Prediction of Sea Level Oscillations: Comparison of Regression-based Approach

Ahmad Fitri Jamali, Aida Mustapha, *Member, IAENG,* and Salama A. Mostafa, *Member, IAENG*

Abstract—**In Malaysia, especially in the east coast region on the peninsula, rely heavily of the sea level reading to alert for protecting the low-lying residential regions along the coastal areas. Because recent climate change has driven the rise of sea level globally, it is imperative that the government has the capacity to estimate any increase in sea level with sufficient lead time in the case of natural disaster. This study primarily aims to investigate the validity and effectiveness of four regression models, which are the Decision Tree Regression (DTR), Decision Forest Regression (DFR), Linear Regression (LR), and Bayesian Linear Regression (BLR) for predicting the monthly variation of the mean sea level. Variations of the regression models are used because these techniques have not been explored to predict mean sea level in coastal areas. The input dataset is sourced from Kerteh, Tioman Island, and Tanjung Sedili in Malaysia from January 2007 to December 2017. The performance of all algorithms are measured and compared based on the class Mean Absolute Error (MAE), Root Mean Square Errors (RMSE), Relative Absolute Error (RAE), Relative Squared Error (RSE), and Coefficient of Determination (R^2). The results are hoped to model predictions of the mean sea level as part of activity in performing sea level data analysis. Therefore, this research will be able to alert other government and authorities to make an early strategy to handle the problems.**

Index Terms—**hydrology, sea-level, regression, machine-learning.**

I. INTRODUCTION

MALAYSIA is a country comprising Peninsular Malaysia, Sabah, and Sarawak separated by the South China Sea. It covers 14 countries which are Perlis, Kedah, Penang, Perak, Selangor, Negeri Sembilan, Pahang, Melaka, Johor, Kelantan, Terengganu, Sabah, and Sarawak. Malaysia also has one Federal Government consisting of three territories, which are the Federation of Kuala Lumpur, Federal Territory of Labuan and Federal Territory of Putrajaya. The northern border of Malaysia is Thailand while the southern border is Singapore.

In Malaysia, the low-lying coastal regions host very large cities with a dense population. This poses a high security risk when associated with any sudden increase in sea level. In general, the contemporary sea level rise are due to increase of ocean temperature, melting of ice sheets, as well as varying ocean circulation, El Nino–Southern Oscillation (ENSO) and Pacific Decadal Oscillation [1]. The shoreline at Peninsular Malaysia's east coast (ECPM) is directly exposed to impact from extreme rain and storms especially during the times of northeast monsoon [2].

Mean sea level rise is in relation with the sea level rise. The increase in sea levels is one of the most alarming and expensive consequences of the climate change phenomena. The direct impact to the coastal areas include coastal flooding, tidal flooding, coastal erosion, and inflow of salt water. Adding to the impact are the severity from other events that leads to the compounded effects of storm surges, high spring tides and surface waves, as well as flooding of the river [3]. These phenomenon calls for a consistent analysis on sea level rise so as to protect the low-lying residential regions and coastal areas.

The sea level data is a complex data that contains space dimension. To predict the sea level rise, this research explores four variations of regression algorithms, which are Decision Tree Regression (DTR), Decision Forest Regression (DFR), Linear Regression (LR), and Bayesian Linear Regression (BLR). Regression techniques have been widely used in initialization [4], optimization [5], outlier detection [6], prediction [7] as well as multi-class classification problems [8].

The sea level data under study is sourced from the Kerteh, Tioman Island, and Tanjung Sedili in Malaysia within the period of January 2007 to December 2017. Following [2], this study will predict the sea levels for five different time periods in 1, 5, 10, 20, and 40 years. The expected result for this project is to compare the monthly mean sea level oscillations of the algorithm for prediction across all algorithms.

The remaining of this paper is organized as follows. Section 2 deals with the literature review as it will explain in detail the process or software used and make a comparison between the methods. Next, Section 3 will address the methods used to forecast monthly mean sea level variability. In addition, Section 4 reveals the evaluation of the dataset, discussion about the result and the assessment of the system of comparison. Finally, Section 5 concludes the paper.

II. RELATED WORK

The rise in sea level is very impactful to ecosystem and living things. Various incidents can occur when the country is experiencing rise in sea level such as loss of assets and human lives, stall in economic activities, shock in mental health on top of loss on plants, and animals. The severity of such unfortunate events highly depends on how extreme the event is, how much are is being exposed to the event, and how vulnerable is the area under the event [9]. The sea level rise in Malaysia usually happen along the Peninsular Malaysia as well as Sabah and Sarawak coastlines. Table I provides comparison of dataset, evaluation metrics, and prediction algorithms with regards to sea level dataset.

Manuscript received September 28, 2020; revised June 18, 2021. This research is supported by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2020/ICT02/UTHM/02/4).

A. Mustapha is an Associate Professor of Mathematics and Statistics Department, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, KM1, Jalan Panchor, 84600 Panchor, Johor, Malaysia (e-mail: aidam@uthm.edu.my).

S.A. Mostafa is a Senior Lecturer of Software Engineering Department, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400, Batu Pahat, Johor, Malaysia (e-mail: salama@uthm.edu.my).

A.F. Jamali is a technician of Mesiniaga Berhad, Malaysia (email: fitri97jamali@gmail.com)

TABLE I
RELATED WORK

Reference	Dataset	Evaluation Metrics	Prediction Algorithms
[10]	Hourly sea level measurements were obtained from SEAFRAME (Sea-Level Fine Resolution coustic Measurement Equipment) station deployed at Hillarys Boat Harbor.	Root Mean Square Error and Correlation Coefficient Scatter Index	Artificial Neural Networks
[11]	Obtained from the U.K. National Tide Gauge Network. Mean sea level data from the Newlyn tide auge.	Coefficient of Determination and Root Mean Square Error	Multi-linear regression, Feed forward Back Propagation, Radial Basis Function, Generalized Regression NN
[12]	Hourly sea-level records from a Sea-level Fine Resolution Acoustic Measuring Equipment (SEAFRAME). Station deployed at Cocos (Keeling) Islands	Root Mean Square Error, Correlation Coefficient and Scatter Index	Artificial Neural Network
[13]	Based on the data obtained from a tide gauge installed at Port Resolution by our research team and the monthly sea level at Port Vila	Coastal slope estimation, Mean wave height estimation, Mean tidal range and Relative sea level rise	Spatial Bayesian Networks
[14]	Hourly sea-level records from SEAFRAME for Darwin Harbor, Australia.	Coefficient of Determination, Root Mean Square Error and Variance	Artificial Neural Networks, Adaptive and Neuro-fuzzy Inference System

Existing research on predicting sea level variation at the Hillarys Boat Harbour, Western Australia using the Artificial Neural Networks (ANN). The results reported that the feasibility of neural sea level forecasts in terms of the correlation coefficient (0.7 to 0.9), root mean square error (10% of range), and scatter index (0.1 to 0.2) [10]. [11] adopted various ANN models to estimate daily mean sea level heights, which are Feed-Forward Back Propagation (FFBP), Radial Basis Function (RBF), and Generalized Regression Neural Network (GRNN) algorithms. The results showed that the ANN and MLR models yielded comparatively better results than the standard method used to measure sea level, at least square estimates. FFBP, RBF, and MLR algorithms produced significantly better results than the GRNN method while the best performance was obtained by the FFBP algorithm.

In 2008, [12] predicted sea-level variations at the Cocos (Keeling) Islands, again using the ANN algorithm. ANN demonstrated reliable results in terms of the correlation coefficient (0.85 to 0.95), root mean square error (80 to 100mm), and scatter index (0.1 to 0.2) when compared with actual observations. This findings suggested that ANN is very useful for prediction of sea level variations at site-specific forecasts.

In another local at Darwin Harbor, Australia, [2] reported results of Neuro-Fuzzy and Neural Networks for forecasting sea level. The method, called the Adaptive Neuro Fuzzy Inference System (ANFIS), was found to be optimal for prediction with its adaptive learning rate. ANFIS models was optimal for predictions while the adaptive learning rate and Levenberg–Marquardt were best suited for training the ANN models. Consequently, for all prediction intervals, ANFIS and ANN models provided similar predictions and performed better than the ARMA models developed for the same purpose.

In the most recent work, [13] used Spatial Bayesian Networks for predicting sea level rise induced coastal erosion in a small Pacific Island. The results reported in this work supported adaptation planning based on risks and incorporated high-resolution coastal process models to support planning for local land use.

III. METHODOLOGY

The prediction experiments performed in this study is based on the Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM is an abstract, high-level model for data mining and it is also general enough to be used for other data analysis needs [15]. The CRISP-DM project addressed parts of these issues by defining a process model that provides a framework for the implementation of data mining projects that are independent of both the industry and the technology used. The CRISP-DM process model aims to make large data mining projects, less costly, more reliable, more repeatable, more manageable, and faster (Wirth, 2000). The CRISP-DM Process model for Data Mining consists out six phases, which are visualized in Fig. 1. The methodology starts with business understanding phase, data understanding, data preparation, modeling, assessment and deployment [16].

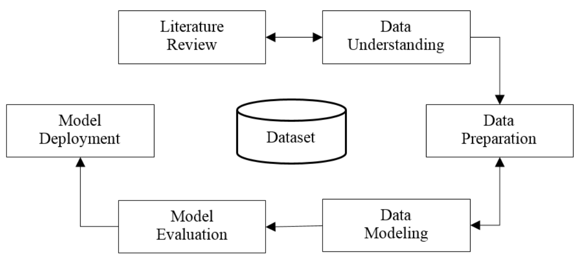

Fig. 1. CRISP-DM Process Model for Data Mining

The experiments were carried out using the Microsoft Azure tool with validation method for training and testing. Kerteh datasets are divided into which are 80% of the obtained data as training data (from January 1, 2007 to December 31, 2015). The remaining 20% was used as testing data (from January 1, 2016 to December 31, 2017) for the DTR, DFR, BLR and LR models [2].

A. Dataset

There are three (3) datasets, which are from Kerteh, Tioman Island, and Tanjung Sedili that consists an average of 125 data of each place. The datasets are described by monthly mean sea level, daily rainfall, mean cloud cover, temperature data will be analyze. All the datasets were combined into one datasets because the data for data sets were too small and features in the data is related to sea level are similar.

Table II, Table III, and Table IV show the excerpts of data collected from January 2007 to December 2017 for each location, respectively. This research scope is limited to the Kerteh, Tioman Island, and Tanjung Sedili in Malaysia. As previously mentioned, the data is organized based on the best algorithm that be analyzed using Microsoft Azure tool.

TABLE II
DATA OF SEA LEVEL AT KERTEH

Mean Sea Level (mm)	Sea Surface Temperature (C)	Rainfall Amount (mm)	Mean Cloud Cover (Okta)
7230	26.60	474.6	7.27
7039	27.50	16.0	6.92
7067	28.15	201.2	6.97
7033	28.85	160.2	6.94
6930	29.95	122.0	7.09
6880	30.15	198.1	7.08
6836	29.55	274.2	7.01
6847	29.15	232.6	7.02
7057	29.27	269.6	7.13

The historical monthly mean sea level (MMSL) was obtained from the Department of Survey and Mapping Malaysia (DSMM), while the historical monthly rainfall data were obtained at a temporal resolution of 3h with a spatial resolution of a 0.25° latitude–longitude grid from the Tropical Rainfall Measuring Mission (TRMM) satellite. The mean cloud cover was obtained from the Malaysia Meteorological Department. The monthly sea surface temperature (SST) at a spatial grid resolution of 1.0° in latitude–longitude and a temporal resolution of one day were obtained from the website of the National Weather Service, Climate Prediction Centre of Oceanic and Atmospheric Administration [2].

TABLE III
DATA OF SEA LEVEL AT PULAU TIOMAN

Mean Sea Level (mm)	Sea Surface Temperature (C)	Rainfall Amount (mm)	Mean Cloud Cover (Okta)
7227	27.5	880.4	7.2
7227	27.3	16.4	6.9
7090	28.6	183.2	7.0
7052	29.8	37.0	6.7
6950	29.7	196.2	6.8
6911	29.9	176.3	6.9
6878	29.4	249.6	7.0
6880	29.0	214.6	7.0
7076	29.2	92.6	7.0

The comparison of prediction model performance was performed based on the Mean Absolute Error (MAE), Root Mean Square Errors (RMSE), Relative Absolute Error (RAE), Relative Squared Error (RSE), and Coefficient of Determination (R^2).

B. Algorithms

This study employed four regression algorithms that have not been tested in the literature, which are Decision Tree Regression, Decision Forest Regression, Linear Regression, and Bayesian Linear Regression.

TABLE IV
DATA OF SEA LEVEL AT TANJUNG SEDILI

Mean Sea Level (mm)	Sea Surface Temperature (C)	Rainfall Amount (mm)	Mean Cloud Cover (Okta)
7307	27.8	454.2	7.2
7054	27.5	23.9	6.9
7067	28.7	574.2	7.0
7034	29.9	40.8	6.7
6939	29.8	40.8	6.8
6907	29.9	174.3	6.9
6872	29.5	200.8	7.0
6881	29.1	40.8	7.0
7051	29.3	40.8	7.0

Decision Tree Analysis is a general, predictive modelling tool that has applications across a variety of areas. Decision trees are generally built through an algorithmic approach which finds ways to divide a set of data based on different conditions. It is a type of supervised learning that is commonly used for classification and regression tasks. The goal is to create a model that predicts the value of a class variable by learning from the features extracted from simple rules of judgment as shown in Eq. 1.

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i \tag{1}$$

In training, Decision Forests (DF) build many individual decision trees. Predictions from all trees are pooled to make the final prediction; class mode for classification or mean regression prediction. They are referred to as ensemble techniques, as they use a collection of results to make a final decision. This means a Decision Forest aggregates hundreds or thousands of individual Decision Trees and iteratively dividing the nodes in each tree due to limited number of features. The final prediction for DF is then performed by multiplying each tree's predictions as shown by Eq. 2.

$$H(x) = \arg\max \sum_{i=1}^{K} W_j I(h_j(x) = Y) \tag{2}$$

Next, Linear Regression is a type of regression analysis that focus on linear relationship between the independent (X) and the dependent (y) variables. The motive of the linear regression algorithm is to find the best values that would give the data points the best fit line. The cost function as shown in Eq. 3, where the model parameter, β, is multiplied with the the input matrix, X, with small value of error, ϵ, that are possibly caused by random sampling noise or latent variables. The output from a Linear Regression is one single point estimate for the "best" model parameters by minimizing the sum of squared errors given the actual data. The model parameters are then used to predict new data points.

$$y = \beta^T X + \epsilon \tag{3}$$

Bayesian Linear Regression is formulated using the probability distributions rather than point estimations in a Linear

Regression model. The response variable, y, is not estimated as a single value, but is derived from the probability distribution as shown in Eq. 4 [17]. The Bayesian Linear Regression model with the answer sampled from the normal distribution as shown in Eq. 5 has the objective to determine the posterior probability distribution for the model parameters, given the inputs, X, and outputs, y.

$$y \sim N(\beta^T X, \sigma^2) \tag{4}$$

$$P(\beta|y,x) = \frac{P(y|\beta,X) \times P(\beta|X)}{P(y|X)} \tag{5}$$

The transpose of the weight matrix multiplied by the predictor matrix will produce the mean value for a Linear Regression. Next, the standard deviation, ρ, multiplied by the identity matrix will produce the variance. Multiplication with the identity matrix is need because the model produced a multi-dimensional formulation.

C. Evaluation Metrics

In this project, the main measure is the performance between Boosted Decision Tree Regression (DTR), Decision Forest Regression (DFR) Linear Regression (LR) and Bayesian Linear Regression (BLR). The performance of the prediction is evaluated in terms of Root Mean Square Errors, Coefficient of Determination, Relative Squared Error, Relative Absolute Error and Mean Absolute Error. The formula for the evaluation measurements are described as follows.

Mean Absolute Error (MAE): MAE measures how similar the predicted values are to the actual values, therefore a lower score is better. The formula for MAE is shown in Eq. 6.

$$\text{MAE} = \frac{1}{n}\sum |y - y| \tag{6}$$

Root Mean Square Errors (RMSE): RMSE also measures how similar the predicted values are to the actual values, therefore a lower score is better. The formula for RMSE is shown in Eq. 7.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (MSL_p - MSL_0)^2}{N}} \tag{7}$$

. Relative Absolute Error (RAE): RAE is the value of relative absolute difference between the predicted against the real values. The formula for RAE is shown in Eq. 8.

$$\text{RAE} = \frac{[\sum_{i=1}^{n} (P_i - A_i)^2]^{\frac{1}{2}}}{X} \tag{8}$$

Relative Squared Error (RSE): RSE is the normalized value of the total squared error from the estimated values. It is calculated by dividing the real square error by the expected square error. The formula for RSE is shown in Eq. 9.

$$\text{RSE} = \frac{1}{n}\sum_{j=1}^{n} T_j \tag{9}$$

Coefficient of Determination (R^2): R^2 represents the predictive power of the model as between 0 and 1. Zero means that the configuration is random means explain nothing. 1 means that the match is fine. The formula for calculating R^2 is shown in Eq. 10.

$$R^2 = 1 - \frac{\text{MSE (model)}}{\text{MSE (baseline)}} \tag{10}$$

IV. Results and Discussion

This paper is set to compare the performance of various regression-based approachs, which are Decision Tree Regression (DTR), Decision Forest Regression (DFR), Linear Regression (LR), and Bayesian Linear Regression (BLR). The experiments were carried out using Azure Machine Learning, a suite of Machine Learning software involving different techniques that produce the results consisting of Mean Absolute Error (MAE), Root Mean Square Errors (RMSE), Relative Absolute Error (RAE), Relative Squared Error (RSE), and Coefficient of Determination (R^2).The experiments were conducted using data splitting method. The dataset that used after combine all three datasets that is dataset Kerteh, Tanjung Sedili and Pulau Tioman. Table V shows the prediction results based on the four regression algorithms.

TABLE V
Experimental Results

Metrics	Decision Tree Regression	Decision Forest Regression	Linear Regression	Bayesian Linear Regression
MAE	82.304	75.866	82.348	120.763
RMSE	104.135	96.07	101.25	158.287
RAE	0.745	0.686	0.745	1.093
RSE	0.61	0.519	0.576	1.409
R^2	0.39	0.481	0.424	-0.409

Based on the table, for the DTR input design on the combination of the variable Mean Sea Level, Sea Surface Temperature, Rainfall, and Mean Cloud Cover, it is showed the highest Mean Absolute Error for Bayesian Linear Regression was 120.763mm and the lowest MAE for Decision Forest Regression is 75.866mm. Fig. 2 shows the comparison of results for MAE.

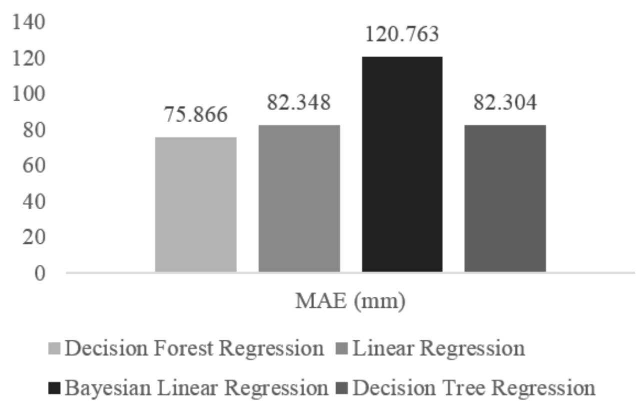

Fig. 2. Comparison for Mean Absolute Error

In terms of Root Mean Square Errors, Bayesian Linear Regression is the highest was 158.287mm and the lowest is Decision Forest Regression was 96.07mm as shown in Fig. 3. Both MAE and RAE are meant to measure the distance between the predicted values and the actual values. The difference is that the individual error in MAE are weighted equally in the average while the errors in RMSE are aquared before they are averaged. This means RMSE produce a relatively high weight to large errors. In the context of the sea level prediction, the smaller the error is the more desireable the result will be, therefore Decision Forest Regression is the best model when assessed based on MAE and RMSE.

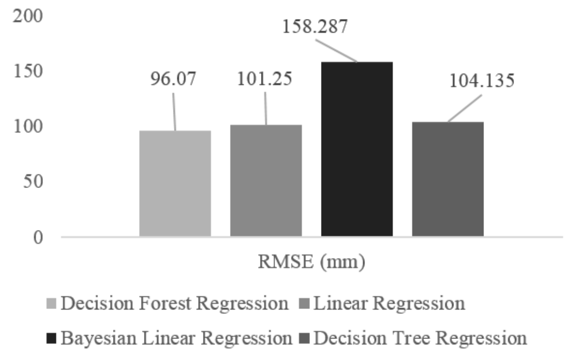

Fig. 3. Comparison for Root Mean Square Error

Relative Absolute Error, like the Relative Squared Error, ranges from 0 to infinite, being 0 the best value. For the RAE model performance, the highest is Bayesian Linear Regression with 1.093 and the lowest RAE is algorithms Decision Forest is 0.686. This is consistent with the results in terms of RSE, whereby highest results is 1.409, also from Bayesian Linear Regression and the lowest value is 0.519 from the Decision Forest Regression algorithm. This means the Decision Forest regression model is indeed the mest prediction model for sea level data, while Bayesian Linear Regression is not suitable for modeling this dataset. Fig. 4 shows the comparison of results for RAE while Fig 5 shows the comparison of results for RSE.

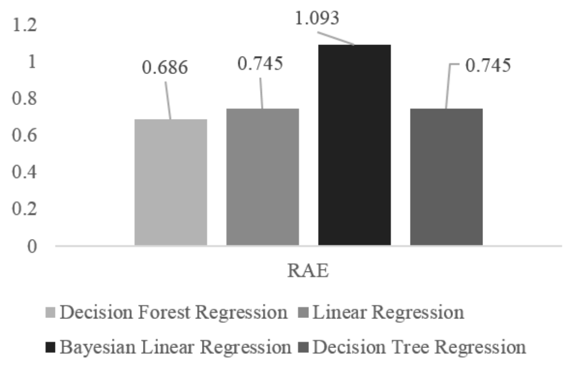

Fig. 4. Comparison for Relative Absolute Error

Recall that the measure of Coefficient of Determination (R^2) shows how well a regression model fits the observed data by comparing the fit of the regression model with that of a horizontal straight line, which is the null hypothesis. This means the higher the value of the R^2, the better fit the regression model is. Based on the results in Table V, the highest percentage among the regression models is the Decision Forest Regression with 48.1%, followed by Linear Regression and Decision Tree Regression. Note that the result for Bayesian Linear Regression is -0.409, which indicate the model is random. Being random means the

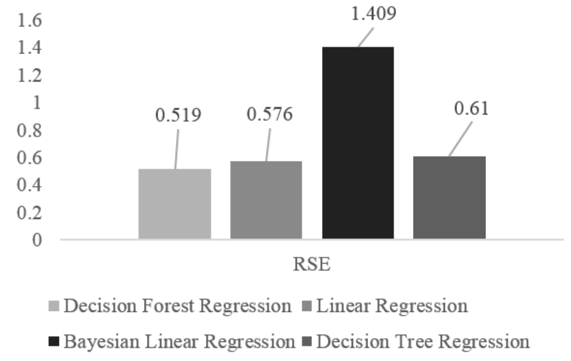

Fig. 5. Comparison for Relative Squared Error

chosen model does not follow the trend of the data, so the fit is worse than a horizontal line. Again, this result is consistent with the performance with measurement metrics used. The results for R^2 are shown in Fig. 6.

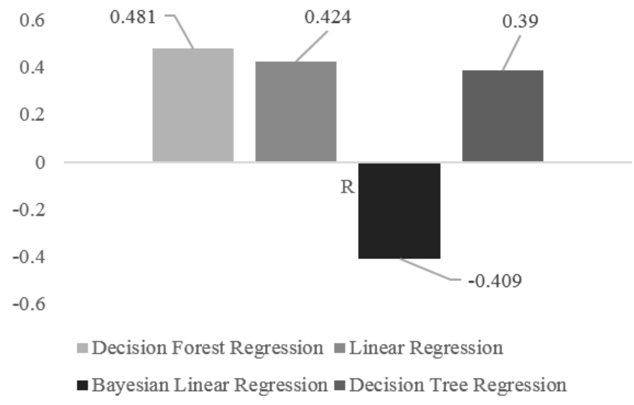

Fig. 6. Comparison for Coefficient of Determination (R^2)

Overall, the results indicated that the Decision Forest Regression constitutes suitable technique to predict the mean sea level prediction and obtained an effectively higher performance because all the average errors results get lowest value when compared with all the algorithms. Decision Forest are better than Decision Trees since decision boundary for Decision Forest is more accurate since it aggregates the results from multiple Decision Trees. This will lead to lower redistribution of error rate.

V. Conclusion

This mean sea level dataset has been used for prediction experiment for coastal area in Kerteh, Pulau Tioman and Tanjung Sedili. The main challenge in this study is the size of dataset and exploring a suitable algorithm for numerical dataset. The results showed that the best regression-based prediction algorithm is the Decision Forest Regression and the least effective algorithms for predict the mean sea level is Bayesian Linear Regression. The findings of this study is hoped to be useful in helping government and authorities to make an early strategy to handle any disastrous events.

References

[1] J. A. Church, N. J. White, C. M. Domingues, D. P. Monselesan, and E. R. Miles, "Sea-level and ocean heat-content change," in *International Geophysics*. Elsevier, 2013, vol. 103, pp. 697–725.

[2] V. Lai, A. N. Ahmed, M. A. Malek, H. Abdulmohsin Afan, R. K. Ibrahim, A. El-Shafie, and A. El-Shafie, "Modeling the nonlinearity of sea level oscillations in the malaysian coastal areas using machine learning algorithms," *Sustainability*, vol. 11, no. 17, p. 4643, 2019.

[3] S. Ehsan, R. Ara Begum, N. Ghani Md Nor, and K. Nizam Abdul Maulud, "Current and potential impacts of sea level rise in the coastal areas of malaysia," in *IOP Conference Series: Earth and Environmental Science*, vol. 228, no. 1. IOP Publishing, 2019, p. 012023.

[4] R. Nakano and S. Satoh, "Weak dependence on initialization in mixture of linear regressions," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, 2018.

[5] P. D. Michailidis, "A preliminary performance study on nonlinear regression models using the jaya optimisation algorithm," *IAENG International Journal of Applied Mathematics*, vol. 48, no. 4, pp. 424–428, 2018.

[6] P. I. Dalatu, A. Fitrianto, and A. Mustapha, "A comparative study of linear and nonlinear regression models for outlier detection," in *International Conference on Soft Computing and Data Mining*. Springer, 2016, pp. 316–326.

[7] T. L. Todelo and C. J. G. Aliac, "Predictability of earthquake occurrence using auto regressive integrated moving average (arima) model," in *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists*, 2019, pp. 13–15.

[8] J. T. Lalis, "A new multiclass classification method for objects with geometric attributes using simple linear regression," *International Journal of Computer Science*, vol. 43, no. 2, pp. 198–203, 2016.

[9] J. Feng, D. Li, T. Wang, Q. Liu, L. Deng, and L. Zhao, "Acceleration of the extreme sea level rise along the chinese coast," *Earth and Space Science*, vol. 6, no. 10, pp. 1942–1956, 2019.

[10] O. Makarynskyy, D. Makarynska, M. Kuhn, and W. Featherstone, "Predicting sea level variations with artificial neural networks at hillarys boat harbour, western australia," *Estuarine, Coastal and Shelf Science*, vol. 61, no. 2, pp. 351–360, 2004.

[11] E. Sertel, H. Cigizoglu, and D. Sanli, "Estimating daily mean sea level heights using artificial neural networks," *Journal of Coastal Research*, vol. 24, no. 3 (243), pp. 727–734, 2008.

[12] D. Makarynska and O. Makarynskyy, "Predicting sea-level variations at the cocos (keeling) islands with artificial neural networks," *Computers & Geosciences*, vol. 34, no. 12, pp. 1910–1917, 2008.

[13] O. Sahin, R. A. Stewart, G. Faivre, D. Ware, R. Tomlinson, and B. Mackey, "Spatial bayesian network for predicting sea level rise induced coastal erosion in a small pacific island," *Journal of environmental management*, vol. 238, pp. 341–351, 2019.

[14] S. Karimi, O. Kisi, J. Shiri, and O. Makarynskyy, "Neuro-fuzzy and neural network techniques for forecasting sea level in darwin harbor, australia," *Computers & Geosciences*, vol. 52, pp. 50–59, 2013.

[15] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. C.-D. Wirth, "1.0: step-by-step data mining guide. 2000. 78p," 2005.

[16] C. Shearer, "The crisp-dm model: the new blueprint for data mining," *Journal of data warehousing*, vol. 5, no. 4, pp. 13–22, 2000.

[17] G. Bernstein and D. R. Sheldon, "Differentially private bayesian linear regression," in *Advances in Neural Information Processing Systems*, 2019, pp. 525–535.