

## Research Paper

# Applying project management knowledge and hybrid algorithm in predicting the time and cost of completing dam projects

Reza Bakhshi<sup>1</sup>, Sina Fard Moradinia<sup>\*1,2</sup>, Rasool Jani<sup>1,2</sup>, Ramin Vafaei Poor Sorkhabi<sup>1,2</sup>

1. Department of Civil Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran

2. Robotics and Soft Technologies Research Center, Tabriz Branch, Islamic Azad University, Tabriz, Iran

Received: 2022/12/10

Revised: 2023/01/16

Accepted: 2024/01/05

Use your device to scan and read the article online



DOI:

[10.30495/wej.2024.31317.2374](https://doi.org/10.30495/wej.2024.31317.2374)

### Keywords:

Hybrid algorithm, EDAC, Kalman filter, Cost prediction, Dam construction

### Abstract

**Introduction:** Precisely predicting the time and cost of completing projects is vital because the lack of a proper estimation will be accompanied by an irrational upsurge in the exact execution costs compared to the set budget. Using the earned value method (EVM) to predict the time and cost of projects is prevalent. However, using this method alone highlights good accuracy in predicting time and cost of projects. Consequently, models based on EVM were developed.

**Methods:** The present article was developed using the EVM method and hybrid algorithms to predict the time and cost of completing projects. To attain this goal, the data from four dams, A, B, C, and D, were used to build models, and the data of the under-construction dam E were used to validate the models resulting from the modeling stage. To this end, the parameters earned schedule (Month), earned value (\$), actual progress (%), and actual cost (%) are used as inputs for predicting time and for predicting cost, as well as these parameters, time is also defined as input of hybrid algorithms.

**Findings:** Comparing the consequences of the hybrid algorithms in the training and test stage designates the high accuracy of the LSSVM-PSO model compared to the LSSVM-GA. The low variance in the error values of these two stages for this model suggests its high generalization ability on unseen data. The use of these hybrid models in forecasting the time for the E dam gave a prior warning for the delay in the completion of the project in the first month. Likewise, in cost predicting, the LSSVM-PSO and LSSVM-GA models issued an early warning in the seventh and ninth months, respectively, for the non-conformity of the project cost with the planned cost. This is while the Kalman filter stated the primary warning to predict the project's completion time in the seventh month, and this model gave no warning regarding the planned cost. Comparing these results with the periodical reports of the E dam construction project designates the excellent performance of hybrid models, particularly the LSSVM-PSO model.

**Citation:** Reza Bakhshi, Sina Fard Moradinia, Rasool Jani, Ramin Vafaei Poor Sorkhabi. Applying project management knowledge and hybrid algorithm in predicting the time and cost of completing dam projects. Water Resources Engineering Journal. 2024; 17 (61): 65- 85.

**Corresponding author:** Sina Fard Moradinia

**Address:** 1. Department of Civil Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

2. Robotics and Soft Technologies Research Center, Tabriz Branch, Islamic Azad University, Tabriz, Iran

**Tell:** +989143142619, <https://orcid.org/0000-0003-3539-7311>

**Email:** [fardmoradinia@iaut.ac.ir](mailto:fardmoradinia@iaut.ac.ir)

## Extended Abstract

### Introduction

It is hard to complete a project on time with a planned cost. Project monitoring and control include project progress processes so that potential problems are recognized over time and, if indispensable, operations are considered to resolve them. To this end, it is essential to frequently measure and evaluate the project performance and determine its variance from the planned baseline. The primary prediction before the start of the project is considered the basic plan for completing the project in the indicated time and budget. From the time the project starts, the project performance is assessed in line with time and cost compared to the initial plan. The challenges in construction projects have obliged project managers to use techniques and models to control and predict their cost and time. According to this forecast, the project manager determines that the project needs minor alteration or significant adjustment and problem-solving. Though, advanced techniques are not used for this aim. This leads to not correctly forecasting the cost and completion time of the projects and, as a result, inappropriate control of the project. Consequently, numerous researchers have used statistical analysis methods such as multiple regression, methods based on probabilities such as Monte Carlo simulation, and methods based on artificial intelligence.

### Materials and Methods

To carry out this study, information on four dam projects' timing and costs built in different regions of Iran was collected. Because of the confidentiality of their information, these dams were named A, B, C, and D. These dams have a total of 383 data points that were considered for the training of intelligent hybrid models and 306 data points (80% of the total data) were considered for the training stage and 77 data points (20% of the total data) for the hybrids model test. Data from another dam (dam E) will also validate the models. Fifteen months have passed since the construction of the E dam began, and the progress report for the

fifteenth month has been submitted. Consequently, there are 15 data points available for this dam. Based on the plan and preliminary design presented, the complete construction of this dam will take about 31 months and cost about 590,146.82 dollars. Accompanied by LSSVM-GA and LSSVM-COA hybrid models, the Kalman filter will also be used to forecast the completion time and cost of the E project. To apply these models, it is essential to calculate the obtained value indices for the collected data.

### Findings

Cumulative distribution functions (CDFs) have been used in the article to designate the input and output data (time and cost forecast). CFD for Earned Schedule (ES) is about 22% for  $ES < 6.6$  and about 53% for  $6.6 < ES < 37.8$ , 28% for  $EV < 209323$ , and about 56% for  $209323 < EV < 14020973$  and for the rest of the data this value,  $EV > 14020973$  is about 16%. The CFD value for Time (t) as  $t < 14$  is about 14%, and for  $14 < t < 85$ , it is about 54%, and for the rest of the data, this value for  $t > 85$  is about 32%. The value of CFD for Cost Ratio (CR) as  $CR < 0.039$  is about 18%, and for  $0.039 < CR < 0.666$ , it is about 65%, and for the rest of the data, this value of  $ES > 0.666$  is about 17%. Regarding this analysis, CFDs are usually distributed for four variables ES, EV, t, and CR.

The CFD value for Actual Progress (AP) in the form of  $AP < 0.0027$  is about 33%, and for  $0.0027 < AP < 0.037$ , it is about 60%, and for the rest of the data, this value  $AP > 0.037$  is about 7%. The value of CFD for Actual Cost (AC) as  $AC < 0.0015$  is about 35%, and for  $0.0015 < AC < 0.058$ , it is about 79%, and for the rest of the data, this value  $AC > 0.058$  is about 4%. According to this analysis, CFDs are not generally distributed for these two parameters.

### Discussion

To do modeling, first, the type of kernel function must be determined for the LSSVM algorithm. Applying simple LSSVM with diverse kernel functions based on the RMSE criterion disclosed that LSSVM with RBF kernel function has higher accuracy in estimating cost ratio and EDAC than other

kernel functions. On the other hand, it can be understood that the error of models with RBF and MLP kernel functions is meaningfully less than models with Polynomial and Linear kernel functions. This shows that there is a complex nonlinear relationship between inputs and outputs.

By setting the controllable parameters of PSO and GA optimization algorithms and applying hybrid LSSVM with these algorithms on the training data, it was shown that both optimization algorithms converged to the solution in less than 50 iterations. The cross-diagram evaluation of LSSVM-PSO and LSSVM-GA algorithms revealed that both models have high performance in the training and testing stages. The results attained by applying these algorithms to the validation data also have good accuracy. However, the dispersion of data points in this stage is more than in the training and test stages. Comparing the RMSE values for LSSVM-PSO and LSSVM-GA algorithms in estimating the completion time of projects by stages designates that the error of LSSVM-PSO in all three stages of training, testing, and validation is lower than LSSVM-GA. Furthermore, the slight difference between the error values in this model's two training and testing stages compared to LSSVM-GA shows the ability to generalize this model to other projects or unseen data. The higher accuracy of the LSSVM-PSO model compared to LSSVM-GA in the validation stage approves this claim.

The predicted EDAC profile based on LSSVM-PSO and LSSVM-GA hybrid models for the validation data (E dam project) indicated that the lower limit values predicted by both algorithms are more than planned. This shows that the project did not proceed according to the plan in the first month. Examining the reports of this project has also designated the same. Both algorithms mutually cover the project's completion time of 52 to 62 months from its initiation. Indeed, the output value of LSSVM-PSO and LSSVM-GA models show the project completion time as 57 and 59 months, respectively.

Scrutinizing the distribution of data points in the cross diagram of the actual and predicted cost ratio values of LSSVM-PSO and LSSVM-

GA hybrid models for training, testing, and validation stages revealed that both models perform well in target parameter prediction. However, the LSSVM-GA model performed better in the test phase. The comparison of the error values of these models by stages shows that the LSSVM-PSO model has a better performance than the other model. The slight variance between the error values of the training and testing stages in the LSSVM-PSO model compared to the LSSVM-GA model is indicative of the ability to generalize the model to other projects, and the excellent performance of this model in the validation stage confirms this statement. The projected cost profile for the E dam construction project based on EDAC was calculated with the help of LSSVM-PSO and LSSVM-GA models. Regarding the lower error of the LSSVM-PSO model compared to the LSSVM-GA model in these two stages, the range of changes between the lower and upper limits is smaller. Based on these calculations, the LSSVM-PSO model has issued an initial warning in the seventh month for the project cost to be far from the planned cost. This is while the results of the LSSVM-GA model issued this initial warning in the ninth month. The continuance of the forecasting process likewise reveals that the tools used by the project manager could not issue this initial warning to revise the operational plan. According to the construction reports review of this dam, the managers have confirmed the increase in the costs of the termination of the operation in the 13th month since the beginning of the project.

The probability profile of EDAC and the cost obtained from the Kalman model for project E disclosed that regarding this model's output, the project's completion time up to the fifth month from the start time was according to the plan. Nonetheless, over time, the predicted completion time is significantly different from the planned time for the project, so in the seventh month, the model issues an initial warning for taking more time to complete the project than the planned time. The output of the Kalman model for the cost of completing the project is in line with the planned cost, but from the tenth month onwards, the cost projected by the model gradually deviates from the

planned cost. This model has not issued any warning for irretrievable deviation of the project completion cost from the planned cost.

### **Conclusion**

The current study used LSSVM-PSO and LSSVM-GA hybrid algorithms to predict project completion time and cost. The data from four built dams were used to develop these models. These data were divided into training data (80% of the total data) and test data (20% of the total data). The data of another dam, whose progress report was submitted until the 15th month, was used to validate the developed models. Next, the Kalman filter was used for the validation data to evaluate the hybrid models in time and cost prediction.

Using hybrid algorithms revealed that they have good accuracy in the model training and testing phase to predict the time and cost of completing the projects. Likewise, the low difference between the LSSVM-PSO model in these two stages indicated the high generalizability of this model compared to the LSSVM-GA model, which was confirmed by the excellent performance of this model in the validation stage. Regarding the results of applying these two models on the E dam for EDAC, it was found that the LSSVM-PSO and LSSVM-GA models predicted the impossibility of completing the project within the planned period by announcing the early warning in the first month. This finding is in line with the periodic reports recorded for this project. The result of applying these two models for cost prediction also disclosed the better performance of the LSSVM-PSO model in the training and test stages. The low error variance between the training and test stages also showed the high generalization capability of this model compared to the LSSVM-GA model on unseen data. Applying these models to the E dam data strongly confirmed this claim. The early warning for the deviation of the project completion cost from the planned cost for the E dam in the LSSVM-PSO and LSSVM-GA models was issued in the seventh and ninth months, respectively. Based on the output of the

LSSVM-PSO and LSSVM-GA model, the completion time of the E dam project is predicted to be 57 and 59 months, respectively, and the completion cost is 1014000 and 1054000 dollars, respectively. In the meantime, the result of applying the Kalman filter in the time prediction on the validation data announces the initial warning for schedule delay in the seventh month. However, it does not announce a warning for increasing the project cost more than the planned program's cost. The Kalman model predicts the completion time and cost of the E dam project in the 15th month as 52 months and 623 thousand dollars, respectively.

Comparing the hybrid models with the Kalman model discloses that the hybrid models have more acceptable results with the periodic reports provided for the E project due to the training based on several dams' data. Unquestionably, it should be noted that regarding the delays that were used in several projects in the training phase of hybrid models, the resulting models are very conservative. This is while the Kalman model was developed only based on the project's limited data and past trends. Consequently, using hybrid models, particularly LSSVM-PSO, is exceedingly recommended to predict the time and cost of project completion.

### **Ethical Considerations compliance with ethical guidelines**

The cooperation of the participants in the present study was voluntary and accompanied by their consent.

### **Funding**

No funding.

### **Conflicts of interest**

The authors declared no conflict of interest.

## مقاله پژوهشی

## بکارگیری دانش مدیریت پروژه و الگوریتم هیبریدی در پیش‌بینی زمان و هزینه تکمیل پروژه‌های سدسازی

رضا بخشی<sup>۱</sup>، سینا فرد مرادی نیا<sup>۲\*</sup>، رسول جانی<sup>۳</sup>، رامین وفائی‌پور سرخابی<sup>۴</sup>  
 ۱- گروه مهندسی عمران، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران  
 ۲- مرکز تحقیقات رباتیک و فناوری‌های نرم، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران

## چکیده

**مقدمه:** یکی از وظایف مدیر پروژه، کنترل پروژه با استفاده از دو فاکتور زمان و هزینه می‌باشد. این دو اهرم از مهمترین کلیدهای کنترلی پروژه به منظور استفاده بهینه از منابع می‌باشد.

**روش:** در این تحقیق روش جدیدی با بکارگیری ارزش کسب شده و الگوریتم‌های هیبریدی برای پیش‌بینی زمان و هزینه تکمیل پروژه‌های سدسازی توسعه داده شد. برای این منظور گزارشات مالی و زمان‌بندی پنج سد A، B، C، D و E گردآوری شدند. ارزش کسب شده، پیشرفت واقعی، برنامه کسب شده و هزینه واقعی برای هر ماه از گزارشات این سدها استخراج شد. سپس از پارامترهای چهار سد (A تا D) به عنوان ورودی در توسعه مدل‌هایی برای پیش‌بینی زمان با استفاده از الگوریتم حداقل مربعات ماشین بردار پشتیبان (LSSVM) و تلفیق شده با الگوریتم‌های بهینه‌سازی ازدحام ذرات (PSO) و ژنتیک (GA) استفاده گردید.

**یافته‌ها:** مقایسه نتایج حاصل از این مدل‌های در مرحله آموزش نشان داد که الگوریتم LSSVM-PSO از دقت بالاتری برخوردار است. در ادامه برای توسعه مدل پیش‌بینی‌کننده هزینه با استفاده از الگوریتم‌های هیبریدی علاوه بر پارامترهای ورودی استفاده شده در مرحله، زمان پیش‌بینی شده توسط مدل‌ها نیز به عنوان ورودی الگوریتم‌ها تعریف شد. مقایسه نتایج این مرحله نیز نشان داد که الگوریتم LSSVM-PSO از دقت بالاتری نسبت به سایر مدل‌ها برخوردار است. در ادامه برای اعتبارسنجی مدل‌های توسعه داده شده، این مدل‌ها برای پیش‌بینی زمان و هزینه در پروژه ساخت سد E استفاده شد. نتایج نشان داد که الگوریتم LSSVM-PSO نسبت به سایر الگوریتم‌های هیبریدی از دقت بالاتری در تخمین زمان اتمام پروژه برخوردار است که نشان دهنده قابلیت تعمیم بالای این مدل برای پیش‌بینی زمان سایر پروژه‌های ساخت سد است. این در حالی است که برخلاف انتظار این مدل از دقت کمتری در پیش‌بینی هزینه اتمام پروژه ساخت سد E نسبت به مدل MLP-PSO برخوردار است که نشان می‌دهد پیش‌بینی هزینه از پیچیدگی بالاتری برخوردار است و بکارگیری مدل‌ها در این مسئله بایستی با احتیاط بیشتری صورت گیرد. برای ارزیابی بیشتر نتایج مدل‌های هیبریدی از روش احتمالاتی فیلتر کالمن نیز برای پیش‌بینی زمان و هزینه پروژه استفاده شد که نتایج آن حاکی از دقت بالاتر مدل‌های حاصل از الگوریتم‌های هیبریدی است.

**نتیجه‌گیری:** مدل LSSVM-PSO می‌تواند با دقت بالاتری نسبت به روش‌های مرسوم به پیش‌بینی زمان و هزینه پروژه بپردازد.

تاریخ دریافت: ۱۴۰۱/۰۹/۱۹

تاریخ داوری: ۱۴۰۱/۱۰/۲۶

تاریخ پذیرش: ۱۴۰۲/۱۰/۱۵

از دستگاه خود برای اسکن و خواندن مقاله به صورت آنلاین استفاده کنید



DOI:

[10.30495/wej.2024.31317.2374](https://doi.org/10.30495/wej.2024.31317.2374)

## واژه‌های کلیدی:

الگوریتم هیبریدی، ارزش کسب شده، پیش‌بینی هزینه، سدسازی

\* نویسنده مسئول: سینا فرد مرادی نیا

نشانی: ۱- گروه مهندسی عمران، واحد تبریز، دانشگاه آزاد اسلامی تبریز، ایران.

۲- مرکز تحقیقات رباتیک و فناوری‌های نرم، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران

تلفن: ۰۹۱۴۳۱۴۲۶۱۹

پست الکترونیکی: [fardmoradina@iaut.ac.ir](mailto:fardmoradina@iaut.ac.ir)

## مقدمه

تکمیل یک پروژه در زمان‌بندی مناسب با هزینه برنامه‌ریزی شده بسیار دشوار است و این امر مستلزم استفاده از مدل‌های پیش‌بینی‌کننده دقیق و بروزرسانی زمان‌بندی با پیشرفت پروژه است. وجود هر گونه خطا در پیش‌بینی اولیه زمان تکمیل پروژه می‌تواند بودجه انجام پروژه را تحت شعاع قرار دهد. برای این منظور لازم است عملکرد پروژه به صورت منظم اندازه‌گیری و مورد ارزیابی قرار گیرد و واریانس آن از خط مبنای برنامه‌ریزی شده مشخص گردد. پیش‌بینی اولیه قبل از آغاز پروژه به عنوان طرح مینا برای تکمیل پروژه در زمان و بودجه مشخص لحاظ می‌شود و از زمان شروع پروژه، عملکرد پروژه به لحاظ زمانی و هزینه نسبت به طرح اولیه مورد ارزیابی قرار می‌گیرد.

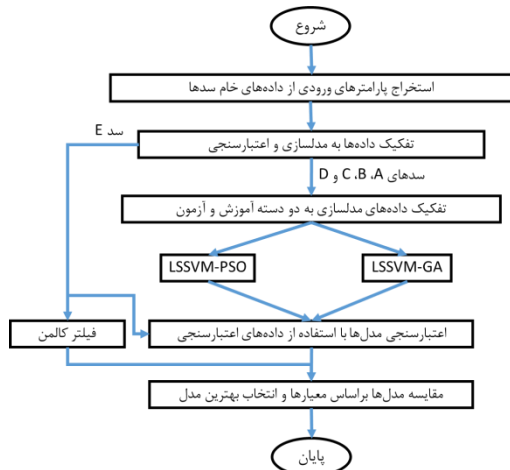
چالش‌های موجود در پروژه‌های ساخت و ساز، مدیران پروژه‌ها را مجبور به استفاده از تکنیک‌ها و مدل‌هایی برای کنترل و پیش‌بینی هزینه و زمان آن‌ها نموده است. براساس این پیش‌بینی، مدیر پروژه تعیین می‌کند که پروژه نیاز به اصلاح جزئی دارد یا نیاز به اصلاح اساسی و حل مشکلات دارد [۱]. محققان بسیاری به استفاده از روش‌های تجزیه و تحلیل آماری نظیر رگرسیون چندگانه، روش‌های مبتنی بر احتمالات مانند شبیه‌سازی مونت کارلو و روش‌های هوش مصنوعی روی آورده‌اند. تداوم در این پژوهش‌ها و ارائه مدل‌های جدید نشان می‌دهد که هنوز به یک مدل جامع و با دقت مناسب که برای کارفرما و پیمانکار رضایت بخش باشد حاصل نشده است. در جدول ۱ به برخی از تحقیقات انجام گرفته اشاره شده است.

جدول ۱- مطالعات انجام شده برای پیش‌بینی زمان و هزینه پروژه

محققین	روش مورد استفاده	شاخص‌های پیش‌بینی	نوع پیش‌بینی
آتالا و هگز [۲]	شبکه‌های عصبی و رگرسیون	متغیرهای محیطی	هزینه ساختمان
کیم و همکاران [۳]	روش رگرسیونی، شبکه‌های عصبی و استنتاج مبتنی بر مورد	ویژگی‌های ساختمان	هزینه ساختمان
ویلموت و می [۴]	شبکه‌های عصبی	مختصات برنامه اولیه پروژه	هزینه اتوبان
کیم و رینشمیت [۵]	فیلتر کالمن و روش بیزین	شاخص‌های ارزش کسب شده	زمان پروژه با داده‌های مصنوعی
سونمز [۶]	شبکه‌های عصبی	متغیرهای مربوط به ساختمان، مکان و شرایط ساخت وساز	هزینه ساختمان
وانگ و همکاران [۷]	ماشین بردار پشتیبان و شبکه عصبی	۶۴ متغیر مربوط به وضعیت ابتدایی پروژه	احتمال موفقیت هزینه و زمان ساختمان
چنگ و همکاران [۸]	ماشین بردار پشتیبان، منطق فازی و الگوریتم ژنتیک	شاخص‌های ارزش کسب شده	هزینه ساختمان
واترز و وانگوک [۹]	روش رگرسیون بردار پشتیبان	شاخص‌های ارزش کسب شده	زمان و هزینه
کیم [۱۰]	روش بیزین	شاخص‌های ارزش کسب شده	هزینه پروژه
مرتاجی و همکاران [۱۱]	تحلیل نقطه تغییر	شاخص‌های ارزش کسب شده	زمان و هزینه پروژه
اخباری [۱۲]	شبیه‌سازی مونت کارلو و شبکه عصبی	شاخص‌های ارزش کسب شده	زمان و هزینه پروژه با داده‌های مصنوعی
جیانگ [۱۳]	شبکه عصبی پسرو به همراه انتخاب ویژگی	۶ متغیر از وضعیت ابتدایی پروژه	هزینه پروژه
نجفی و همکاران [۱۴]	اعداد خاکستری	شاخص‌های ارزش کسب شده	زمان و هزینه پروژه
سلطان و اشرفی [۱۵]	روش‌های آماری	شاخص‌های ارزش کسب شده	زمان و هزینه پروژه
خلف و همکاران [۱۶]	الگوریتم بهینه‌سازی ازدحام ذرات	۸ متغیر مربوط به وضعیت ابتدایی پروژه	زمان و هزینه پروژه

کاهش ریسک تخمین زمان و هزینه پروژه	بکارگیری مفهوم طبقه‌بند	تصمیم‌گیری چند معیاره	اسدآبادی و زویکیل [۱۷]
زمان و هزینه پروژه	شاخص‌های ارزش کسب شده	شبکه عصبی پرسپترون چند لایه هیبریدی با الگوریتم‌های بهینه‌سازی ابتکاری	بخشی و همکاران [۱۸]

E بهره برده خواهد شد. در ادامه به تشریح آماده‌سازی داده‌ها و الگوریتم‌های هیبریدی مورد استفاده در این مطالعه، مطابق با روش انجام کار نشان داده شده شکل ۱، پرداخته خواهد شد.



شکل ۱- فرآیند انجام مطالعه.

### آماده‌سازی داده‌ها

برای مدل‌سازی زمان اتمام پروژه از ورودی‌های ارزش کسب شده (EVM)، پیشرفت واقعی، برنامه‌ریزی کسب شده (ESM) و هزینه واقعی استفاده خواهد شد. بنابراین برای اطلاعات گردآوری شده از عملیات ساخت پروژه‌های سد، بصورت ماهانه این داده‌ها استخراج شدند. پیشرفت و هزینه واقعی مورد استفاده در این مطالعه بصورت درصدی از زمان و هزینه برنامه‌ریزی شده اولیه در نظر گرفته شده است.

روش زمان‌بندی کسب شده روی سه متغیر عملکردی پایه- ارزش کسب شده (EV)، هزینه واقعی (AC) و ارزش برنامه‌ریزی شده (PV)، به منظور ارزیابی اینکه پروژه کجاست و کجا قرار بوده باشد- تکیه دارد. با استفاده از این متغیرها کارایی یک پروژه از نظر شاخص عملکرد زمان‌بندی ( $SPI=EV/PV$ ) و شاخص عملکرد هزینه ( $GPI=EV/AC$ ) اندازه‌گیری می‌شود. پیش‌بینی ارزش کسب شده براساس این اصل بنیادی است که عملکرد گذشته، بهترین شاخص در دسترس برای قضاوت در مورد عملکرد آینده پروژه است. برای نمونه، زمانی که کارایی آینده مشابه کارایی گذشته در نظر گرفته شود، تخمین در زمان تکمیل ( $EAC(t)$ ) با مشخص بودن بودجه در زمان اتمام (BAC) و کارایی هزینه گزارش شده در زمان  $t$ ، به صورت رابطه ذیل محاسبه می‌شود [۲۳-۲۴-۲۵]:

ارزیابی مطالعات پیشین نشان می‌دهد که در پژوهش‌های اندکی به پیش‌بینی همزمان زمان و هزینه اتمام پروژه پرداخته شده است. علاوه بر این مقایسه روش‌های آماری و هوش مصنوعی در پیش‌بینی زمان و هزینه پروژه‌ها، حاکی از دقت بالای مدل‌های هوشمند است. از میان روش‌های هوشمند، شبکه عصبی و ماشین بردار پشتیبان بیشتر مورد استفاده قرار گرفته است. هیبرید این الگوریتم‌ها با الگوریتم‌های بهینه‌سازی فراابتکاری در سایر مطالعات نشان داده است که استفاده از الگوریتم‌های فراابتکاری به عنوان الگوریتم‌های آموزش دهنده موجب بهبود دقت مدل‌های نهایی می‌گردد [۱۹-۲۰-۲۱]. بخشی و همکاران [۱۸] از مدل‌های هیبریدی شبکه عصبی پرسپترون چند لایه با الگوریتم ژنتیک و ازدحام ذرات برای پیش‌بینی زمان و هزینه پروژه استفاده کردند و نشان دادند که هیبرید الگوریتم‌های پیش‌بینی کننده و بهینه‌سازی می‌تواند عملکرد آن‌ها را در فرآیند پیش‌بینی بهبود ببخشد. براین اساس، در این مطالعه از هیبرید حداقل مربعات ماشین بردار پشتیبان ( $LSSVM^1$ ) با الگوریتم‌های بهینه‌سازی ازدحام ذرات ( $PSO^2$ ) و ژنتیک ( $GA^3$ ) برای پیش‌بینی زمان و هزینه پروژه‌های سدسازی با بکارگیری شاخص‌های ارزش کسب شده استفاده خواهد شد که براساس دانش ما تاکنون برای پیش‌بینی زمان و هزینه پروژه بکار نرفته است. بنابراین می‌توان گفت نوآوری تحقیق در استفاده از مدل هیبریدی LSSVM-PSO و LSSVM-GA در پیش‌بینی زمان و هزینه پروژه است [۲۲].

### روش انجام مطالعه

به منظور انجام این مطالعه اطلاعات مربوط به زمان‌بندی و هزینه‌های انجام پروژه‌های چهار سد ساخته شده در نواحی مختلف ایران گردآوری شد که به دلیل محرمانه بودن اطلاعات آن‌ها، این سدها A, B, C و D نامگذاری شدند. این سدها در مجموع دارای ۳۸۳ نقطه داده هستند که برای آموزش مدل‌های هیبریدی هوشمند در نظر گرفته شدند و از این تعداد ۳۰۶ نقطه داده (۸۰٪ کل داده) برای مرحله آموزش و ۷۷ نقطه داده (۲۰٪ کل داده) برای آزمون مدل‌های هیبریدی لحاظ شدند. از داده‌های سد دیگر (یعنی سد E) نیز برای اعتبارسنجی مدل‌ها استفاده خواهد شد. از شروع عملیات احداث سد E پانزده ماه می‌گذرد و گزارش پیشرفت ماه پانزدهم آن ارائه شده است؛ بنابراین تعداد ۱۵ نقطه داده برای این سد موجود است. براساس برنامه و طرح اولیه ارائه شده احداث کامل این سد حدود ۳۱ ماه زمان و در حدود ۵۹۰۱۴۶/۸۲ دلار هزینه دربر خواهد داشت. در این مطالعه برای اعتبارسنجی نتایج مدل‌های هیبریدی، از فیلتر کالمن نیز برای پیش‌بینی زمان و هزینه تکمیل پروژه

<sup>3</sup> Genetic Algorithm

<sup>1</sup> Least Square Support Vector Machine

<sup>2</sup> Particle Swarm Optimization

که در آن  $z_k$  بردار مشاهدات،  $H$  ماتریس مشاهدات و  $v_k$  بردار نوفه مشاهدات است که فرض می‌شود دارای توزیع گاوسی با میانگین صفر و کوواریانس  $R$ ، یعنی  $v_k \sim \mathcal{N}(0, R)$  باشد.

### حداقل مربعات ماشین بردار پشتیبان

ماشین‌های بردار پشتیبان ابزارهای محاسباتی قدرتمند برای یادگیری نظارت شده هستند که برای حل مسائل طبقه‌بندی و تخمین توابع غیرخطی توسط سوپکین و وندل‌وال معرفی شده‌اند [۳۰]. روش بردار پشتیبان برای رگرسیون به منظور حل یک مسئله بهینه‌سازی محدب، بلخص یک مسئله برنامه‌ریزی کوادراتیک، فرموله شده است. این امر در نتیجه بکارگیری تابع  $\varepsilon$ -insensitive loss و پنینگ حاصل می‌شود که در آن، مسئله بهینه‌سازی به عنوان نامعادله مقید فرموله و در شرایط Mercer به منظور ارتباط فضای ویژگی‌های غیرخطی نگاشت شده به تابع کرنل انتخاب شده، بهره‌برداری می‌شود. یکی از معایب رگرسیون بردار پشتیبان هزینه محاسباتی بالای آن به دلیل حل یک مسئله برنامه‌ریزی کوادراتیک است. معرفی نسخه حداقل مربعات رگرسیون بردار پشتیبان یک پیشرفت بزرگ برای رفع این عیب بوده است. در واقع، در این نسخه نیاز به حل یک سیستم خطی به جای مسئله برنامه‌ریزی کوادراتیک در رگرسیون بردار پشتیبان استاندارد است.

با در نظر گرفتن یک دیتاست آموزش  $\{x_i, y_i\}, i = 1, 2, \dots, N$ ، که در آن داده‌های ورودی دارای  $n$  بعد  $x_i \in R^n$  و داده خروجی یک بعدی  $y_i \in R$ ، در یک الگوریتم رگرسیون بردار پشتیبان، رگرسیونی از فرمول بردارهای ماشین پشتیبان به صورت رابطه (۶) شکل می‌گیرد.

$$y = w^T \phi(x) + b \quad (6)$$

که در آن  $w$  بردار وزن،  $b$  ترم بایاس و  $\phi(x)$  تابع کرنلی است داده‌های ورودی را به فضای ویژگی با ابعاد بالاتر نگاشت می‌کند. تابع کرنل می‌تواند خطی، چند جمله‌ای، تابع شعاعی پایه و پرسپترون چند لایه باشد. تابع هزینه  $C$  نیز به صورت رابطه (۷) کمینه‌سازی می‌شود.

$$\min C(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (7)$$

$$\text{Subject to: } y = w^T \phi(x) + b + e_i$$

در اینجا  $\gamma$  پارامتر تنظیم است که برای بدست آوردن مدلی با دقت بالا نیاز است مقدار بهینه آن تعیین شود. این پارامتر بین کاهش خطای یادگیری و صافی سازش ایجاد می‌کند. این بهینه‌سازی معمولاً با الگوریتم لاگرانژین فرموله شده به شکل رابطه (۸) حل می‌شود.

$$L(w, b, e, \alpha) = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i \{w^T \phi(x) + b + e_i - y_i\} \quad (8)$$

$$EAC(t) = BAC/CPI(t) \quad (1)$$

روش ارزش برنامه‌ریزی روی ماتریس عملکردی، مشابه ارزش کسب شده، تکیه دارد؛ اما برنامه‌ریزی کسب شده روی واریانس برنامه‌ریزی روی بعد زمان بجای بعد ارزش منحنی پیشرفت تجمعی تمرکز دارد. ارزش برنامه‌ریزی در زمان  $t$  ( $ES(t)$ ) به عنوان زمان برنامه‌ریزی شده برای تکمیل ارزش حاصل شده ( $EV(t)$ ) گزارش شده در زمان  $t$  گزارش‌گیری تعریف می‌شود. با مشخص بودن منحنی ارزش برنامه‌ریزی شده ( $PV(t)$ ) و ارزش کسب شده ( $EV(t)$ ) در زمان  $t$ ،  $ES(t)$  با رابطه (۲) محاسبه می‌شود [۲۳-۲۴-۲۵]:

$$PV(ES(t)) = EV(t) \quad \text{or} \quad ES(t) = PV[EV(t)]^{-1} \quad (2)$$

زمان تخمینی در اتمام پروژه ( $EDAC(t)$ ) پیش‌بینی شده در زمان  $t$ ، براساس ترم مدت برنامه‌ریزی شده پروژه ( $PD$ ) و برنامه‌ریزی کسب شده در زمان  $t$  محاسبه می‌شود (رابطه (۳)) [۲۳].

$$EDAC(t) = t + \frac{PD - ES(t)}{ES(t)/t} = \frac{PD}{ES(t)/t} \quad (3)$$

### الگوریتم‌های پیش‌بینی‌کننده

به دلیل دقت بالا، سرعت بالای حل مسئله و همچنین مصرف کمتر منابع محاسباتی [۲۶] الگوریتم حداقل مربعات خطا ماشین بردار پشتیبان، در این تحقیق از این الگوریتم پیش‌بینی‌کننده استفاده خواهد شد. البته برای ارزیابی نتایج این روش‌ها از فیلتر کالمن نیز بهره برده خواهد شد. در ادامه به تشریح هر یک از این الگوریتم‌ها پرداخته خواهد شد.

### فیلتر کالمن

فیلتر کالمن یک الگوریتم بازگشتی است که برای تخمین وضعیت حقیقی، اما پنهان سیستم دینامیکی در حضور مشاهدات نوفه‌دار استفاده می‌شود. فیلتر کالمن فرم نسبتاً ساده‌ای دارد و نیازمند قدرت محاسباتی کوچکی است. از اینرو به صورت گسترده در مسائل پیش‌بینی و کنترل مورد استفاده قرار گرفته است [۲۷].

فیلتر کالمن برای تخمین حالت‌ها<sup>۴</sup> مبتنی بر سیستم‌های دینامیکی خطی در فرمت فضای حالت استفاده می‌شود. فرآیند مدل ارزیابی حالت از زمان  $k-1$  تا زمان  $k$  به صورت ذیل تعریف می‌شود [۲۸]:

$$x_k = Fx_{k-1} + Bu_{k-1} + w_{k-1} \quad (4)$$

که در آن  $F$  ماتریس انتقال حالت است که روی بردار حالت قبلی، یعنی  $x_{k-1}$  اعمال می‌شود و  $B$  ماتریس کنترل ورودی است که روی بردار کنترل  $u_{k-1}$  اعمال می‌شود و  $w_{k-1}$  بردار نوفه فرآیند است که فرض می‌شود دارای توزیع گاوسی با میانگین صفر و کوواریانس  $Q$ ، یعنی  $w_{k-1} \sim \mathcal{N}(0, Q)$  است.

رابطه بین حالت و مشاهدات در گام فعلی زمان با جفت شدن مدل فرآیند با مدل مشاهدات به صورت رابطه (۵) توصیف می‌شود [۲۹].

$$z_k = Hx_k + v_k \quad (5)$$

4 states



طبیعی توسعه داده شده‌اند. این الگوریتم‌ها بر روی ساختارهای رشته‌ای، مانند ساختارهای بیولوژیکی، عمل می‌کنند که با گذر زمان مطابق قواعد بقای بهترین‌ها با استفاده از نرخ تغییر ساختار اطلاعات تصادفی تغییر می‌یابند. از اینرو، در هر نسل، یک رشته جدید با استفاده از بخشی از مناسب‌ترین اعضای مجموعه قدیمی ایجاد می‌شود.

الگوریتم ژنتیک همانند سایر الگوریتم‌های تاملی با ایجاد یک جمعیت اولیه از کروموزم‌ها که به صورت تصادفی تولید شده‌اند آغاز می‌شود. سپس این کروموزم‌ها با تابع هدف ارزیابی می‌شوند و به عنوان کروموزم‌های والدین براساس مقدار تابع هزینه‌ای که دارند در فرآیند تولید مثل شرکت می‌کنند. شانس کروموزم‌هایی که دارای مقدار هزینه کمتری هستند نسبت به سایر کروموزم‌ها بیشتر است. در ادامه از عملگر جهش برای ترکیب جواب‌های فعلی استفاده می‌شود. با توجه به ماهیت تصادفی این عملگر احتمال گیر افتادن در نقاط بهینه محلی را به حداقل می‌رساند. استفاده از نخبه‌گرایی می‌تواند تضمین نماید که با اعمال این فرآیند بهترین جمعیت بعدی حداقل به خوبی جمعیت فعلی خواهد بود [۳۲].

الگوریتم بهینه‌سازی ازدحام ذرات (PSO) یکی از مهم‌ترین الگوریتم‌های بهینه‌سازی هوشمند در حوزه هوش ازدحامی است که توسط کندی و ابرهات ارائه شد. این الگوریتم متاهیورستیک برای بهینه‌سازی توابع پیوسته غیرخطی مناسب است [۳۳]. در الگوریتم بهینه‌سازی ازدحام ذرات اعضای جمعیت به صورت مستقیم با هم در ارتباط هستند و از طریق اشتراک اطلاعات با یکدیگر و یادآوری بهترین جواب خود، به حل مسئله می‌پردازند [۲۰].

هدف الگوریتم بهینه‌سازی ازدحام ذرات تعیین موقعیت مکانی است که به ازای آن تابع هزینه دارای کمترین مقدار است. این الگوریتم متشکل از تعدادی ذرات است که هر کدام دارای پنج ویژگی سرعت (V)، موقعیت (x)، هزینه، بهترین موقعیت (Pb) و کمترین هزینه هستند. در اولین تکرار، سرعت و موقعیت هر ذره به صورت تصادفی تعیین و براساس موقعیت آن، مقدار هزینه مشخص می‌شود. هزینه بدست آمده در تکرار اول به عنوان کمترین هزینه و موقعیت فعلی هر ذره به عنوان بهترین موقعیت آن در نظر گرفته می‌شوند. براساس مقایسه کمترین هزینه هر ذره با سایر ذرات، بهترین موقعیتی ( $G_B$ ) که دارای کمترین هزینه در اجتماع ذرات است مشخص می‌شود. در تکرار بعدی (t+1) سرعت ذره نام براساس موقعیت، سرعت و بهترین موقعیت خود و همچنین بهترین موقعیت اجتماع در تکرار قبلی (t) با استفاده از رابطه (۱۷) تعیین می‌شود.

$$V_i(t+1) = wV_i(t) + c_1r_1(Pb_i(t) - x_i(t)) + c_2r_2(G_B(t) - x_i(t)) \quad (17)$$

در این رابطه: w وزنی است که به عنوان اینرسی شناخته شده و برای تنظیم سرعت ذرات در هر تکرار استفاده می‌شود. c1 و c2 بترتیب ضریب یادگیری فردی و اجتماعی هستند و،  $r_1$  و  $r_2$  اعداد تصادفی یکنواخت در بازه [۰، ۱] می‌باشند [۳۳].

که در آن  $\alpha_i$  ضرب لاگرانژی هستند که می‌توانند مثبت یا منفی باشند. این‌ها با محدودیت‌های برابری کاهن-تاکلر که در روابط (۹) تا (۱۴) لیست شده‌اند تعیین می‌شوند [۳۱].

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i \varphi(x_i) \quad (9)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \quad (10)$$

$$\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i, \quad i = 1, 2, \dots, N \quad (11)$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - \gamma = \gamma e_i, \quad i = 1, 2, \dots, N \quad (12)$$

بنابراین بردار وزن به صورت زیر حاصل خواهد شد:

$$w = \sum_{i=1}^N \alpha_i \varphi(x_i) = \sum_{i=1}^N \gamma e_i \varphi(x_i) \quad (13)$$

وزن‌ها به عنوان ترکیب‌های خطی از مضارب لاگرانژی برای داده‌های ورودی آموزش تعریف می‌شوند جایگذاری رابطه (۱۳) در رابطه (۶)، منجر به ایجاد رابطه (۱۴) می‌شود که در آن  $K(x_i, x)$  تابع کرنل می‌باشد. عملکرد حداقل مربعات رگرسیون بردار پشتیبان بسیار وابسته به تابع کرنل مورد استفاده و مقادیر هابیر پارامترها است [۲۰].

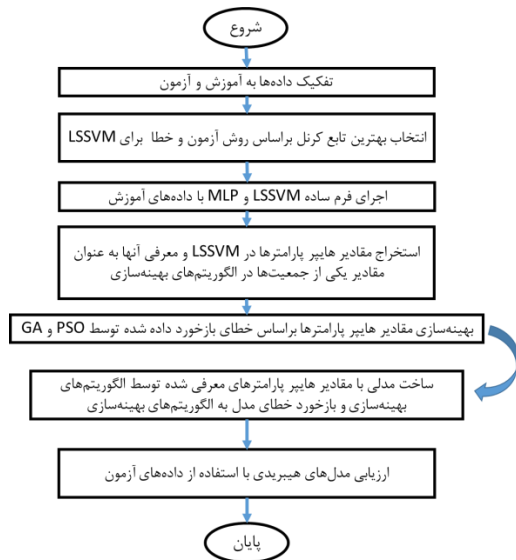
$$y = \sum_{i=1}^N \alpha_i \varphi(x_i)^T \varphi(x) + b = \sum_{i=1}^N \alpha_i K(x_i, x) + b \quad (14)$$

### الگوریتم‌های بهینه‌سازی

الگوریتم‌های بهینه‌سازی را می‌توان به دو دسته الگوریتم‌های دقیق و الگوریتم‌های تقریبی طبقه‌بندی نمود. هرچند الگوریتم‌های دقیق قادر به یافتن جواب بهینه به صورت دقیق هستند اما در مورد مسائل بهینه‌سازی سخت کارایی لازم را ندارند. در اینگونه مسائل استفاده از الگوریتم‌های بهینه‌سازی فرابتکاری از دسته الگوریتم‌های تقریبی نشان داده است که می‌تواند از دقت و کارایی بالایی برخوردار باشد. در این مطالعه، به دلیل نیاز به تنظیمات کمتر، راحتی پیاده‌سازی و همچنین کارایی بالای محاسباتی الگوریتم بهینه‌سازی ازدحام ذرات [۳۱] از این الگوریتم برای آموزش الگوریتم حداقل مربعات ماشین بردار پشتیبان بهره برده خواهد شد. برای قیاس نتایج این هیبریدسازی، الگوریتم حداقل مربعات ماشین بردار پشتیبان با الگوریتم ژنتیک، که یکی از پرکاربردترین الگوریتم‌های بهینه‌سازی است، نیز هیبرید خواهد شد. در ادامه به تشریح هر یک از این الگوریتم پرداخته خواهد شد.

### الگوریتم ژنتیک

الگوریتم‌های ژنتیک الگوریتم‌های جستجوی تصادفی هستند که براساس تلاش‌های انجام شده برای کپی مکانیزم انتخاب و ژنتیک



شکل ۲- نحوه تعیین مقادیر بهینه هایپر پارامترهای الگوریتم‌های پیش‌بینی کننده با استفاده از الگوریتم‌های بهینه‌سازی فراابتکاری

### نتایج

برای انجام مدلسازی هوشمند به منظوره پیش‌بینی زمان و هزینه تکمیل پروژه، ابتدا براساس داده‌های خام مربوط به زمان‌بندی و هزینه‌های انجام شده در هر ماه، پارامترهای زمان‌بندی کسب شده (ماه)، ارزش کسب شده (دلار)، پیشرفت واقعی (%، هزینه واقعی (%، زمان (ماه) و نسبت هزینه محاسبه شدند. در جداول ۲ و ۳ برخی شاخص‌های آماری توصیفی برای پارامترهای ورودی و هدف به ترتیب برای پیش‌بینی زمان و هزینه با الگوریتم‌های هیبریدی ارائه شده است. با توجه به ناهمگونی مقیاس و واحد پارامترهای ورودی، این داده‌ها با استفاده از رابطه (۱۹) بین صفر و یک نرمالایز شدند. البته بجای هزینه در این مطالعه از نسبت هزینه- که نسبت هزینه انجام شده تجمعی تا ماه مورد محاسبه به مقدار هزینه لازم برای اتمام پروژه در ماه‌های آتی براساس برنامه اولیه می‌باشد- استفاده شد.

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (19)$$

در این رابطه  $X_n$  مقدار نرمالایز شده،  $X$  مقدار اندازه‌گیری شده یا محاسبه شده و  $X_{max}$  و  $X_{min}$  به ترتیب مقادیر کمینه و بیشینه پارامتر مورد بررسی هستند.

موقعیت جدید ذره در تکرار فعلی به اندازه سرعت جدید از موقعیت در تکرار قبلی فاصله خواهد داشت. این فرآیند در رابطه (۱۸) به وضوح نشان داده شده است.

$$x_i(t+1) = x_i(t) + V_i(t+1) \quad (18)$$

براساس موقعیت جدید، هزینه برای هر ذره مجدد محاسبه و با مقایسه هزینه جدید با کمترین هزینه خود، در صورتی که کمترین هزینه بیشتر از هزینه در تکرار فعلی باشد مقدار کمترین هزینه و بهترین موقعیت ذره بروزرسانی خواهد شد. همزمان با اعمال این عملیات هزینه فعلی هر ذره نیز با کمترین هزینه اجتماع در تکرار قبلی مقایسه خواهد شد و در صورتی که مقدار هزینه فعلی ذره کمتر از هزینه اجتماع باشد مقدار بهترین موقعیت اجتماع نیز به روزرسانی خواهد شد. این عملیات تا رسیدن به شرط خاتمه ادامه می‌یابد.

### هیبرید الگوریتم‌های بهینه‌ساز و پیش‌بینی کننده

علاوه بر، تعیین نوع تابع کرنل مناسب در LSSVM، اختصاص مقادیر مناسب به هایپر پارامترها نیز عملکرد آن‌ها را به شدت تحت شعاع قرار می‌دهد. بنابراین همراه با تعیین نوع تابع کرنل مناسب، لازم است مقادیر بهینه‌ای به هایپر پارامترها تخصیص داده شود. از طرفی مطالعات انجام شده در مقایسه الگوریتم‌های بهینه‌سازی کلاسیک و متاهوریستیک نشان داد که الگوریتم‌های متاهوریستیک از قدرت بالایی در دستیابی به جواب بهینه در مسائل غیرخطی و پیچیده برخوردارند. بنابراین از الگوریتم‌های متاهوریستیک GA، PSO و برای بهینه‌سازی مقادیر هایپر پارامترها استفاده خواهد شد.

در این مطالعه ابتدا، نوع تابع کرنل LSSVM براساس روش سعی و خطا تعیین خواهد شد. براین اساس تعداد هایپر پارامترها یا به عبارتی تعداد متغیرهای تصمیم‌گیری در الگوریتم‌های بهینه‌سازی مشخص می‌شود. در هر تکرار، الگوریتم‌های بهینه‌سازی مقادیری را برای هایپر پارامترها تعیین خواهند کرد. براساس مقادیر تعیین شده، مدلسازی با داده‌های آموزش انجام می‌شود. با مقایسه خروجی مدل با مقادیر اندازه‌گیری شده خطا محاسبه می‌شود. این مقدار خطا به الگوریتم متاهوریستیک بازخورد داده می‌شود تا براساس تابع هدف تعریف شده (کمینه‌سازی خطا) مقادیر هایپر پارامترها در تکرار بعدی بهبود یابد. این حلقه تا محقق شدن شرط خاتمه در جریان است. براساس مقادیر هایپر پارامترها در آخرین تکرار، مدل ایجاد شده و با داده‌های آزمون مورد ارزیابی قرار خواهد گرفت. شکل ۲. Error! Reference source not found. نحوه هیبرید الگوریتم‌های بهینه‌سازی با الگوریتم‌های پیش‌بینی کننده را در قالب فلوجارت نشان می‌دهد.

جدول ۲- توصیف آماری متغیرهای ورودی برای پیش بینی زمان.

ویژگی هدف	ویژگی‌های ورودی				شاخص آماری
	زمان	هزینه واقعی	پیشرفت واقعی	ارزش کسب شده	
T	AC	AP	EV	ES	علائم
ماه	درصد	درصد	دلار	ماه	واحد
۶۳/۴۲	۰/۰۱	۰/۰۱	۵/۴۹ میلیون	۲۴/۲۸	میانگین
۴۴/۵۹	۰/۰۲	۰/۰۲	۸/۵۴ میلیون	۲۱/۶۴	انحراف استاندارد
۱/۰۰	۰	۰	۰	۰/۰۳	کمینه
۱۶۷/۰۰	۰/۳۰	۰/۱۶	۳۲/۱۰ میلیون	۸۱/۲۹	بیشینه
۰/۵۲۷۰	۷/۱۲۳۴	۳/۶۸۶۲	۱/۶۰۳۰	۰/۹۸۵۰	چولگی
-۰/۷۸۷۹	۷۴/۲۸۴۴	۱۷/۶۵۰۴	۱/۲۷۳۱	-۰/۲۱۱۵	کشیدگی

جدول ۳- توصیف آماری متغیرهای ورودی برای پیش بینی هزینه.

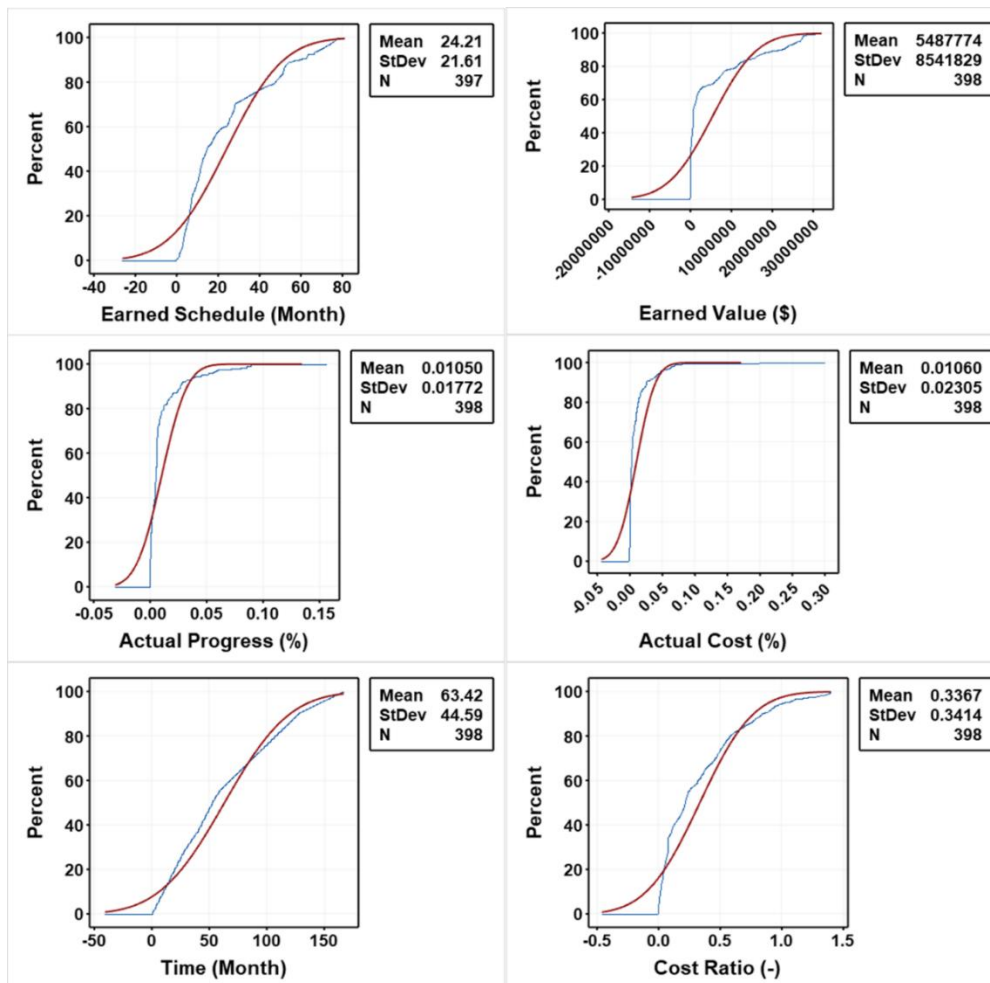
ویژگی هدف	ویژگی‌های ورودی				شاخص آماری
	نسبت هزینه	زمان	هزینه واقعی	پیشرفت واقعی	
CR	T	AC	AP	EV	ES
-	ماه	درصد	درصد	دلار	ماه
۰/۳۴	۶۳/۴۲	۰/۰۱	۰/۰۱	۵/۴۹ میلیون	۲۴/۲۸
۰/۳۴	۴۴/۵۹	۰/۰۲	۰/۰۲	۸/۵۴ میلیون	۲۱/۶۴
۰	۱/۰۰	۰	۰	۰	۰/۰۳
۱/۴۰	۱۶۷/۰۰	۰/۳۰	۰/۱۶	۳۲/۱۰ میلیون	۸۱/۲۹
۱/۲۱۰۰	۰/۵۲۷۰	۷/۱۲۳۴	۳/۶۸۶۲	۱/۶۰۳۰	۰/۹۸۵۰
۰/۷۹۰۰	-۰/۷۸۷۹	۷۴/۲۸۴۴	۱۷/۶۵۰۴	۱/۲۷۳۱	-۰/۲۱۱۵

تابع توزیع تجمعی برای زمان به صورت  $t < 14$  حدود ۱۴٪ و برای  $t < 85$  حدود ۳۲٪ می‌باشد. مقدار تابع توزیع تجمعی برای نسبت هزینه به صورت  $CR < 0.039$  حدود ۱۸٪ و برای  $CR < 0.666$  حدود ۶۵٪ و برای باقیماندهی داده‌ها، این مقدار  $CR > 0.666$  در حدود ۱۷٪ می‌باشد. براساس شکل ۳ میزان تابع توزیع تجمعی برای چهار پارامتر متغیر شامل برنامه کسب شده، ارزش کسب شده، زمان و نسبت هزینه به طور نرمال توزیع می‌شوند. مقدار تابع توزیع تجمعی برای پیشرفت واقعی به صورت  $AP < 0.0027$  حدود ۳۳٪ و برای  $AP\% < 0.037$  حدود ۶۰٪ و برای باقیماندهی داده‌ها این مقدار  $AP\% > 0.037$  در حدود ۷٪ می‌باشد. مقدار تابع توزیع تجمعی برای هزینه واقعی به صورت  $AC\% < 0.0015$  حدود ۳۵٪ و برای  $AC\% < 0.058$  حدود ۷۹٪ و برای باقیماندهی داده‌ها این مقدار  $AC\% > 0.058$  در حدود ۴٪ می‌باشد. براساس شکل ۴ میزان تابع توزیع تجمعی برای این دو پارامتر به طور نرمال توزیع نمی‌شوند.

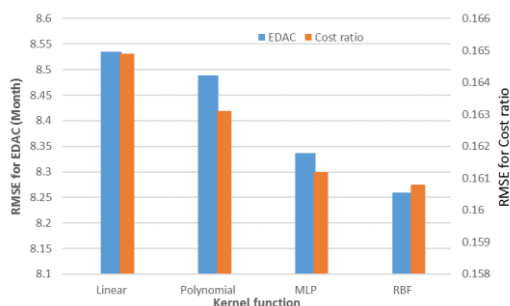
به منظور توصیف داده‌های ورودی و خروجی (پیش‌بینی زمان و هزینه) در مقاله از توابع توزیع تجمعی استفاده شده است. در شکل ۳ نمودار توزیع ۳۹۸ مجموعه داده ورودی و خروجی استفاده شده در این مقاله نشان داده شده است. مقدار تابع توزیع تجمعی با استفاده از رابطه (۲۰) محاسبه شد.

$$F_X(x) = P(X \leq x), \text{ for all } x \in R \quad (20)$$

در این رابطه  $x$  دامنه مقدار متغیر داده؛  $X$  مقدار متغیر  $x$  در یک رکورد داده خاص و  $R$  مجموعه داده سوابق داده‌ها می‌باشد. با توجه به اطلاعات موجود در شکل ۳، مقدار تابع توزیع تجمعی برای برنامه کسب شده به صورت  $ES < 6.6$  حدود ۲۲٪ و برای  $ES < 37.8$  حدود ۵۳٪ و برای باقیماندهی داده‌ها این مقدار  $ES > 37.8$  در حدود ۱۵٪ می‌باشد. مقدار تابع توزیع تجمعی برای ارزش کسب شده به صورت  $EV < 209323$  حدود ۲۸٪ و برای  $EV < 14020973$  حدود ۵۶٪ و برای باقیماندهی داده‌ها این مقدار  $EV > 14020973$  در حدود ۱۶٪ می‌باشد. مقدار



شکل ۳- تابع توزیع تجمعی متغیرهای ورودی و مقادیر خروجی برای پیش‌بینی هزینه و زمان: برنامه کسب شده، ارزش کسب شده، پیشرفت واقعی، هزینه واقعی، زمان و نسبت هزینه (خط آبی) و برای توزیع‌های طبیعی که با میانگین صفر و انحراف معیار یک تعریف می‌شوند (خط قرمز).



شکل ۴- مقایسه عملکرد LSSVM ساده با توابع کرنل مختلف در تخمین EDAC و Cost ratio براساس RMSE

با توجه به مشخص شدن نوع تابع کرنل در الگوریتم LSSVM، الگوریتم PSO و GA بایستی مقادیر بهینه  $\gamma$  و  $\sigma^2$  را، با هدف کمینه‌سازی خطای بین مقادیر پیش‌بینی و واقعی، مشخص نمایند. بنابراین تعداد متغیرهای تصمیم در این الگوریتم‌های بهینه‌سازی برابر دو خواهد بود. پارامتر  $\gamma$  برای افزایش عملکرد مدل روی داده‌های آموزش و کاهش پیچیدگی

برای انجام مدلسازی لازم است ابتدای کار، نوع تابع کرنل مشخص گردد. بدین منظور الگوریتم LSSVM با توابع کرنل مختلف به صورت جداگانه برای ساخت مدل با استفاده از داده‌های مدلسازی بکار برده شد. در شکل ۴ عملکرد LSSVM ساده با توابع کرنل مختلف براساس معیار RMSE نشان داده شده است. همانطور که در این شکل مشاهده می‌شود LSSVM با تابع کرنل RBF نسبت به سایر توابع کرنل از دقت بالاتری در تخمین EDAC و cost ratio برخوردار است. از طرفی مشاهده می‌شود که خطای مدل‌هایی با توابع کرنل RBF و MLP نسبت به مدل‌هایی با توابع کرنل Linear و Polynomial بصورت چشمگیری کمتر است. این امر نشان می‌دهد که رابطه غیرخطی پیچیده‌ای بین ورودی‌های و خروجی‌های وجود دارد.

اعتبارسنجی در تخمین زمان اتمام پروژه برای مدل‌های LSSVM-PSO و LSSVM-GA در شکل ۶ نشان داده شده است. همانطور که در این شکل مشاهده می‌شود هر دو مدل در مراحل آموزش و آزمون از عملکرد بالایی برخوردار بوده‌اند. به نظر می‌رسد نتایج حاصل از اعمال این الگوریتم‌ها روی داده‌های اعتبارسنجی نیز از دقت خوبی برخوردار باشد هرچند پراکندگی نقاط داده در این مرحله نسبت به مراحل آموزش و آزمون بیشتر است. مقایسه مقادیر کمی معیارهای ارزیابی برای الگوریتم‌های LSSVM-PSO و LSSVM-GA در تخمین زمان اتمام پروژه‌ها به تفکیک دیتاست‌ها در جدول ۵ ارائه شده است. همانطور که در این جدول مشاهده می‌شود خطای LSSVM-PSO در هر سه مرحله آموزش، آزمون و اعتبارسنجی نسبت به LSSVM-GA کمتر است. علاوه بر این، اختلاف کم مقادیر خطا در دو مرحله آموزش و آزمون این مدل نسبت به LSSVM-GA، نشان دهنده قابلیت تعمیم این مدل بر روی سایر پروژه‌ها یا داده‌های دیده نشده است. دقت بالاتر مدل LSSVM-PSO نسبت به LSSVM-GA در مرحله اعتبارسنجی تأیید کننده این ادعا است.

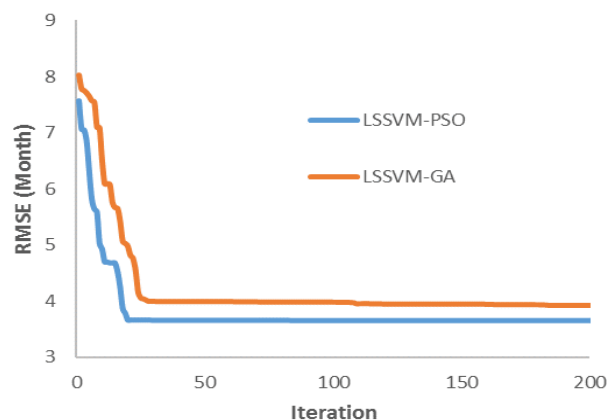
مدل استفاده می‌شود. مقدار بالا برای این پارامتر نشان دهنده little regularization است که منجر به حصول یک مدل غیرخطی تر خواهد شد. پارامتر  $\sigma^2$  روی تعداد همسایگی‌ها در مدل تأثیر می‌گذارد؛ بنابراین، مقدار بالای این پارامتر به معنی همسایه‌های بیشتر در مدل است از اینرو یک مدل غیرخطی تر بدست خواهد آمد. بنابراین عملکرد و رفتار مدل بسیار وابسته به مقادیر این پارامترها است.

برای تنظیم پارامترهای قابل کنترل در الگوریتم PSO و GA از روش آزمون و خطا استفاده شد. در جدول ۴ نتایج تحلیل حساسیت روی پارامترهای قابل تنظیم این الگوریتم‌ها نشان داده شده است. تعداد تکرارهای الگوریتم‌ها برابر ۲۰۰ در نظر گرفته شد؛ هرچند بررسی‌ها نشان داد که در تکرارهای کمتر الگوریتم‌ها به جواب بهینه همگرا می‌شوند.

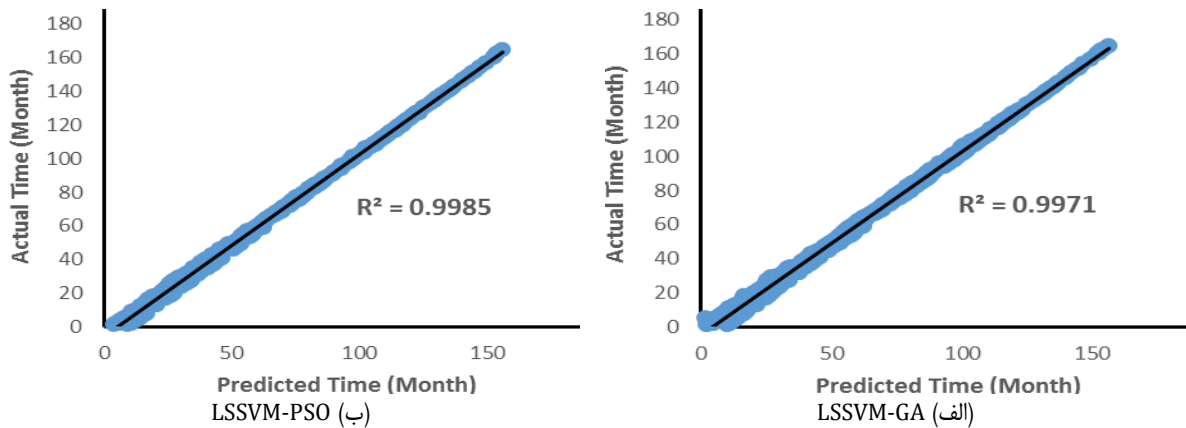
در شکل ۵ روند کاهش خطا در تکرارهای مختلف این الگوریتم‌ها برای تخمین زمان اتمام پروژه در مرحله آموزش نشان داده شده است. همانطور که در این شکل مشاهده می‌شود هر دو الگوریتم در تعداد تکرارهای کمتر از ۵۰ به جواب همگرا شده است. نمودارهای متقاطع مقادیر پیش‌بینی شده و واقعی برای مراحل آموزش، آزمون و

جدول ۴- مقادیر پارامترهای کنترلی الگوریتم‌های بهینه‌سازی GA و PSO برای تخمین زمان و هزینه پروژه‌ها

مقادیر پارامترهای قابل کنترل الگوریتم‌های بهینه‌سازی به تفکیک الگوریتم	پارامترها	الگوریتم بهینه‌سازی
<b>LSSVM</b>		
۱۰۰	اندازه ازدحام	الگوریتم ازدحام ذرات (PSO)
۲/۰۵	ضریب یادگیری فردی	
۲/۰۵	ضریب یادگیری جمعی	
۰/۹۷	ضریب اطلاق‌کاک	
۱۱۰	جمعیت	الگوریتم ژنتیک (GA)
چرخ رولت	روش انتخاب	
یکنواخت (p = 1)	تقاطع	
یکنواخت (p=0.07)	جهش	
۰/۱۰	نرخ جهش	
۲	فشار انتخاب	



شکل ۵- نمودار تغییرات خطا در تکرارهای مختلف برای الگوریتم‌های LSSVM-GA و LSSVM-PSO



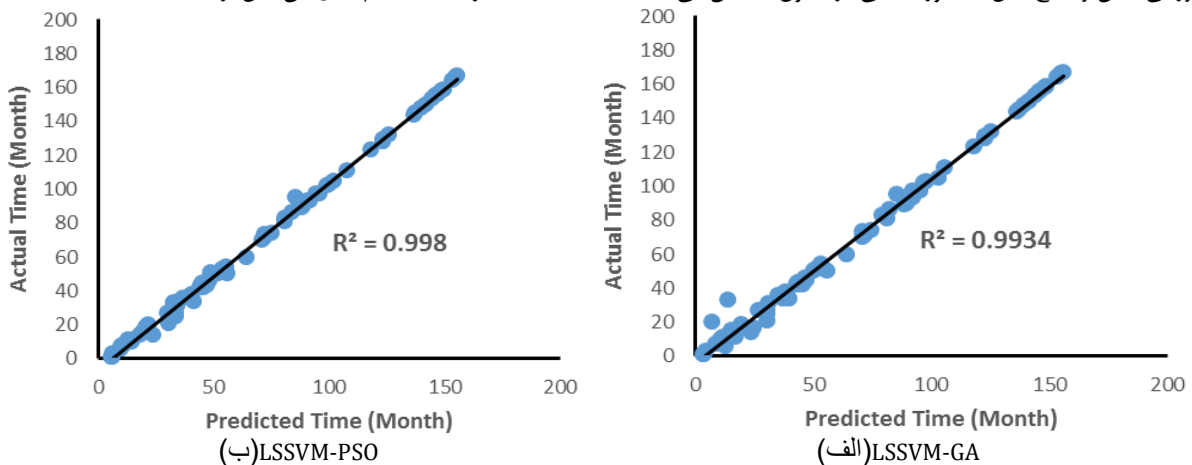
شکل ۶- نمودارهای متقاطع مقادیر پیش‌بینی شده و واقعی در تخمین زمان اتمام پروژه برای الگوریتم‌های LSSVM-PSO، LSSVM-GA، در مرحله آموزش

جدول ۵- نتایج ارزیابی مدل‌های LSSVM-PSO، LSSVM-GA در مرحله آموزش براساس معیارهای مختلف در تخمین زمان اتمام پروژه‌ها

نام مدل	APD (%)	AAPD (%)	SD	RMSE	R-square
LSSVM-PSO	-۱۸/۱۸۵۶	۲۰/۲۸۷۸	۳/۶۴۳۸	۳/۶۶۱۴	۰/۹۹۸۵
LSSVM-GA	-۱۶/۰۳۰۱	۱۹/۱۳۳۱	۳/۹۲۴۹	۳/۹۲۰۳	۰/۹۹۷۱

در مرحله آزمون نیز مدل‌های هیبریدی LSSVM-PSO نسبت به نوع متناظرش LSSVM-GA از دقت بالاتری برخوردار بودند. این امر نشان می‌دهد الگوریتم LSSVM-PSO از توانایی بالاتری در مدل‌سازی مسائل غیرخطی و پیچیده دارا است. الگوریتم هیبریدی LSSVM-PSO نسبت به LSSVM-GA در این مرحله نیز از دقت بالاتری برخوردار است که حکایت از قابلیت تعمیم بالای این مدل دارد.

نتایج بکارگیری مدل‌های آموزش دیده برای پیش‌بینی زمان اتمام پروژه بر روی داده‌های آزمون در شکل ۷ نشان داده شده است. همانطور که در این شکل مشاهده می‌شود نقاط داده در مدل‌های هیبریدی مربوط به الگوریتم LSSVM-PSO از پراکندگی کمتری نسبت به LSSVM-GA برخوردار هست و تمامی آن‌ها نزدیک به خط برازش شده قرار دارند؛ بررسی دقیق‌تر نتایج مدل‌ها بصورت کمی در جدول ۶ نشان می‌دهد

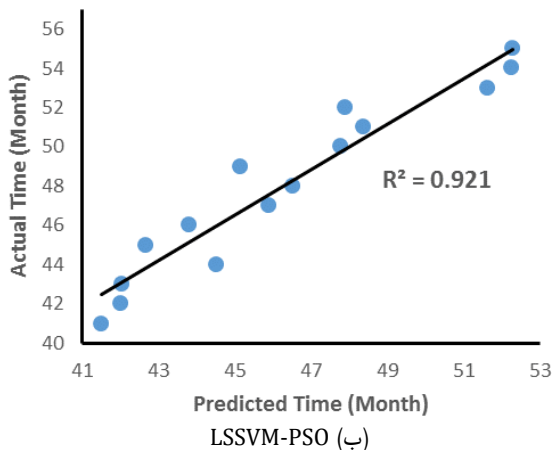


شکل ۷- نمودارهای متقاطع مقادیر پیش‌بینی شده و واقعی در تخمین زمان اتمام پروژه برای الگوریتم‌های LSSVM-PSO، LSSVM-GA در مرحله آزمون

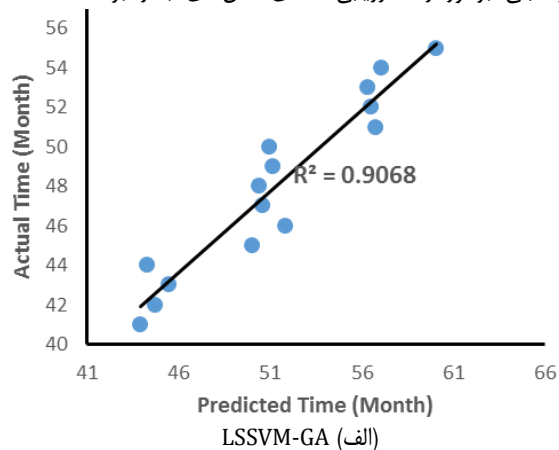
جدول ۶- نتایج ارزیابی مدل‌های LSSVM-PSO، LSSVM-GA در مرحله آزمون براساس معیارهای مختلف در تخمین زمان اتمام پروژه‌ها

نام مدل	APD (%)	AAPD (%)	SD	RMSE	R-square
LSSVM-PSO	-۲۲/۸۶۰۹	۲۶/۲۱۳۳	۵/۲۴۳۶	۵/۲۱۱۴	۰/۹۹۸۰
LSSVM-GA	-۱۲/۹۴۷۳	۱۹/۹۲۲۸	۵/۷۸۵۶	۵/۹۰۰۳	۰/۹۹۳۴

بصورت کمی در جدول ۷ ارائه شده است. همانطور که در این جدول مشاهده می‌شود مدل هیبریدی LSSVM-PSO نسبت به مدل هیبریدی LSSVM-GA از خطای کمتری برخوردار هست که نشان دهنده قابلیت تعمیم بالاتر این مدل دارد.



در شکل ۸ نمودارهای متقاطع زمان پیش‌بینی شده برای اتمام پروژه با زمان واقعی برای اتمام آن برای داده‌های اعتبارسنجی (سد E) نشان داده شده است. همانطور که در این شکل مشاهده می‌شود همه مدل‌ها از ضریب تعیین بالایی در پیش‌بینی زمان اتمام پروژه برای داده‌های اعتبارسنجی برخوردارند. ارزیابی خطای مدل‌های بکار برده شده



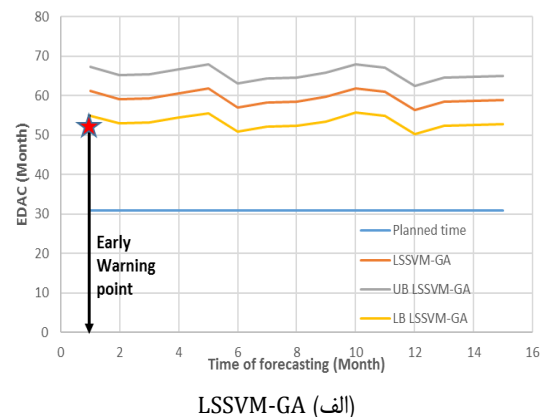
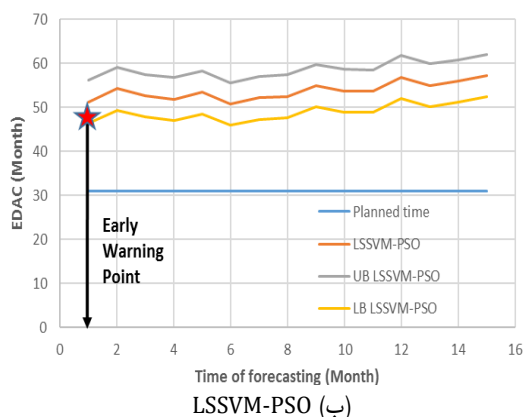
شکل ۸- نمودارهای متقاطع مقادیر پیش‌بینی شده و واقعی در تخمین زمان اتمام پروژه برای الگوریتم‌های LSSVM-PSO، LSSVM-GA در مرحله اعتبارسنجی

جدول ۷- نتایج ارزیابی مدل‌های LSSVM-PSO، LSSVM-GA در مرحله اعتبارسنجی بر اساس معیارهای مختلف در تخمین زمان اتمام پروژه‌ها

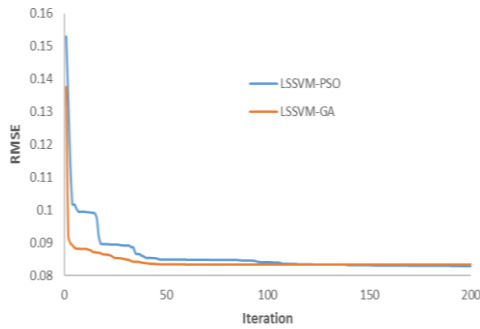
R-square	RMSE	SD	AAPD (%)	APD (%)	نام مدل
۰/۹۲۱۰	۲/۹۳۸۲	۱/۹۶۳۶	۵/۰۹۰۴	۴/۴۲۸۸	LSSVM-PSO
۰/۹۰۶۸	۳/۷۲۳۷	۱/۶۴۹۴	۶/۹۸۴۶	-۶/۹۸۴۶	LSSVM-GA

در این شکل مشاهده می‌شود مقادیر LB پیش‌بینی شده با همه مدل‌ها بیشتر از برنامه‌ریزی شده می‌باشد. این امر نشان می‌دهد که پروژه در ماه اول طبق برنامه پیش‌نرفته است. مقدار خروجی مدل‌های LSSVM-PSO، LSSVM-GA و به ترتیب زمان خاتمه پروژه را برابر ۵۷ و ۵۹ ماه نشان می‌دهند.

در شکل ۹ پرو فایل EDAC پیش‌بینی شده بر اساس مدل‌های هیبریدی LSSVM-PSO، LSSVM-GA برای داده‌های اعتبارسنجی (پروژه سد E) نشان داده شده است. در این شکل بازه تغییرات EDAC بین مرز پایین (LB) و بالا (UB) بر اساس خطایی که هر دو مدل در مرحله آموزش و آزمون مدل داشته‌اند در نظر گرفته شده است. همانطور که

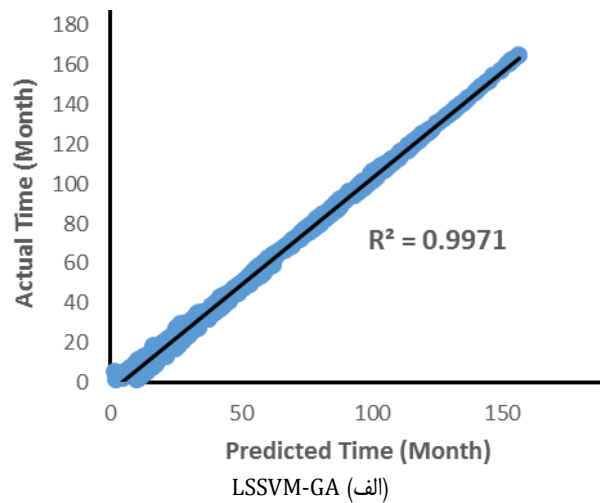
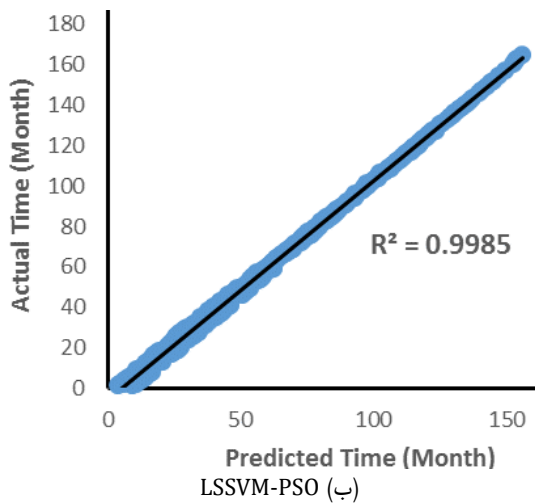


شکل ۹- پرو فایل EDAC بر اساس مدل‌های هیبریدی برای پروژه ساخت سد E



شکل ۱۰- نمودار تغییرات خطا در تکرارهای مختلف برای الگوریتم‌های LSSVM-GA و LSSVM-PSO

تغییرات خطا در تکرارهای مختلف الگوریتم‌های LSSVM-PSO، LSSVM-GA در پیش‌بینی نسبت هزینه برای مرحله آموزش در شکل ۱۰ نشان داده شده است. همانطور که در این شکل مشاهده می‌شود الگوریتم‌های بهینه‌سازی برای الگوریتم LSSVM تقریباً در تکرار حدود ۱۲۰ با مقدار خطای کمتری نسبت به الگوریتم‌های مشابه همگرا می‌شوند. نمودارهای متقاطع مقادیر واقعی و پیش‌بینی شده نسبت هزینه برای مدل‌های هیبریدی بکار برده شده در مرحله آموزش در شکل ۱۱ نشان داده شده است. بررسی پراکندگی نقاط داده در همه این مدل‌ها نشان از عملکرد خوب مدل‌های هیبریدی LSSVM در پیش‌بینی پارامتر هدف دارد.



شکل ۱۱- نمودارهای متقاطع مقادیر پیش‌بینی شده و واقعی در تخمین هزینه اتمام پروژه برای الگوریتم‌های LSSVM-PSO، PSO، LSSVM-GA در مرحله آموزش

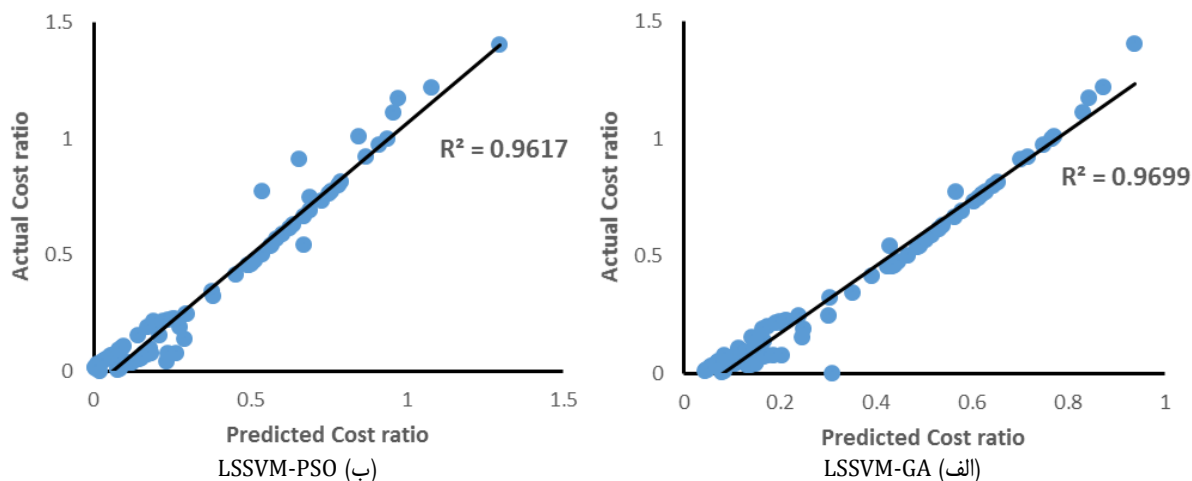
جدول ۸- نتایج ارزیابی مدل‌های LSSVM-PSO، LSSVM-GA در مرحله آموزش براساس معیارهای مختلف در تخمین هزینه اتمام پروژه‌ها

R-square	RMSE	SD	AAPD (%)	APD (%)	نام مدل
۰/۹۹۸۵	۰/۰۸۲۹	۰/۰۸۳۰	۱۰۳/۲۴۹۷	-۸۹/۹۷۸۴	LSSVM-PSO
۰/۹۹۷۱	۰/۰۸۳۳	۰/۰۸۲۹	۱۵۱/۰۵۸۰	-۱۴۲/۳۲۳۰	LSSVM-GA

هزینه، مقادیر را کم برآورد نموده است در حالی که در مدل LSSVM-PSO این مسئله چندان محسوس نیست. مقایسه کمی مدل‌ها براساس معیارهای خطای ارائه شده در جدول ۹ نشان می‌دهد که مدل LSSVM-PSO از دقت بالاتر و خطای کمتر در مقایسه با الگوریتم LSSVM-GA در این مرحله برخوردار بوده است.

در شکل ۱۲. Error! Reference source not found. نمودارهای متقاطع مقادیر واقعی و پیش‌بینی شده نسبت هزینه با استفاده از مدل‌های هیبریدی برای داده‌های آزمون نشان داده شده است. همانطور که در این شکل مشاهده می‌شود مدل LSSVM-GA در نسبت‌های هزینه پایین مقادیر را بیش برآورد و در مقادیر بالاتر نسبت





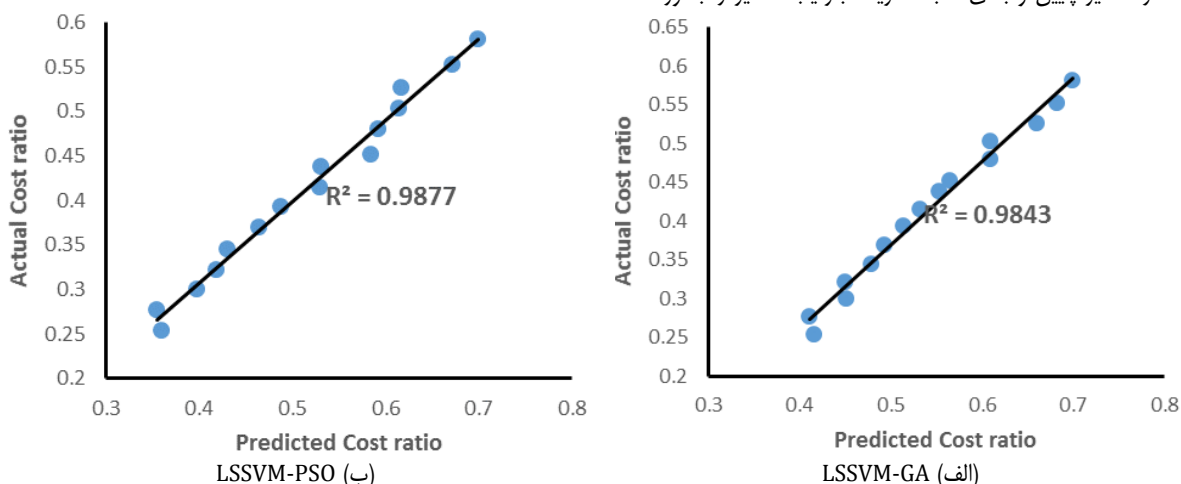
شکل ۱۲- نمودارهای متقاطع مقادیر پیش‌بینی شده و واقعی در تخمین هزینه اتمام پروژه برای الگوریتم‌های LSSVM-PSO، LSSVM-GA در مرحله آزمون

جدول ۹- نتایج ارزیابی مدل‌های LSSVM-PSO، LSSVM-GA در مرحله آزمون براساس معیارهای مختلف در تخمین هزینه اتمام پروژه‌ها

نام مدل	APD (%)	AAPD (%)	SD	RMSE	R-square
LSSVM-PSO	-۱۳۳/۳۰۰	۱۴۶/۴۰۶۱	۰/۰۸۲۴	۰/۰۸۳۲	۰/۹۶۱۷
LSSVM-GA	-۷۹۴/۶۴۶۰	۸۱۰/۱۱۳۵	۰/۱۲۶۱	۰/۱۲۸۵	۰/۹۶۹۹

مقایسه نمودارهای متقاطع مقادیر واقعی و پیش‌بینی نسبت هزینه با مدل‌های آموزش دیده بر روی داده‌های سد E در شکل ۱۳ نشان داده شده است. همانطور که در این شکل مشاهده می‌شود مدل LSSVM-GA در مقادیر پایین و بالای نسبت هزینه، بترتیب مقادیر را بصورت

محسوسی بیش‌برآورد و کم برآورد نموده‌اند. وجود این مسئله برای مدل LSSVM-PSO نیز بصورت کم‌رنگتری قابل تشخیص است. خطای مدل‌ها در مرحله اعتبارسنجی در جدول ۱۰ نشان داده شده است.

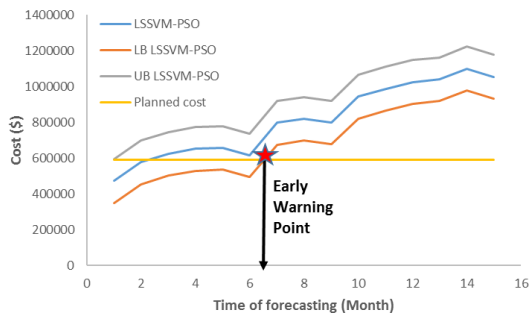


شکل ۱۳- نمودارهای متقاطع مقادیر پیش‌بینی شده و واقعی در تخمین هزینه اتمام پروژه برای الگوریتم‌های LSSVM-PSO، LSSVM-GA در مرحله اعتبارسنجی

جدول ۱۰- نتایج ارزیابی مدل‌های LSSVM-PSO، LSSVM-GA در مرحله اعتبارسنجی براساس معیارهای مختلف در تخمین هزینه اتمام پروژه‌ها

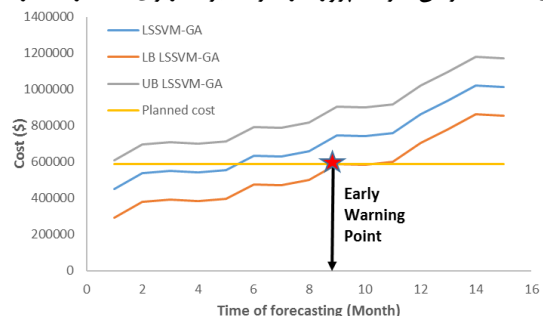
نام مدل	APD (%)	AAPD (%)	SD	RMSE	R-square
LSSVM-PSO	-۲۶/۰۰۴۰	۲۶/۰۰۴۰	۰/۱۰۴۷	۰/۱۰۴۰	۰/۹۸۷۷
LSSVM-GA	-۳۳/۵۶۴۱	۳۳/۵۶۴۲	۰/۱۲۴۷	۰/۱۲۸۸	۰/۹۸۴۳

کرده‌اند. مدل LSSVM-PSO هشدار اولیه سریع‌تری را برای بازبینی هزینه کرده‌ها ارائه داده است. ادامه روند پیش‌بینی نیز گویای این مطلب است که ابزارهای مورد استفاده توسط مدیر پروژه نتوانسته است این هشدار اولیه را صادر نماید تا در برنامه عملیاتی تجدید نظرگردد. براساس بررسی گزارش‌های ساخت این سد، مدیران در ماه ۱۳م از زمان آغاز پروژه نیز افزایش هزینه‌های خاتمه عملیات را تأیید کرده‌اند.



LSSVM-PSO (ب)

پروفایل هزینه پیش‌بینی شده برای پروژه ساخت سد E بر مبنای EDAC و با کمک چهار مدل LSSVM-GA، LSSVM-PSO در شکل ۱۴ نشان داده شده است. در این شکل نیز محدود احتمالی هزینه پیش‌بینی شده با استفاده از خطای مدل‌ها در مرحله آموزش و آزمون محاسبه شده است. همانطور که در این شکل مشاهده می‌شود مدل‌های LSSVM-PSO، LSSVM-GA، PSO بترتیب در ماه‌های هفتم، هشتم هشدار اولیه را برای فاصله گرفتن هزینه پروژه از هزینه برنامه‌ریزی شده را صادر

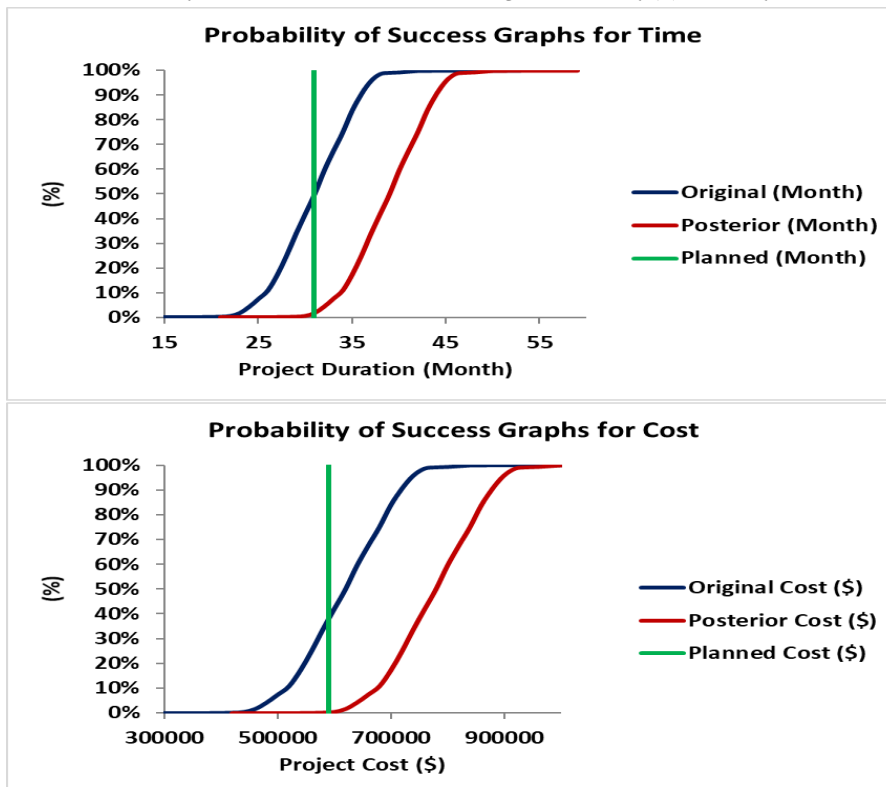


LSSVM-GA (الف)

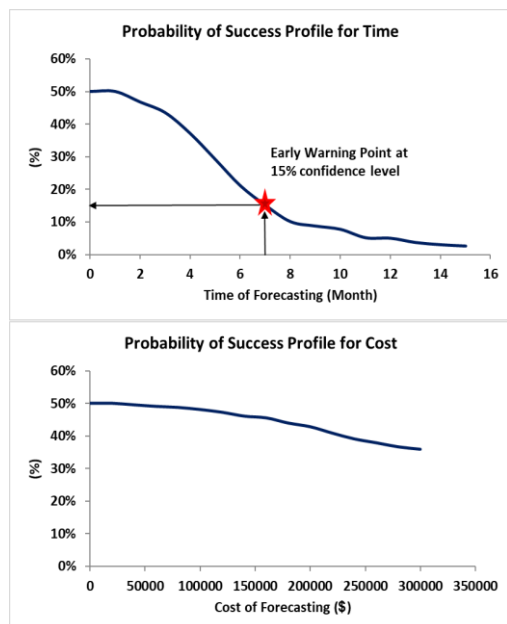
شکل ۱۴- پروفایل تخمین هزینه بر مبنای EDAC و با استفاده از مدل‌های هیبریدی برای پروژه ساخت سد E

مشخص و با هزینه در نظر گرفته شده با سطح اطمینان ۹۰٪ باشد. همانطور که در این شکل مشخص است انجام این پروژه با زمان و هزینه برنامه‌ریزی شده برای این پروژه که به ترتیب ۳۱ ماه و ۵۹۰۱۴۶٫۸۲ دلار می‌باشد براساس خروجی مدل فیلتر کالمن با احتمال موفقیت ۵۰٪ همراه است. این در حالی است که احتمال اتمام پروژه با هزینه برنامه‌ریزی شده حدود ۴۰٪ است.

در ادامه فیلتر کالمن روی داده‌های اعتبارسنجی (یعنی سد E) اعمال شد. در شکل ۱۵ نمودار احتمال موفقیت حاصل از فیلتر کالمن بترتیب برای پیش‌بینی زمان و هزینه نشان داده شده است. در این شکل منحنی‌های احتمالاتی original و posterior احتمال موفقیت پروژه در زمان و هزینه مشخص را نشان می‌دهند. این نمودارها می‌تواند پاسخگوی سؤالاتی نظیر احتمال موفقیت اتمام پروژه در یک زمان



شکل ۱۵- نمودار احتمال موفقیت برای پیش‌بینی زمان و هزینه با استفاده از فیلتر کالمن



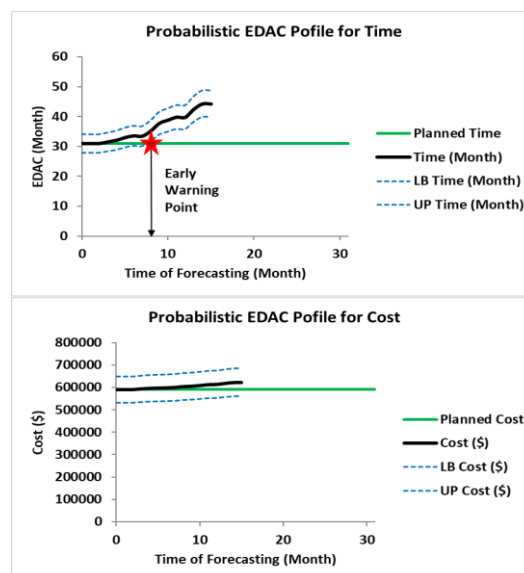
**شکل ۱۷- احتمال موفقیت در پروفایل زمان و هزینه با استفاده از فیلتر کالمن**

### نتیجه گیری

در این مطالعه الگوریتم‌های هیبریدی LSSVM-GA و LSSVM-PSO با کمک از مفاهیم ارزش زمانی برای پیش‌بینی زمان خاتمه و هزینه انجام پروژه‌ها استفاده شد. برای توسعه این مدل‌ها از داده‌های چهار سد ساخته شده استفاده شد. این داده‌ها به دو گروه داده‌های آموزش (80% کل داده‌ها) و آزمون (20% از کل داده‌ها) تفکیک شدند. برای اعتبارسنجی مدل‌های توسعه داده شده نیز از داده‌های سد دیگری که تا ماه ۱۵ گزارش پیشرفت آن ارائه شده بود استفاده گردید. در ادامه فیلتر کالمن نیز برای ارزیابی مدل‌های هیبریدی در پیش‌بینی زمان و هزینه، بر روی داده‌های اعتبارسنجی اعمال شد.

نتایج بکارگیری الگوریتم‌های هیبریدی نشان داد که مدل‌های حاصل از دقت خوبی در مراحل آموزش و آزمون برای پیش‌بینی زمان اتمام پروژه‌ها برخوردار هستند. اختلاف کمتر خطای مدل LSSVM-PSO نسبت به مدل دیگر در پیش‌بینی زمان اتمام پروژه در این دو مرحله حاکی از قابلیت تعمیم بالای این مدل دارد؛ که این امر با عملکرد خوب این مدل در مرحله اعتبارسنجی مورد تأیید قرار گرفت. بکارگیری تمامی مدل‌های آموزش دیده روی داده‌های اعتبارسنجی نیز نشان داد که براساس پیشرفت حاصل شده در ماه اول، مدت زمان انجام پروژه بسیار بیشتر از برنامه زمان‌بندی اولیه خواهد بود و هشدار اولیه توسط تمامی مدل‌های هیبریدی در ماه اول فعال شد. این در حالی است که هشدار اولیه براساس پیش‌بینی فیلتر کالمن، در ماه هفتم فعال می‌شود. مدت زمان اتمام پروژه ساخت سد E برای مدل‌های هیبریدی LSSVM-PSO و LSSVM-GA بترتیب برابر با ۵۷ و ۵۹ است. این در حالی است که نتیجه اعمال فیلتر کالمن در پیش‌بینی زمان بر روی داده‌های

در شکل ۱۶ پروفایل احتمالی EDAC و هزینه حاصل از مدل کالمن برای پروژه E نشان داده شده است. بازه LB و UB در این شکل با سطح اطمینان ۹۰٪ مشخص شده است. براساس خروجی این مدل زمان اتمام پروژه تا ماه پنجم از زمان شروع مطابق برنامه‌ریزی انجام شده بوده است. اما با گذشت زمان، زمان خاتمه پیش‌بینی شده از زمان برنامه‌ریزی شده برای اتمام پروژه به صورت چشمگیری فاصله می‌گیرد. گونه‌ای که مدل در ماه هفتم هشدار اولیه را برای بیشتر زمان بردن انجام پروژه از زمان برنامه‌ریزی شده را صادر می‌کند. خروجی مدل کالمن برای هزینه اتمام پروژه مطابقت خوبی با هزینه برنامه‌ریزی دارد؛ اما از ماه دهم به بعد هزینه پیش‌بینی شده توسط مدل به صورت آرامی از هزینه برنامه‌ریزی شده فاصله می‌گیرد. این مدل هیچ‌گونه هشدار را برای فاصله گرفتن غیرقابل جبران هزینه اتمام پروژه از هزینه برنامه‌ریزی شده را صادر نکرده است.



**شکل ۱۶- پروفایل احتمالی EDAC برای پیش‌بینی زمان و هزینه با استفاده از فیلتر کالمن**

شکل ۱۷ به منظور تعیین احتمال موفقیت در پروفایل هزینه و زمان با روش فیلتر کالمن مورد استفاده قرار می‌گیرد. با استفاده از این منحنی می‌توان درصد احتمال موفقیت پروژه در مدت زمان برنامه‌ریزی شده به هنگام اعلام هشدار اولیه را پیش‌بینی کرد. برای نمونه نتایج حاصله از این منحنی در ماه هفتم نشان می‌دهد که احتمال موفقیت به منظور دستیابی به مدت زمان برنامه‌ریزی شده برای این پروژه، ۱۵٪ می‌باشد. به هنگام اعلام هشدار اولیه در تأخیر برنامه زمان‌بندی، تدابیر لازم را برای جبران آن و تجدید برنامه‌ریزی انجام شده در نظر گرفت.

داشت مدل‌های حاصل بسیار محافظه‌کارانه عمل می‌کنند. این در حالی است که مدل کالمن تنها براساس داده‌های محدود خود پروژه و براساس روند گذشته آن توسعه داده شده است. بنابراین استفاده از مدل‌های هیبریدی بویژه LSSVM-PSO برای پیش‌بینی زمان اتمام پروژه به شدت توصیه می‌شود.

### پیشنهادات

با توجه به دقت و قابلیت تعمیم بالای الگوریتم‌های یادگیری عمیق در پیش‌بینی، پیشنهاد می‌شود در مطالعات آتی از این الگوریتم‌ها بهره گرفته شود. البته لازم به ذکر است که بکارگیری این الگوریتم‌ها نیازمند حجم داده بیشتری است.

### ملاحظات اخلاقی پیروی از اصول اخلاق پژوهش

همکاری مشارکت‌کنندگان در تحقیق حاضر به صورت داوطلبانه و با رضایت آنان بوده است.

### حامی مالی

هزینه تحقیق حاضر توسط نویسندگان مقاله تامین شده است.

### مشارکت نویسندگان

طراحی و ایده‌پردازی: سینا فرد مرادی نیا؛ رضا بخشی  
روش‌شناسی و تحلیل داده‌ها: سینا فرد مرادی نیا؛ رضا بخشی؛ رسول جانی، رامین وفایی پور  
نظارت و نگارش نهایی: سینا فرد مرادی نیا؛ رضا بخشی. رسول جانی، رامین وفایی پور سرخابی

### تعارض منافع

بنابر اظهار نویسندگان، مقاله حاضر فاقد هرگونه تعارض منافع بوده است.

اعتبارسنجی، هشدار اولیه برای تأخیر در برنامه‌زمانبندی را در ماه هفتم اعلام می‌کند. لازم به ذکر است که این مدل زمان اتمام پروژه را برابر ۵۲ ماه پیش‌بینی می‌کند. براساس ارزیابی‌های انجام شده توسط مدیران پروژه، زمان اتمام پروژه حدود ۵۵ ماه خواهد بود که نتایج مدل‌های هیبریدی نسبت به مدل کالمن به واقعیت نزدیک‌تر بوده است.

آموزش الگوریتم‌های هیبریدی برای پیش‌بینی نسبت هزینه پروژه‌ها در مراحل آموزش و آزمون نیز نشان داد که الگوریتم LSSVM-PSO از دقت بالاتری برخوردار است. همچنین مدل LSSVM-GA در مقادیر پایین و بالای نسبت هزینه بطور چشمگیری، بترتیب، مقادیر را بیش برآورد و کم برآورد می‌نمایند. هر چند این مسئله در نتایج مدل LSSVM-PSO نیز مشاهده گردید؛ ولی نسبت به مدل دیگر از شدت کمتری برخوردار است. با توجه به دقت پایین‌تر مدل‌ها در پیش‌بینی هزینه‌ها نسبت به پیش‌بینی زمان، می‌توان گفت بکارگیری مدل‌های هیبریدی در پیش‌بینی هزینه پروژه‌ها نیازمند احتیاط بیشتری است. این مسئله می‌تواند ناشی از متأثر بودن هزینه‌های پروژه به عوامل مختلفی باشد که در ورودی‌های مدل‌ها بکار نرفته است. هشدار اولیه برای فاصله گرفتن هزینه عملیات ساخت سد از هزینه برنامه‌ریزی شده برای مدل‌های LSSVM-PSO و LSSVM-GA بترتیب در ماه‌های هفتم و هشتم صادر شده است. این در حالی است که نتیجه اعمال فیلتر کالمن در پیش‌بینی هزینه بر روی داده‌های اعتبارسنجی، هیچ‌گونه هشدار اولیه‌ای برای فاصله گرفتن هزینه کردها از هزینه‌های برنامه‌ریزی شده اعلام نمی‌کند.

مقایسه مدل‌های هیبریدی با مدل کالمن نشان می‌دهد که مدل‌های هیبریدی با توجه به آموزشی که براساس داده‌های چند سد ساخته شده داشتند نتایج قابل قبول‌تری با گزارشات دوره‌ای ارائه شده برای انجام پروژه E دارند. البته لازم به ذکر است که با توجه به تأخیرهایی که در چند پروژه استفاده شده در مرحله آموزش مدل‌های هیبریدی وجود

1242.

## References

- Gowan, J. A., Mathieu, R. G., & Hey, M. B. 2006. Earned value management in a data warehouse project. *Information management & computer security*, 14(1), 37-50.
- Attalla, M., & Hegazy, T. (2003). Predicting cost deviation in reconstruction projects: Artificial neural networks versus regression. *Journal of construction engineering and management*, 129(4), 405-411.
- Kim, G. H., An, S. H., & Kang, K. I. 2004. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and environment*, 39(10), 1235-
- Wilmot, C. G., & Mei, B. 2005. Neural network modeling of highway construction costs. *Journal of construction engineering and management*, 131(7), 765-771.
- Kim, B. C., & Reinschmidt, K. F. 2010. Probabilistic forecasting of project duration using Kalman filter and the earned value method. *Journal of Construction Engineering and Management*, 136(8), 834-843.
- Sonmez R. 2011. Range estimation of construction costs using neural networks with bootstrap prediction intervals. *Expert Syst Appl.*; 38(8):9913-7.
- Wang Y-R, Yu C-Y, Chan H-H. 2012. Predicting construction cost and schedule success using artificial neural networks

- ensemble and support vector machines classification models. *Int J Proj Manag.*30(4):470-8.
8. Cheng M-Y, Hoang N-D, Roy AF V, Wu Y-W. 2012. A novel time-dependend evolutionary fuzzy SVM inference model for estimating construction project at completion. *Eng Appl Artif Intell.* 25(4):744-52.
  9. Wauters M, Vanhoucke M. 2014. Support vector machine regression for project control forecasting. *Autom Constr.*47:92-106.
  10. Kim B-C. 2015. Integrating risk assessment and actual performance for probabilistic project cost forecasting: a second moment Bayesian model. *IEEE Trans Eng Manag.* 62(2):158-70.
  11. Mortaji STH, Noorossana R, Bagherpour M. 2015. Project completion time and cost prediction using change point analysis. *J Manag Eng.* 31(5):4014086.
  12. Akhbari M. 2018. Project time and cost forecasting using monte carlo simulation and artificial neural networks. *Int J Ind Eng Prod Res.* 29(2):231-9.
  13. Jiang Q. 2019. Estimation of construction project building cost by back-propagation neural network. *J Eng Des Technol.*
  14. Najafi S, Moosavirad SH, Ariafar S. 2019. Predicting the project time and costs using EVM based on gray numbers. *Eng Constr Archit Manag.*
  15. Soltan S, Ashrafi M. 2020. Predicting project duration and cost, and selecting the best action plan using statistical methods for earned value management. *J Proj Manag.* 5(3):157-66.
  16. Khalaf TZ, Çağlar H, Çağlar A, Hanoon AN. 2020. Particle swarm optimization based approach for estimation of costs and duration of construction projects. *Civ Eng J.*6(2):384-401.
  17. Asadabadi MR, Zwikael O. 2021. Integrating risk into estimations of project activities' time and cost: A stratified approach. *Eur J Oper Res.* 291(2):482-90.
  18. Bakhshi R, Moradnia SF, Jani R, Poor RV. 2022. Presenting a Hybrid Scheme of Machine Learning Combined with Metaheuristic Optimizers for Predicting Final Cost and Time of Project. *KSCE J Civ Eng.*1-16.
  19. Anemangely M, Ramezanzadeh A, Tokhmechi B, Molaghab A, Mohammadian A. 2018. Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network. *J Geophys Eng.*15(4):1146-34.
  20. Mehrad M, Bajolvand M, Ramezanzadeh A, Neycharan JG. 2020. Developing a new rigorous drilling rate prediction model using a machine learning technique. *J Pet Sci Eng.*192:107338.
  21. Sabah M, Mehrad M, Ashrafi SB, Wood DA, Fathi S. 2021. Hybrid machine learning algorithms to enhance lost-circulation prediction and management in the Marun oil field. *J Pet Sci Eng.*198:108125.
  22. Kennedy J, Eberhart R. 1995. Particle swarm optimization. In: *Proceedings of ICNN'95-international conference on neural networks.* IEEE. p. 1942-8.
  23. Czarnigowska A. 2008. Earned value method as a tool for project control. *Bud i Archit.*3(2):15-32.
  24. Lipke W. 2014. Introduction to earned schedule. *PM World J.*3(11):8-9.
  25. Kim B-C, Reinschmidt KF. 2011. Combination of project cost forecasts in earned value management. *J Constr Eng Manag.* 137(11):958-66.
  26. Wang, H. and Hu, D., 2005. October. Comparison of SVM and LS-SVM for regression. In *2005 International conference on neural networks and brain (Vol. 1, pp. 279-283).* IEEE.
  27. Welch G, Bishop G. 1995. An introduction to the Kalman filter.
  28. Suykens JAK, Vandewalle J. 1999. Least squares support vector machine classifiers. *Neural Process Lett.* 9(3):293-300.
  29. Vapnik V. 2013. *The nature of statistical learning theory.* Springer science & business media.
  30. Duan K, Keerthi SS, Poo AN. 2003. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing.* 51: 41-59.
  31. Couceiro, M., Ghamisi, P., Couceiro, M. and Ghamisi, P., 2016. Particle swarm optimization. *Fractional order darwinian particle swarm optimization: Applications and evaluation of an evolutionary algorithm,* pp.1-10.
  32. Coello CAC, Lamont GB, Van Veldhuizen DA. 2007. *Evolutionary algorithms for solving multi-objective problems.* Vol. 5. Springer.
  33. Clerc M. 2010. *Particle swarm optimization.* Vol. 93. John Wiley & Sons.