The standard quadratic programming algorithm is used to obtain the optimal Lagrange multipliers; the regression function is rewritten as follows:

$$f(x) = \sum_{i=1}^{N_d} \alpha_i^* \varphi(x_i)^T \varphi(x) + b = \sum_{i=1}^{N_d} \alpha_i^* K(x_i, x) + b \quad (4)$$

Where:

$K(x_i, x)$: kernel function, which can simplify the mapping.

By using the kernel function, the data can be mapped implicitly into a feature space (i.e. without full knowledge of φ) [22]. Commonly used kernel functions include (1) linear kernel function; (2) polynomial kernel function; (3) radial basis kernel function; (4) Sigmoid kernel function and (5) spline kernel function.

α_i^*: The optimal Lagrange multipliers.

Radial basis function or RBF kernel is used in this research, it is a common kernel function used in various kernelized learning algorithms. In certain, it is commonly used in support vector machine classification [23]. The RBF kernel on two samples x_i and x_j, represented as feature vectors in some input space, is defined as [24]:

$$k(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2, \gamma > 0} \quad (5)$$

3.2 Radial Basis Function

A Radial Basis Function (RBF) neural network comprises three layers; an input layer, a hidden layer and an output layer as shown in Fig. 4. Hidden neurons use RBF as activation function. Different types of radial basis functions can be used, but the most convenient is the Gaussian function [15]. An RBF network positions one or more RBF neurons in the space described by the predictor variables. The space has as many dimensions according to the predictor variables. The Euclidean distance is calculated from the center of each neuron to the point being evaluated; by using a radial basis function for the distance to determine the weight for each neuron. The radial basis function is so named because the radius distance is the argument to the function. The node in the hidden layer is a p-multivariate Gaussian function, given as follows:

$$h_j(x) = exp\left[\frac{-1}{2\sigma_i^2} \|x - c_j\|^2\right], \quad j = 1,2, \dots, M \quad (6)$$

Where:

$\|.\|$: The Euclidean norm.

x: The input vector.

c_j: The center vector of the j^{th} hidden neuron.

σ: The width of the hidden neurons.

The best predicted value for the new point is found by summing the output values of the RBF functions multiplied by weights (w_i) computed for each neuron as shown below:

$$F(x) = \sum_{i=1}^{N} w_i(h_j(x)) \quad (7)$$

The training algorithm developed by Chen et al.(2005)[25] is used in this research. The optimal center points and spreads for each neuron are determined by using this algorithm. It also determines when to stop adding neurons to the network by monitoring the estimated leave-one-out (LOO) error and terminating when the LOO error beings to increase due to over fitting. Ridge regression is used for computation the optimal weights between the neurons in the hidden layer and the summation layer.

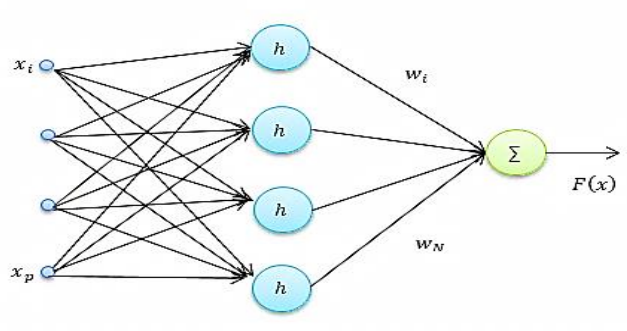

Figure 4: Basic Radial Basis Function Architecture.

Decision Tree Forest

Decision Tree Forest (DTF) is an ensemble (collection) of decision trees whose classifications or regressions is combined to make the overall classification or regression for the forest [26]. Decision Tree Forest models often can provide greater predictive efficiency and accuracy than single-tree models. In general, the more trees in the forest mean the forest is more robust. In the same way in the DTF classifier, The higher number of trees in the forest gives the high accuracy results. When developing DTF model, several parameters must be selected to adjust the behavior of this model, the first one is the number of tree in forest; this parameter specifies how many trees are to be constructed in the decision tree forest. It is recommended that a minimum value of 100 be used [15]. Another one parameter is minimum size node to split; a node in a tree in the forest will not be split if it has fewer than this number of rows in it. The third parameter is maximum tree levels; some research suggests that it is better to grow very large trees, therefore; a greater value should be set to the maximum tree levels and the minimum node size control would limit the size of the trees. The DTF algorithm to construct a decision tree forest is as follows [15].

i. Choose a random sample of N observations from the data set with replacement (this is called "bagging"). There are observations that can be selected again, while; others data will not be selected. By sampling process (2/3) of the rows will be selected. The data outside of the selection are called the "out of bag (OOB)" rows. A new random selection of rows is made for each constructed tree.

ii. Using the selected rows from step 1 for constructing a decision tree as shown in Fig. 5. The tree does not prune when it's constructed to the maximum size. Only

allow a subset of the sum of a predictors set of variables to consider possible splitters for each node. Select the set of predictors to be represented as a selected random subset of the total set. Carry out a new random selection for each split. The predictor (possibly the best one) excluded from one split may be used for another split in the same tree.

iii. Repeat steps i and ii (try again for several times to build a thick forest of trees).

iv. To "score" a row, run the row through each tree in the forest and record the predicted value (i.e., terminal node). Determining the average score predicted of the trees for a prediction analysis.

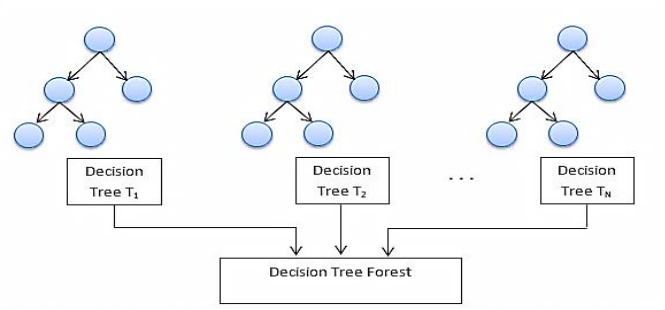

Figure 5: Basic Decision Tree Forest Architecture.

APPLICATION OF DATA MINING TECHINQUES

The daily averages of stage and discharge data for the Gharraf River were used in this present study for the period from January 2015 to December 2015. Three data mining techniques are used here, support vector machine (SVM), radial basis function (RBF) neural network, and decision tree forest (DTF). The performances of each model for both training and forecasting data are evaluated according to coefficient of correlation (R) (Eq. 8), root mean squared error (RMSE) (Eq. 9), and mean absolute percentage error (MAPE) (Eq. 10).

$$R = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2(y_i-\bar{y})^2}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2\sum_{i=1}^{N}(y_i-\bar{y})^2}} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i-x_i)^2}{N}} \quad (9)$$

$$MAPE = \frac{100}{N}\times\sum_{i=1}^{N}\left|\frac{x_i-y_i}{x_i}\right| \quad (10)$$

Where:

x_i : The observed values at the i^{th} time step.

y_i : The predicted values at the i^{th} time step.

N : The number of observations.

$\bar{x}$ and $\bar{y}$:The mean value of the observations and predictions, respectively.

Various combinations of river stage S_t with specified lag times (S_{t-1}) for antecedent stage for one day and the antecedent discharge for one day (Q_{t-1}) are examined as inputs to the data mining models to determine the evaluation of the effect of each input variables on target variable Q_t (t is current time). Three models for each data mining model with different inputs combination are employed. The information and input variables for these models are shown in Table 2.

Table 2: Different Inputs Combination of Models with Their Data Mining Techniques.

Model No.	Input variables	Target variable	Data mining technique
1	S_t	Q_t	SVM
2	S_t, S_{t-1}	Q_t	
3	S_t, Q_{t-1}	Q_t	
4	S_t	Q_t	RBF
5	S_t, S_{t-1}	Q_t	
6	S_t, Q_{t-1}	Q_t	
7	S_t	Q_t	DTF
8	S_t, S_{t-1}	Q_t	
9	S_t, Q_{t-1}	Q_t	

The observed data are usually subdivided into two parts: training and testing. Training data are used to determine the architectures of data mining models. The performance of the trained data mining models is then tested by the remaining data (i.e., testing data) that are not used in the training step. However, different selections of training and testing events may yield different results and sometimes lead to different conclusions. To overcome this problem and to reach the same conclusions, cross validations are conducted in this research [27]. When cross validation is used to evaluate the performance of a data mining model, a random set of rows is selected to each validation fold after stratifying on the target variable. When observations that are clustered in a small number of groups, in this situation a cross validation control variable is useful here.

SVM is used for modeling stage-discharge in the study area. By using a kernel function, SVM model transforms the input data into an n-dimensional space where a hyper plane can be constructed to partition the data. Radial Basis Function (RBF) is used in this research. There is no way in advance to find which kernel function would be better to apply, but the RBF function has been found to achieve the best job in most cases [15]. The RBF kernel non-linearly maps samples into a higher dimensional space, so that non-linear relationships between target groups and prediction features can be addressed. By applying SVM, stopping criteria (Epsilon) must be specified; this parameter is equal to (0.001), it is a tolerance factor that controls the iterative optimization process. The accuracy of an SVM model is depend on the choice of the model parameters such as C, γ, P, etc. For large values of C, the optimization will select a smaller-margin hyperplane if that hyperplane does a better job of getting all the training points classified correctly. There are two methods for finding optimal parameter values, a grid search and a pattern search. Grid and pattern search are

used here to obtain the optimal parameter values, once the grid search finishes, a pattern search is carried out over a narrow search region surrounding the best point that obtained by the grid search.

RBF neural network models by fitting Gaussian functions to the training data are applied for modeling stage-discharge. There are several parameters such as (maximum neurons, absolute tolerance, minimum radius, maximum radius, minimum lambda, maximum lambda, population size, max. generations, max. gen. flat, and boosting tolerance) must be specified for applying this model. The input values for these parameters are shown in Table 3. Extra neuron is not expiring when the training algorithm detects that over fitting may occur. The training algorithm stops when the residual mean squared error (MSE) is smaller than the absolute tolerance value. A larger radius for the training algorithm is required. If the validation error is significantly greater than the training error, in this case the increasing of the maximum radius value is required [15]. Lambda regularization parameters are used when computing weights that added to the network. During the training algorithm, a population size of candidate neurons is created with random centers and spreads (minimum and maximum radius). The population size parameter controls how many candidate neurons are created. It is suitable to increase the population size of candidate neurons if there are many predictor variables. DTREG used for constructing RBF neural networks an evolutionary method named Repeating Weighted Boosting Search (RWBS). Maximum generations controls the maximum number of generations of candidate neurons to be created by the RWBS evolutionary algorithm. Maximum generations flat controls the successive generations of RWBS. Candidate neurons are "mated" through each RWBS generation, and the refinement in estimated leave-one-out error is calculated. If the estimated error is smaller than the boosting tolerance value, the boosting operation stops and the next generation start.

Table 3: RBF Neural Network Parameters.

Parameter	value	Parameter	value
Maximum Neurons	100	Maximum Lambda	10
Absolute Tolerance	1×10^{-6}	Population Size	200
Minimum Radius	0.01	Maximum Generations	20
Maximum Radius	400	Maximum Generations Flat	5
Minimum Lambda	0.001	Boosting Tolerance	1×10^{-4}

There are three important parameters for constructing decision tree forest model (number of trees in forest, minimum size node to split, and maximum tree levels). The default values for these parameters are shown in Table 4. Number of trees in forest controls the number of constructed trees in the decision tree forest. In general, the more trees that be used for constructing (DTF) model lead to get better results, in contrast; the refinement in the results decreases with the increase in the number of trees after a specific number of tree. At a certain point the interest in regression performance from learning extra constructed trees will be lower than the cost of calculation time for learning theses additional trees. A node in a tree will not be split if it has value less the value of minimum size node to split. This is a suitable process to limit the growing of the tree. When a leaf contains too little predictors, further splitting will result in overfitting. Maximum tree levels control the maximum limits (depth) in which each tree can be planted in the forest.

Table 4: The Default Values of DTF Parameters.

Parameter	value
Number of Trees in Forest	200
Minimum Size Node to Split	2
Maximum Tree Levels	50

RESULTS AND DISCUSSION

Three data mining techniques (SVM, RBF neural network, and DTF) are used for stage-discharge modeling of Gharraf River, southern Iraq. SVM grid and pattern searches found optimal values for parameters as shown in Table 5. RBF neural network has an input layer, a hidden layer and an output layer. The node in the hidden layer is a p-multivariate Gaussian function whose outputs are inversely proportional to the distance from the center of the neuron. The optimum parameter values of RBF neural network models are presented in Table 6. Table 7 show the model size summary of DTF models.

Table 5: Optimum Value of SVM Parameters.

Model No.	Epsilon	C	Gamma	P
1	0.001	98.831	3.283	4.626
2	0.001	34.770	3.570	5.079
3	0.001	2111.455	3.233	0.220

Table 6: Optimum value of RBF Neural Network Parameters.

Model No.	Number of Neurons	Minimum Radius	Maximum Radius	Minimum Lambda	Maximum Lambda	Regularization Lambda for Final Weights
4	3	1.057	5.881	1.617	5.159	1.7204×10^{-4} after 5 iterations
5	3	1.057	373.853	0.412	5.159	8.2637×10^{-5} after 4 iterations
6	8	0.784	373.853	0.083	2.247	2.9889×10^{-8} after 4 iterations

Table 7: The Model Size Summary of DTF Models.

Model No.	Number of Trees in Forest	Maximum Depth of Any Tree in the Forest	Average Number of Group Splits in Each Tree
7	200	9	21.1
8	200	14	45.1
9	200	17	84.6

Table 6 show that the optimum number of neuron in model (3) is higher than other models (1 & 2), the number of neurons increased by increasing the number of input variables, also; the diversity of input variables (S_t, Q_{t-1}) is led to increase the number of neurons from 3 to 8. The maximum radius increased dramatically when using another input variables. From summary results of Table 7, the maximum depth of any tree in the forest and average number of group splits in each tree are increased when using two input variables, also; the diversity of input variables is affected clearly on these parameters.

Table 8 shows the results of applying three data mining techniques with regression accuracy results in validation stage. There is no significant improvement in the results of all models when adding the variable (S_{t-1}) to the input variable (S_t) as can be seen from the statistical results of models (2, 5, and 8). The lack for the improvement of results due to that the observations specially (stage variable) that are clustered in a small number of groups, in other words, the range of this data is very little, so that the difference between maximum value and minimum value is equal to 0.95. The best improvement occurred in the DTF models, increasing the model's performances by reducing RMSE 3.47% and increasing R^2 by 1.73% for models No. (8) Compared with model No. (7). DTF have two stochastic (randomizing) elements, the selection of the data type used as input for each tree, and a set of predictive variables that are considered a candidate for each node division. This randomization combined with the combination of predictions of trees greatly improves overall predictive accuracy. The input combination variables (S_t, Q_{t-1}) increased the model's performance by reducing RMSE (41.68%, 35.54% and 37.31%) and increasing R^2 (14.27%, 12.83% and 14.67%) for models No. (3, 6 and 9) respectively compared with the models (1, 2, 3) have one input variable (S_t). It is clear that the Q_t is mostly depended on the antecedent discharge values. Results illustrated that the SVM models are slightly better than RBF and DTF models. Figures (6 to 14) present the details of the predicted and actual target value of average daily stream flow for the developed models. A summary for all predicted results of developed models, data mining models could be used by the Iraqi Ministry of Water Resources to obtain results close to reality and to give an approximation of expected flow values of Gharraf River, southern Iraq, better than other high-cost models such as conceptual models.

Table 8: R, RMSE, and MAPE of Data Mining Techniques in Validation Stage.

Model No.	Technique	R	RMSE	MAPE
1	SVM	0.827	6.348	7.260
2		0.829	6.318	7.196
3		0.945	3.702	3.463
4	RBF	0.826	6.365	7.295
5		0.828	6.337	7.283
6		0.932	4.103	3.870
7	DTF	0.811	6.606	7.513
8		0.825	6.377	7.279
9		0.930	4.141	4.128

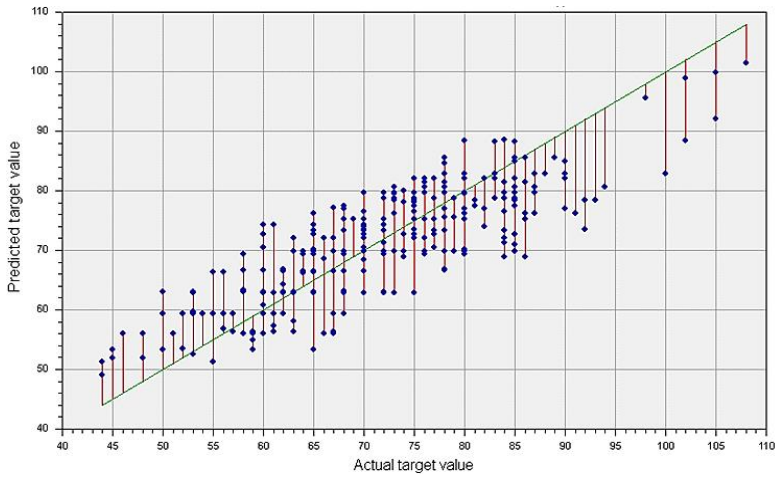

Figure 6: Predicted and Actual River Discharge for Model No. (1).

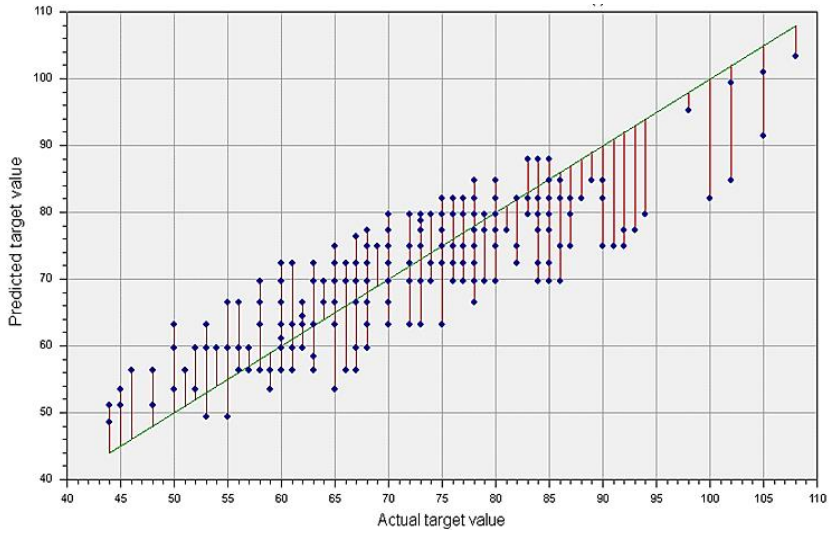

Figure 7: Predicted and Actual River Discharge for Model No. (2).

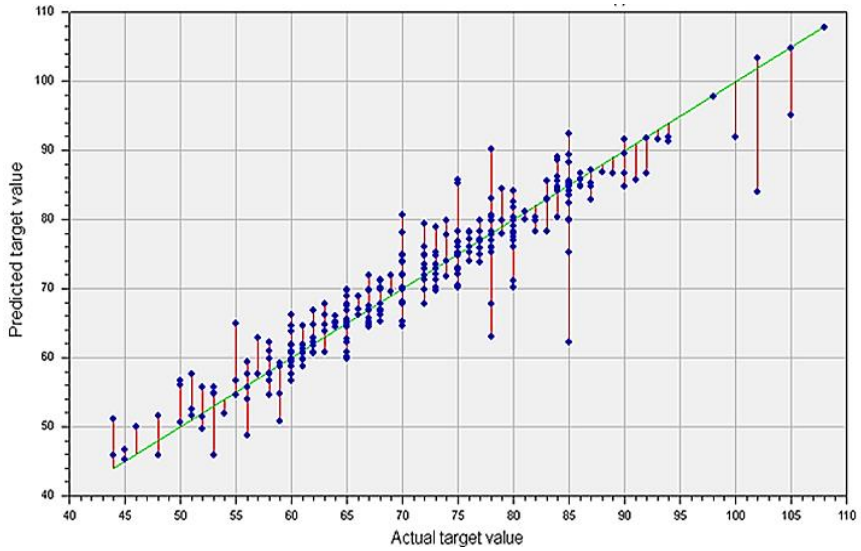

Figure 8: Predicted and Actual River Discharge for Model No. (3).

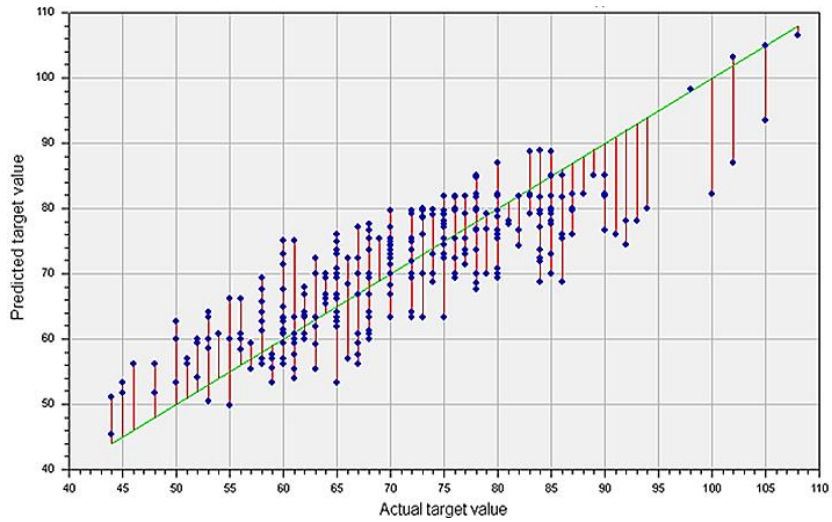

Figure 9: Predicted and Actual River Discharge for Model No. (4).

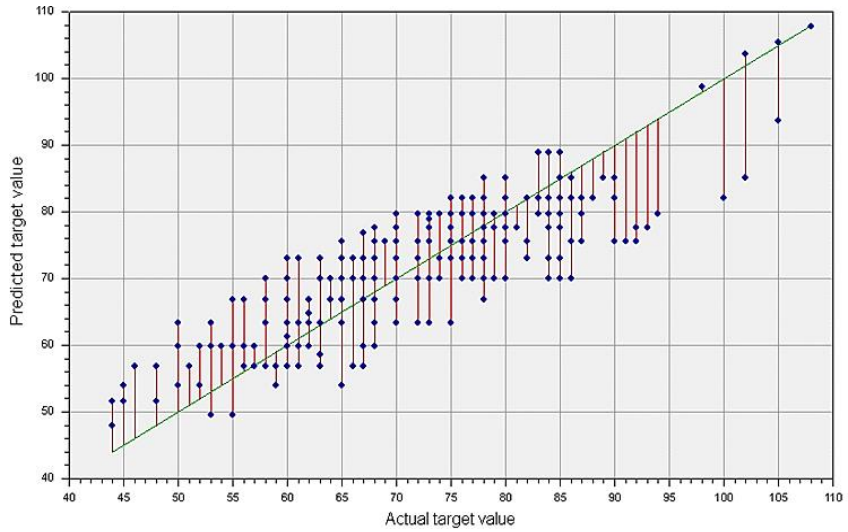

Figure 10: Predicted and Actual River Discharge for Model No. (5).

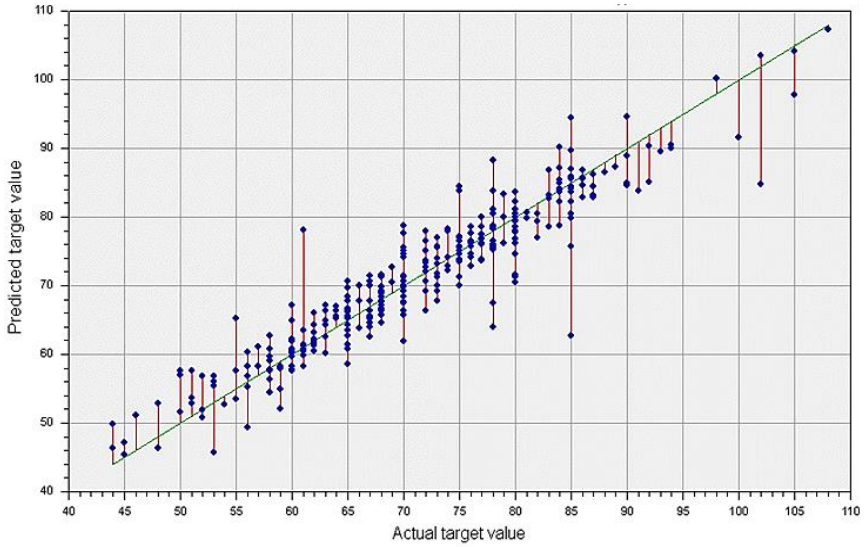

Figure 11: Predicted and Actual River Discharge for Model No. (6).

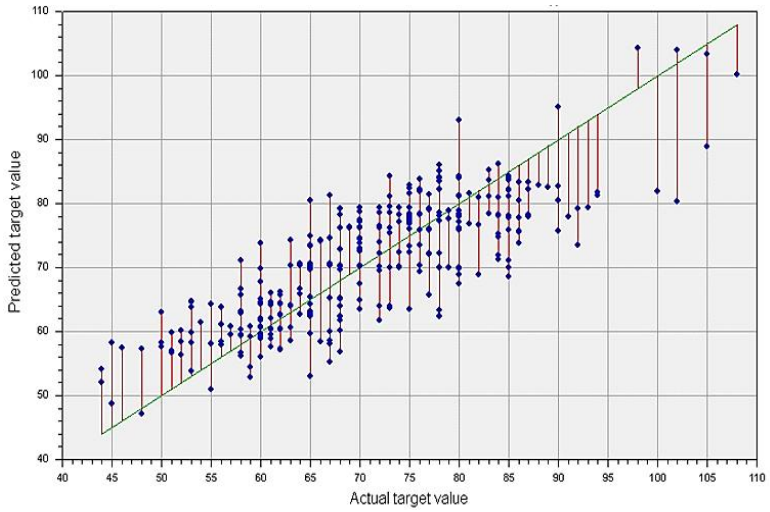

Figure 12: Predicted and Actual River Discharge for Model No. (7).

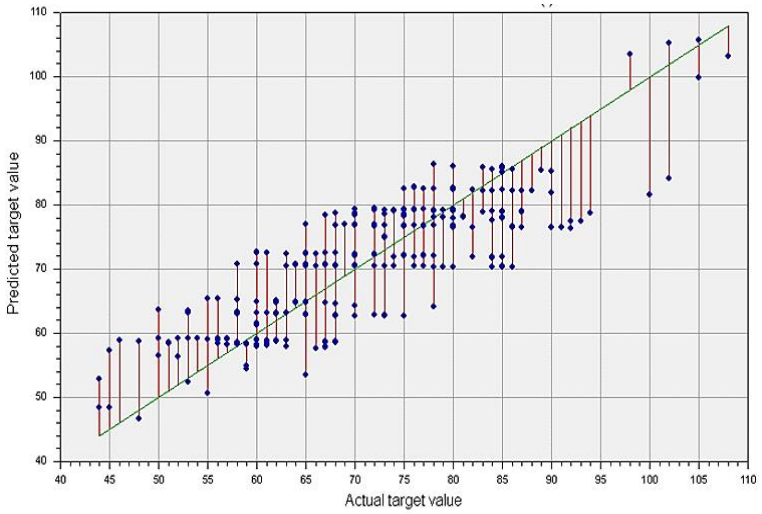

Figure 13: Predicted and Actual River Discharge for Model No. (8).

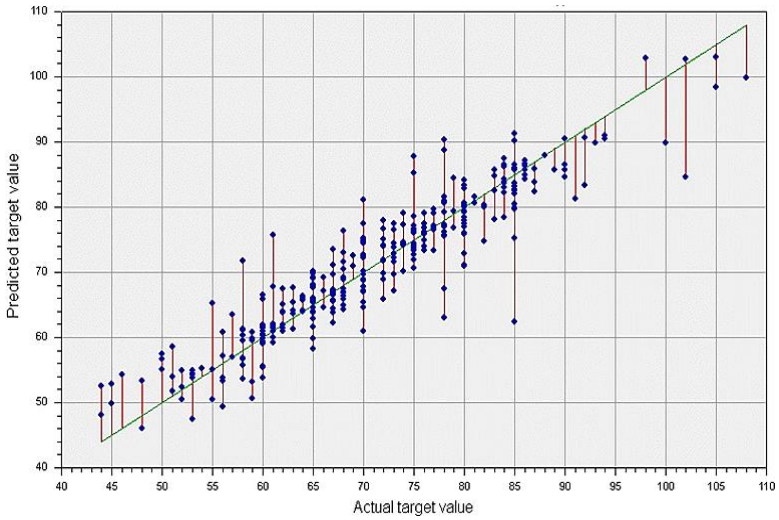

Figure 14: Predicted and Actual River Discharge for Model No. (9).

CONCLUSIONS

Three data mining techniques (support vector machine, radial basis function neural network, and decision tree forest) are presented for modeling the relationship between stage and discharge of Gharraf River, southern Iraq. The daily averages of stage and discharge data during the period from January 2015 to December 2015 were used in this present study. The performances of each model for both training and validation data are evaluated according to coefficient of correlation, root mean squared error, and mean absolute percentage error. Different combinations of river stage S_t with specified lag times (S_{t-1}) and the antecedent discharge (Q_{t-1}) are tested as inputs to the data mining models to evaluate the effect of each input variables on target variable Q_t. Cross validation method is used to evaluate the performance of the data mining models. It is a useful method when the data observations are clustered in a small number of groups such as stage of Gharraf River. There is no noticeable improvement in the results of all models when using the input combination variables (S_t, S_{t-1}) compared with the predicted results of models have only one input variable (S_t). Lack of improved predictability due to that the observations specially (stage variable) that are clustered in a small number of groups. The range of the observed stage river is very little compared to the observed discharge river. The input combination variables (S_t, Q_{t-1}) increased the model's performance by reducing RMSE (41.68%, 35.54% and 37.31%) and increasing R^2 (14.27%, 12.83% and 14.67%) for models No. (3, 6 and 9) respectively compared with the model's performance that have only one input variable (Model No. 1, 2, &3). It is clear that the Q_t is mostly depended on the antecedent discharge values. The performance of SVM models is slightly better than RBF and DTF models. Data mining techniques (SVM, RBF, and DTF) are powerful techniques for modeling stage-discharge relationship; these techniques could be used to obtain results close to reality and to give an approximation of expected flow from antecedent stage and discharge.

REFRENCES

[1] Tawfik, M.; Ibrahim, A.; Fahmy, H.: Hysteresis sensitive neural network for modeling rating curves. Journal of Computing in Civil Engineering. 11, 201-

211(1997).

[2] Bhattacharya, B.; Solomatine , D.P. : Application of neural network in stage discharge relationship. Proceedings of the international conference of hydroinformatics, Iowa, USA (2000).

[3] Bhattacharya, B.; Solomatine, D.P. : Neural networks and M5 model trees in modeling water level-discharge relationship for an Indian river. ESAN'2003 Proceedings-European Symposium on Artificial Neural Network Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 407-412(2003).

[4] Bhattacharya, B. ; Solomatine, D.P. : Neural networks and M5 model trees in modelling water level-discharge relationship. Neurocomputing.63, 381-396 (2005).

[5] Sudheer, K.P.; Jain, S.K. : Radial basis function neural network for modeling rating curves. Journal of Hydrologic Engineering. 8, 161-162 (2003).

[6] Lohani, A.K.; Goel, N.K.; Bhatia, K.K : Takagi-Sugeno fuzzy inference system form modeling stage-discharge relasionship. Journal of Hydrology .331, 146-160 (2006).

[7] Bisht, D.C.; Raju, M.M.; Joshi, M.C. : ANN Based River Stage – Discharge Modelling for Godavari River, India. Computer Modeling with New Technologies. 14, 48-62 (2010).

[8] Ghimire ,B.N.; Reddy, M.J. : Development of stage-discharge rating curve in river using genetic algorithm and model tree. International Workshop Advanced in Statistical Hydrology. Italy (2010).

[9] Ajmera, T.K.; Goyal, M.K.: Development of stage-discharge rating curve using model tree and neural networks: an application of Peachtree Creek in Atlanta. Expert Systems with Applications. 39, 5702-5710 (2012).

[10] Al-Aboodi A. H.; Dakheel A. A.; Ibrahim H. T.: Comparison of Data-Driven Modelling Techniques for Predicting River Flow in an Arid Region. International Journal of Applied Engineering Research. 12 (11), pp. 2647-2655 (2017).

[11] Dorado, J.; Rabunal, J.R.; Pazos, A.; Rivero, D.; Santos, A. ; Puertas, J. : Prediction and modelling of the rainfall-runoff transformation of a typical urban basin using ANN and GP. Applied Artificial Intelligence. 17(4), 329-343 (2003).

[12] Cheng, C.; Wu, X.; Chau, K.W.: Multiple criteria rainfall-runoff model calibration using a parallel genetic algorithm in a cluster of computers. Hydrological Sciences Journal. 50(6), 1069-1087 (2005).

[13] Hazi Mohammad Azamathulla.; Aminuddin Ab. Ghani ; Cheng Siang Leow ; Chun Kiat Chang ; Nor Azazi Zakaria: Gene-Expression Programming for the Development of a Stage-Discharge Curve of the Pahang River. Water Resource Management. 25, 2901–2916 (2011).

[14] Zahiri, A.R.; Eghbali, P.: Gene Expression Programming for Prediction of Flow Discharge in Compound Channels. Journal of Civil Engineering and Urbanism. 2 (4), 164-169 (2012).

[15] Phillip H., Sherrod : DTREG Predictive Modeling Software (2003).

[16] Al-Sahaf, M.: Iraq Hydrology and Hydrometry. Extended Abstract of Cand. Sci. (Geogr.) Dissertation, Moscow: Mosk. Gos. Univ (1965).

[17] U.S. Department of Agriculture: Fact Sheet-USDA at Work for Agriculture in Iraq (2009) at http://www.fas.usda.gov/icd/iraq/Iraqfactsheet.asp.

[18] Saleh, D.K. : Stream gage descriptions and stream flow statistics for sites in the Tigris and Euphrates Rivers Basins, Iraq. USGS data series 540:146 (2010).

[19] Atiaa, A.M. : Modeling of stage–discharge relationship for Gharraf river, southern Iraq by using data driven techniques. Water Utillity Journal . 9:31–46 (2015).

[20] Vapnik, V.N: The Nature of Statistical Learning Theory. Springer: Berlin, Germany,(1995).

[21] Ethem Alpaydin: Introduction to Machine Learning. MIT Press, Cambridge (2004).

[22] Jian-Yi Lin; Chun-Tian Cheng ; Kwok-Wing Chau: Using support vector machines for long-term discharge prediction. Hydrological Sciences Journal. 51 (4),599-612 (2006).

[23] Yin-Wen Chang; Cho-Jui Hsieh; Kai-Wei Chang; Michael Ringgaard ; Chih-Jen Lin : Training and testing low-degree polynomial data mappings via linear SVM. The Journal of Machine Learning Research 11(3):1471–1490 (2010).

[24] Bernhard Schölkopf; Koji Tsuda ; Jean-Philippe :Kernel Methods in Computational Biology (2004).

[25] Sheng Chen; Xia Hong; Chris, J. Harris : Orthogonal Forward Selection for Constructing the Radial Basis Function Network with Tunable Nodes. International Conference on Intelligent Computing. Part 1, LNCS 3644, pp.777-786 (2005).

[26] Hastie; Trevor; Tibshirani; Robert, Friedman; Jerome: The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Second Edition. Springer. ISBN 0-387-95284-5 (2008).

[27] Ming-Chang Wu; Gwo-Fong Lin: An Hourly Streamflow Forecasting Model Coupled with an Enforced Learning Strategy. Water . 7, 5876-5895 (2015).