Jurnal Kejuruteraan 34(3) 2022: 475-484
https://doi.org/10.17576/jkukm-2022-34(3)-14

Predictive Analysis of Azure Machine Learning for the Rheological Behaviour of Unaged and Polymer Modified Bitumen

Abdalrhman Milad[a*], Sia Zhen Xian[b], Sayf A. Majeed[c], Abobaker G. F. Ahmeda[d], Munder Bilema[e] , Naeem Aziz Memon[f], Ahmed Elmesh[g], Atef Abu Salim[h], Qadir Bux alias Imran Latif[a] & Nur Izzi Md Yusoff[b]

[a]Department of Civil and Environmental Engineering, College of Engineering, University of Nizwa, P.O. Box 33, Nizwa PC 616, Ad-Dakhliyah, Oman
[b] Department of Civil Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Selangor, Malaysia,
[c] Technical Computer Engineering Department, Al-Hadba University College, 41001, Mosul, Nineveh, Iraq,
[d]Department of Civil and Structural Engineering, Higher Institute of Science and Technology Aljufra (HIST), Sokna, Libya
[e] Faculty of Civil Engineering and Built Environment, Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia
[f]Department of Civil Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan
[g]High Institute of Comprehensive Professions, Civil Engineering, Tripoli, Libya
[h]Department of Electrical and Computer Engineering, College of Engineering, University of Nizwa, P.O. Box 33, Nizwa PC 616, Ad-Dakhliyah, Oman

**Corresponding author: a.milad@unizwa.edu.om*

Received 1 May 2021, Received in revised form 28 July 2021
Accepted 15 September 2021, Available online 30 May 2022

ABSTRACT

Rheology can be defined as the primary measurement associated with bitumen flow and deformation characteristics. In the long term, DSR testing consumes a long time, expensive cost and skilled labour to operate equipment or machines in the laboratory. The complex modulus, G and phase angle, δ, are essential parameters for characterising and predicting the rheological behaviour of unaged bitumen (UB) and polymer-modified bitumen (PMB) in the model. This study developed three regression models using Azure machine learning (AML) to predict the rheological behaviour of UB and PMB. There are three types of data used as input data to develop the regression model: temperature, frequency, and modified material content. Regression models were developed with three processes or steps that need to be prioritised: data collection, model preparation, and model validation. Algorithms used in model development are decision tree regression (DFR), boosted decision tree regression (BDTR) and linear regression (LR). The results show G* and δ values. The R^2 values in the G* and δ predictions obtained from the DFR models are 0.8199 and 0.9480, respectively. Moreover, the R^2 values in the G* and δ predictions obtained from the LR models are 0.4219 and 0.7836, respectively.*

Keywords: Rheology; complex modulus, phase angle, δ and Azure machine learning, regression models

INTRODUCTION

Bitumen rheology can be defined as the primary measurement associated with bitumen flow and deformation characteristics. Therefore, understanding the flow and deformation (rheological properties) of bitumen in asphalt mixtures is vital to determining pavement performance (Yusoff 2011). Bitumen is a type of elastic material characterised by different loads, frequencies and temperature domains (Burger et al. 2001). The rheological properties of bitumen are measured using standard tests, including softening point, viscosity (at 65 and 135 ^{o}C), elastic recovery (at 25 ^{o}C using a ductilometers), storage stability (penetration and softening point) and thin-film oven test (softening point, viscosity, elastic recovery). Nevertheless, these measurements are insufficient to accurately describe the rheological behaviour, while these failures need to be linked to the bitumen rheological properties of the asphalt-mixing performance.

Through the Strategic Highway Research Program campaign, known as SHRP, dynamic shear ratios (DSR) have been introduced to characterise bitumen's rheological properties in the viscous elastic region (Airey 1997). Moreover, this DSR measurement is a highly complex instrument for determining various parameters such as complex modulus, G* and phase angle, δ. Thus, the output parameters generated by the DSR can be used to predict the main types of disturbances in the pavement, namely, the effects of rutting, fatigue and cracking.

DSR has restrictions on high frequency or low temperature and results in data being exposed to test errors from the rheometer (Yusoff 2011). In the long term, DSR testing consumes plenty of time, expensive cost and skilled labour to operate equipment or machines in the laboratory (Zeghal 2008). The model's use can be a valuable tool to describe the rheological bitumen binders and asphalt mixtures (Mohammad et al., 2005). Jongepier and Kuilman (1969) developed an empirical algebraic equation, a regular log, to predict rheology behaviour by predicting the bitumen relaxation spectrum. These models are based on algebraic equations where parameters are general and have no physical meaning. This makes it difficult to understand the rheological behaviour of bitumen (Behzadfar & Hatzikiriakos, 2013).

The data obtained analysed by recalculation procedure to obtain the elastic modulus values for each pavement layer where the elastic modulus value impact the life span (Khamis et al., 2018). Therefore, the analytical model offers a basic understanding of rheological behaviour and is more attractive to use. Oeser and Freitag (2009) applied artificial neural networks (ANN) to develop models to show the properties of asphalt and demonstrate that it can replace empirical rheological models. According to Negnevitsky (2005), ANN is part of the machine of learning involving appropriate mechanisms that enable computers to learn from experience, learn by example, and learn by analogy over time. Thus, AML is a proposed tool or method for predicting UB and PMB's rheological properties.

AML is a network platform running experiments with various algorithms (Mittal et al., 2021). It is straightforward to operate and provide much analysis. AML can build predictive analysis models by using data from one or more sources (Karthikeyan 2021). Using machine learning can save time, cost, and energy and obtain accurate data (Kelleher et al., 2015). This model is believed to assist in improving the quality material of bitumen binders. The results have also modified the performance of roads in the field of transportation. In order to enhance the optimal prediction for treatment techniques by AML techniques for flexible pavement maintenance in tropical regions to aid decision-maker in taking the right action for asphalt deterioration, Overall, this study provides a framework for integrating computational intelligence in AML modelling towards more effective, efficient, and reliable pavement maintained compared with previously available solutions (Milad et al., 2020).

Algorithms used in model development are DFR, BDTR and LR. According to Barga et al. (2015), LR is one of the oldest predictive techniques in statistics. It originates in the work of Carl Friedrich Gauss in 1795. LR fits a linear model between reaction and independent variables to predict the results of a set of observed variables. The formula with the simple LR model structure is shown in Equation (1).

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + \cdots \varepsilon \quad (1)$$

Y is the response variable (the outcome you are trying to predict). X_1, X_2. X_3, etc., are the independent variables used to predict the output. B_0 is a constant that is the intercept of the regression line B_1, B_2, B_3, etc., which are the coefficients of the independent variables. ε These refer to the partial slopes of each variable. is the error or noise associated with the response variable that the independent variables cannot explain X_1, X_2. X_3,. The LR model has two components: a deterministic portion $B_1X_1 + B_2X_2 + ...$ and a random portion (the error, e). These two components affect the result as the signal and noise in the model.

The BDTR is a combination of the decision tree (DT), whereas DT is built with a boosting method. Boosting is used to enhance accuracy. DT is a structure consisting of a root node and several other branches and leaves. The tree is a transverse root node at least one of the leaves, with the trail being determined by the solution to the question related to each subsequent node. Each tree is converted to the forest and classified by most trees within the application phase (Sjunnebo 2013). The DFR is generated by the DT combination, whereas DT is constructed using the bagging method. Bagging is used to reduce DT variance. Although DT is easy to interpret and intuitive, its predictions are inaccurate compared to other regressions such as DFR.

Furthermore, the tree structure is susceptible to the data provided. This explains that small changes in the data can have drastic effects on the structure of the tree. DFR is used to solve this problem. Different trees are designed to train different data and are randomly selected with replacements to make them more specific (Fazeli 2017). This research investigates the rheological behaviour of unaged bitumen and polymer-modified bitumen- by DSR testing consumes a long time, expensive cost and skilled labour to operate equipment or machines in the laboratory. Hence, to offer bitumen and paving engineers with machine learning. However, this study uses three regression models DFR, BDTR and LR, to predict the rheological behaviour of unaged bitumen (UB) and polymer-modified bitumen (PMB).

METHOD AND DATA PREPARATION FOR PREDICTION

This study involves only developing regression models to predict UB and PMB's rheological behaviour in terms of complex modulus, G* and phase angle, δ using existing data from experiments conducted in the laboratory. Azure learning engines are used to develop regression models such as DFR, BDTR and LR. The flowchart for developing the regression model is shown in Figure 1. It is frequently necessary to identify the specific variables required to develop the model in most regression tasks. The machine learning studio offers two feature selection modules to choose the most suitable variable for modelling purposes, such as linear discriminant analysis and feature selection based on the filter. Specified aims influence the choice of a suitable approach. Here, it is predictively represented by regression models to obtain the greatest accuracy and descriptive mining (Milad et al., 2020).

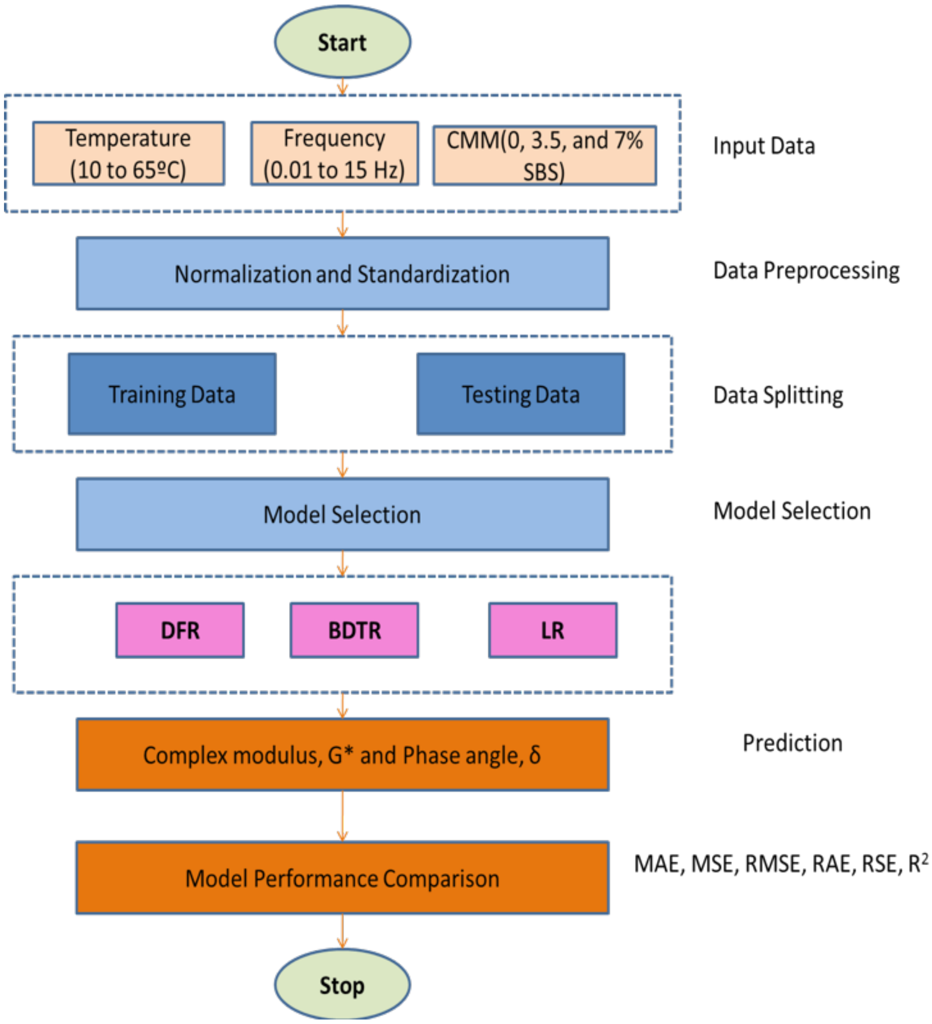

FIGURE 1. Flow Chart for Developing the Regression Model

The data used in this study are part of the data obtained from a study conducted by a group of researchers at Nottingham Transportation Engineering Centre, University of Nottingham. Additionally, data for the experiment and sample preparation and the target data are the results obtained from experiments, namely G* and δ values, used as the input data to develop the DFR, BDTR LR models regression. Input data and target data used to develop the regression model to facilitate the model's development. There are three types of data used as input data to develop the three regression models:

1. Temperature (10 to 65ºC)
2. Frequency (0.01 to 15 Hz)
3. Content of the modified material (CMM) (0, 3.5, and 7% SBS)

The total number of data sets used for each model is 98 sets (Figure 2). The data were divided into two sections. The data are 69 sets, about 70% will be used as training sets, and the remaining 29 sets of approximately 30% used as test sets. Two models are to be developed that produce one output for G* and another output for the value δ. Before the regression model is developed, the existing dataset needs to be rearranged to meet AML requirements. The rearranged dataset is used as inputs. Then, the format of the dataset must be changed from XLSX to CSV. After the data are uploaded to this experiment, the data dragged and dropped in the workplace. Select Columns in The Data Set are also selected and dragged, and dropped in the workplace. Two sections are connected by drawing a line from the data output to the input Select Columns in The Data Set. In the Select Columns in The Data Set, the data must be selected, such as temperature, frequency and phase angle, to perform the δ value prediction.

AML support data sets for two logical sets based on the preferred ratio. The split data option is dragged and dropped into the experiment. Then, the two parts are connected by drawing a line from the data output from the Select Column in The Data Set to the split data input. In split data, the row fraction is defined as 0.7, and the random seed is 12345. This indicates that the machine will randomly split the data into two sets: from the start point 12345 to move 70% of the data to the training area. The remaining 30% of the data will be devoted to testing the model in experiments.

The training model means teaching the model to evaluate the data and algorithms to perform the training task. The training model is dragged and dropped into the experimental workspace, and it connects the output part 1 of the split data to the input part 2 of the trained model. The phase angle as the prediction target is selected in the model training section. The DFR, BDTR and LR algorithms are selected to perform this experiment. Algorithms are the most widely used method in machine learning, such as mining data and statistics. The algorithm is dragged and dropped into the experimental workspace and connected to the input part 1 of the trained model. The default setting is used for these three regressions.

The AML depicts the score model by comparing the input data to data that generated thought model predictions. The scoring model is dragged and dropped into the experimental workspace. Input part 1 will be connected to the train model's output, whereas input part 2 will be connected to the output part 2 of split data. Finally, the evaluation model is dragged, dropped into the workspace, and links to the score model. Then, the 'run' button is pressed to evaluate until it is completed. The model validation process is necessary to ensure accuracy in the development of the model. In this study, the coefficient of determination, R^2, evaluates the developed model's accuracy. Meanwhile, the mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), relative absolute error (RAE) and relative square root error (RSE). Thus, as shown, AML to determine the most accurate regression model to predict the complex modulus, G* and phase angle, δ of UB and PMB.

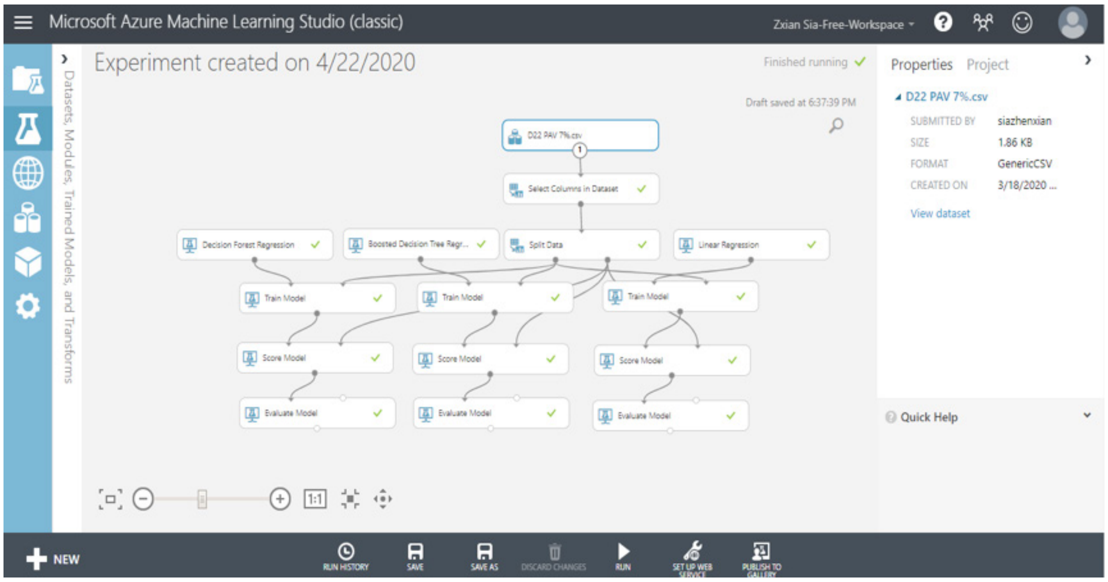

FIGURE 2. Rheological Behaviour Data Process in the Azure ML Studio

RESULT AND DISCUSSION

Three types of regression models developed in this study are DFR, BDTR, and LR. These models are developed to predict UB and PMB's rheological behaviour in complex modulus (G*) and phase angle (δ).

PREDICTION FOR COMPLEX MODULUS, G*VALUE

Thirty-six models are developed and divided into three groups. These three groups include Russian 80 penetration grade bitumen with unaged, Russian 80 penetration grade bitumen with short term ageing and Russian 80 penetration grade bitumen with long term ageing. All of these three groups contain styrene-butadiene-styrene (SBS) as a modification factor to modify bitumen quality. The percentages of SBS used are 0, 3, 5 and 7% in each model type. AML has been used to analyse the data provided to predict G* values.

UNAGED BITUMEN MODEL

DSR test is conducted on a Russian 80 penetration grade bitumen to obtain a UB dataset. This dataset is used to obtain the output value of G* by using the regression model. The entire bitumen dataset obtained will be trained, tested and validated. For the UB model, there are four different types of content material, namely 1) Russian 80 penetration grade bitumen, 2) 97% of Russian 80 penetration grade bitumen containing 3% SBS, 3) 95% Russian 80 penetration grade bitumen containing 5 % SBS, and 4) 93% of Russian 80 penetration grade bitumen containing with 7% SBS. From Table 1, it is found that the DFR model is the best because it has the highest R^2 and the lowest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen materials. Nonetheless, the LR model is flawed because it has the lowest R^2 values and the highest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen content materials.

TABLE 1. Result on the complex modulus value of the UB

Material	Algorithm	MAE	MSE	RMSE	RAE	RSE	R^2
Russian 80	DFR	1.27x10$_6$	7.53x10^{12}	2.74x10$_6$	0.34	0.18	0.8199
penetration grade	BDTR	1.38x10$_6$	9.36x10^{12}	3.06x10$_6$	0.37	0.22	0.7760
bitumen	LR	2.72x10$_6$	2.42x10^{13}	4.91x10$_6$	0.73	0.58	0.4219
97% Russian 80	DFR	1.07x10$_6$	5.71x10^{12}	2.39x10$_6$	0.29	0.15	0.8546
penetration grade	BDTR	1.32x10$_6$	8.81x10^{12}	2.97x10$_6$	0.36	0.22	0.7756
bitumen with 3% SBS	LR	2.66x10$_6$	2.27x10^{13}	4.76x10$_6$	0.72	0.58	0.4225
95% Russian 80	DFR	7.99x10$_5$	4.86x10^{12}	2.20x10$_6$	0.22	0.12	0.8766
penetration grade	BDTR	1.16x10$_6$	7.84x10^{12}	2.80x10$_6$	0.31	0.20	0.8808
bitumen with 5% SBS	LR	3.02x10$_6$	2.49x10^{13}	4.99x10$_6$	0.84	0.63	0.3682
93% Russian 80	DFR	8.01x10$_5$	2.90x10^{12}	1.71x10$_6$	0.33	0.18	0.8232
penetration grade	BDTR	9.67x10$_5$	3.27x10^{12}	1.81x10$_6$	0.39	0.20	0.8004
bitumen with 7% SBS	LR	1.71x10$_6$	8.85x10^{13}	2.98x10$_6$	0.70	0.54	0.4593

THE SHORT-TERM AGEING BITUMEN MODEL

DSR tests are conducted on RTFOT aged Russian 80 penetration grade bitumen to obtain short-term ageing bitumen datasets. This dataset is used to obtain the output value of G* by using the regression model. The entire bitumen dataset obtained trained, tested and validated. For short-term ageing bitumen models, there are four different types of content material, namely 1) RTFOT aged Russian 80 penetration grade bitumen, 2) 97% of RTFOT aged Russian 80 penetration grade bitumen containing 3% SBS, 3) 95% of RTFOT aged Russian 80 penetration grade bitumen containing with 5% SBS, and 4) 93% of RTFOT aged Russian 80 penetration grade bitumen containing with 7% SBS. From Table 2, it is found that the DFR model is the best because it has the highest R^2 values and the lowest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen materials. However, the LR model is inadequate because it has the lowest R^2 values and the highest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen content material.

TABLE 2. Result on the complex modulus value of the short-term ageing bitumen

Material	Algorithm	MAE	MSE	RMSE	RAE	RSE	R^2
RTFOT aged Russian 80	DFR	1.46x10$_6$	9.26x10^{12}	3.04x10$_6$	0.34	0.18	0.8174
penetration grade	BDTR	1.52x10$_6$	1.06x10^{13}	3.25x10$_6$	0.35	0.21	0.7916
bitumen	LR	3.05x10$_6$	2.84x10^{13}	5.33x10$_6$	0.70	0.56	0.4405
RTFOT aged 97% Russian 80	DFR	1.79x10$_6$	1.33x10^{13}	3.65x10$_6$	0.33	0.17	0.8268
penetration grade	BDTR	2.08x10$_6$	1.47x10^{13}	3.84x10$_6$	0.38	0.19	0.8086
bitumen with 3% SBS	LR	3.79x10$_6$	4.16x10^{13}	6.45x10$_6$	0.69	0.54	0.4595
RTFOT aged 95% Russian 80	DFR	1.57x10$_6$	9.99x10^{12}	3.16x10$_6$	0.32	0.17	0.8339
penetration grade	BDTR	1.84x10$_6$	1.23x10^{13}	3.35x10$_6$	0.38	0.19	0.8133
bitumen with 5% SBS	LR	3.35x10$_6$	3.21x10^{13}	5.67x10$_6$	0.69	0.53	0.4663
RTFOT aged 93% Russian 80	DFR	1.06x10$_6$	4.88x10^{12}	2.21x10$_6$	0.30	0.16	0.8379
penetration grade	BDTR	1.31x10$_6$	5.44x10^{12}	2.33x10$_6$	0.38	0.18	0.8195
bitumen with 7% SBS	LR	2.35x10$_6$	1.58x10^{13}	3.97x10$_6$	0.67	0.52	0.4757

LONG TERM AGEING BITUMEN MODEL

DSR tests are conducted on PAV aged Russian 80 penetration grade bitumen to obtain short-term ageing bitumen datasets. This dataset is used to obtain the output value of G* by using the regression model. The entire bitumen dataset obtained trained, tested and validated. For short-term ageing bitumen models, there are four different types of content material, namely 1) PAV aged Russian 80 penetration grade bitumen, 2) 97% of PAV aged Russian 80 penetration grade bitumen containing with 3% SBS, 3) 95% of PAV aged Russian 80 penetration grade bitumen containing with 5% SBS, and 4) 93% of PAV aged Russian 80 penetration grade bitumen containing with 7% SBS. From Table 3, it is found that the DFR model is the best because it has the highest R^2 values and the lowest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen materials. Nonetheless, the LR model is flawed because it has the lowest R^2 values and the highest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen content materials.

TABLE 3. Result on the complex modulus value of the long-term ageing bitumen

Material	Algorithm	MAE	MSE	RMSE	RAE	RSE	R^2
PAV aged Russian 80	DFR	2.41x10$_6$	2.29x10^{13}	4.79x10$_6$	0.27	0.13	0.8661
penetration grade	BDTR	3.14x10$_6$	2.68x10^{13}	5.17x10$_6$	0.36	0.16	0.8436
bitumen	LR	5.71x10$_6$	8.57x10^{13}	9.26x10$_6$	0.65	0.50	0.4993
PAV aged 97% Russian 80	DFR	1.89x10$_6$	1.37x10^{13}	3.70x10$_6$	0.25	0.12	0.8825
penetration grade	BDTR	2.75x10$_6$	1.92x10^{13}	4.39x10$_6$	0.37	0.17	0.8344
bitumen with 3% SBS	LR	4.68x10$_6$	5.72x10^{13}	7.56x10$_6$	0.63	0.49	0.5088
PAV aged 95% Russian 80	DFR	1.80x10$_6$	1.27x10^{13}	3.57x10$_6$	0.26	0.13	0.8721
penetration grade	BDTR	2.44x10$_6$	1.54x10^{13}	3.93x10$_6$	0.36	0.15	0.845
bitumen with 5% SBS	LR	4.32x10$_6$	4.87x10^{13}	6.98x10$_6$	0.63	0.49	0.5103
PAV aged 93% Russian 80	DFR	2.37x10$_6$	2.02x10^{13}	4.50x10$_6$	0.26	0.12	0.8778
penetration grade	BDTR	3.18x10$_6$	2.44x10^{13}	4.94x10$_6$	0.36	0.15	0.8529
bitumen with 7% SBS	LR	5.62x10$_6$	7.84x10^{13}	8.85x10$_6$	0.63	0.47	0.5268

PREDICTION FOR PHASE ANGLE, Δ VALUE

Thirty-six models are developed and divided into three groups. These three groups include Russian 80 penetration grade bitumen with unaged, Russian 80 penetration grade bitumen with short term ageing and Russian 80 penetration grade bitumen with long term ageing. All of the three groups contain styrene-SBS as a modification factor to modify bitumen quality. The percentages of SBS used are 0, 3, 5 and 7% in each model type. AML have been used to analyse the data provided to predict δ values.

UNAGED BITUMEN MODEL

DSR test is conducted on a Russian 80 penetration grade bitumen to obtain a UB dataset. This dataset is used to obtain the output value of δ by using the regression model. The entire bitumen dataset obtained will be trained, tested and validated. For the UB model, there are four different types of content material, namely 1) Russian 80 penetration grade bitumen, 2) 97% of Russian 80 penetration grade bitumen containing 3% SBS, 3) 95% Russian 80 penetration grade bitumen containing 5 % SBS, and 4) 93% of Russian 80 penetration grade bitumen containing with 7% SBS. From Table 4, it is found that the DFR model is the best due to the highest R^2 and the lowest MAE, MSE, RMSE, RAE and RSE values in all of the four types of bitumen materials. Nevertheless, the LR model is flawed because it has the lowest R^2 values and the highest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen content materials.

TABLE 4. Result on the phase angle value of the UB

Material	Algorithm	MAE	MSE	RMSE	RAE	RSE	R^2
Russian 80	DFR	2.41	9.82	3.15	0.21	0.05	0.948
penetration grade	BDTR	2.56	10.65	3.26	0.22	0.06	0.9442
bitumen	LR	4.86	41.32	6.43	0.42	0.22	0.7836
97% Russian 80	DFR	2.28	8.43	2.90	0.26	0.07	0.9317
penetration grade	BDTR	3.30	26.30	5.13	0.39	0.21	0.7868
bitumen with 3% SBS	LR	4.02	32.15	5.67	0.45	0.26	0.7394
95% Russian 80	DFR	2.39	11.62	3.41	0.32	0.14	0.8618
penetration grade	BDTR	3.02	13.76	3.71	0.40	0.16	0.8364
bitumen with 5% SBS	LR	5.38	48.84	6.99	0.71	0.58	0.4193
93% Russian 80	DFR	1.93	6.89	2.62	0.26	0.09	0.9076
penetration grade	BDTR	2.19	8.88	2.98	0.29	0.12	0.8809
bitumen with 7% SBS	LR	6.16	59.65	7.72	0.83	0.80	0.1997

THE SHORT-TERM AGEING BITUMEN MODEL

DSR tests are conducted on RTFOT aged Russian 80 penetration grade bitumen to obtain short-term ageing bitumen datasets. This dataset is used to obtain the output value of δ by using the regression model. The entire bitumen dataset obtained will be trained, tested and validated. For short-term ageing bitumen models, there are four different types of content material, namely 1) RTFOT aged Russian 80 penetration grade bitumen, 2) 97% of RTFOT aged Russian 80 penetration grade bitumen containing 3% SBS, 3) 95% of RTFOT aged Russian 80 penetration grade bitumen containing with 5% SBS, and 4) 93% of RTFOT aged Russian 80 penetration grade bitumen containing with 7% SBS.

From Table 5, it is found that the DFR model is the best because it has the highest R^2 values and the lowest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen materials. The LR model, however, is a flawed model due to the lowest R^2 values and the highest MAE, MSE, RMSE, RAE and RSE values in all four types of bitumen content material.

TABLE 5. Result on the phase angle value of the short-term ageing bitumen

Material	Algorithm	MAE	MSE	RMSE	RAE	RSE	R^2
RTFOT aged Russian 80	DFR	2.22	8.85	2.97	0.16	0.04	0.9646
penetration grade	BDTR	2.54	11.52	3.39	0.19	0.05	0.9540
bitumen	LR	4.91	40.76	6.38	0.36	0.16	0.8371
RTFOT aged 97% Russian 80	DFR	1.86	5.69	2.37	0.16	0.03	0.9698
penetration grade	BDTR	2.67	10.73	3.28	0.23	0.06	0.9423
bitumen with 3% SBS	LR	3.83	27.97	5.29	0.34	0.15	0.8496
RTFOT aged 95% Russian 80	DFR	1.89	5.43	2.33	0.18	0.03	0.9668
penetration grade	BDTR	2.26	8.28	2.88	0.22	0.05	0.9494
bitumen with 5% SBS	LR	3.51	22.78	4.74	0.34	0.14	0.8627
RTFOT aged 93% Russian 80	DFR	1.42	3.21	1.79	0.15	0.03	0.9749
penetration grade	BDTR	1.97	5.88	2.42	0.21	0.05	0.9541
bitumen with 7% SBS	LR	2.91	16.21	4.03	0.31	0.13	0.8734

LONG TERM AGEING BITUMEN MODEL

DSR tests are conducted on PAV aged Russian 80 penetration grade bitumen to obtain short-term ageing bitumen datasets. This dataset is used to obtain the output value of δ by using the regression model. The entire bitumen dataset obtained trained, tested and validated. For short-term ageing bitumen models, there are four different types of content material, namely 1) PAV aged Russian 80 penetration grade bitumen, 2) 97% of PAV aged Russian 80 penetration grade bitumen containing with 3% SBS, 3) 95% of PAV aged Russian 80 penetration grade bitumen containing with 5% SBS, and 4) 93% of PAV aged Russian 80 penetration grade bitumen containing with 7% SBS.

From Table 6, it is found that the DFR model is due to the highest R^2 values and the lowest MAE, MSE, RMSE, RAE and RSE values in all of the four types of bitumen materials. Nonetheless, the LR model is a flawed model due to the lowest R^2 values and the highest MAE, MSE, RMSE, RAE and RSE values in all of the four types of bitumen content materials.

TABLE 6. Result on the phase angle value of the long-term ageing bitumen

Material	Algorithm	MAE	MSE	RMSE	RAE	RSE	R^2
PAV aged Russian 80	DFR	2.37	7.84	2.80	0.14	0.02	0.9764
penetration grade	BDTR	2.43	9.74	3.12	0.15	0.03	0.9706
bitumen	LR	4.98	37.56	6.13	0.31	0.11	0.8868
PAV aged 97%	DFR	1.85	4.86	2.20	0.14	0.02	0.9806
Russian 80 penetration grade	BDTR	2.02	6.59	2.57	0.15	0.03	0.9737
bitumen with 3% SBS	LR	3.78	22.78	4.77	0.28	0.09	0.9089
PAV aged 95%	DFR	1.89	5.46	2.34	0.15	0.02	0.9757
Russian 80 penetration grade	BDTR	1.98	6.30	2.51	0.16	0.03	0.9720
bitumen with 5% SBS	LR	3.82	24.23	4.92	0.30	0.11	0.8923
PAV aged 93%	DFR	1.94	5.66	2.38	0.16	0.03	0.9720
Russian 80 penetration grade	BDTR	1.95	6.74	2.60	0.16	0.03	0.9666
bitumen with 7% SBS	LR	3.67	22.57	4.75	0.31	0.11	0.8882

DISCUSSION THE TYPES OF REGRESSION

Table 1 to 3 show that the predicted value is G*, whereas Table 4-6 show that the predicted value is δ. In both types of predictions, three types of regression models are used: the DFR, BDTR and LR models. The DFR and BDTR models can predict the rheological behaviour of UB and PMB with good result. Nonetheless, the prediction results of using LR are poor compared to the DFR and BDTR models. This can be demonstrated by the LR model having lower R^2 values and higher MAE, MSE, RMS, RAE and RSE values obtained from the G* and δ values predictions.

DFR and BDTR have similar properties because of the combination of the decision tree (DT), and they differ from LR. LR is the most widely used statistic method to predict continuous variable values because of their simple interpretation. Tree regression is a regression alternative that does not require predictions of the data to be analysed and is a simple method of interpreting results. Comparison of LR prediction level with tree regression is performed through simulation. Generally, the LR prediction error value is always lower than the tree regression when the LR model is positively correlated to the data. Nevertheless, the LR model is low correlated with the data and resulted in a lower regression tree error value than LR when it has a large data number (Díaz & Correa 2013).

Additionally, the LR model is a linear model that works well when data is linear properties. Nonetheless, linear models are limited to use when data are non-linear (Billings & Coca 1999). Tree regression is a regression allowing data to be non-linear and non-parameters (Gardner & Dorling 2000). Thus, large numbers of data and non-linear data lead to poor LR models. At the same time, tree regression not affected by these factors.

The DFR model obtained R^2, MAE, MSE, RMSE, RAE and RSE values, which are approximately the same as the BDTR model; however, the results obtained from the DFR are better than BDTR. This is evidenced by the DFR having higher R^2 values and lower MAE, MSE, RMSE, RAE and RSE values obtained from the G* and δ values predictions. Although the DFR and BDTR models comprise multiple decision trees (DT) using ensemble learning techniques, the two models differ in the method for predicting results. DT is a tree-based model for predicting the value of a target variable based on input variables, and a single DT has a significant variance in performance. This is because DT prediction data have slight bias and significant variance, overfitting problems and are difficult to use in general (Lee & Lee 2015).

Therefore, the DFR and BDTR models are developed by combining several DTs to solve the problem. DFR extracts data from random input data to build a single DT. Forming a single DT is repeated several times, and the final DFR model is determined based on the well-defined DT weights generated from the repeated processes. Since DFR adds random factors to sample variables, it maximises the advantages of ensemble learning techniques to produce high accuracy of predictions and classifications (Belgiu & Dragut 2016). BDTR extracts the data arbitrarily from the input data to build the DT. The DT formation process is repeated several times. In the sampling process, data that are not classified correctly in the previous DT formation process selected as the next step.

Accordingly, the difference between the two models is that BDTR should consider the model's performance during the previous DT generation process when extracting the sample and developing it. In contrast, the DFR does not deal with this problem (Lee et al., 2017). DFR works with the bagging method, and BDTR works with boosting method. Sometimes, poor boosting performance is caused by an overfitting training set (Freund & Schapire 1996). Simultaneously, the DFR can correct overfitting training sets (Lee et al., 2017). The DFR model can obtain higher prediction results compared to the BDTR model in this case.

COMPARISON AGAINST PAST LITERATURE RESEARCH

A study conducted by Alhamali in 2017 involved the engineering properties of bitumen modified polymers by nano-silica through experimentation and a model approach. In the study, the algorithms used to predict the values of G* and δ are LM, SCG and GDA. All three algorithms have the same architectural structure as [3 5 2], [3 7 2], [3 9 2] and [3 11 2]. Nevertheless, a structured algorithm [3 11 2] is used in this discussion. Table 7 shows the output values of G* and δ on a UB model. The materials used are Russian 80 penetration grade bitumen and modified bitumen polymers by nano-silica. DFR and LM [3-11-2] * give the best results in predicting G* and δ values, respectively. The DFR, however, gives poor results compared to the LM algorithm [3-11-2] *. In the G* prediction, the MAE value obtained from the DFR is 7.53x1012 Pa, while the LM algorithm [3-11-2] * gives 0.015655 Pa, and the percent difference value is 200%. The R^2 value obtained from the DFR is 0.8199, while the LM algorithm [3-11-2] * gives 0.97952, and the percent difference value is 17.7%. In the δ prediction, the MAE value obtained from the DFR is 9.82 Pa, whereas the LM algorithm [3-11-2] * gives 0.038406 Pa, and the percent difference value is 198%. The R^2 value obtained from the DFR is 0.9480, whereas the LM algorithm [3-11-2] * gives 0.95576 value, and the percent difference value is 0.8%. Table 8 shows the results of G* and δ output values for the short-term ageing bitumen model. The materials used are RTFOT aged Russian 80 penetration grade bitumen and RTFOF aged bitumen modified polymer by nano-silica. DFR and SCG [3-11-2] give the best results in predicting G* and δ values, respectively. The DFR yields better results than the SCG algorithm [3-11-2] in the G* value prediction, whereas the DFR produces better results than the SCG algorithm [3-11-2] δ value prediction. In the G* prediction, the MAE value obtained from the DFR is 9.26x1012 Pa, whereas the SCG algorithm [3-11-2] gives 0.03504 Pa, and the percent difference value is 200%. The R^2 value obtained from DFR is 0.8174, whereas the SCG algorithm [3-11-2] gives a 0.91667 value, and the percent difference value is 11.4%. In the δ prediction, the MAE value obtained from the DFR is 8.85 Pa, whereas the SCG algorithm [3-11-2] gives 0.034124 Pa, and the percent difference value is 192%. The R^2 value obtained from the DFR is 0.9646, whereas the SCG algorithm [3-11-2] gives 0.93176, and the percent difference value is 3.5%. Table 4.9 shows the results of G* and δ output values for long-term ageing bitumen models.

TABLE 7. Result on the complex modulus and phase angle value of the UB

Output	MSE	R^2
Prediction of complex modulus, *G** values		
DFR	7.53×10^{12}	0.81990
BDTR	9.36×10^{12}	0.77600
LR	2.42×10^{13}	0.42190
LM [3-11-2]*	0.015655	0.97952
SCG [3-11-2]	0.058060	0.71835
GDA [3-11-2]	0.033397	0.90681
Prediction of phase angle, δ values		
DFR	9.82	0.94800
BDTR	10.65	0.94420
LR	41.32	0.78360
LM [3-11-2]*	0.038406	0.95576
SCG [3-11-2]	0.053409	0.91445
GDA [3-11-2]	0.058828	0.89621

The materials used are RTFOT aged Russian 80 penetration grade bitumen and PAV aged modified bitumen polymer by nano-silica. DFR and LM [3-11-2] * give the best results in predicting G* and δ values. The DFR, However, yields poor results compared to the LM algorithm [3-11-2] *. In the G* prediction, the MAE value obtained from the DFR is 2.29x1013 Pa, whereas the LM algorithm [3-11-2] * gives 0.026374 Pa, and the percent difference value is 200%. The R^2 value obtained from the DFR is 0.8661, whereas the LM algorithm [3-11-2] * gives 0.94946, and the percent difference value is 9.2%. In the δ prediction, the MAE value obtained from the DFR is 7.84 Pa, while the LM algorithm [3-11-2] * gives 0.018643 Pa, and the percent difference value is 199%. The R^2 value obtained from the DFR is 0.9764, whereas the LM algorithm [3-11-2] * gives 0.98799. The percent difference value is 1.2%TABLE 9 Result on the complex modulus and phase angle value of the long-term ageing bitumen from the comparisons made, and it is found that the MAE values for this study are higher and R2 values are lower compared to those conducted by Alhamali (2017). The existence of lower MAE values and R2 values are better for a developed model.

TABLE 8. Result on the complex modulus and phase angle value of the short-term ageing bitumen

Output	MSE	R^2
Prediction of complex modulus, *G** values		
DFR	2.29×10^{13}	0.86610
BDTR	2.68×10^{13}	0.84360
LR	8.57×10^{13}	0.49930
LM [3-11-2]*	0.026374	0.94946
SCG [3-11-2]	0.044512	0.85603
GDA [3-11-2]	0.044860	0.85377
Prediction of phase angle, δ values		
DFR	7.84	0.97640
BDTR	9.74	0.97060
LR	37.56	0.88680
LM [3-11-2]*	0.018643	0.98799
SCG [3-11-2]	0.034766	0.95825
GDA [3-11-2]	0.035378	0.95677

CONCLUSION

In this study, opportunities for azure machine learning techniques for the rheological behaviour of unaged bitumen and polymer-modified bitumen have been examined. Thus, the three types of regression models, the DFR model, are the best due to their prediction data closest to the actual data. The DFR model shows the highest R^2 values in the UB model and short-term ageing bitumen and long-term ageing bitumen. Simultaneously, the LR model is less favourable than the other models because it has the lowest R^2 value in the standard bitumen model, short-term ageing bitumen and long-term ageing bitumen. This study has further strengthened the evidence that AML can solve various problems, especially predicting complicated things. Additionally, in empirical models developed in previous studies, the current study demonstrates the suitability of machine-learning methods; these models are weak but can be used as an alternative to predicting the mechanical properties of standard bitumen and PMB.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Universiti Kebangsaan Malaysia (UKM) for the financial support for this work (FGRS/1/2019/TK08/UKM/02/1 and GUP-2020-003).

DECLARATION OF COMPETING INTEREST

None

REFERENCES

Airey, G. D., 1997. Rheological Characteristics of Polymer Modified and Aged Bitumens. PhD Thesis, University of Nottingham.

Barga, R., Fontama, V. & Tok, W. H. 2015. *Predictive Analytics with Microsoft Azure Machine Learning. Predictive Analytics with Microsoft Azure Machine Learning*. Berkely, CA: Apress.

Behzadfar, E. & Hatzikiriakos, S. G. 2013. Viscoelastic properties and constitutive modelling of bitumen. *Fuel* 108: 391–399.

Belgiu, M. & Drăgu, L. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114: 24–31.

Billings, S. A. & Coca, D. 1999. Discrete wavelet models for identification and qualitative analysis of chaotic systems. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering* 9(7): 1263–1284.

Burger, A. F., Van De Ven, M. F. C., Muller, J. & Jenkins, K. J. 2001. Rheology of Polymer Modified Bitumen: A Comparative Study of Three Binders and Three Binder/Filler Systems Senior Scientist SASOL Oil Sasolburg. *20th South African Transport Conference* (July): 16–20.

Díaz Sepúlveda, J. F. & Correa, J. C. 2013. Comparación entre árboles de regresión CART y regresión lineal. *Comunicaciones en Estadística* 6(2): 175.

Fazeli, N. 2017. Machine Learning to Uncover Correlations Between Software Code Changes and Test Results. Master's thesis. University of Gothenburg.

Freund, Y., & Schapire, R.E. 1996. Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*, 148–156

Gardner, M. W. & Dorling, S. R. 2000. Statistical surface ozone models: An improved methodology to account for non-linear behaviour. *Atmospheric Environment* 34(1): 21–34

Jongepier R. & Kuilman B. 1969. Characteristics of the rheology of bitumens. *Journal of the Association of Asphalt Paving Technologists* 38: 98–122.

Karthikeyan, S. A. (2021). *Azure Well-Architected Framework: What and Why?. In Demystifying the Azure Well-Architected Framework* (pp. 1-13). Apress, Berkeley, CA.

Kelleher, J. D. & Arcy, A. D. 2015. FMLPDA_ Sample Chapter_ Information Based Learning 1–31

Khamis, N., Yazid, M. R. M., Hamim, A., Rosyidi, S. A. P., Yusoff, N. I. M. & Borhan, M. N. 2018. *Jurnal Teknologi Bitumen-Filler Mastic Using Artificial* 1: 71–78.

Lee, Saro & Lee, C. W. 2015. Application of decision-tree model to groundwater productivity-potential mapping. *Sustainability* (Switzerland) 7(10): 13416–13432.

Lee, Sunmin, Kim, J. C., Jung, H. S., Lee, M. J. & Lee, S. 2017. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk* 8(2): 1185–1203.

Milad, A., Yusoff, N.I.M., Majeed, S.A., Ibrahim, A.N.H., Hassan, M.A. and Ali, A.S.B., 2020, February. Using an Azure Machine Learning Approach for Flexible Pavement Maintenance. *In 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)* (pp. 146-150). IEEE.

Milad, A., Majeed, S.A. and Yusoff, N.I.M., 2020. Comparative Study of Utilising Neural Network and Response Surface Methodology for Flexible Pavement Maintenance Treatments. *Civil Engineering Journal* 6(10), pp.1895-1905.

Mittal, R., Arora, S., Kuchhal, P., & Bhatia, M. P. S. (2021). An Insight into Tool and Software Used in AI, Machine Learning and Data Analytics. AI and Machine Learning Paradigms for Health Monitoring System: *Intelligent Data Analytics*, 86, 45.

Mohammad, L.N., Wu, Z., Myres, L., Cooper, S. & Abadie, C. 2005. A practical look at simple performance tests: Louisiana's experience. *Journal of the Association of Asphalt Paving Technologists* 74: 557–600.

Negnevitsky, M. 2005. *Artificial Intelligence A Guide to Intelligent Systems*. Essex: Harlow Pearson Education.

Oeser, M. & Freitag, S. 2016. Fractional derivatives and recurrent neural networks in rheological modelling - Part I: Theory. International *Journal of Pavement Engineering* 17(2): 87–102.

Sjunnebo, J. 2013. Application of the Boosted Decision Tree Algorithm to Waveform Discrimination

Yusoff, N. I. M., Shaw, M. T. & Airey, G. D. 2011. Modelling the linear viscoelastic rheological properties of bituminous binders. *Construction and Building Materials* 25(5): 2171–2189.

Zeghal, M. 2008. Visco-elastic portrayal of bituminous materials: artificial neural network approach. *Geotechnical Special Publication* (178): 1041–1048.