

Evaluation of Nonparametric Decision Tree Models for Predicting Scour Depth of Bridges

Mehdi Komasi¹ , Rezvan Dalvand² 

1. Department of Civil Engineering, Faculty of Engineering, Ayatollah Boroujerdi University, Boroujerd, Iran. E-mail: Komasi@abru.ac.ir.
2. Corresponding author, Master's graduate, Department of Civil Engineering, Faculty of Engineering, Ayatollah Boroujerd University, Boroujerd, Iran. E-mail: rezvandalvand1992@gmail.com.

Article Info

Article type:
Research Article

Article history:

Received 25 Sep 2024
 Revised 20 Dec 2024
 Accepted 08 Jan 2025
 Published 25 Mar 2025

Keywords:

Scour,
 Decision Tree,
 Gradient Boosting,
 Random Forest,
 Extreme Gradient Boosting.

ABSTRACT

The complex nature of bridge pier scour and the influence of various parameters on its estimation highlight the necessity of using a comprehensive nonlinear model. In the present study, four decision-tree-based models along with data from different bridges are employed. The input variables used for modeling the decision trees include upstream flow velocity, median bed particle diameter, upstream flow depth at the pier, pier width, angle of attack of the flow to the pier, pier length, particle diameter for which 84% of the particles are finer, and the pier shape factor, while the output variable is the local scour depth. In this study, modern decision tree methods are applied to calculate and compare bridge pier scour depth. The results indicate that the Extreme Gradient Boosting (XGB) and Random Forest (RF) models, with coefficients of determination of 0.76 and 0.73, respectively, exhibit higher accuracy compared to the four investigated models. The Gradient Boosting (GB) model ranks next with a coefficient of determination of 0.67, while the Decision Tree (DT) model shows the lowest performance. Furthermore, sensitivity analysis performed on the models reveals that pier width and pier length have the greatest influence on scour depth. After these two parameters, upstream flow depth has the most significant effect on scour depth.

Cite this article: Komasi M, Dalvand R. Evaluation of nonparametric decision tree models for predicting scour depth of bridges. *Water Resources and Climate Change*. (2025); 1(1): 39-48.
<https://doi.org/10.22091/wrcc.2025.11363.1005>.

Evaluation of Nonparametric Decision Tree Models for Predicting Scour Depth of Bridges

Mehdi Komasi¹, Rezvan Dalvand²

1. Department of Civil Engineering, Faculty of Engineering, Ayatollah Boroujerdi University, Boroujerd, Iran. E-mail: Komasi@abru.ac.ir.
2. Corresponding author, Master's graduate, Department of Civil Engineering, Faculty of Engineering, Ayatollah Boroujerd University, Boroujerd, Iran. E-mail: rezvandaland1992@gmail.com.

Extended Abstract

Background and Objective

Water distribution networks represent a critical component of urban infrastructure, ensuring the delivery of safe water at adequate pressure. Rapid urbanization and aging pipelines have increasingly led to failures, pressure deficits, leakage, and reduced network reliability, highlighting the need for efficient design and rehabilitation strategies. The design problem is inherently multi-objective, requiring a balance between minimizing construction costs and maintaining sufficient hydraulic performance. Traditional design approaches are largely heuristic and unable to guarantee near-optimal solutions, whereas evolutionary algorithms have shown superior performance in complex optimization tasks. However, many of these algorithms suffer from drawbacks such as premature convergence, instability, or sensitivity to parameter tuning. The Multi-Objective Grey Wolf Optimizer (MOGWO), inspired by cooperative hunting behavior, has emerged as a robust alternative due to its effective balance between global exploration and local exploitation. This study proposes a novel MOGWO-based framework that minimizes network cost while maximizing resilience, and evaluates its effectiveness on the real-world D-zone distribution network in Mashhad.

Methodology

The present study uses four decision tree-based models and data (and information) from multiple bridges. The information used in this study to model the decision trees include the upstream flow velocity, average bed particle diameter, upstream flow depth and base width, angle of water intrusion into the base, base length, and diameter of particles of which 84% are smaller than its diameter, base shape factor, as input variables and local scour depth as output in the model. In this study, new decision tree methods are used to calculate and compare the bridge base scour depth. In this study, data from the US Federal Highway Administration was used to build models and validate decision trees. It was extracted from statistics on several bridges in the US. All information was collected in the field and included information related to the depth of scour around bridge piers in different locations.

Findings

The results show that the Extreme Gradient Boosting (XGB) and Random Forest (RF) models with coefficients of determination of 0.76 and 0.73 had higher accuracy than the four models examined, and the Gradient Boosting (GB) model was in the second place after these two models with coefficients of determination of 0.67, and the Decision Tree (DT) model was

in the last place. Also, with the sensitivity analysis performed on the models, it was observed that the base width and the base length have the greatest effect on the scour depth. After these two parameters, the upstream flow depth has the greatest effect on the scour depth.

Conclusion

Bridge foundation scour is one of the most important issues in bridge safety. Considering the importance of this issue, in the present study, an attempt has been made to use new decision tree-based models to evaluate this issue. Decision tree methods are one of the best evaluation methods due to their simple understanding and ability to work with large and complex data. In future study, it is possible to combine new decision tree-based methods with other data mining or machine learning methods. The results show that the Extreme Gradient Boosting (XGB) and the Random Forest (RF) models had higher accuracy than the four models examined, and the Gradient Boosting (GB) model and the Decision Tree (DT) model were in last place. Also, with the sensitivity analysis performed on the models, it was observed that the base width and the base length have the greatest effect on the scour depth. After these two parameters, the upstream flow depth has the greatest effect on the scour depth.

Keywords: Scour, Decision Tree, Gradient Boosting, Random Forest, Extreme Gradient Boosting

Cite this article: Komasi M, Dalvand R. Evaluation of nonparametric decision tree models for predicting scour depth of bridges. *Water Resources and Climate Change*. (2025); 1(1): 39-48. <https://doi.org/10.22091/wrcc.2025.11363.1005>.

Declarations

- **Ethical Approval**

■ The paper is not currently being considered for publication elsewhere. All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

- **Competing interests**

■ Conflict of Interest - None

- **Availability of data and materials**

■ Data will be made available on the request.

- **Authors Contributions**

Rezvan Dalvand developed the theory and performed the computations. Mehdi Komasi encouraged Rezvan Dalvand to investigate a specific aspect. Mehdi Komasi supervised the findings of this work. All authors discussed the results and contributed to the final manuscript. Rezvan Dalvand wrote the manuscript with support from Mehdi Komasi. Mehdi Komasi conceived the original idea.

- **Acknowledgements**

- The authors would like to thank all participants in the present study

- **Funding**

- This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ارزیابی مدل‌های مبتنی بر درخت تصمیم غیرپارامتریک برای پیش‌بینی عمق آبستگی پایه‌های

پل

مهدی کاماسی^۱، رضوان دالوند^۲

۱. گروه مهندسی عمران، دانشکده فنی و مهندسی، دانشگاه آیت‌الله العظمی بروجردی (ره)، بروجرد، ایران. رایانامه: Komasi@abru.ac.ir
 ۲. نویسنده مسئول، دانش‌آموخته کارشناسی ارشد، گروه مهندسی عمران، دانشکده فنی و مهندسی، دانشگاه آیت‌الله العظمی بروجردی (ره)، بروجرد، ایران. رایانامه: rezvandalvand1992@gmail.com

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی تاریخ دریافت: ۱۴۰۳/۰۷/۰۴ تاریخ بازنگری: ۱۴۰۳/۰۹/۳۰ تاریخ پذیرش: ۱۴۰۳/۱۰/۱۹ تاریخ انتشار: ۱۴۰۴/۰۱/۰۵ کلیدواژه‌ها: آبستگی، درخت تصمیم، تقویت گرادیان، جنگل تصادفی، تقویت گرادیان شدید.	ماهیت پیچیده آبستگی پایه پل و تأثیرات پارامترهای مختلف در برآورد آن ضرورت استفاده از یک مدل غیرخطی و جامع را بیش‌تر نمایان می‌سازد. در پژوهش حاضر از چهار مدل مبتنی بر درخت تصمیم و همچنین از آمار پل‌های مختلف بهره‌گیری می‌شود. اطلاعات مورد استفاده در این پژوهش برای مدل‌سازی درخت‌های تصمیم‌گیری شامل سرعت جریان در بالادست، قطر متوسط ذرات بستر، عمق جریان در بالادست پایه و عرض پایه، زاویه هجوم آب به پایه، طول پایه، قطر ذراتی که ۸۴ درصد از آنها از قطر آن کوچک‌تر هستند، ضریب شکل پایه، به‌عنوان متغیرهای ورودی و عمق آبستگی موضعی به‌عنوان خروجی در مدل می‌باشد. در این پژوهش از روش‌های جدید درخت تصمیم برای محاسبه و مقایسه عمق آبستگی پایه پل استفاده می‌شود. نتایج نشان می‌دهند که مدل تقویت گرادیان شدید^۲ (XGB) و جنگل تصادفی^۳ (RF) با ضریب تبیین ۰/۷۶ و ۰/۷۳ دارای دقت بیش‌تری نسبت به چهار مدل بررسی‌شده بودند و مدل تقویت گرادیان^۴ (GB) در رده بعد از این دو مدل با ضریب تبیین ۰/۶۷ قرار دارد و مدل درخت تصمیم^۵ (DT) در رده آخر قرار دارد. هم‌چنین با تحلیل حساسیت صورت‌گرفته بر روی مدل‌ها مشاهده شد که عرض پایه و همچنین طول پایه بیش‌ترین تأثیر را بر روی عمق آبستگی دارند. پس از این دو پارامتر عمق جریان در بالادست بیش‌ترین تأثیر را بر روی عمق آبستگی دارد.

استناد: کاماسی مهدی، دالوند رضوان. ارزیابی مدل‌های مبتنی بر درخت تصمیم غیرپارامتریک برای پیش‌بینی عمق آبستگی پایه‌های پل. منابع آب و تغییر اقلیم. ۱۴۰۴؛ (۱): ۳۹-۴۸. <https://doi.org/10.22091/wrcc.2025.11363.1005>

²- XGBoosting
³- Random Forest
⁴- Gradient Boosting
⁵- Decision Tree

۱- مقدمه

آبشستگی پایه پل‌ها یکی از اصلی‌ترین دلایل تخریب پل‌ها در سراسر جهان است. محاسبه خطرپذیری پایه‌های پل برای مدیریت صحیح و اختصاص منابع برای نگهداری و کاهش آلودگی امری ضروری می‌باشد [۱]. تخریب پل‌ها و خسارات وارده علاوه بر تلفات جانی و ضررهای مالی سبب قطع راه‌های ارتباطی شده که بدین اصل حمل‌ونقل را نیز مختل می‌کند. آبشستگی پدیده طبیعی است که به واسطه جریان آب رودخانه‌ها و یا سیلاب‌ها به وجود می‌آید و نتیجه آن عمل فرسایش به وسیله جریان آب است که موجب انتقال و فرسایش مواد از کف، سواحل و اطراف پایه‌های پل می‌شود و در بسیاری از موارد منجر به تخریب پل و سازه‌های هیدرولیکی می‌گردد. با کنترل و محافظت از پایه‌های پل در برابر آبشستگی و ارائه روش‌های مناسب پیش‌بینی آبشستگی، می‌توان از وارد آمدن این خسارات پیشگیری کرد. روش‌های محاسبه عمق آبشستگی به سه روش معادلات تجربی، آزمایشگاهی و داده‌کاوی تقسیم می‌شوند.

تاکنون روابط تجربی متفاوتی در زمینه برآورد آبشستگی در اطراف پایه‌های پل ارائه شده است. از جمله معادلات ارائه شده در این زمینه می‌توان به مراجع [۲، ۳ و ۴] اشاره کرد. یکی از معروف‌ترین معادلات ارائه شده مربوط به دانشگاه کلرادو بوده که به HEC-18 معروف است و می‌تواند بیشینه عمق آبشستگی اطراف پایه‌های پل را در دو حالت آب زلال و بستر متحرک محاسبه کند [۵]. این معادله براساس داده‌های آزمایشگاهی آبشستگی اطراف پایه‌های استوانه‌ای استفاده شده است. روش دیگر محاسبه عمق آبشستگی پایه‌های پل روش‌های آزمایشگاهی می‌باشد. در این زمینه به عنوان نمونه در پژوهشی به بررسی کاربرد صفحات مستغرق در کنترل آبشستگی اطراف پایه‌های پل مستطیلی با دماغه

گرد پرداختند که یکی از روش‌های کاهش آبشستگی موضعی در اطراف پایه‌های پل استفاده از صفحات مستغرق بود [۱].

در پژوهشی با عنوان عملکرد هندسه طوق در آبشستگی موضعی در پایه پل، بیان شده است که استفاده از هر دو نوع طوق مربعی و طوق دایره‌ای به‌ویژه در زیر بستر، باعث کاهش عمق آبشستگی می‌شود. نوع مربعی نسبت به دایره‌ای، عملکرد بهتری در کاهش عمق آبشستگی داشته است [۶].

پیچیدگی پدیده آبشستگی در اطراف پایه‌های پل به دلیل ماهیت سه‌بعدی جریان و ایجاد گردابه‌های متعدد در چاله، آبشستگی را با مقداری خطا محاسبه می‌کند. در پژوهشی دیگر در این زمینه به مطالعه عمق آبشستگی در پایه شیب‌دار پل‌ها در طول مسیر مستقیم توسط مطالعات آزمایشگاهی پرداخته شده است که هدف از این مطالعه بررسی تأثیر زاویه شیب پایه پل با انجام آزمایش در طول یک کانال مستقیم در آزمایشگاه بوده که در این آزمایش، پایه‌های استوانه‌ای با چهار زاویه مختلف در مجرای کنار کانال مستقیم قرار گرفتند و آزمایش‌ها در چهار سرعت مختلف تحت شرایط آب زلال انجام شد. نتایج به دست آمده نشان دادند که بیشینه و کمینه عمق آبشستگی در زاویه ۰ و ۱۵ درجه رخ داده است. مشابه عمق آبشستگی، ابعاد حفرات آبشستگی با افزایش زاویه پایه پل، کاهش یافته است [۷].

بر همین اساس پژوهشگران در کنار معادلات تجربی و روش‌های آزمایشگاهی از روش‌های داده‌کاوی استفاده می‌کنند. در پژوهشی در این زمینه اقدام به پیش‌بینی سطح آب زیرزمینی براساس طبقه‌بندی و درخت رگرسیون شد که در پژوهش ایشان رابطه‌ای بین عوامل تأثیرگذار مانند باران و سطح مخزن و تغییر تراز آب زیرزمینی، برقرار شد. سپس مدل طبقه‌بندی و رگرسیون درختی^۵ مدل‌سازی شد و سپس نتایج حاصله با ماشین بردار پشتیبانی^۶

^۵- Classification And Regression Tree

^۶- Support Vector Machines

XGB به‌طور قابل ملاحظه‌ای سریع‌تر از راه‌حل‌های فعلی عمل می‌کند و هم‌چنین به میلیاردها نمونه در یک دستگاه مقیاس می‌شود. همان‌طور که از نام آن مشخص است، عملکرد این الگوریتم برای تقویت می‌باشد و قادر است طبقه‌بندی‌های ضعیف را به دسته‌های قوی با استفاده از راه‌بردهای آموزشی و ادغام تبدیل کند [۱۳].

با رویداشت به پیشینه پژوهش و پژوهش‌های انجام شده در این زمینه به‌نظر می‌رسد که مقایسه روش‌های جدیدتر درخت تصمیم تا به حال در پژوهشی انجام نگرفته است و در این پژوهش سعی بر این موضوع شده که مدل‌های درخت تصمیم برای سنجش عمق آب‌سستگی پایه پل‌ها مورد بررسی قرار گیرند و مقایسه‌ای بین این روش‌ها صورت گیرد و در کم‌تر پژوهشی به این چهار روش جدید مدل مبتنی بر درخت تصمیم برای ارزیابی عمق آب‌سستگی به‌صورت یکجا پرداخته شده است. هم‌چنین در ادامه تأثیر هشت پارامتر مؤثر در آب‌سستگی به‌تفکیک در تمام مدل‌ها، بررسی خواهد شد.

۲- مواد و روش‌ها

۲-۱- معرفی پارامترها و محدوده مورد مطالعه

شناخته‌شده‌ترین الگوریتم در صنعت، الگوریتم توسعه‌یافته توسط مؤسسه روسکویین^۷ است که به‌طور عمده برای تولید درخت تصمیم استفاده می‌شود [۱۴].

در این پژوهش برای ساخت مدل‌ها و صحت‌سنجی درخت‌های تصمیم‌گیری از داده‌های سازمان فدرال معابر آمریکا استفاده شده است که از آمار تعدادی از پل‌های آمریکا استخراج شده که کلیه اطلاعات به‌صورت میدانی برداشت شده و مربوط به عمق آب‌سستگی اطراف پایه‌های پل در نقاط مختلف را شامل می‌شود [۱۵].

این اطلاعات مورد استفاده شامل سرعت جریان در بالادست (V)، قطر متوسط در ذرات بستر (D_{50})، عمق جریان در بالادست پایه (Y) و عرض پایه (b)، زاویه هجوم

(SVM)، مقایسه شدند. نتایج نشان دادند که مدل طبقه‌بندی و رگرسیون درختی نه‌تنها توانایی و قابلیت تعمیم بهتری داشتند، بلکه مزایای زیادی در تحلیل ویژگی‌های دینامیک زمین‌لرزه و غربالگری متغیرهای مهم داشتند [۸]. هم‌چنین در پژوهشی اقدام به کاربرد روش داده‌کاوی درخت تصمیم در تعیین مؤثرترین عامل‌های نامطلوب آب آشامیدنی در دشت کازرون واقع در غرب استان فارس شد. نتایج تحقیق ایشان با استفاده از مدل‌سازی با بهره‌گیری از روش داده‌کاوی درخت تصمیم CART حاصل شد. نتایج نشان دادند که دو عامل مجموع املاح محلول و کلسیم، تأثیر بیش‌تری بر عدم کیفیت آب آشامیدنی در این دشت داشتند [۹].

مدل RF یک روش یادگیری گروهی است که چندین درخت تصمیم را برای حل مسائل رگرسیون ادغام می‌کند [۱۰]. تنظیم پارامترها یکی از مراحل کلیدی مدل درختان تصادفی RF می‌باشد. مدل جنگل تصادفی براساس یادگیری ماشین نیازی به مفروضات ندارد و مدل را براساس ویژگی‌های داده‌ها می‌سازد. هم‌چنین این مدل قادر است که داده‌ها با ابعاد بالا را تدقیق کند و دقت پیش‌بینی را بدون افزایش قابل توجه محاسبات آن، بهبود ببخشد.

الگوریتم XGB برای کنترل دقیق وزن‌های مختلف و روش GB به‌عنوان نظم‌دهنده شناخته می‌شود. هم‌چنین عملکرد طبقه‌بندی را به‌وسیله کاهش بیش از حد بیش‌برازش با استفاده از انطباق مدل منظم‌تر بهبود می‌بخشد [۱۱].

در سال‌های اخیر این الگوریتم جذابیت بیش‌تری پیدا کرده است و برای اهداف مختلف، تأثیرگذار بوده است و بسیاری از دانشمندان به‌منظور تولید نتایج پیشرفته از آن استفاده می‌کنند. یکی از دلایل مهم موفقیت الگوریتم XGB، مقیاس‌پذیری در چالش‌ها و سناریوهای مختلف می‌باشد [۱۲].

^۷- Rosquin

به رنگ بنفش تیره می‌باشد که نشان از همبستگی کامل دو پارامتر است.

هم‌چنین سلول به رنگ روشن‌تر نشان از همبستگی کم‌تر بین پارامترها می‌باشد.

۲-۳- مدل‌های تصمیم‌گیری غیرپارامتریک درختی

درخت تصمیم‌گیری از نسل جدید روش‌های داده‌کاوی به‌شمار می‌آید که در دو دهه اخیر توسعه زیادی یافته است. از این فن برای کشف و استخراج دانش از یک پایگاه داده و هم برای ایجاد مدل‌های پیش‌بینی می‌توان استفاده نمود.

هم‌چنین درخت تصمیم‌گیری یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد.

درخت تصمیم قادر به تولید توصیفات قابل درک برای انسان بوده و از روابط موجود در یک مجموعه داده تشکیل یافته و می‌تواند برای وظایف دسته‌بندی و پیش‌بینی به کار رود.

این ساختار تصمیم‌گیری می‌تواند به شکل روابط ریاضی و محاسباتی که به توصیف دسته‌بندی و ساده‌سازی یک مجموعه از داده‌ها کمک می‌کنند نیز معرفی شود [۱۷].

الگوریتم DT یک الگوریتم یادگیری نظارت‌شده است که می‌تواند برای حل رگرسیون و طبقه‌بندی استفاده شود، اما بیش‌تر برای حل طبقه‌بندی ترجیح داده می‌شود. در این روش می‌توان از داده‌های بزرگ و پیچیده نیز استفاده کرد [۱۸].

آب به پایه (α) ، طول پایه (L) ، قطر ذراتی که ۸۴ درصد از آنها از قطر آن کوچک‌تر هستند (D_{84}) ، ضریب شکل پایه (k) ، به‌عنوان متغیرهای ورودی و عمق آبشستگی موضعی (d_s) به‌عنوان خروجی در مدل معرفی شدند.

تحلیل داده‌های مورد استفاده در این پژوهش در جدول ۱ نشان داده شده است.

برای مقایسه بهتر کارایی مدل‌های ارائه‌شده از ضریب تبیین (R^2) و خطای میانگین مربعات (MSE) استفاده شده است. نحوه محاسبه این ضرایب در روابط (۱) و (۲) نشان داده شده است:

$$R^2 = 1 - \frac{\sum (d_s^{measured} - d_s^{computed})^2}{\sum (d_s^{measured} - \bar{d}_s^{measured})^2} \quad (1)$$

$$MSE = \frac{\sum (d_s^{measured} - d_s^{computed})^2}{N} \quad (2)$$

که در آن:

MSE = خطای میانگین مربعات

R^2 = ضریب تبیین

$d_s^{measured}$ = عمق آبشستگی مشاهداتی

$d_s^{computed}$ = عمق آبشستگی محاسباتی

$\bar{d}_s^{measured}$ = میانگین اعماق آبشستگی محاسباتی

N = تعداد کل نمونه‌ها

۲-۲- تحلیل حساسیت پارامترها

نقشه حرارتی^۹ و میزان همبستگی بین داده‌ها در شکل ۱ نشان داده شده است.

همان‌طور که در شکل ۱ نشان داده شده است، هرچه رنگ سلول ماتریس تیره‌تر باشد، میزان همبستگی بین پارامترها بیش‌تر بوده است، به‌طوری که روی قطر ماتریس که همبستگی بین دو پارامتر یکسان، نشان داده شده است،

^۹- Heat map

^۸- Mean Squared Error

ساخت درختان رگرسیون جنگل تصادفی به‌طور معمول در مسائل رگرسیون استفاده می‌شود که در آن روابط بسیار غیرخطی بین خروجی داده شده و مجموعه‌ای از ورودی‌های داوطلب شده می‌توان انتظار داشت. یکی از جنبه‌های مهم RF این است که MSE را با حذف هر ورودی محاسبه می‌کند؛ بنابراین این یک روش کمی‌سازی هدفمند است که از اهمیت هر ورودی برای پیش‌بینی خروجی استفاده می‌کند.

مدل RF یک روش یادگیری گروهی است که چندین درخت تصمیم را برای حل مسائل رگرسیون ادغام می‌کند [۱۰].

تنظیم پارامترها یکی از مراحل کلیدی مدل RF می‌باشد. مدل جنگل تصادفی براساس یادگیری ماشین نیازی به مفروضات پیشین ندارد و مدل را براساس ویژگی‌های داده‌ها می‌سازد.

هم‌چنین این مدل قادر است که داده‌ها با ابعاد بالا را دقیق کند و دقت پیش‌بینی را بدون افزایش قابل توجه محاسبات آن بهبود ببخشد [۱۹].

نوع دیگر مدل مبتنی بر درخت تصمیم GB است که این روش یک روش یادگیری ماشین برای مسائل رگرسیون و طبقه‌بندی است. مدل GB ترکیبی خطی از یک سری مدل‌های ضعیف است که به‌صورت تناوبی برای ایجاد یک مدل نهایی قوی ساخته شده است [۲۰].

این روش به خانواده الگوریتم‌های یادگیری گروهی، تعلق دارد و عملکرد آن از الگوریتم‌های اساسی بهتر است. GB یک روش یادگیری ماشین است که مدل‌های درخت تصمیم ضعیف متعدد را به‌منظور ساختن یک مجموعه قوی ترکیب می‌کند. این فرآیند به‌صورت تکراری با هدف کاهش خطاها انجام می‌شود. این فرآیند در شکل ۲ نشان داده شده است [۲۱].

جدول ۱- مشخصات داده‌های مورد استفاده

Table 1. Specifications of the data used

Data Specifications	Maximum value	Minimum value	Standard deviation	Average
Base shape factor (k)	3.00	1.00	0.82	1.79
Base width (b) (ft)	18.01	0.95	16.60	5.00
Bridge base length (L) (ft)	90.00	8.00	149.03	32.42
Water inrush angle to base (α) (deg)	85.00	0.00	96.04	5.48
Flow depth upstream of base (Y) (ft)	65.70	0.40	140.99	14.92
Flow velocity upstream (V) (ft/s)	13.40	0.29	6.89	4.47
Average diameter of bed particles (D_{50}) (ft)	108.00	0.15	356.86	8.11
The diameter of particles that 84% of them are smaller than (D_{84}) (ft)	0.354	0.001	0.004	0.002
Local scour depth (d_s) (ft)	25.10	0.30	19.11	3.80

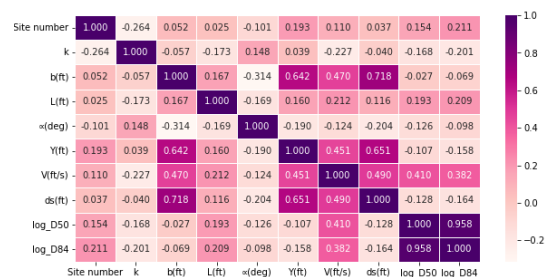


Figure 1. Heat map and correlation between parameters between data

شکل ۱- نقشه حرارتی و میزان همبستگی بین پارامترها بین داده‌ها

مدل RF یک الگوریتم یادگیری ماشین محبوب است که برای مسائل طبقه‌بندی و رگرسیون استفاده می‌شود. این مدل شامل بسیاری از درختان می‌شود که به‌عنوان یک مجموعه عمل می‌کنند و هم‌چنین این الگوریتم از تصادفی بودن ویژگی‌ها استفاده می‌کند.

که مشاهده می‌شود با توجه به مقادیر ضریب تبیین R^2 ، مدل XGB با مقدار ۰/۷۶ نسبت به مدل‌های دیگر برتری دارد که در ادامه با توجه به مقادیری که مدل‌های مختلف اختیار می‌کنند، به ترتیب مدل RF، GB و DT در رده‌های بعدی قرار دارند. همچنین با توجه به مقادیر خطای میانگین مربعات خطا MSE ، مدل XGB مقدار ۰/۱۶ را اختیار کرده که نشان از دقت بیش‌تر این مدل نسبت به مدل‌های دیگر دارد، در ادامه به ترتیب مدل RF، GB و DT در رده‌های بعدی قرار دارند.

در ادامه، مقادیر محاسبه‌شده ارتفاع آبشستگی توسط مدل‌های مختلف با مقادیر واقعی آبشستگی مقایسه و نتایج در شکل ۵ نمایش داده شده است.

جدول ۲- مقادیر ضریب تبیین و خطای میانگین مربعات در مدل‌ها

Table 2. Values of coefficient of determination and mean square error in Models

Model	Mean square error (MSE)	Coefficient of determination (R^2)
DT	0.27	0.60
RF	0.18	0.73
GB	0.22	0.67
XGB	0.16	0.76

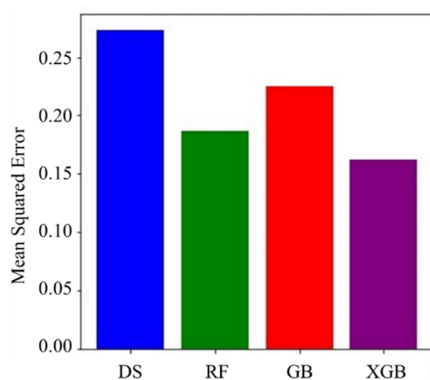


Figure 3. Histogram of the Mean Square Error (MSE) of the models

شکل ۳- هیستوگرام خطای میانگین مربعات (MSE) مدل‌ها

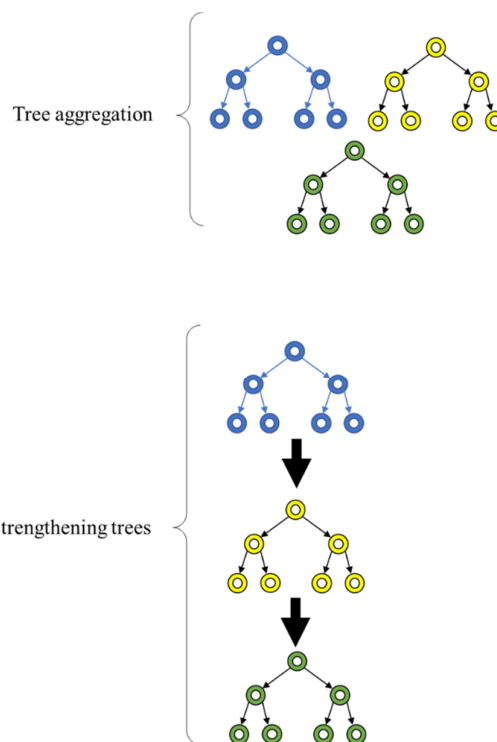


Figure 2. Example of a boosting tree

شکل ۲- نمونه‌ای از درخت تقویت

۳- بحث و بررسی نتایج

مدل‌های درخت تصمیم DT، RF، GB و همچنین XGB در این پژوهش استفاده شده است. برای تعیین عمق آبشستگی پایه پل پس از بررسی متغیرهای مؤثر بر پدیده آبشستگی در این مدل، سرعت جریان در بالادست (V)، قطر متوسط در ذرات بستر (D_{50})، عمق جریان در بالادست پایه (Y) و عرض پایه (b)، زاویه هجوم آب به پایه (α)، طول پایه (L)، قطر ذراتی که ۸۴ درصد از آنها از قطر آن کوچک‌تر هستند (D_{84})، ضریب شکل پایه (k)، به‌عنوان متغیرهای ورودی و عمق آبشستگی موضعی (d_s) به‌عنوان خروجی در مدل معرفی شدند. در ادامه به مقایسه مدل‌های مختلف با استفاده از R^2 و MSE در جدول ۲ پرداخته شده است.

در ادامه با استفاده از نمودار هیستوگرام برای مقایسه بهتر این مقادیر در شکل‌های ۳ و ۴ به مقایسه مدل‌های مختلف براساس R^2 و MSE پرداخته شده است. همان‌طور

۳-۱- تحلیل حساسیت مدل‌ها

به منظور سنجش میزان حساسیت هریک از مدل‌ها به متغیرهای ورودی از پارامتر β استفاده می‌گردد. کاربرد ضریب β هنگامی است که بیش‌تر از یک متغیر مستقل در مدل وجود داشته باشد. در این صورت مقدار β کمک می‌کند که سهم نسبی هر متغیر در پیش‌بینی متغیر وابسته مقایسه شود و به عبارتی بتوان تعیین کرد که کدام متغیرها بیش‌ترین تأثیر را بر متغیر وابسته دارند.

در این تحلیل، متغیر وابسته را مقادیر عمق آبشستگی محاسبه‌شده توسط مدل‌ها و متغیرهای مستقل را سایر پارامترها در نظر گرفته شده است و این ضریب توسط رابطه (۳) محاسبه می‌گردد [۲۲]:

$$\beta = \frac{\text{cov}(x_i, y_i)}{\text{var}(y_i)} \quad (3)$$

که در آن:

x_i = مقادیر پارامترهای مستقل

y_i = مقادیر پارامتر وابسته

ضریب β نسبت به مقدار عدد یک سنجیده می‌شود. بدین معنا که ضریب β برابر با یک، بزرگ‌تر از یک و کوچک‌تر از یک، به ترتیب نشان‌دهنده هم‌شدت بودن، شدیدتر بودن و ملایم‌تر بودن حساسیت پارامترهای موردنظر با پارامتر عمق آبشستگی محاسباتی مدل‌ها می‌باشد.

با توجه به جدول ۳ در چهار مدل ارائه‌شده برای برآورد عمق آبشستگی، عرض پایه (b)، طول پایه (L) بیش‌ترین تأثیر را بر روی عمق آبشستگی دارد. پس از این دو پارامتر، عمق جریان در بالادست پایه (Y) بیش‌ترین تأثیر را بر روی عمق آبشستگی دارد.

ضریب حساسیت برخی از پارامترها با علامت منفی نشان‌دهنده اثر کم رنگ این پارامترها بر روی عمق آبشستگی دارد.

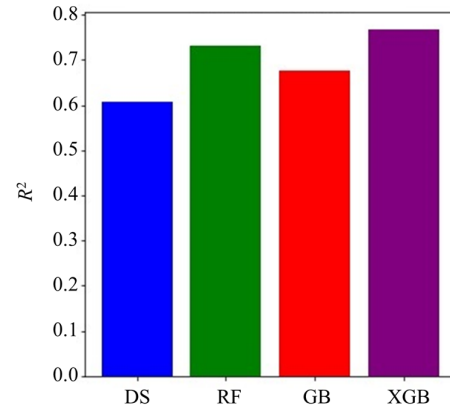


Figure 4. Histogram of the coefficient of determination (R^2) of the models

شکل ۴- هیستوگرام ضریب تبیین (R^2) مدل‌ها

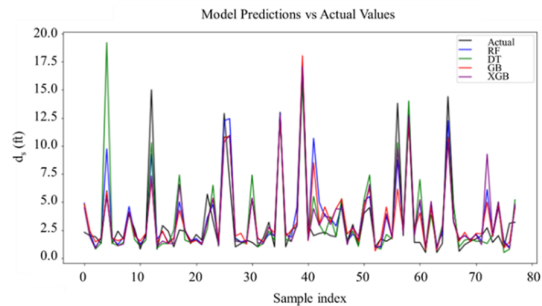


Figure 5. Comparison of calculated scour height values by different models with actual scour values

شکل ۵- مقایسه مقادیر محاسبه‌شده ارتفاع آبشستگی توسط مدل‌های مختلف با مقادیر واقعی آبشستگی

همان‌طور که در شکل ۵ نشان داده شده است مقادیر مختلفی از آبشستگی توسط مدل‌های مختلف درخت تصمیم به‌دست آمده است که با مقادیر آبشستگی مشاهداتی به‌صورت نمودار مقایسه شده است که در محور عرضی این نمودار تعداد ۲۵ درصد از نمونه‌ها به‌صورت تصادفی انتخاب شده و در محور طولی مقادیر آبشستگی محاسبه‌شده این نمونه‌ها، ترسیم شده است. همان‌طور که مشاهده می‌شود مقادیر با تقریب مناسبی با مقادیر مشاهداتی نزدیک می‌باشند که نشان از عملکرد خوب مدل‌های درخت تصمیم مورد استفاده در این پژوهش می‌باشد.

XGB و RF دارای مقادیر قابل قبول و بهتری بودند و بعد از این دو مدل، مدل GB و DT مقادیر بهتری را اختیار کرده‌اند. همچنین در این پژوهش از تحلیل حساسیت پارامترها با استفاده از ضریب β انجام شد که بعد از به دست آوردن این ضریب برای پارامترهای مختلف در مدل‌ها مشاهده شد که عرض پایه (b) و همچنین طول پایه (L)، بیش‌ترین تأثیر را بر روی عمق آبشستگی دارد. پس از این دو پارامتر، عمق جریان در بالادست پایه (Y)، بیش‌ترین تأثیر را بر روی عمق آبشستگی دارد و همچنین ضریب حساسیت برخی از پارامترها با علامت منفی نشان‌دهنده اثر کم‌رنگ این پارامترها بر روی عمق آبشستگی است. روش‌های درخت تصمیم به علت فهم ساده و کارکردن با داده‌های بزرگ و پیچیده یکی از بهترین روش‌های ارزیابی می‌باشد. در ادامه در پژوهش‌های بعدی می‌توان از ترکیب روش‌های جدید مبتنی بر درخت تصمیم با روش‌های دیگر داده‌کاوی و یادگیری ماشین استفاده کرد.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده و این موضوع مورد تأیید ایشان است.

مشارکت نویسندگان

رضوان دالوند نظریه را توسعه داد و محاسبات را انجام داد. مهدی کماسی، رضوان دالوند را تشویق کرد تا یک جنبه خاص را بررسی کند. مهدی کماسی بر یافته‌های این پژوهش نظارت داشت. همه نویسندگان درباره نتایج بحث کردند و در تهیه نسخه نهایی مقاله مشارکت داشتند. رضوان دالوند مقاله را با حمایت مهدی کماسی نگارش کرد. ایده اصلی توسط مهدی کماسی مطرح شد.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

جدول ۳- درصد حساسیت (β) متغیرهای مؤثر بر روی عمق

آبشستگی در مدل‌های مختلف درختی

Table 3. Percentage sensitivity (β) of variables affecting scour depth in different tree models

Specifications of data	DT (%)	RF (%)	GB (%)	XGB (%)
Base shape factor (k)	-3.5	-3.2	-5.3	-4
Base width (b) (ft)	72	81.2	78	86
Bridge base length (L) (ft)	45.5	43.8	46	36.3
Water inrush angle to base (α) (deg)	66	-64	-73	-77
Flow depth upstream of base (Y) (ft)	45.9	40.3	28.5	41.5
Flow velocity upstream (V) (ft/s)	33.7	34.5	38	37.1
Average diameter of bed particles (D_{50}) (ft)	0.4	17	-0.1	-9.1
The diameter of particles that 84% of them are smaller than (D_{84}) (ft)	-1.0	-58	-0.4	-0.3

۴- نتیجه‌گیری

آبشستگی پایه پل یکی از مهم‌ترین مسائل در ایمنی پل‌ها می‌باشد. با توجه به اهمیت این موضوع در پژوهش حاضر سعی بر این شده است از مدل‌های جدید مبتنی بر درخت تصمیم برای ارزیابی این مهم استفاده گردد. در این پژوهش از چهار مدل DT، RF، GB و XGB برای مدل‌سازی استفاده شده است که نتایج مقادیر ارزیابی‌شده توسط مدل‌ها به مقادیر مشاهداتی تا حدود بسیار زیادی مشابهت داشته که نشان از عملکرد خوب این مدل می‌باشد. همچنین این مقادیر عملکرد بهتری نسبت به مدل‌های قدیمی‌تر درخت تصمیم که در پژوهش‌های قبلی از آن استفاده شده، داشته است. همچنین در این پژوهش برای مقایسه بهتر کارایی مدل‌های ارائه‌شده از تحلیل‌های آماری، مانند R^2 و MSE استفاده شده است که با توجه به بررسی پارامترهای فوق در مدل‌های بررسی‌شده، مدل

حامی مالی

سپاسگزاری

این پژوهش هیچ‌گونه کمک هزینه خاصی دریافت
از داوران محترم به‌خاطر ارائه نظرهای ساختاری و
نکرده است. علمی سپاسگزاری می‌شود.

References

- [1] Azad M, Mashkov M. Multi-stage optimization of decision and inhibitory trees for decision tables with many-valued decisions. *European Journal of Operational Research*. 2017 Dec 16; 263(3): 910-921. doi: **10.1016/j.ejor.2017.06.026**.
- [2] Lim S, Cheng N. Prediction of live-bed scour at bridge abutments. *Journal of Hydraulic Engineering*. 1998 Jun 1; 124(6): 635-638. doi: **10.1061/(ASCE)0733-9429(1998)124:6(635)**.
- [3] Melville B, Sutherland AJ. Design method for local scour at bridge piers. *Journal of Hydraulic Division*. 1988 Oct 1; 114(10). doi: **10.1061/(ASCE)0733-9429(1988)114:10(1210)**.
- [4] Mia MF, Nago H. Design method of time-dependent local scour at circular bridge pier. *Journal of Hydraulic Engineering*. 2003 May 15; 129(6): 420-427. doi: **10.1061/(ASCE)0733-9429(2003)129:6(420)**.
- [5] Arneson L, Zevenbergen L, Lagasse F, Clopper F. Evaluating Scour at Bridges. Federal Highway Administration Hydraulic Engineering Circular No. 18: Federal Highway Administration Publication No. FHWA-HIF-12-003 HEC-18.2012. 340 p.
- [6] Shariati H, Khodashenas SR, Esmaili K. Effect of collar shape on the local scouring around the bridge piers, *Journal of Hydraulics*. 2010 Feb; 4(4): 1-11. doi: **10.30482/jhyd.2010.85424 [In Persian]**.
- [7] Karimi N, Heidarnajad M, Masjedi A. Scour depth at inclined bridge piers along a straight path: A laboratory study. *Engineering Science and Technology, an International Journal*. 2017 Aug; 20(4): 1302-1307. doi: **10.1016/j.jestech.2017.07.004**.
- [8] Yao S, Kronenburg A, Shamooni A, Stein OT, Zhang W. Gradient boosted decision trees for combustion chemistry integration. *Applications in Energy and Combustion Science*. 2022 Sep; 11, 100077. doi: **10.1016/j.jaecs.2022.100077**.
- [9] Sulaimanpour S, Mesbah S, Hedayati B. Application of K-Means and CART data mining algorithms in determining the most effective factors of drinking water quality in Noorabad Plain, Fars Province. 11th National Seminar on Watershed Management. Yasouj. University of Yasouj. 2016 Apr 19 [In Persian].
- [10] Grömping U. Variable importance assessment in regression: Linear regression versus random forest. *The American Statistician*. 2009 Jan 01; 63(4): 308-319. doi: **10.1198/tast.2009.08199**.
- [11] Carmona P, Climent F, Momparler A. Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*. 2019 May; 61: 304-323. doi: **10.1016/j.iref.2018.03.008**.
- [12] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016 Jan 01; pp. 785-794. doi: **10.1145/2939672.2939785**.
- [13] Adnane M, El M, El Fkihi S, Haj Thami RO. Prediction demand for classified ads using machine learning: An experiment study. In: *Proceedings of the 2nd International Conference on Networking, Information systems & security*. 2019 Mar 27; 39: 1-6. doi: **10.1145/3320326.332037**.
- [14] An Y, Zhou H. Short-term effect evaluation model of rural energy construction revitalization based on ID3 decision tree algorithm. *Energy Reports*. 2020 Jul; 8(4): 1004-1012. doi: **10.1016/j.egyr.2022.01.239**.
- [15] Mueller DS, Wanger CR. Field Observations and Evaluations of Streambed Scour at Bridges. US Department of Transportation, Federal Highway Admin. Turner-Fairbank Highway Research Center, McLean, Va 22101-2296. 2005 May.
- [17] Tsai, Ch-F, Chiou Y-J. Earnings management prediction: A pilot study of combining neural networks and decision trees. *Expert Systems with Applications*. 2009 Apr; 36(3): 7183-7191. doi: **10.1016/j.eswa.2008.09.025**.

- [18] Dilmi S. A combined water quality classification model based on kernel principal component analysis and machine learning techniques. *Desalination and Water Treatment*. 2022 Dec; 279: 61-67. **doi: 10.5004/dwt.2022.29069.**
- [19] Zahiri J. Riprap Design for Bridge Piers Using Nonparametric Models. *Water and Soil Science*. 2016 Mar; 5(4): 187-199 [In Persian].
- [20] Hastie T. The elements of statistical learning, Data Mining, Inference, and Prediction. Second Edition. New York: Springer; 2009 Aug. **doi: 10.1007/978-0-387-84858-7.**
- [21] Yao S, Kronenburg A, Shamooni A, Stein OT, Zhang W. Gradient boosted decision trees for combustion chemistry integration. *Applications in Energy and Combustion Science*. 2022 Sep; 11, 100077. **doi: 10.1016/j.jaecs.2022.100077.**
- [22] Mia MF, Nago H. Design method of time-dependent local scour at circular bridge pier. *Journal of Hydraulic Engineering*. 2003 May 15; 129(6): 420-427. **doi: 10.1061/(ASCE)0733-9429(2003)129:6(420).**