Cash flow prediction using artificial neural network and GA-EDA optimization

Mohsen Sadegh Amalnik[a], Hossein Iranmanesh[a*], Atabak Asghari[a], Ali Mollajan[a], Vahed Fadakar[b] and Reza Daneshazarian[c]

[a]Department of Industrial Engineering, College of Engineering, University of Tehran(U.T), Tehran, Iran
[b]Faculty of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran
[c]Renewable Energy Department, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran

C H R O N I C L E	A B S T R A C T
	Cash flow models are one of the spotlights for evaluating a project. The actual data should be modeled then it could be used for the prediction process. In this paper, 996 airplane maintenance basis data are used as a database, and 119 similar data are chosen after clustering. The project is divided into 20 equal periods and first three periods are used for simulating the next point. The predicted data for each point is achieved by using of previous points from the beginning. The model is based on artificial neural network, and it is trained by three algorithms which are Genetic Algorithm (GA), Estimation of Distribution Algorithm (EDA), and hybrid GA-EDA method. Two dynamic ratios are used which are dividing the population into two halves, and the other is a ratio without dividing. The ratio would give a proportion to GA and EDA models in the hybrid algorithm, and then the hybrid algorithm could model the system more accurately. For each algorithm, three main errors are calculated which are mean absolute percentage error (MAPE), mean square error (MSE), and root means square error (RMSE). The best result is achieved for hybrid GA-EDA model without dividing the population and the MAPE, RMSE, and MSE values are %0.022, 28944.59 Dollars, and 837789503.79 Dollars, respectively.

Nomenclature

ACF	Actual data Cash Flow	NET	Overall input signal
e	Exponential function	PCF	Prediction Cash Flow
$EHNN$	Evolutionary Hybrid Neural Network	w	Weight
$Error$	The algorithm's error	x	Input neuron
$F(NET)$	The activation function	*Superscript*	
gen	Chromosome	EDA	Estimation of Distribution Algorithm
I. R	Incremental ratio	GA	Genetic Algorithm
M	Mid-point	*Subscript*	
$Mean\_error$	The average error	i	The period
MSE	Mean Square Error	tr	Train
N	The number of population	ts	Test

* Corresponding author. Tel.: +98-9123855616
E-mail address: h.iranmanesh@ut.ac.ir (H. Iranmanesh)

1. Introduction

One of the critical parameters in the designing of a system is considering its cash flow. More than 60% of the failures in the construction section is caused by economic factors (Russell, 1991). Cash flow model would have a significant influence on the project. Dynamic cash flow models could forecast the crucial parameters, and it would have an effect on the business plans. Almond and Remer (1979) presented sixth various cash flow models for an industrial economic applications. The two levels were modified, and it was shown that the cash flow model for the project level would be easier than company level. Chen et al. (2005) presented a cost-schedule integration (CSI) by combining pattern-matching logic and factorial experiments. They used the payment lags, separate tracking of material and payment frequency. Khosrowshahi and Kaka (2007) represented the cash flow model which included a mathematical model and estimating models. Their result showed that financial estimating models should be used for determination of the cash flow of the project. S curves are used for forecasting the cash flow, and it is the primary method in the cash flow projects (Touran, et al., 2004).

Artificial intelligence networks are used widely by the scholars. It is an alternative method for controlling the project costs. Cheng et al. (2009) used artificial intelligence methods for forecasting the cash flow models. They used K-mean clustering, genetic algorithm and artificial neural network (ANN). Blanc and Seltzer (2015) analyzed the cash flow for a corporation, and their results represented rectifiable biases. Also, they used debasing by employing selected statistical correction models for enhancing the estimating accuracy. Li et al. (2015) investigated the cash flow for South African firms which are listed on the Johannesburg Stock Exchange. According to their results, the inclusion of explanatory variable does not enhance the models, and they represented the application of moving the average model. Son et al. (2012) investigated the hybrid principal components analysis (PCA) and super vector machine (SVM) for the cash flow estimation in the commercial building projects. Their model was based on 84 data sets from commercial buildings. The hybrid PCA-SVM method was more accurate than single PAC and SVM methods. Kao et al. (2013) studied the cash flow for the stock price using nonlinear independent component analysis and SVM. They used two data sets including Japan, China, Shanghai Exchange Stock and Nikkei 225 stock indexes. Hwee and Tiong (2002) represented the cash flow model for contracting firms. They analyzed five various risk factors such as under/over measurement risk, duration, variation risk, and material cost variances, and estimated the cash flow considering these risks. Bec and Mogliani (2015) proposed a study for nowcasting French GDP using forecast combination and information pooling. The information pooling results were accurate regarding the forecast combination scheme. Ghysels and Ozkan (2015) estimated the US federal government budget. They used frequency data regression method for predicting the annual federal expenditures and receipts.

Artificial neural network is one of the prediction methods which are in the spotlight of the researchers of this field. Andrawis et al. (2011) introduced a new forecasting model for NN5 forecasting competition. The competition considered the estimation of 111-time series which represented the daily cash withdrawal values at ATMs. Xiong et al. (2013) performed a study on the crude oil price over long periods. They used a revised hybrid model empirical mode decomposition (EMD) and feed-forward neural network (FNN). They used the weekly report from West Texas Intermediate crude oil prices. Venkatesh et al. (2014) used the clustering and artificial neural networks for cash demand estimation in ATMs. After clustering the ATMs, they used four different neural networks on them which are general regression neural network (GRNN), multi-layer feed-forward neural network (MLFF), group method of data handling (GMDH), and wavelet neural network (WNN). The GRNN has the best outlet among the four networks with the yield of 18.44%. Cheng and Roy (2010, 2011) represented the cash flow model by time-dependent support vector machines (TDSVM). They used the evolutionary fuzzy decision model for the SVM method by focusing on the time series management. Cheng et al. (2010) developed a fuzzy neural network for improving the cash flow prediction. Namazi et al. (2016) analyzed the cash flow risk factors by using artificial neural network. Their results showed that the profitability is the highest risk, and the second and third elements are debt policy and company size.

In this paper genetic algorithm (GA), estimation distribution algorithm (EDA), and hybrid GA-EDA method are used for training the artificial neural network (ANN) to predict the cash flow model. The data of 996 airplane maintenance basis are used for the database. First, the data is clustered and the similar data is chosen. The 996 data were decreased to 119 similar data. Then this data is used for the algorithms - GA, EDA, and hybrid GA-EDA - which are used for training the ANN method. In the previous studies the data was divided into two halves, then they combined the results from different algorithms, but in the present study the whole data is investigated in the GA, EDA, and hybrid GA-EDA methods and the best result is chosen in every step. Using this approach would enhance the training path in the ANN method, and this functionalized method leads to precise results.

2. Methodology

The mixed method of the trained neural network and the hybrid genetic algorithm (GA) – estimation of distribution algorithm (EDA) is used for the forecasting the cash flow model. The clustering is used by the fuzzy C-means clustering method for preventing the data tolerances and increasing the accuracy then the final clustered data is utilized for the neural network input. The clustering has been done in three steps and 996 airplane maintenance bases construction projects by considering nine parameters; in this case, 119 projects are selected according to their similarity to training the neural network. This approach would lead to enhance the performance of the network, and the outliers are not used. After selecting the data which is used for training the neural network, the data is used as input for the hybrid GA-EDA method. The scheme of the proposed method is shown in Fig. 1.
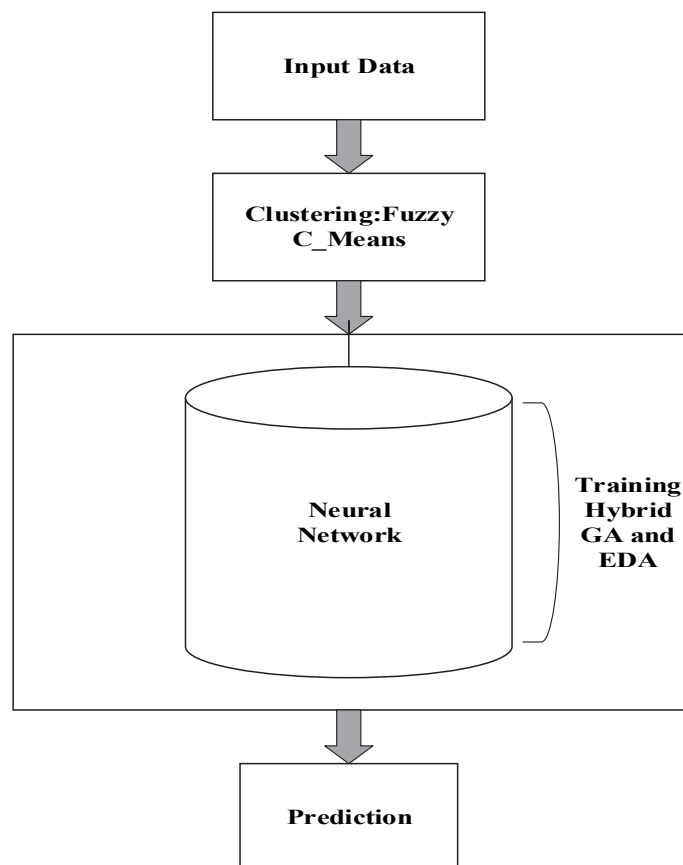

Fig. 1. Schematic of the prediction method

2.1. GA

Genetic algorithms are widely used in the optimization projects by the Darwinian fundamentals of natural selection and genetic replication (Holland & Goldberg, 1989) and this algorithm is heuristic

one. This algorithm includes an individual population and they represent a path to the results. The individuals play the chromosome role and they are series of bits. In the problem optimizations, the representations would be complicated. The results should show the entire search to examine. The size of the representation must be optimized so the performance of the genetic algorithm (GA) would decrease.

2.2. EDA

Estimation of distribution algorithms uses stochastic and heuristic search procedure (Larrañaga & Lozano, 2001; Mühlenbein, 1997). The main difference between the GA and EDA methods are in their search processes, GA uses this strategy in which the evolution and reproduction in one step would be finished up to next one and it would require probability distribution for the fittest individuals. Mutation operators would be ignored by this strategy and the required parameter for EDA would decrease. EDA use the following strategy which contains four fundamental levels:

a. The initial population P0 of N individuals are created at the start point.
b. A number of individuals are chosen and the fittest ones would be considered.
c. The *n*-parametric probabilistic graphical model explains the degrees of freedom and the reliance of the *n* variables.
d. The new population would have N new individuals and it would be obtained by modeling the probability learned distribution.

2.3. Hybrid GA-EDA

The aim of hybrid GA-EDA algorithm is to use the advantages of both Genetic Algorithm (GA) and Estimation of Distribution Algorithm (EDA). The EDA and GA algorithms do not have any absolute privilege on each other. The main difference between the methods is on the creating crossovers. The GA method selects the best population by considering crossover and mutation, on the other hand, the EDA process creates a probability distribution model then the new population would be available by reviewing the individual which has the best fittest with the model. In the hybrid model, a participation function should be used which employs the gained population from both EDA and GA models. In each step, this function identifies the fraction of the GA and the EDA algorithms. This population ratio could be dynamic or constant. The value of the participation function could be achieved by considering the following situations:

- Constant ratio: In this approach, the ratio has a specific value and it is given to each GA and EDA methods. It would not change in the next steps.
- Odd and even ratio: The even ratio would be given to one of EDA or GA algorithms, and the other algorithm would obtain the odd ratio.
- Incremental ratio: This ratio correlation is as follow:

$$\text{I. R.} = \frac{gen}{M + gen},$$
(1)

where; gen is the iteration number and the M is the mid-point. The use of mid-point is for preventing the growth of ratio be more than 50%. In this equation, the amount of ratio would be increased in each iteration.
- Dynamic ratio: In this method, the ratio between the GA and EDA algorithms would be obtained by considering the both algorithm average scores and each algorithm would gain its score.
- Exact value: This ratio is used to reduce the tolerance of the calculated predicted value. The calculated value for each step is different because as it shown in Eq. (23) this ratio is obtained by errors during the training period.

In this paper, two different kinds of the dynamic ratio are used with and without dividing the population. The dynamic ratio by dividing the population, the population is divided into half for both GA and EDA algorithms. The ratio would change in each step would be modified and improved. This process would be continued to the end of the process.

$$The\ EDA\ ratio\ of\ population = \frac{MSEtr_{i+3}^{EDA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}} \times population \tag{2}$$

$$The\ GA\ ratio\ of\ population = \frac{MSEtr_{i+3}^{GA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}} \times population \tag{3}$$

$$PCFtr_{i+3} = \left(\frac{MSEtr_{i+3}^{EDA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) * PCFtr_{i+3}^{GA} + \left(\frac{MSEtr_{i+3}^{GA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) * PCFtr_{i+3}^{EDA} \tag{4}$$

The second dynamic ratio is without the dividing the population, and in this approach, the population would be integrated, and each algorithm would advance the training of the ANN individually. At the end of the training session in each iteration, the effectiveness would be achieved from the artificial neural network coefficients.

$$PCFts_{i+3} = \left(\frac{MSEtr_{i+3}^{EDA}}{MSEtr_{i+3}^{EDA} + MSEts_{i+3}^{GA}}\right) * PCFts_{i+3}^{GA} + \left(\frac{MSEtr_{i+3}^{GA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) * PCFts_{i+3}^{EDA} \tag{5}$$

2.4. Training the ANN

For training the ANN the three algorithms as mentioned above - GA, EDA, and hybrid GA-EDA methods – could be used. The ratios which are obtained from the training would be used as final ratios in the ANN algorithm if each of the algorithms used individually. But in the hybrid algorithm, each model would have a fraction of the dynamic ratio which could be changed according to their accuracy. The mean square error (MSE) method has been used for the portion of each algorithm. After training the ANN model and achieving the optimum ratios, the next step would be prediction the new data according to their similarity with the ANN input data and put them in one cluster.

In this paper, the project is divided into 20 equal parts, and the cash flow model prediction would be achieved by using of artificial neural network method in these 20 parts. The cost of the project in three first periods (5%, 10%, and 15% physical progress) would be considered to the start point. The cost of these first periods is used for the estimation of the cost of 20% physical progress. In the next step, for prediction 25% point, the prior four points (5%, 10%, 15%, and 20%) are used. In each progress, the previous points from the beginning are considered in the calculations. It would lead to precise prediction results, and it would have lower errors in comparison with previous prediction models. The prediction algorithm is presented and i parameter is used as a counter for the cost of each part of the project:

$$PCF_{i+3}^{GA} = EHNN^{GA}(ACFtr_1, ACFtr_2, \dots, ACFtr_{i+2}) \tag{6}$$

$$PCFtr_{i+3}^{GA} = EHNN^{GA}(ACFtr_1, ACFtr_2, \dots, ACFtr_{i+2}) \tag{7}$$

$$PCFts_{i+3}^{GA} = EHNN^{GA}(ACFtr_1, ACFtr_2, \dots, ACFtr_{i+2}) \tag{8}$$

$$MSEtr_{i+3}^{GA} = (\sum_{i=1}^{tr}((ACFtr_{i+3} - PCFtr_{i+3}^{GA})/ACFtr_{i+3})^2)/N_{tr}) \tag{9}$$

$$Error\_PCFtr_{i+3}^{GA} = PCFtr_{i+3}^{GA} - ACFtr_{i+3} \tag{10}$$

$$PCF_{i+3}^{EDA} = EHNN^{EDA}(ACFtr_1, ACFtr_2, \dots, ACFtr_{i+2}) \tag{11}$$

$$PCFtr_{i+3}^{EDA} = EHNN^{EDA}(ACFtr_1, ACFtr_2, \dots, ACFtr_{i+2}) \tag{12}$$

$$PCFts_{i+3}^{EDA} = EHNN^{EDA}(ACFtr_1, ACFtr_2, \dots, ACFtr_{i+2}) \tag{13}$$

$$MSEtr_{i+3}^{EDA} = (\sum_{i=1}^{tr} ((ACFtr_{i+3} - PCFtr_{i+3}^{EDA})/ACFtr_{i+3})\text{^}2)/N_{tr}) \tag{14}$$

$$Error\_PCFtr_{i+3}^{EDA} = PCFtr_{i+3}^{EDA} - ACFtr_{i+3} \tag{15}$$

$$PCFtr_{i+3} = \left(\frac{MSEtr_{i+3}^{EDA}}{MSEtr_{i+3}^{EDA}+MSEtr_{i+3}^{GA}}\right) * PCFtr_{i+3}^{GA} + \left(\frac{MSEtr_{i+3}^{GA}}{MSEtr_{i+3}^{EDA}+MSEtr_{i+3}^{GA}}\right) * PCFtr_{i+3}^{EDA} \tag{16}$$

$$PCFts_{i+3} = \left(\frac{MSEtr_{i+3}^{EDA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) * PCFts_{i+3}^{GA} + \left(\frac{MSEtr_{i+3}^{GA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) \\ * PCFts_{i+3}^{EDA} \tag{17}$$

$$Error_{PCFtr_{i+3}} = \left(\frac{MSEtr_{i+3}^{EDA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) * Error_{PCFtr_{i+3}}^{GA} + \left(\frac{MSEtr_{i+3}^{GA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) \\ * Error_{PCFtr_{i+3}}^{EDA} \tag{18}$$

$$Mean\_error\_PECFtr_{i+3} = (\sum_{tr} Error\_PECFtr)/Ntr \tag{19}$$

$$PCF\_New\_Data_{i+3} = \left(\frac{MSEtr_{i+3}^{EDA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) * EHNN^{GA}(ACF_1^{new\ data}, ACF_2^{new\ data}, \dots, ACF_{i+2}^{new\ data})$$

$$+ \left(\frac{MSEtr_{i+3}^{GA}}{MSEtr_{i+3}^{EDA} + MSEtr_{i+3}^{GA}}\right) * EHNN^{EDA}(ACF_1^{new\ data}, ACF_2^{new\ data}, \dots, ACF_{i+2}^{new\ data}) \tag{20}$$

$$P\_ACF_{i+3} = PCF\_New\_Data_{i+3} + Mean\_error\_PCFtr_{i+3} \tag{21}$$

2.5. ANN

The artificial neural network moles are based on the human's biological neural system. The human's body neural system has single nodes which are neurons. Each of these neurons receives the information from the other neurons or external environment, then they process the data with an activation function, and a processes output would be provided for next neurons or nodes. This property of the information processing system in the neural networks makes them be a critical method for training from examples which could be used in the future case even not represented ones.

Assume that $x_1, x_2, \dots, x_n$ are the input neurons and the $w_1, w_2, \dots, w_n$ are their weights. These inputs with their weight could be used as vectors. The overall input signal is called as *NET* input and it defines as follows:

$$NET = \sum w_i x_i \tag{22}$$

The activation process on the *NET* would be done by the activation function (F) and the output signal would be as $F(NET)$. One of the activation functions which is used in this paper is sigmoid function and it is illustrated as follow equation:

$$F(NET) = \frac{1}{1 + e^{-net}}. \tag{23}$$

The other activation functions are hyperbolic tangential function($F(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$), sinusoidal/cosine function($F(x) = \sin x \ or \ F(x) = \cos x$), and liner function($F(x) = x$).
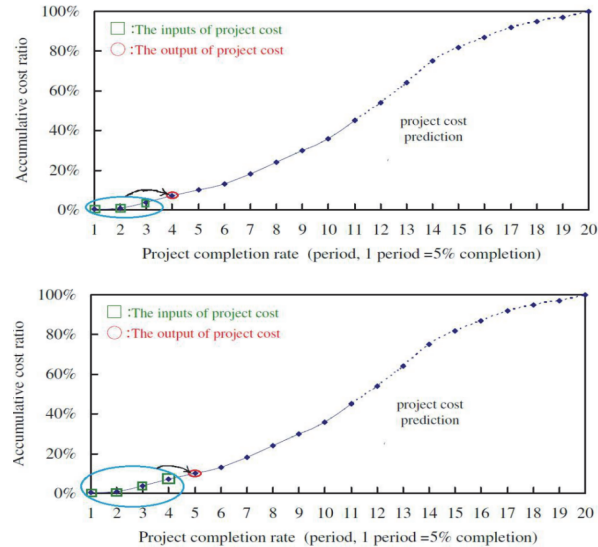
Fig. 2. The data input and output mapping in the cash flow

Table 1
The algorithm's parameters

Parameters	Values
No. of output neurons	1
No. of hidden layers neurons	10
Activation function slope	ICA
Crossover rate	0.9
Mutation rate for GA	0.05
Recome percent for GA	0.05
Population size	150
Iteration set	100
Max Generations	20
Learning Rate for EDA	0.3
Amount of mutation to affect the probability vector for EDA	10
Participation Function for GA-EDA	dynamic

3. Results and discussion

The cash flow model is carried out by four different methods for training the artificial neural network. The methods are the genetic algorithm (GA), estimation of distribution algorithm (EDA), hybrid GA-EDA model and the population is divided into these three methods, and the last one is hybrid GA-EDA without dividing the population. The error values are low and for better analyzing the accuracy of each method the logarithm value of the errors is shown in their related figures (Figs. 5, 8, 11, and 14)

3.1 Cash flow model by ANN method trained with the Genetic Algorithm

The cash flow estimation chart has two columns, the first one is the cash flow of the system and the second column is the hybrid algorithm with dividing the population. The cash flow prediction is shown in Fig. 3. The horizontal axis is the physical progress percentage of the project, and the vertical axis is the cost of the project. The mean absolute percentage error for the 17 periods is %0.040 and the mean absolute percentage error for the last period of the system is %0.046. The prediction error for the various percentage of physical progress is shown in Fig. 4. This point should be mentioned that each period error would have an effect on the other period's error. The calculated error for the first three periods is zero because these three points were the real values and they were used for the cash flow prediction. The logarithmic value of estimation errors is shown in Fig. 5.
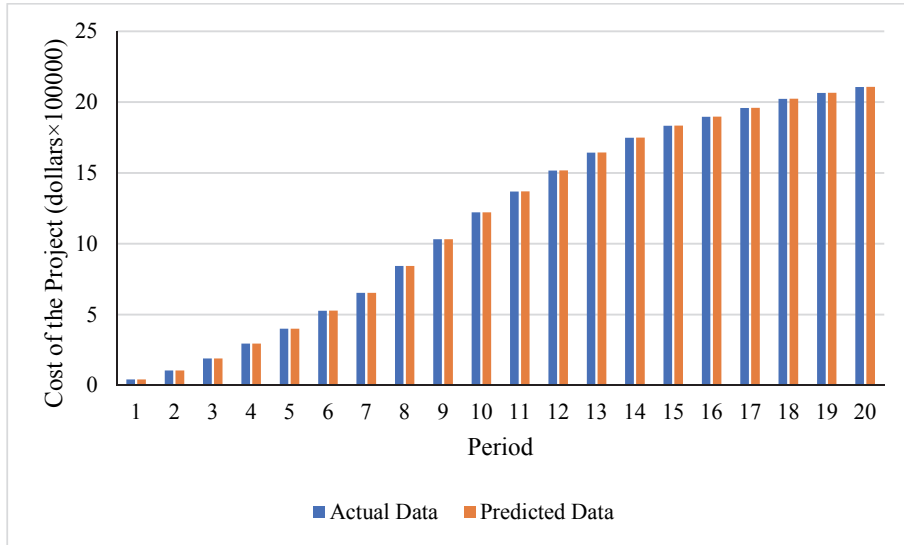
50

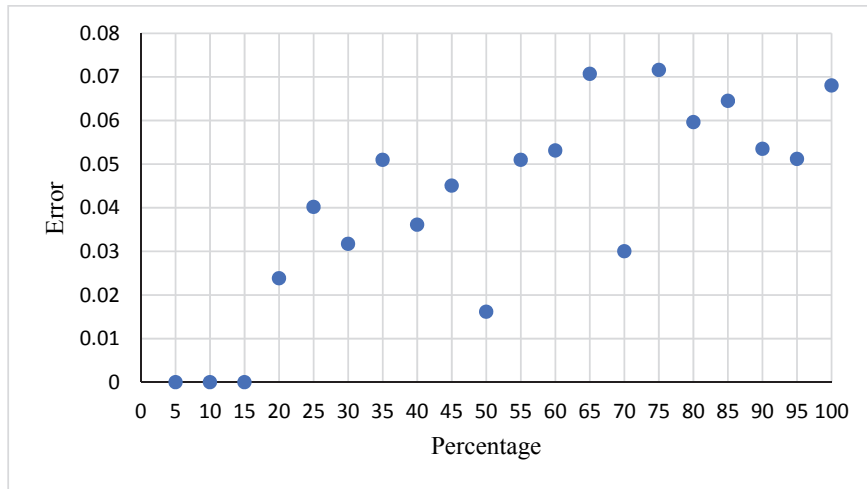**Fig. 3.** The cost of the project for each period using GA

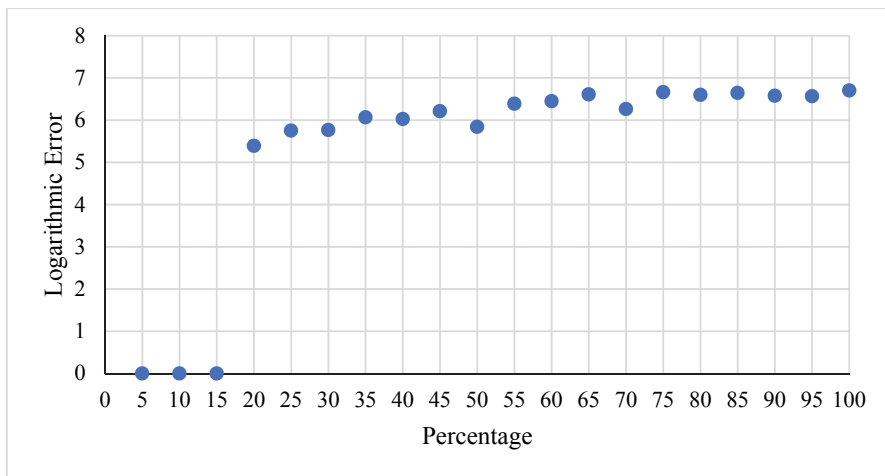**Fig. 4.** Cash flow prediction error using GA

Fig. 5. Logarithmic error of cash flow prediction using GA

3.2. Cash flow model by ANN method trained with EDA

The cash flow model is represented by estimation of distribution algorithm with dividing the population.
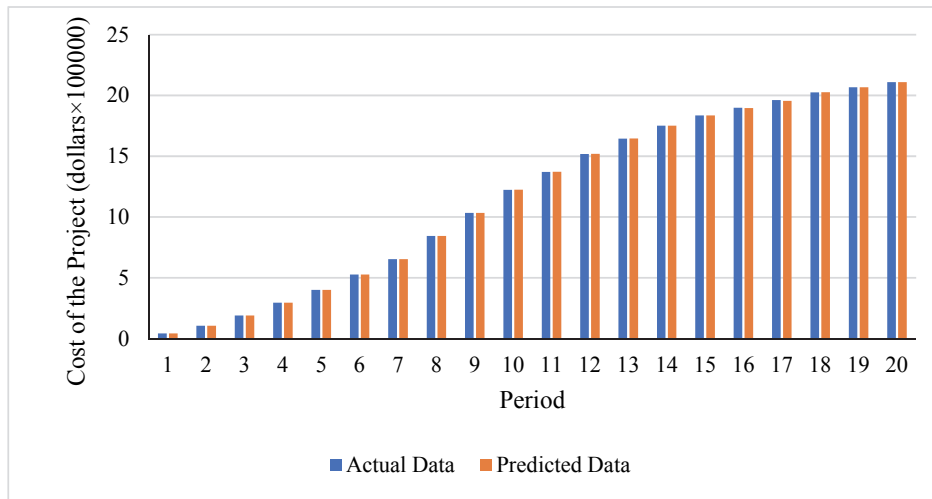

Fig. 6. The cost of the project for each period using EDA

In Fig. 6, the prediction of the cash flow and the real cost of the project is shown for 20 portions of the physical progress. The mean absolute percentage error for the last 17 points is %0.059 and for the whole project is %0.051. The mean absolute percentage error and the percentage of the physical progress of the project are shown in Fig. 7 and Fig. 8.
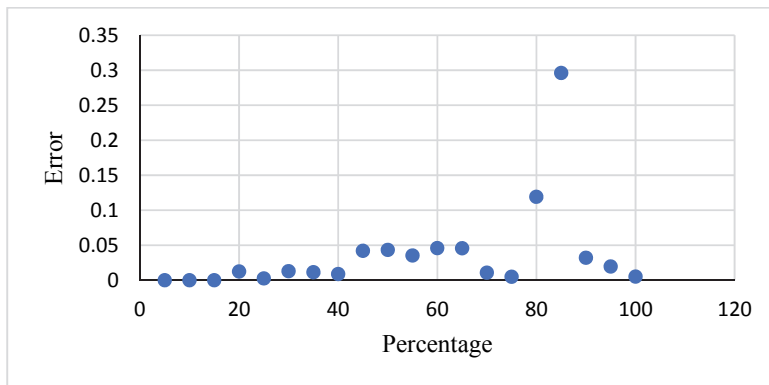

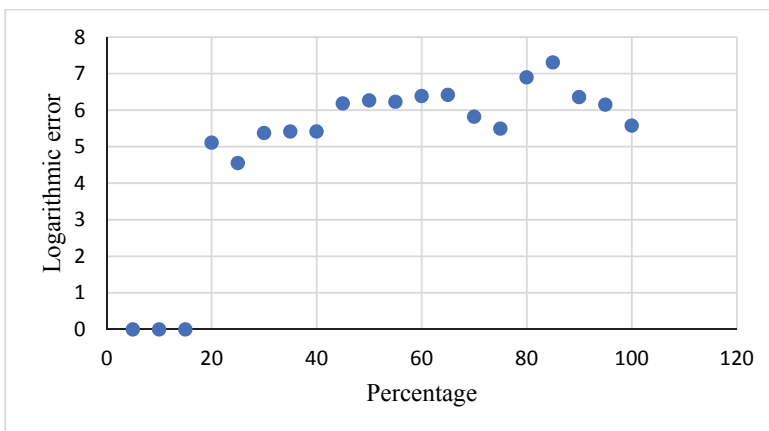**Fig. 7.** Cash flow prediction error using EDA

Fig. 8. Logarithmic error of cash flow prediction using EDA

3.3. Cash flow model by ANN method trained with hybrid GA-EDA with diving the population

In Fig. 9, the results of the hybrid GA-EDA prediction against and the cost of the project for 20 parts are shown. The population is divided and the average estimation error is %0.047 for the last 17 points and the mean error for the cost of the project is %0.049. The estimation errors are shown in Fig. 10. and the logarithmic value of the prediction errors are shown in Fig. 11.
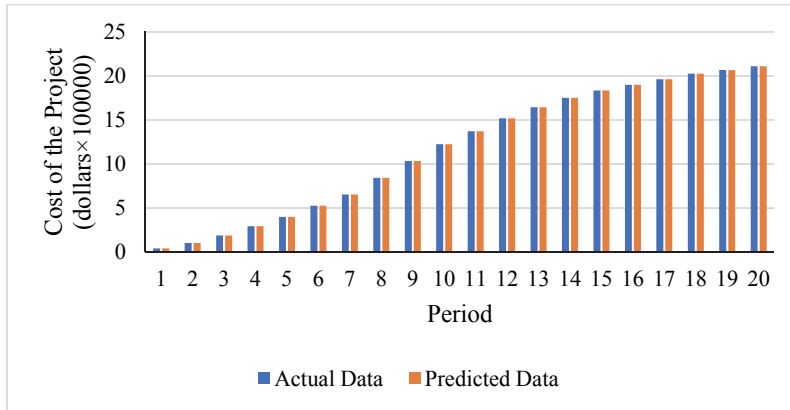

Fig. 9. The cost of the project for each period using hybrid GA-EDA with dividing the population
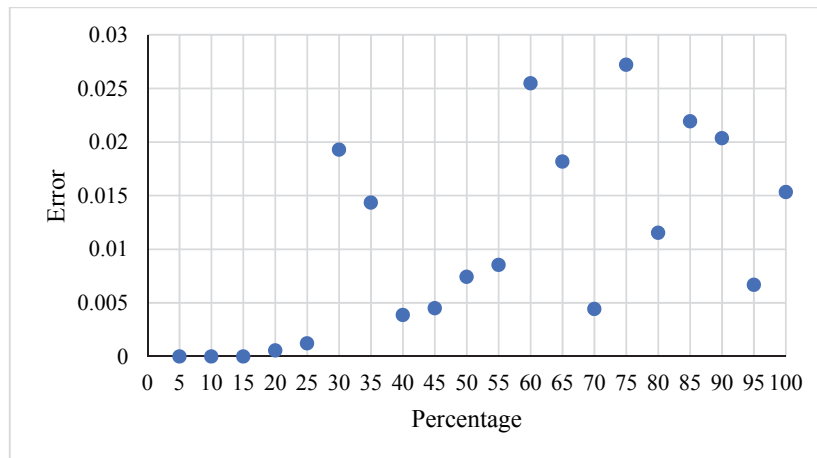

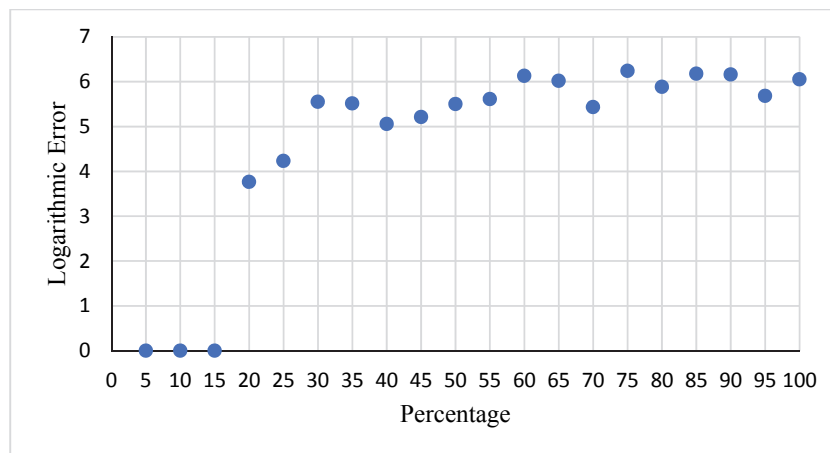**Fig. 10.** Cash flow prediction error using hybrid GA-EDA with dividing the population

Fig. 11. Logarithmic error of cash flow prediction using hybrid GA-EDA with dividing the population

3.4. Cash flow model by ANN method trained with hybrid GA-EDA without diving the population

The hybrid GA-EDA algorithm is used for the prediction of the cash flow model in this section. The population is being considered without dividing it. The results of the hybrid algorithm are shown in Fig. 12, for 20 studied percentages. The mean estimation error for the cost of the 17 points is 0.022 and for the whole project is %0.035. The estimation errors are shown in Fig. 13 and Fig. 14 Modeling the cash flow without dividing the population has the best results, and it would be more realistic for further studies.
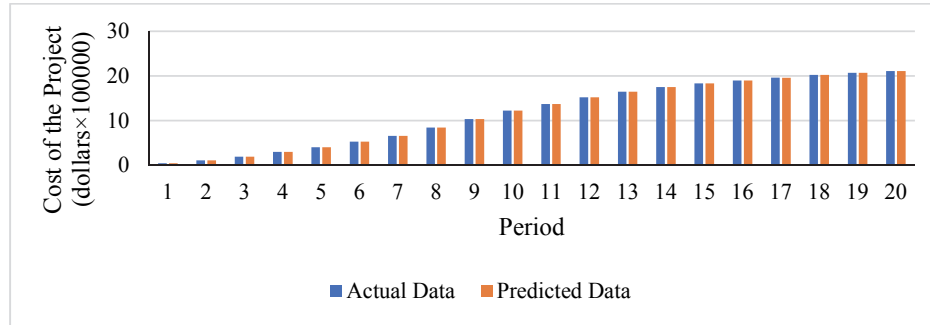

Fig. 12. The cost of the project for each period using hybrid GA-EDA without dividing the population
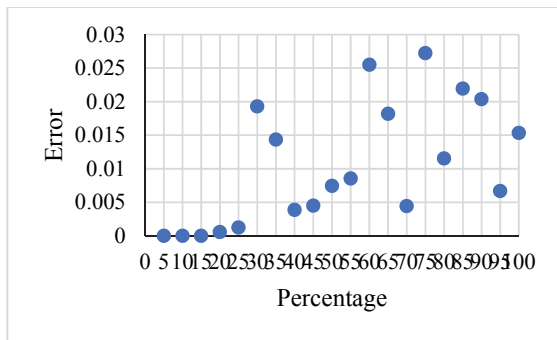

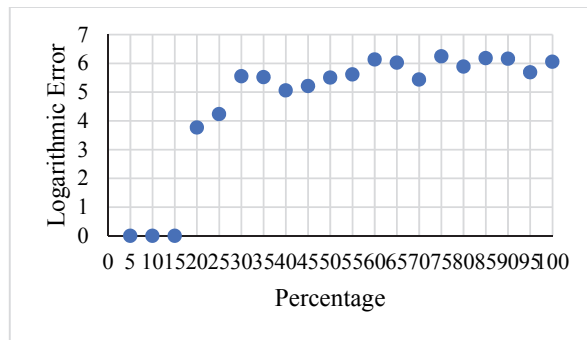**Fig. 13.** Cash flow prediction error using hybrid GA-EDA with dividing the population

Fig. 14. Logarithmic error of cash flow prediction using hybrid GA-EDA with dividing the population

The results of the hybrid GA-EDA trained ANN method without dividing the population has the best performance by considering three main criteria which are mean absolute percentage error (MAPE), mean square error (MSE), and root means square error (RMSE). The results of the ANN method which is trained by GA model are in the second-place due to the aforementioned three parameters. The results of the EDA trained ANN method to have the lowest accuracy among the four training methods. The results of the MAPE, MSE, and RMSE are given in Tables 2-7.

Table 2
The MAPE, MSE, and RMSE parameters for GA (without exact value) trained ANN

	ANN_GA(without exact value)	Criteria		
		MAPE	MSE	RMSE
Data test	Datatest_1	0.041679842	3546858110	59555.50445
	Datatest_2	0.142588934	45553927688	213433.6611
	Datatest_3	0.062991577	9680497589	98389.51971
	Datatest_4	0.189402537	94598447310	307568.6059
	Datatest_5	0.18182084	97847359773	312805.6262
	Datatest_6	0.085736957	22305525403	149350.3445
	Datatest_7	0.174533491	76739030392	277018.1048
	Datatest_8	0.136675119	49941710495	223476.4204
	Datatest_9	0.129224219	58885879172	242664.1283
	Datatest_10	0.101219845	31655972298	177921.2531
	Datatest_11	0.175507339	75365641334	274528.0338
	Datatest_12	0.187669016	87180517296	295263.471
	average	0.134087476	1.90546E+14	13803848.45

Table 3

The MAPE, MSE, and RMSE parameters for GA (with exact value) trained ANN

	ANN_GA(with exact value)	Criteria		
		MAPE	MSE	RMSE
Data test	Datatest 1	0.048114659	2519589418	50195.51194
	Datatest 2	0.007403851	103494579.7	10173.22858
	Datatest 3	0.022230126	565054432.4	23770.87361
	Datatest 4	0.054239375	2361580306	48596.09352
	Datatest 5	0.047425921	2540511016	50403.48218
	Datatest 6	0.012436251	144887765	12036.93337
	Datatest 7	0.049519164	2685424112	51821.07787
	Datatest 8	0.079470728	5316773570	72916.20924
	Datatest 9	0.06889683	4014650843	63361.27242
	Datatest 10	0.045902995	1801035851	42438.61274
	Datatest 11	0.033582089	1429116227	37803.65362
	Datatest 12	0.0077026	133401510.9	11549.95718
	average	0.039743716	1967959969	44361.69484

Table 4

The MAPE, MSE, and RMSE parameters for EDA (without exact value) trained ANN

	ANN_EDA(without exact value)	Criteria		
		MAPE	MSE	RMSE
Data test	Datatest 1	0.074715978	11861072824	108908.5526
	Datatest 2	0.188113879	45662887489	213688.7631
	Datatest 3	0.224645835	66517000913	257908.9004
	Datatest 4	0.177642177	51780089050	227552.3875
	Datatest 5	0.143838549	39034373908	197571.1869
	Datatest 6	0.181503874	48306741054	219787.9457
	Datatest 7	0.179232608	52902203598	230004.7904
	Datatest 8	0.196756652	63492911221	251977.9975
	Datatest 9	0.169921943	46958848640	216699.9046
	Datatest 10	0.159912574	48151808905	219435.2043
	Datatest 11	0.163669962	42709485414	206662.7335
	Datatest 12	0.163070495	34264400777	185106.458
	average	0.168585377	45970151983	214406.5111

Table 5

The MAPE, MSE, and RMSE parameters for EDA (with exact value) trained ANN

	ANN_EDA(with exact value)	Criteria		
		MAPE	MSE	RMSE
Data test	Datatest 1	0.043923647	8443428378	91888.12969
	Datatest 2	0.048931279	3009258475	54856.70857
	Datatest 3	0.108537175	30777099835	175434.0327
	Datatest 4	0.075667491	8174295083	90411.80832
	Datatest 5	0.048708481	2983159962	54618.3116
	Datatest 6	0.031619535	1229581347	35065.38674
	Datatest 7	0.042647331	4078202830	63860.80824
	Datatest 8	0.048100119	5818894399	76281.678
	Datatest 9	0.125652819	32045479813	179012.513
	Datatest 10	0.02707253	915203943.3	30252.33782
	Datatest 11	0.029127234	2365608157	48637.518
	Datatest 12	0.081257739	9007801686	94909.43939
	average	0.059270448	9070667826	95240.05368

Table 6

The MAPE, MSE, and RMSE parameters for hybrid GA-EDA trained ANN with diving the population with exact value

	ANN_GA-EDA(with dividing and with exact value	Criteria		
		MAPE	MSE	RMSE
Data test	Datatest 1	0.058943566	3750417182	61240.64975
	Datatest 2	0.051069955	3014160827	54901.37364
	Datatest_3	0.024629783	629586290	25091.55814
	Datatest_4	0.043476555	1974109762	44430.955
	Datatest_5	0.035170143	1057787699	32523.64831
	Datatest_6	0.048416327	1830647896	42786.07128
	Datatest_7	0.058804285	2619115099	51177.29085
	Datatest 8	0.037469463	1819133056	42651.29607
	Datatest_9	0.043193119	1877564252	43330.86951
	Datatest_10	0.094134291	6504039686	80647.62666
	Datatest_11	0.029435327	906796466.5	30113.06139
	Datatest_12	0.034613112	1249885728	35353.72298
	average	0.046612994	2269436995	47638.60824

Table 7
The MAPE, MSE, and RMSE parameters for hybrid GA-EDA trained ANN without diving the population with exact value

ANN_GA-EDA(without dividing and with exact value)		Criteria		
		MAPE	MSE	RMSE
Data test	Datatest_1	0.012404422	219431197.7	14813.21024
	Datatest_2	0.028038437	866619729.1	29438.40568
	Datatest_3	0.028523653	956693514.9	30930.46257
	Datatest_4	0.026807543	844264231.9	29056.22535
	Datatest_5	0.012429593	183168032.8	13533.95851
	Datatest_6	0.024143924	1086956116	32969.01751
	Datatest_7	0.011249668	179079926.1	13382.0748
	Datatest_8	0.027985553	1371326488	37031.42569
	Datatest_9	1.15922E-05	467.7679126	21.62794287
	Datatest_10	0.035414872	1974921815	44440.09243
	Datatest_11	0.039348992	2073246005	45532.91123
	Datatest_12	0.015659642	297766521.3	17255.91265
	average	0.021834824	837789503.8	28944.59369

4. Conclusion

The cost of every project should be evaluated, and the spending features should be determined. The project could be influenced by cash flow estimation. The cash flow prediction is gained by studying the actual data and simulation it with appropriate algorithms such as the genetic algorithm. In this paper, artificial neural network (ANN) is chosen to be used in the forecasting path. Three algorithms are used for training the ANN model which are the genetic algorithm (GA), estimation distribution algorithm (EDA), and hybrid GA-EDA method. 996 airplane maintenance bases data are used as actual data, and after clustering them, 119 similar data is chosen. The best results from three algorithms are chosen to improve the performance. In the hybrid GA-EDA method, there should be a ratio between two algorithms to improve the model. The dynamic ratio is chosen in two forms. First one is the ratio with dividing the population into half and the second one is without dividing. For the second one, the ratio could be enhanced in each step. The project is divided into 20 equal parts, and the first three steps are used for predicting the fourth one. For instance, to predict the 30% point, five previous points are used (5%, 10%, 15%, 20%, and 25%). The best result is chosen by calculating three error amounts which are MAPE, MSE, and RMSE. The output of hybrid GA-EDA without dividing the population has the best result, and the values of MAPE, MSE, and RMSE are %0.022, 28944.59 Dollars, and 837789503.79 Dollars, respectively. The accuracy of the GA, hybrid GA-EDA with dividing the population, and the EDA are lower than hybrid GA-EDA without dividing the population, respectively, by determining the three error amounts.

References

Almond, B., & Remer, R. S. (1979). Models for present-worth analysis of selected industrial cash flow patterns. *Engineering and Process Economics*, *4*(4), 455-466.

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. *International Journal of Forecasting*, *27*(3), 672-688.

Bec, F., & Mogliani, M. (2015). Nowcasting French GDP in real-time with surveys and "blocked" regressions: Combining forecasts or pooling information? *International Journal of Forecasting*, *31*(4), 1021-1042.

Blanc, S. M., & Setzer, T. (2015). Analytical debiasing of corporate cash flow forecasts. *European Journal of Operational Research*, *243*(3), 1004-1015.

Chen, H. L., O'Brien, W. J., & Herbsman, Z. J. (2005). Assessing the accuracy of cash flow models: the significance of payment conditions. *Journal of Construction Engineering and Management*, *131*(6), 669-676.

Cheng, M. Y., & Roy, A. F. (2011). Evolutionary fuzzy decision model for cash flow prediction using time-dependent support vector machines. *International Journal of Project Management*, *29*(1), 56-65.

Cheng, M. Y., Tsai, H. C., & Liu, C. L. (2009). Artificial intelligence approaches to achieve strategic control over project cash flows. *Automation in Construction*, *18*(4), 386-393.

Cheng, M. Y., Tsai, H. C., & Sudjono, E. (2010). Evolutionary fuzzy hybrid neural network for project cash flow control. *Engineering Applications of Artificial Intelligence*, *23*(4), 604-613.

Ghysels, E., & Ozkan, N. (2015). Real-time forecasting of the US federal government budget: A simple mixed frequency data regression approach. *International Journal of Forecasting*, *31*(4), 1009-1020.

Holland, J. H., & Goldberg, D. (1989). *Genetic algorithms in search, optimization and machine learning*. Massachusetts: Addison-Wesley.

Hwee, N. G., & Tiong, R. L. (2002). Model on cash flow forecasting and risk analysis for contracting firms. *International Journal of Project Management*, *20*(5), 351-363.

Kao, L. J., Chiu, C. C., Lu, C. J., & Yang, J. L. (2013). Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. *Neurocomputing*, *99*, 534-542.

Khosrowshahi, F., & Kaka, A. P. (2007). A decision support model for construction cash flow management. *Computer-Aided Civil and Infrastructure Engineering*, *22*(7), 527-539.

Larrañaga, P., & Lozano, J. A. (Eds.). (2001). *Estimation of distribution algorithms: A new tool for evolutionary computation* (Vol. 2). Springer Science & Business Media.

Li, Y., Moutinho, L., Opong, K. K., & Pang, Y. (2015). Cash flow forecast for South African firms. *Review of Development Finance*, *5*(1), 24-33.

Mühlenbein, H. (1997). The equation for response to selection and its use for prediction. *Evolutionary Computation*, *5*(3), 303-346.

Namazi, M., Shokrolahi, A., & Maharluie, M. S. (2016). Detecting and ranking cash flow risk factors via artificial neural networks technique. *Journal of Business Research*, *69*(5), 1801-1806.

Russell, J. S. (1991). Contractor failure: analysis. *Journal of Performance of Constructed Facilities*, *5*(3), 163-180.

Son, H., Kim, C., & Kim, C. (2012). Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables. *Automation in Construction*, *27*, 60-66.

Touran, A., Atgun, M., & Bhurisith, I. (2004). Analysis of the United States Department of Transportation prompt pay provisions. *Journal of Construction Engineering and Management*, *130*(5), 719-725.

Venkatesh, K., Ravi, V., Prinzie, A., & Van den Poel, D. (2014). Cash demand forecasting in ATMs by clustering and neural networks. *European Journal of Operational Research*, *232*(2), 383-392.

Xiong, T., Bao, Y., & Hu, Z. (2013). Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Economics*, *40*, 405-415.