Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 6, No. 4, November 2021, Pp. 329-336

A hybrid tabu search and genetic algorithm imputation approach for incomplete data

Bain Khusnul Khotimah[1], Yeni Kustiyahningsih[2], Miswanto[3]
Department of Informatics Engineering, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Indonesia[1]
Department of Information System, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Indonesia[2]
Department of Mathematics, Faculty of Science of Technology, University of Airlangga Surabaya, Indonesia[3]

Article Info
Keywords:
Missing Value, Genetic Algorithm, Tabu Search, Customer Segmentation, Imputation

Article history:
Received: November 13, 2021
Accepted: November 27, 2021
Published: November 30, 2021

Cite:
Khusnul Khotimah, B., Kustiyahningsih, Y., & Miswanto, M. (2021). A Hybrid Tabu Search and Genetic Algorithm Imputation Approach for Incomplete Data. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 6(4).
https://doi.org/10.22219/kinetik.v6i4.1340

*Corresponding author.
Bain Khusnul Khotimah
E-mail address:
bain@trunojoyo.ac.id

Abstract
The common problem for data collection is happening missing value during the data collection and processing process that the quality of the data testing is decreased. A computational based technique for dealing with missing values, namely Genetic Algorithm Imputation (GAI). The usage was used to estimate the dataset's missing values. GAI generates the optimal set of missing values with the acquisition of information as a function of fitness to measure individual solutions' performance. GAI conducts continuous searching until the missing criteria value is found according to best fitness. So, it is trapped in optimal conditions temporarily. The improvement of GAI with tabu search is known as TS-GAI, that strength is two metaheuristic techniques modified at the mutase stage to distract the local optima's search. In applying missing values, this technique works better when many possible values are used instead of the mixed attribute having missing values. Because the new generation chromosome values generate many opportunities to make up for the missing values. The experimental results show that the TS-GAI shows better performance on 30% MV with a fitness value of 0.212. It converges at 159 iterations. Generally, TS-GAI is a faster iteration than simple GAI and it has a lower RMSE level than other imputation techniques.

1. Introduction

Hybrid genetic algorithms with tabu search have been studied for many years. Because the most of the additional algorithms on GA to explore global candidates by utilizing local optimal points and to solve complex optimization problems [1][2]. The additional algorithm, the form of tabu search has the role of saving superior individuals from previous generations, reusing individuals as elites, maintaining population diversity, and preventing. The method could solve for converging local minimums. Since, the TS-GA will be easier, and more powerful than the conventional hybrid method [3]. The hybrid GA model was developed to contribute to specific application domains, but not much has been developed to improve handling missing values [4]. For example, Hwang (2014) compared the methods of treating missing value in wireless data. Data used to predict the value of imputation in the wireless distribution network. This study will attempt to add the tabu list to store the optimal value to avoid and evaluates candidate solutions never visited before [5].

The problem with mixed data is the difficulty in analyzing data with multiple features, when there is a missing value that spreads to almost all features. Mixed data requires a comprehensive imputation, which is expected to increase precision and efficiency [6]. However, the effect of imputation methods on data depends on differences in information and data types. The genetic algorithm for imputation works with random numbers on the initial chromosomes as replacement values, that allowing the best chromosomes not to be involved in the imputation process. The genetic algorithm uses random numbers in each of the best chromosome selections from several parents, so it takes a long time. The solution of GA is not necessarily the optimal solution because of strongly influenced by the random numbers generated [7]. MOGAImp takes incomplete information into account to estimate a value that is more resistant to noise. MOGAImp is designed to treat mixed attribute data together, which taken into account the relationships between attributes. The flexibility of the MOGAImp coding system allows easy adaptation to various application domains [8].

The customer segmentation identification factor in the retail business is strongly influenced by various risks, including credit relationships with customers, product identification in each customer segment, demand management, and determination of which customers are the most likely. Each customer may be different in requirement, demographics, geography, and preferences, behavior, and others. Grouping is applied as the application of segmentation, in which the customer is further divided into several smaller groups or segments [9]. Each segment member exhibits similar characteristics of market behavior. Segmentation includes geographic segmentation, demographic segmentation, media segmentation, price segmentation, psychographic or lifestyle segmentation,

segmentation distribution, and time segmentation to target profitable customers [10]. Customer segmentation often has data heterogeneity and data loss during the process of machine learning [11].

Customer segmentations have aims to identify the requirement of each customer include very complex task. The previous research used various mathematical models to segment customers without considering the data correlation. The relationship between customer segmentation with promotional activities without regard to the constraints or the dependent variable promotional activities. Vigneau et al. [12] studied customer satisfaction through many questionnaires and found that MV% questionnaires contained many missing scores. Even though they have strict data collection standards, the customer database's data of missing rate is still as high as 30%. The most of the customer data is collected through questionnaires and interviews, which possibly including many missing value (MV) [13][14]. Missing data imputation is the simplest means of preprocessing MV is listwise deletion (LD) [15], which deletes instances from the dataset directly. The correct proposition of value estimation is an important contribution to improving customer value segmentation for usage imputation in missing data.

The GA-TS hybrid model for imputation is developed by evaluating each individual for using objective functions in each iteration. The best individuals store the generation into long-term and short-term tabu lists. When selecting parent candidates by the tournament selection method, we refer to the tabu list to not select individuals of the same genotype via the Hamming distance [16]. Tabu restrictions could only be applied to one parent to produce offspring. The usage of tabu boundaries would be kept away for focusing individuals on local optima. Solutions are gradually being accumulated into the long-term tabu list in a multimodal function, several solutions (respectively, a pareto solution) which obtained simultaneously. The algorithm's basic idea applied the best solutions from each generation into long-term and short-term tabu lists, which prevent individuals from being selected more than n times [17].

This research uses a combination of genetic algorithms with tabu search for imputing missing data by utilizing the value of chromosome results at the best accuracy. The process of finding significant imputation values can improve the best accuracy. The procedure for solving problems with more than one candidate solution, as well as the process of evaluating the candidate solutions to produce the best solution from the existing set of candidate solutions. Filling in missing values in customer segmentation data using genetic algorithms aims to provide an overview of solutions in overcoming missing values so that the data can be complete and can be used for further processing. genetic algorithm to determine estimates in filling in missing values, so that the data becomes complete and can be used for further purposes.

In this paper, the problem of missing value is treated by approaches hybrid of genetic and tabu search algorithm for optimization. That using tabu search based on a set of a random set of valid solutions in several iterations. The algorithm creates a new generation of solutions maintained from the previous generation; the new solution starts with a clear list of tabu. Tabu search is a meta-heuristic search method that uses the local search method used for optimization. Technique of method randomly generates chromosomes from the solution to that neighbor's solution in hopes of finding a better solution. The tabu list suppresses the possibility of local convergence at an early stage of the iteration by exploring new solution spaces for better solutions. The final results are accumulated in the long-term tabu list. This means that several peaks are obtained for the multimodal problem, which is the pareto optimal solution.

2. Research Method

2.1 The Combination model of Tabu Search-Genetic algorithm (TS-GA)

The algorithm combines the two approaches tabu search and genetic algorithm to optimize it based on a set of a random set of valid solutions to the multiple iterations [19]. This algorithm creates a new generation of retained solutions from the previous generation. Tabu search techniques would be improving local search performance by using memory structures that keep track of the solution [3]. If the previous solution in the short-term, it is marked as 'tabu' so that the process does not happen again. Tabu search on a genetic algorithm is a local search algorithm meta heuristics that can solve combinatorial optimization problems. The different to local hill-climbing technique, the tabu search algorithm tends to be a locally optimal solution. Tabu search has fast execution speed on the genetic algorithm does not return to the solution that has been explored the best chromosomes are stored in tabu list [6]. The solution already exists prevented by using a memory called tabu list [15].

Tabu search is used to store a set of solutions in chromosomes that have been evaluated. In each iteration process, the solution to be evaluated will be matched first with the tabu list's contents. The solution already exists on the lists, so tabu solutions will not be re-evaluated in the next iteration. If there are no more solutions that are not members of the tabu list, then the best score that has just been obtained is the real solution. The steps for finding a tabu is shown in Figure 1. The individual xi in generation t is the $p(t)$ individual in the population. The fitness function contains a value that represents the individual population of a particular generation selected by rank. Suitability for individuals based on interpolation ranging from best (rank 1) to worst (rank $n* \leq N$). The average fitness of the same individual ranking will be sampled at the same level [20][21].

The individual x_i in generation t is the $p(t)$ individual in the population. The fitness function contains a value that represents the individual population of a particular generation selected by rank. The suitability for individuals based on

interpolation ranging from best (rank 1) to worst (rank $n* \leq N$). The average fitness of the same individual ranking will be sampled at the same level. The outline TS-GA for multiobjective functions is shown in Figure 1.

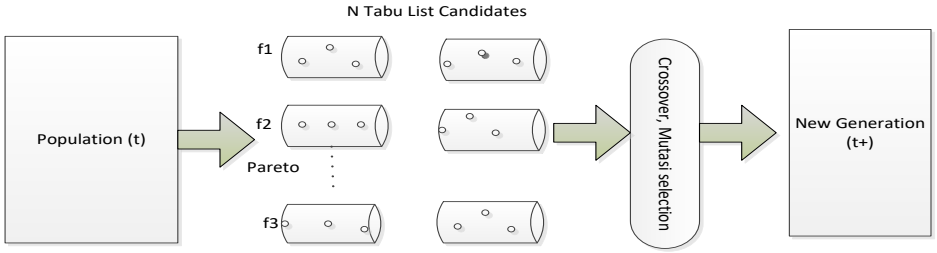

Figure 1. Diagram Tabu Search-GA

The number of the tabu list is m+1, where m is the number of the objective Functions. The Pareto f_n optimals are evaluated by the ranking method. Each tabu list evaluates each offspring. The individuals are selected by the tournament selection utilizing each objective for Pareto optima. The short-term tabu list is replaced with FIFO method. So, the long-term tabu list is replaced with the rank of tabu list. The latest best fitness individual of Genetic operations are applied for individuals.

2.2 Performance Algorithm

Individual evaluation is an individual representation, that is carried out after the population of generation. This stage aims to determine the quality of the optimal level of each chromosome in a population. Chromosome is the most optimal that will be chosen as the solution to this problem. The evaluation process is carried out to find the fitness value. After the random population is generated the fitness value. An individual must be evaluated based on the value of a specific objective function. Genetic algorithms can be solved complex problems by minimizing the fitness value. Evaluate each target data's fitness value from the objective function by comparing the actual value with the imputed from the lowest using RMSE. The final RMSE to test the TS-GA performance is as follows Equation 1.

$$\mathrm{F(x)} = \mathrm{RMSE} \tag{1}$$

Each chromosome will be evaluated using a fitness function with different weights for each conflict based on the GA Process of class-dependent imputation [23]. It considers class labels to estimate the value to be calculated. Therefore, the GA for imputation can be considered as the best trade-off between the three evaluation measures analyzed, with the advantage that this solution is class-independent. Tabu search based Genetic Algorithm uses the basic structure of the Genetic Algorithm. The genetic algorithm's crossover function is coupled with tabu searches to bring more diversity to the population, because each population chromosome is calculated based on RMSE.

2.3 Genetic Algorithm Imputation for Missing Data

GAI has optimized the chromosome vector component which contains a set of fitness functions [17]. This set of points is known as the pareto-optimal set. There will be no improvement in one cost vector component for every optimal point, that does not cause degradation in the other remaining components. Every element in the pareto optimal is a non-inferior solution to the missing value problem. Multicriteria target vector optimization is used in combination with genetic algorithms [22]. GAI is a new data imputation method based on a genetic algorithm. GAI takes into account information from incomplete instances based on chromosome diversity. Let x_i be a set of all n available training examples. If the training data is denoted by x_t, then for the attribute set is denoted by $\{a_1, a_2, ..., a_n\}$ and the class is denoted by the set of values $\{C_1, C_2, ..., C_n\}$.

Several attributes have a certain percentage missing value and x_m becomes a set of examples which includes the sample set with the missing attribute values. This experiment is limited to mixed attribute values. The domain value for a particular attribute is considered the population. However, a set of solutions calculates missing value (MV) of all possible values calculated from the GA operation as follows:

GA Imputation Algorithm

Step 1 Determine the attribute with the missing value and the domain value position of the missing attribute.
Step 2 Replace missing attributes with all available values from the domain continuously for all attributes with the missing values.
Step 3 Perform genetic crossover, mutation and selection operations on the selected instance.
Step 4 Calculate the fitness value using RMSE.

Step 5 If the attribute example has been replaced with the attribute value, the calculated value is successful, but the fitness value RMSE> 1.
Step 6 Repeat this procedure for Select experts, in step 3
Step 7 The end

3. Research Methodology

This algorithm uses a combination of genetic algorithm and tabu search which imputations have been constructed that conform to certain limitations. The proposed algorithm (GA-TS) applies neighboring TS to generate a portion of the new generation selected in the GA process for imputation.

The main steps of the proposed algorithm are:

Step 1 Initialize TS-GA parameters and generate missing data with a certain percentage
Step 2 Eliminate data with a certain percentage for the testing process
Step 3 Generating a random initial population to generate a viable set of solutions (chromosomes).
Step 4 Initial data Imputation process to obtain complete data
Step 5 Perform the initial imputation with a random chromosome.
Step 6 Calculate the RMSE and evaluate each chromosome based on imputation results with data without imputation.
Step 7 Determine the suitability function of each chromosome in the population.
Step 8 Implementing the GA operator with a crossover, mutation, and selection mechanism to generate a new population.
Step 9 Copying the best solution from the current population to the new population by storing in a tabu list in a specified n. A tabu search algorithm creates new members in the new population as neighbors to randomly selected solutions in the current population.
Step 10 Apply the crossover operator to equip new population members.
Step 11 Apply the mutation operator to the new population.
Step 12 Let the current population become the new population.
Step 13 The imputation process is carried out using a new population, but if the chromosome value in the new population has not the convergence criteria, stop. Otherwise, go to step 2.
Step 14 Storing the best chromosomes for imputation until the smallest RMSE is obtained.

3.1 Preparation Parameters

The selection process for the best chromosomes, the generation update is carried out that tabu list is checked. If the same result is already in the tabu list, then the genetic algorithm process will be repeated until it finds a result. Hybrid modl GA-TS avoid local optimum. The number of tabu list has used *m+1*, where *m* is the number of objective functions. Pareto optima are evaluated using a rating method. Each tabu list evaluates individuals which selected by tournament selection through their respective objectives, and all specified parameters are used in Table 1.

Table 1. Parameters of Tabu Search-GAI

Parameters	Values
Population Size	100
Generation	100-1000
Chromosome Size	Number of fitur
Fitness Function	One Point Crossover
Selection	Tournament
Tour Size	10-
Crossover	Keep-Best
Probability of Crossover	0.5-1
Probability of Mutation	0.3-0.5
Elitism	Keep-best
Runs	10

This parameter setting has conducted several experiments with the TS-GA parameter, which set through calibration testing with the customer segmentation data set. The results of the evaluation are obtained from the simulation to the convergence of the proposed methods and depend on the complexity of the dataset. The dataset is grouped by dimensions, the number of missing values, and the time required for modeling.

3.2 Imputation simulation data

Customer segmentation requirement based on characteristics by age and background, that there are often having a missing value. Segmentation of customers with missing value such as age, homeownership, education level often

occurs due to remiss survey timing. Customer missing value will drastically reduce the amount of data to be analyzed, which increased the risk of costly errors. Missing data can also lead to misleading results by introducing bias. The simulation of missing data is shown in Table 2.

Table 2. Sample of Customer Segmentation Attributes

Attribute1	Attribute2	Attribute3	Attribute4
Gender	Age	Annual Income	Spending Score
Male	19	Low	3
Male	34	Medium	12
Female	20	Medium	6
Male	45	High	45
Female	31	Low	10

The result of solutions was built for each attribute which presents data by some attribute values are missing. So, missing values are sorted into an attribute array file, which consists of mixed attributes, namely continuous and category. Table 3. (a) Process depicts a sample data structure consisting of attributes Att1, Att2, Att3, and Att4.; (b) Data has missing values Data has missing values across multiple attribute sets.

Table 3. Sample Atribut of Missing Value

a. Complete data				b. Incomplete Data			
Attribute1	Attribute2	Attribute3	Attribute4	Attribute1	Attribute2	Attribute3	Attribute4
Gender	Age	Annual Income	Spending Score	Gender	Age	Annual Income	Spending Score
Male	19	Low	3	?	19	Low	3
Male	34	Medium	12	Male	34	?	12
Female	20	Medium	6	Female	?	Medium	6
Male	45	High	31	Male	45	High	?
Female	30	Low	10	Female	31	?	10

Table 4 showed process strategy of codification, that was developed by considering the objectives can handle continuous attributes tableand categorical attributes. The data structure with the application of genetic operators is translated into an index as a description of a multi-purpose strategy.

Table 4. Process Index on Data

Attribute	Value indexing for data continue					
Attr2 (Age)	10-20	21-30	31-40	41-50	51-60	61-100
Index	1	2	3	4	5	6
Attr4 Spending Score	1-10	11-20	21-30	31-40	41-50	51-100
Index	1	2	3	4	5	5

The grouping of some data features into indexes for identification was referred in Table 5. The index is used to facilitate the position of missing data for imputing new data as a replacement. The usage of indexing helps the imputation process, which used the final chromosome at the best fitness. The test is carried out as many as n iterations to produce more consistent imputation results. The number of iterations is determined by storing the best chromosome values in the taboo which will be stopped until convergence.

Table 5. The sequence of Numbers as an Index is the Position of the Imputation Replacement Value

Attribute	Value indexing for data category				
Attr1	Male (g_{11})	Female (g_{12})			
Index (Gender)	1	2			
Attr2	19 (g_{21})	20 (g_{22})	30 (g_{23})	34 (g_{24})	45 (g_{25})

Index (Age)	1	1	2	3	4
Attr3	Low (g_{31})	Medium (g_{32})	High (g_{33})		
Index (Annual Income)	1	2	3		
Attr4	3 (g_{41})	6 (g_{42})	10 (g_{43})	12 (g_{44})	30 (g_{45})
Index (Spending Score)	1	1	1	2	3

The decoding process maps the genotype index to the solution for each attribute. The value in the attribute column is indicated by the index used in the phenotype. For example, the genotypes are shown in Table 6. Index of 1, 1, 2, 1, are mapped in consultation with the solution set of their respective attributes, so that "1" is mapped “Male” and “Medium”, "2" represents "Female" and “Low”.

Table 6. Final Indexing for One Record Data

Code	Final indexing				
Phenotype	Male	20	Medium	Low	30
Index	1	1	2	1	
Ghenotype	(g_{11})	(g_{22})	(g_{32})	(g_{31})	(g_{45})

4. Results and Discussion

The customer segmentation data set contains 5 attributes and has 200 tuples, representing the data for 200 customers. Attributes in data set have CustomerId, gender, age, annual income (k$), expense score on a scale (1-100). To determine the performance of the method, by eliminating the value on features with different percentages. The missing value consists of 4 variables from 10 variables, namely age, gender, age, annual income. The data is used for analysis to gain a competitive advantage over their rival companies, demonstrating better knowledge of customer requirements. The test will be carried out 10 times on each parameter combination at each time change.

4.1 Comparison of The Number of Generations and Time

The GAI-TS method uses the right value search technique to replace the values in the missing value variables. GA technique has a chromosome coding that is flexible to the data, where each chromosome corresponds to the value for imputation, so that the value corresponds to the value of the original data.This test will focus on changing parameters and attribute of missing value. Parameters of generation and time was changed to find out performance. So, two algorithms are compared using the best parameters of the genetic algorithm. Each test is carried out 10 times to get the best valuable that shown the effectiveness of TS-GA method with the number parameter tabu n = 10 for procentage of missing 10%. The performance has shown more optima with the tabu-GA strategy than GAI. When iteration in ten executions, performance shown simple GAI with elite cromosome better than Tabu-GA for solving a certain parameter function but requires more generation.

Table 7. Differences in The Performance of TS-GAI and GAI Based on Different Parameters

Missing Value %	Pm	Pc	Algorithm	Execution Time (s)	Generation of Converge	Fitness
10	0.01	0.1	GAI	3648	574	0.276
			TS-GAI	3940	120	0.285
20	0.05	0.5	GAI	4671	592	0.249
			TS-GAI	4954	131	0.109
30	0.1	0.8	GAI	5409	630	0.301
			TS-GAI	5998	159	0.212

Table 7 shown the fitness based on execution time that is better the fitness results on TS-GAI with the same parameters and the same percentage of missing values than GAI. The usage parameters are value of p_m = 0.05 and a value of p_c = 0.05.

4.2 Analysis of TS-GAI performance

TS-GAI algorithm is applied a the same of standard parameters of GA in the case of mixed attribute data containing missing values. The experimental analysis of customer segmentation data, the number of constraints defined by the user is set to n in a list for ranking. TS-GAI are built with generations and population changes until the best results

are obtained. Table 8 shown the test on the genetic algorithm, that having repeated fitness values before getting a better fitness value. However, each generation fitness of TS-GAI is always different because the error value tends to decrease due to checking the results of genetic operations. Tabu list that the same value will not be used as parents. Based on the test results, TS-GAI is used to find the right chromosome from a set of taboos for imputation for missing values. replacement value is expected to produce a better fitness value in each generation. Since, GAI loses to convergence because it requires more iterations to find the optimal solution in almost every population size. Whereas TS-GAI with changes in the number of *n* tabu lists also affects the number of generations when converging. The usage elitist TS-GAI strategy finds the optimal solution. Since, many iterations are needed to find the optimal solution for almost every population size. The change in the number of *n* tabu lists also affects the number of generations when converging. It is more *n* tabus than the iteration increases with decreasing error.

Table 8. The Performance of TS-GAI and GAI Based on The Number of Tabu and Population

Tabu List Size	Population size	Fitness	
		TSGAI	GAI
10	20	0.726	0.590
	50	0.653	0.856
20	20	0.638	0.825
	50	0.395	0.492
30	20	0.760	0.642
	50	0.382	0.259

Statistical data on the comparison between algorithms shown in the table shows the taboo-based approach proposed using a genetic-based approach to gradually intensify the search procedure. The risk of getting stuck at local optima is eliminated by using backtracking guided by taboo lists. The more taboos and the number of populations, the more varied solutions. The number of taboos affects the amount of memory and provides the opportunity for more and more values to be stored. The results of the more the number of taboos and the population the more shows this algorithm is optimal.

Table 9. The Differences in the Performance of TS-GAI and GAI Based on the Number of Tabus and Population

	RMSE					
Missing Value (MV) %	TS-GAI	GAI	Mean filled	Min filled	Max filled	Zero filled
10	0.220	0.282	0.795	0.690	0.861	0.681
20	0.125	0.280	0.974	0.973	0.675	0.532
30	0.378	0.389	0.573	0.749	0.825	0.603

Table 9 shown that the missing data of imputation using TS-GAI has a lot of variability in values, when entering the missing values. TS-GAI results achieved lower error values at 30% MV differences compared to single imputation and GAI.

5. Conclusion

TS-GAI test shown maximum results when the iterations were smaller than GAI. The iteration stability produces 165 maximum iterations by producing the smallest error of stores individuals into multiple tabu lists. The more population results in diversity and enhancement the method's performance. TS-GAI resolving the missing value can avoid the local optimum that occurs, which result the smaller error with the least generation. Based on the best parameters, TS-GAI only needs 165 generations to get the best results, while GAI on the best parameters requires 657 generations to get the best performance. The future research TS-GAI can be applied to various domains such as regression, grouping, and time series analysis, investigating the adoption of heuristics to generate initial populations to reduce search space. TS-GAi parameter sensitivity, to improve in applying the mixture attribute data and produce individual fitter. In further research, more constraints can be taken into account by the user.

References

[1] C.T. Tran, M. Zhang, P. Andreae, "A genetic programming-based imputation method for classification with missing data," In: European conference on genetic programming. Springer, pp 149–163, 2016. https://doi.org/10.1007/978-3-319-30668-1_10

[2] R. Armina, A. M. Zain, N.A. Ali, R. Sallehuddin, "A Review On Missing Value Estimation Using Imputation Algorithm," Journal of Physics: Conference Series, Volume 892: 012004, The 6th International Conference on Computer Science and Computational Mathematics (ICCSCM 2017), Langkawi, Malaysia, 4–5 May 2017. https://doi.org/10.1088/1742-6596/892/1/012004

[3] S. Alharbi, "A Hybrid Genetic Algorithm with Tabu Search for Optimization of the Traveling Thief Problem," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 9, no. 11, pp.276-287, 2018. https://dx.doi.org/10.14569/IJACSA.2018.091138

[4] C. Leke, B. Twala, T. Marwala,"Modelling of missing data prediction: computational intelligence and optimization algorithms," In: International Conference on Systems, Man and Cybernetics (SMC), IEEE, pp. 1400–1404, 2014. https://doi.org/10.1109/SMC.2014.6974111
[5] W. Shahzad, Q. Rehman, E. Ahmed, "Missing Data Imputation using Genetic Algorithm for Supervised Learning," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 8, no. 3, pp. 438-445, 2017.https://dx.doi.org/10.14569/IJACSA.2017.080360
[6] J. Josse, F. Husson, "missMDA: a package for handling missing values in multivariate data analysis," J Stat Softw vol. 70, no. 1, pp.1–31, 2016. https://doi.org/10.18637/jss.v070.i01
[7] F. Lobato, C. Sales, I. Araujo, V. Tadaiesky, L. Dias, L. Ramos, "Multi-objective genetic algorithm for missing data imputation", *Pattern Recognit. Lett.*, vol. 68, pp. 126-131, 2015. https://doi.org/10.1016/j.patrec.2015.08.023
[8] S. F. Sabbeh, "Machine-Learning Techniques for Customer Retention: A Comparative Study", International Journal of Advanced Computer Science and Applications (IJACSA), 2018.
[9] H. Hwang, T. Jung, E. Suh, "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry," Expert Systems with Applications, vol. 26, pp. 181-188, 2004. https://doi.org/10.1016/S0957-4174(03)00133-7
[10] S. Nabavi, S. Jafar, "Providing a Customer Churn Prediction Model Using Random Forest and Boosted Trees Techniques," (Case Study: Solico Food Industries Group, Journal of Basic and Applied Scientific Research, vol. 3, no. 6, pp. 1018-1026, 2013.
[11] A. Kazemi, M. E. Babaei, "Modelling Customer Attraction Prediction in Customer Relation Management using Decision Tree: A Data Mining Approach," journal of Optimization in Industrial Engineering, 2011.
[12] E. Vigneau, "Segmentation of a panel of consumers with missing data, Food Quality and Preference," vol. 67, July 2018, pp. 10-17. https://doi.org/10.1016/j.foodqual.2017.04.010
[13] A.B. Zorić, "Predicting Customer Churn In Banking Industry Using Neural Networks," Interdisciplinary Description of Complex Systems, vol. 14, no. 2, pp. 116-124, 2016.
[14] G.S. Linoff, M. J. Berry, (2011). Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons, 2011.
[15] N.F. Fauziah, Y.H. Putra, "Scheduling Regular Classrooms using Heuristic Genetic and Tabu Search Algorithms," IOP Conference Series: Materials Science and Engineering: 012116, vol. 407, no. 1, 2018. https://doi.org/10.1088/1757-899X/407/1/012116
[16] P. Delima, A.M. Sison, R.P. Medina, "Variable Reduction-based Prediction through Modified Genetic Algorithm," Allemar Jhone (IJACSA) International Journal of Advanced Computer Science and Applications, vol.10, no.5, pp.356-363, 2019. https://dx.doi.org/10.14569/IJACSA.2019.0100544
[17] O. M. Elzeki, M. F. Alrahmawy, S. Elmougy, "A New Hybrid Genetic and Information Gain Algorithm for Imputing Missing Values in Cancer Genes Datasets," J. Intelligent Systems and Applications, vol. 12, pp. 20-33, 2019. https://doi.org/10.5815/ijisa.2019.12.03
[18] M. Noei, M.S. Abadeh, "A Genetic Asexual Reproduction Optimization Algorithm for Imputing Missing Values," 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), IEEE, Ferdowsi Universty of Mashhad, pp. 214-218, 23 January 2020. https://doi.org/10.1109/ICCKE48569.2019.8964808
[19] F. Glover, J.P. Kelly, M. Laguna, "Genetic algorithms and tabu search: Hybrids for optimization," Computers & Operations Research , vol. 22, no. 1, pp. 111-134, January 1995. https://doi.org/10.1016/0305-0548(93)E0023-M
[20] X. L. L. Gao, "An effective hybrid genetic algorithm and tabu search for flexible job shop scheduling problem," International Journal of Production Economics, vol. 174, pp. 93-110 April 2016. https://doi.org/10.1016/j.ijpe.2016.01.016
[21] M. D. Akbar, R. Aurachmana, "Hybrid genetic–tabu search algorithm to optimize the route for capacitated vehicle routing problem with time window", International Journal of Industrial Optimization, vol. 1. no.1, pp. 15-28, February 2020. https://doi.org/10.12928/ijio.v1i1.1421
[22] B. K. Khotimah, F. Irhamni, T. Sundarwati, "A Genetic Algorithm for Optimized Initial Centers K-Means Clustering in SMEs", Journal of Theoretical and Applied Information Technology (JATIT), vol. 90, no. 1, pp. 23-30, 15 August 2016.