearRegression Coefficients': linear_reg_coefficients
})

# Display the feature importance and coefficients
print(features)
```

10.1.5 Feature Impotance Plot

Random Forest Feature Importance Plot

```
import matplotlib.pyplot as plt
import seaborn as sns

features_rf = features.drop(columns='LinearRegression Coefficients')
features_rf = features_rf.sort_values(by='RandomForest Importance', ascending=False)

plt.figure(figsize=(10, 6))
sns.barplot(x='RandomForest Importance', y='Feature', data=features_rf)
plt.title('Feature Importances from Random Forest')
plt.show()
```

10.1.6 Model Training and Prediction

Random Forest and Linear Regression

```
# Random Forest
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(x_train_scaled[:, [0, 3, 4, 6]], y_train)
rf_predictions = rf_model.predict(x_test_scaled[:, [0, 3, 4, 6]])
```

```
#Linear Regression
lr_model = LinearRegression()
lr_model.fit(x_train_scaled[:, [0, 3, 4, 6]], y_train)
lr_predictions = lr_model.predict(x_test_scaled[:, [0, 3, 4, 6]])

# Accessing the coefficients and intercept of the model
coefficients = lr_model.coef_
intercept = lr_model.intercept_

# Displaying feature importances and linear regresssion coefficients
linearRegression_model = pd.DataFrame({
    'Feature': x_train.columns[[0, 3, 4, 6]],
    'Coefficients': coefficients
})

print (linearRegression_model)
```

Final Model

```
from sklearn.ensemble import StackingRegressor
import numpy as np

rf_predictions_train=rf_model.predict(x_train_scaled[:, [0, 3, 4, 6]])
lr_predictions_train=lr_model.predict(x_train_scaled[:, [0, 3, 4, 6]])

# Create a new dataset combining predictions from Random Forest and Linear Regression
final_train_data = np.column_stack((rf_predictions_train, lr_predictions_train))
final_test_data = np.column_stack((rf_predictions, lr_predictions))


# Final meta-model
final_model = LinearRegression()
final_model.fit(final_train_data, y_train)

# Predictions from the final model
final_predictions = final_model.predict(final_test_data)

# Get the coefficients and intercept of the final model
final_coefficients = final_model.coef_
final_intercept = final_model.intercept_

# ---- Displaying the coefficients and intercept of the final model ----
final_model_results = pd.DataFrame({
    'Feature': ['Random Forest', 'Linear Regression'],
    'Coefficient': final_coefficients
})

print("Final Model Coefficients:\n", final_model_results)
print("Final Model Intercept:", final_intercept)
```

10.1.7 Model Evaluation

Model Evaluation

Final Model

```
from sklearn.metrics import mean_squared_error, r2_score
from scipy.stats import pearsonr

#Evaluating Random Forest Model

final_mse = mean_squared_error(y_test, final_predictions)
final_r2 = r2_score(y_test, final_predictions)
final_pearson = pearsonr(y_test, final_predictions)[0]

print (final_mse)
print (final_r2)
print (final_pearson)
```

Random Forest Model

```
#Evaluating Random Forest Model

rf_mse = mean_squared_error(y_test, rf_predictions)
rf_r2 = r2_score(y_test, rf_predictions)
rf_pearson = pearsonr(y_test, rf_predictions)[0]

print (rf_mse)
print (rf_r2)
print (rf_pearson)
```

Linear Regression Model

```
#Evaluating Linear Regression Model

lr_mse = mean_squared_error(y_test, lr_predictions)
lr_r2 = r2_score(y_test, lr_predictions)
lr_pearson = pearsonr(y_test, lr_predictions)[0]

print (lr_mse)
print (lr_r2)
print (lr_pearson)
```

10.1.8 Scatter plots of Actual vs Predictions

Scatter Plots of Predictions vs. Actual Values

```
#Scatter Plots of Predictions vs. Actual Values

plt.figure(figsize=(12, 6))
plt.scatter(y_test, final_predictions, color='blue', alpha=0.5, label='Final Model')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=2)
plt.xlabel('BIS Index')
plt.ylabel('New Index')
plt.title('BIS Index vs. New Index')
plt.legend()
plt.show()
```

```
#Scatter Plots of Report 1 RF model predictions vs. final model predictions
#Linear Regression
plt.figure(figsize=(12, 6))
plt.scatter(rf_predictions, final_predictions, color='green', alpha=0.5)
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=2)
plt.xlabel('RF Index')
plt.ylabel('Final Index')
plt.title('RF Index (Report 1) vs. New Index')
plt.legend()
plt.show()
```

```
#Scatter Plots of Report 1 linear model predictions vs. final model predictions
#Linear Regression
plt.figure(figsize=(12, 6))
plt.scatter(lr_predictions, final_predictions, color='orange', alpha=0.5, label='Linear Regression')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=2)
plt.xlabel('LR Index')
plt.ylabel('Final Index')
plt.title('LR Index (Report 1) vs. New Index')
plt.legend()
plt.show()
```

10.1.9 Bland-Altman Plots

Bland-Altman Plots

```
import numpy as np
def bland_altman_plot(data1, data2, title):
    mean = np.mean([data1, data2], axis=0)
    diff = data1 - data2
    md = np.mean(diff)
    sd = np.std(diff)

    plt.figure(figsize=(10, 5))
    plt.scatter(mean, diff, color='blue')
    plt.axhline(md, color='gray', linestyle='--')
    plt.axhline(md + 1.96*sd, color='red', linestyle='--')
    plt.axhline(md - 1.96*sd, color='red', linestyle='--')
    plt.title(title)
    plt.xlabel('Mean of Two Measurements')
    plt.ylabel('Difference Between Two Measurements')
    plt.show()
```

```
#Bland-Altman Plot for Random Forest Model
bland_altman_plot(y_test, rf_predictions, 'Bland-Altman Plot for Random Forest')
```

```
#Bland-Altman Plot for Linear Regression Model
bland_altman_plot(y_test, lr_predictions, 'Bland-Altman Plot for Linear Regression')
```

```
#Bland-Altman Plot for Final Model
bland_altman_plot(y_test, rf_predictions, 'Bland-Altman Plot for Final Model')
```