Journal of Industrial Innovations 4(2) (2026) 70 – 89

Industrial Innovations

https://jii.araku.ac.ir

ISSN: 3092-7811

Predictive Maintenance of Vehicles Using Machine Learning and Ensemble Algorithms

Omid Veisi [a*]

[a] Department of Industrial Engineering, Faculty of Engineering and Aviation, Imam Ali University (AS), Tehran, Iran.
omid_vte@yahoo.com

Original Article

Use your device to scan and read the article online

Citation: Veisi O. Predictive Maintenance of Vehicles Using Machine Learning and Ensemble Algorithms. Industrial Innovations. 2026;4 (2):70-89.

https://doi.org/10.66224/jii.4.2.70

KEYWORDS

Machine Learning;
Predictive Vehicle Maintenance;
Failure Prediction;
Ensemble Models;
Vehicle Data Analysis.

ABSTRACT

In recent years, the use of machine learning methods in the field of predictive vehicle maintenance has become an effective approach for reducing unexpected failures, increasing safety, and optimizing repair costs. The aim of this study is to develop and evaluate various machine learning models to predict vehicle maintenance needs based on operational data from vehicle systems. After preprocessing, the collected data were split into 80% training and 20% testing sets and modeled using twelve algorithms, including SVM, LR, NB, KNN, RF, DT, AdaBoost, Bagging, Stacking, XGBoost, CatBoost, and LightGBM. The performance of the models was assessed using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The results indicated that tree-based models and ensemble methods performed superiorly, among which LightGBM achieved the best performance with an AUC of 0.9475 and an F1-score of 0.9613. The findings highlight that machine learning algorithms particularly ensemble models can play a key role in failure prediction and the design of intelligent vehicle maintenance systems.

Extended Abstract

1. Introduction

The literature indicates that maintenance and repair management plays a critical role in improving equipment reliability, reducing operational costs, and enhancing organizational productivity. Recent advancements in predictive maintenance have been driven by technologies such as the Internet of Things (IoT), cloud computing, and big data analytics, which enable real-time monitoring and fault prediction. Machine learning and deep learning techniques, including Random Forest, XGBoost, CNN, and LSTM, have demonstrated strong capabilities in identifying complex failure patterns from sensor and operational data. Previous studies have shown that ensemble learning methods often outperform traditional approaches due to their ability to model nonlinear relationships and improve prediction accuracy. Furthermore, the integration of IoT-based monitoring systems with artificial intelligence has significantly enhanced real-time maintenance decision-making. Research has also highlighted the importance of feature selection and data preprocessing in improving model performance. Overall, the literature confirms that data-driven predictive maintenance is an effective strategy for reducing downtime, extending equipment lifespan, and optimizing maintenance operations.

2. Modeling and formulation

The developed model demonstrated strong predictive capability for maintenance-related classification tasks by effectively capturing complex patterns within the dataset. Among all evaluated algorithms, LightGBM achieved the best overall performance, indicating its suitability for handling high-dimensional and potentially imbalanced data. The results highlight the

* Corresponding author.
E-mail address: omid_vte@yahoo.com
DOI: https://doi.org/10.66224/jii.4.2.70
Received: November 14, 2025; Received in revised form: February 15, 2026; Accepted: February 25, 2026.
Article type: Research paper

advantages of ensemble learning techniques in improving classification accuracy and robustness. Therefore, the proposed modeling framework can serve as a reliable decision-support tool for predictive maintenance applications.

3. Figures and Tables

The results of this study indicate that ensemble-based and gradient boosting models outperform traditional machine learning algorithms. This finding is consistent with previous studies in the field of predictive maintenance, where ensemble methods have demonstrated superior performance due to their ability to reduce variance and effectively model complex nonlinear relationships. However, some prior studies reported XGBoost as the best-performing algorithm compared to other approaches, whereas in the present dataset, LightGBM achieved the highest performance. This discrepancy may be attributed to differences in feature characteristics, class distribution, and dataset dimensionality.

Table 1 Evaluation Results of the Algorithms

Algorithm	Precision	Accuracy	Recall	F1-Score	ROC-AUC
SVM	0.8221	0.8978	0.8804	0.8890	0.8692
LR	0.8452	0.9350	0.8691	0.9009	0.9078
NB	0.8945	0.9475	0.9207	0.9339	0.9464
KNN	0.8426	0.9658	0.8351	0.8957	0.9227
RF	0.9053	0.9545	0.9254	0.9379	0.9350
DT	0.8894	0.9809	0.8785	0.9269	0.9200
Bagging	0.9316	0.9786	0.9347	0.9562	0.9431
XGBoost	0.8109	0.8089	0.9989	0.8939	0.9178
Stacking	0.9373	0.9767	0.9439	0.9600	0.9451
AdaBoost	0.8203	0.9581	0.8102	0.8780	0.9179
CatBoost	0.9048	0.9586	0.9205	0.9392	0.9320
LightGBM	0.9388	0.9706	0.9522	0.9613	0.9475

4. Conclusion

This study demonstrated the effectiveness of machine learning techniques for predicting maintenance requirements and supporting predictive maintenance decision-making. Among the evaluated models, LightGBM achieved the best performance, while Bagging and Naïve Bayes also produced strong results. The findings indicate that tree-based and ensemble learning methods provide higher accuracy, stability, and predictive capability than simpler algorithms. Consequently, these models can serve as valuable decision-support tools for reducing unexpected equipment failures and improving operational efficiency in maintenance management systems.

نوآوری‌های صنعتی

شاپا: ۷۸۱۱-۳۰۹۲

نگهداری و تعمیرات پیش‌بینانه خودرو با استفاده از الگوریتم‌های یادگیری ماشین و الگوریتم‌های ترکیبی

امید ویسی [الف*]

[الف] گروه مهندسی صنایع، دانشکده مهندسی و پرواز، دانشگاه امام علی (ع)، تهران، ایران. omid_vte@yahoo.com

واژگان کلیدی

یادگیری ماشین؛ نگهداری و تعمیرات پیش‌بینانه خودرو؛ پیش‌بینی خرابی؛ مدل‌های ترکیبی؛ تحلیل داده‌های خودرو.

تاریخ دریافت: ۱۴۰۴/۰۸/۲۳

تاریخ بازنگری: ۱۴۰۴/۱۱/۲۶

تاریخ پذیرش: ۱۴۰۴/۱۲/۰۶

چکیده

در سال‌های اخیر، بهره‌گیری از روش‌های یادگیری ماشین در حوزه‌ی نگهداری و تعمیرات پیش‌بینانه خودرو به یکی از راهکارهای کارآمد برای کاهش خرابی‌های ناگهانی، افزایش ایمنی و بهینه‌سازی هزینه‌های تعمیر تبدیل شده است. هدف این پژوهش، توسعه و ارزیابی مدل‌های مختلف یادگیری ماشین برای پیش‌بینی نیاز به تعمیر خودرو بر اساس داده‌های عملکردی سیستم‌های خودرو است. داده‌های گردآوری شده پس از پیش‌پردازش و تقسیم به مجموعه‌های آموزش ۸۰ درصد و آزمون ۲۰ درصد، توسط دوازده الگوریتم شامل SVM، LR، NB، KNN، RF، DT، AdaBoost، Bagging، Stacking، XGBoost، CatBoost و LightGBM مدل‌سازی شدند. عملکرد مدل‌ها با معیارهای دقت، صحت، بازخوانی، F1-score و ROC-AUC ارزیابی شد. نتایج نشان داد مدل‌های مبتنی بر درخت و روش‌های ترکیبی، عملکرد برتری دارند و در میان آن‌ها، LightGBM با مقدار AUC برابر با ۰/۹۴۷۵ و F1-score برابر با ۰/۹۶۱۳ بهترین عملکرد را ارائه داده است. یافته‌ها بیانگر آن است که الگوریتم‌های یادگیری ماشین، به‌ویژه مدل‌های ترکیبی، می‌توانند در پیش‌بینی خرابی و طراحی سامانه‌های نگهداری هوشمند خودرو نقشی کلیدی ایفا کنند.

۱- مقدمه

با گسترش سریع صنعت حمل‌ونقل و افزایش ناوگان‌های شهری، لجستیکی و عمومی، مدیریت کارآمد نگهداری وسایل نقلیه به یکی از مؤلفه‌های راهبردی در کاهش هزینه‌های عملیاتی و ارتقای ایمنی تبدیل شده است. در چنین شرایطی، رویکردهای سنتی مبتنی بر سرویس‌های دوره‌ای یا تعمیر پس از خرابی، پاسخگوی پیچیدگی فنی و حجم بالای عملیات ناوگان‌های مدرن نیستند. این روش‌ها معمولاً بر برنامه‌های زمان‌محور متکی‌اند و قادر به تشخیص به‌موقع خرابی‌های پنهان یا زودهنگام نیستند؛ مسئله‌ای که می‌تواند به افزایش هزینه‌های تعمیراتی، توقف‌های ناخواسته و کاهش قابلیت اطمینان عملیاتی منجر شود. در حوزه‌هایی مانند حمل‌ونقل عمومی و خدمات شهری، چنین اختلالاتی پیامدهایی فراتر از هزینه مستقیم داشته و بر کیفیت خدمات و اعتماد عمومی نیز اثرگذار است. در پاسخ به این چالش‌ها، نگهداری پیش‌بینانه به‌عنوان یک راهکار داده‌محور و هوشمند مطرح شده است. این رویکرد با بهره‌گیری از داده‌های حسگری، سوابق عملیاتی و اطلاعات عملکردی خودروها، امکان ارزیابی وضعیت فنی و پیش‌بینی احتمال خرابی را پیش از وقوع فراهم می‌کند. برخلاف نگهداری زمان‌محور، تصمیم‌گیری در این چارچوب بر اساس وضعیت واقعی تجهیز صورت می‌گیرد و از این‌رو می‌تواند تخصیص منابع تعمیراتی را بهینه سازد. پیش‌بینی به‌موقع خرابی نه‌تنها موجب کاهش هزینه‌های مستقیم نگهداری می‌شود، بلکه از توقف‌های پیش‌بینی‌نشده جلوگیری کرده و

* نویسنده مسئول؛

آدرس پست الکترونیک: omid_vte@yahoo.com

پایداری عملیاتی ناوگان را افزایش می‌دهد [۱].

تحول دیجیتال در صنعت حمل‌ونقل و گسترش زیرساخت‌های داده‌محور، زمینه بهره‌گیری گسترده از الگوریتم‌های یادگیری ماشین و یادگیری عمیق را در این حوزه فراهم کرده است. این الگوریتم‌ها قادرند با تحلیل داده‌های چندبعدی شامل وضعیت ترمز، باتری، الگوهای رانندگی، مصرف سوخت و سوابق تعمیرات، روابط پیچیده و غیرخطی میان متغیرها را مدل‌سازی کرده و الگوهای خرابی را شناسایی کنند [۲]. از این منظر، پیش‌بینی هوشمند خرابی نه صرفاً یک ابزار نگهداری، بلکه بخشی از معماری سیستم‌های حمل‌ونقل هوشمند محسوب می‌شود. با وجود پیشرفت‌های صورت‌گرفته، برخی محدودیت‌ها در مطالعات پیشین مشاهده می‌شود. بسیاری از پژوهش‌ها بر مجموعه‌داده‌های محدود یا مصنوعی تکیه داشته‌اند که تعمیم‌پذیری نتایج را با چالش مواجه می‌کند. افزون بر این، تمرکز برخی مطالعات بر تعداد اندکی از ویژگی‌های فنی، موجب نادیده‌گرفتن سایر عوامل مؤثر در رفتار خرابی خودرو شده است. همچنین در مواردی، استفاده از مدل‌های ساده و منفرد بدون بهره‌گیری از رویکردهای ترکیبی، موجب محدود شدن دقت پیش‌بینی شده است [۳]. این خلأها ضرورت توسعه مدلی جامع با بهره‌گیری از مجموعه‌ای متنوع از ویژگی‌های عملیاتی و استفاده از الگوریتم‌های پیشرفته را آشکار می‌سازد.

بر این اساس، هدف پژوهش حاضر توسعه یک مدل پیش‌بینی هوشمند برای تعیین نیاز به تعمیرات در ناوگان‌های خودرویی با استفاده از ترکیب ویژگی‌های فنی و عملیاتی و به‌کارگیری الگوریتم‌های مختلف یادگیری ماشین است. در این مطالعه، مجموعه‌ای غنی از متغیرها شامل وضعیت ترمز، باتری، سوابق گزارش‌شده، تاریخچه سرویس‌ها، میزان مصرف سوخت و شاخص‌های رفتاری رانندگی مورد استفاده قرار گرفته و عملکرد مدل‌های مختلف به‌صورت نظام‌مند ارزیابی و مقایسه شده است. این پژوهش تلاش دارد با ارائه یک چارچوب داده‌محور و مقایسه تحلیلی مدل‌ها، گامی در جهت بهبود دقت پیش‌بینی و ارتقای تصمیم‌گیری در مدیریت هوشمند ناوگان خودروها بردارد.

۲- مبانی نظری

پیشرفت روزافزون فناوری و ضرورت ارتقای بهره‌وری ماشین‌آلات و تجهیزات، روند حرکت به سمت خودکارسازی را سرعت بخشیده است. امروزه بخش قابل توجهی از زمان و انرژی انسان‌ها صرف آن می‌شود که چگونه می‌توان سیستم‌ها را در حالت عملیاتی نگه داشت و یا در صورت بروز خرابی، در کوتاه‌ترین زمان ممکن آن‌ها را مجدداً راه‌اندازی کرد. با حرکت صنایع به سوی مکانیزاسیون گسترده‌تر و تولید مداوم، فعالیت‌های نگهداری و تعمیرات نیز با پیچیدگی‌های بیشتری همراه شده‌اند. در همین راستا، دانش نگهداری و تعمیرات دچار تحولات چشمگیری گردیده است. در این میان، رویکرد نگهداری و تعمیرات بهره‌ور جامع به‌عنوان شیوه‌ای گروه‌محور برای ارتقای کارایی تجهیزات مطرح است که در آن، اپراتورها به‌طور مستقیم و فردی مسئول انجام وظایف پایه‌ای نگهداری و تعمیرات ماشین‌آلات خود می‌شوند [۴]. از آنجا که نگهداری و تعمیرات یکی از ارکان اصلی بهره‌وری به شمار می‌رود، می‌توان آن را نوعی فرهنگ دانست که هدف آن بهبود شرایط موجود است؛ فرهنگی که به نیروی انسانی می‌آموزد چه مسیری را برگزیند و چگونه عمل کند تا بیشترین بازدهی حاصل شود. دستیابی به اثربخشی در تجهیزات یک سازمان، در گرو توجه به لزوم مدیریت نظام‌هایی متشکل از انسان، تأسیسات، ماشین‌آلات، تجهیزات و مواد است؛ نظامی که هدف آن افزایش طول عمر مفید دارایی‌ها، بهبود بهره‌برداری و استفاده بهینه از منابعی چون بودجه و نیروی انسانی است. تحقق این امر نیازمند یک نظام کارآمد برای برنامه‌ریزی، تحلیل، کنترل و به‌کارگیری روش‌های مناسب مدیریتی است؛ از همین رو ایجاد یک سیستم نگهداری و تعمیرات پویا و اثربخش اهمیت فراوان دارد. نگهداری و تعمیرات را می‌توان نوعی هنر دانست، زیرا هم پیش از بروز خرابی و هم هنگام وقوع آن، امکان انتخاب اقدامات و رویکردهای گوناگون وجود دارد. در این زمینه، نقش مدیران، سرپرستان، کارشناسان و مسئولان نگهداری و تعمیرات بسیار برجسته‌تر از سایر عوامل، حتی «ماهیت خرابی»، است. در واقع، در فرآیند دستیابی به نگهداری اثربخش و ارتقای بهره‌وری تجهیزات، عامل انسانی بیشترین اهمیت را دارد و به همین دلیل در کانون توجه سیستم‌های بهره‌ور قرار می‌گیرد [۴].

دانش برنامه‌ریزی نگهداری و تعمیرات سابقه‌ای طولانی دارد و در طی این مدت مزایای فراوانی از به‌کارگیری آن در صنایع

مختلف نمایان شده است. امروزه این دانش بستری مناسب برای ارتقای کیفیت و در نتیجه افزایش سودآوری در سازمان‌ها فراهم کرده است. در حال حاضر، بخش عمده‌ای از مدیران و مهندسان صنایع و خدمات اهمیت ویژه‌ای برای بهره‌گیری از این دانش قائل هستند و علاقه‌مندی خود را نسبت به اجرای آن نشان داده‌اند. تجربه‌ها نشان داده است که مدیریت و مهندسی نگهداری و تعمیرات می‌تواند هزینه‌های تعمیراتی را به میزان ۱۵ تا ۳۰ درصد از بودجه‌های سالانه کاهش دهد، که این صرفه‌جویی عددی بسیار چشمگیر است. با توجه به نتایج ارزشمند حاصل از به‌کارگیری برنامه‌ریزی نت، استفاده از فناوری‌های نوین و سیستم‌های نرم‌افزاری برای مدیریت مکانیزه در بخش‌های گوناگون سازمانی امری اجتناب‌ناپذیر شده است. به طور کلی، نگهداری و تعمیرات مجموعه‌ای از فعالیت‌ها و اقدامات است که هدف آن حفظ و کنترل تجهیزات، دستگاه‌ها و تأسیسات در شرایط مطلوب، افزایش عمر مفید دارایی‌های فیزیکی یا بازگرداندن آن‌ها به وضعیت قابل قبول مطابق با استانداردها است. تحقق این اهداف مستلزم طراحی و اجرای یک نظام کارآمد نگهداری و تعمیرات است که با بهره‌گیری از راهبردهای مناسب و متناسب با شرایط سازمان و نوع تجهیزات انتخاب می‌شود. در این میان، برخی از راهبردهای رایج در حوزه نت عبارت‌اند از: نگهداری و تعمیرات بر اساس خرابی[۱]، نگهداری اضطراری، نگهداری پیشگیرانه[۲] و نگهداری مبتنی بر شرایط[۳] [۵].

نگهداری و تعمیرات مجموعه‌ای از فعالیت‌ها، عملیات و روش‌های نظام‌مند است که با هدف حفظ و کنترل تجهیزات در وضعیت مطلوب یا بازگرداندن آن‌ها به شرایط مناسب و استاندارد قابل‌قبول و همچنین کاهش استهلاک و خرابی‌ها انجام می‌گیرد. غایت اصلی این فرآیند ایجاد یک چارچوب سازمان‌یافته برای پایش وضعیت تجهیزات و دستگاه‌ها، ارتقای قابلیت‌های عملکردی آن‌ها، کاهش نرخ خرابی و توقف، کاهش هزینه‌های ناشی از تعمیرات و در نهایت افزایش کیفیت و بهره‌وری است. در واقع، زنجیره نگهداری و تعمیرات نقش کلیدی در بهبود کارایی کلی سازمان ایفا می‌کند [۶]. نگهداری و تعمیرات به عنوان دو مقوله اساسی و مکمل، ضامن افزایش طول عمر تجهیزات محسوب می‌شوند. اجرای صحیح این فرآیندها نیازمند به‌کارگیری روش‌های علمی و ترکیبی از فعالیت‌های مدیریتی و مهندسی است که در قالب چرخه اقتصادی عمر دارایی‌های فیزیکی به کار گرفته می‌شوند [۷]. از آنجا که موفقیت سازمان‌ها بیش از پیش به استفاده درست و مؤثر از تجهیزات وابسته شده است، مدیریت بهینه در حوزه نگهداری و تعمیر اهمیت ویژه‌ای دارد. بی‌توجهی به این امر به‌ویژه در صنایع حیاتی مانند نیروی دریایی، می‌تواند منجر به گسترش عیوب در یگان‌های عملیاتی، افزایش مصرف قطعات یدکی، بالا رفتن هزینه‌ها و ایجاد وقفه در مأموریت‌ها شود. ازاین‌رو، مدیریت این حوزه نیازمند بهره‌گیری از دانش فنی و استفاده عملیاتی از توان تجهیزات است تا نهایت بهره‌وری محقق شد. مدیریت، برنامه‌ریزی و کنترل فرآیند نگهداری و تعمیرات با هدف شناسایی نیازها، اتخاذ رویکردهای مناسب و اقتصادی و تأمین منابع موردنیاز انجام می‌شود. در این راستا، وجود یک سیستم کارآمد تحلیل داده‌های مربوط به تعمیر و نگهداری و استفاده از ابزارهای داده‌کاوی، امکان تصمیم‌گیری دقیق‌تر و دستیابی به اهداف یاد شده را تسهیل و مسیر تحقق آن‌ها را هموار می‌سازد [۸].

پیشرفت‌های فناورانه اخیر در حوزه نگهداری پیش‌بینانه عمدتاً تحت تأثیر ظهور فناوری‌هایی همچون اینترنت اشیا، پردازش ابری و تحلیل کلان داده قرار گرفته‌اند. این فناوری‌ها امکان پایش بلادرنگ وضعیت تجهیزات و خودروها را فراهم کرده و حجم عظیمی از داده‌های سنسوری و عملیاتی را برای تحلیل در دسترس قرار می‌دهند. معماری‌های نوین سیستم‌های فیزیکی-سایبری به‌ویژه در بستر Industry 4.0، به ادغام این فناوری‌ها برای بهینه‌سازی فرایندهای نگهداری و کاهش توقف‌های ناخواسته کمک کرده‌اند [۸]. مرورهای جامع نشان می‌دهند که استفاده از داده‌کاوی و روش‌های داده محور، همراه با زیرساخت‌های اینترنت اشیا و کلان‌داده، دقت و قابلیت اطمینان پیش‌بینی خرابی‌ها را به‌طور چشمگیری افزایش داده است [۹].

در کنار این زیرساخت‌ها، استفاده از الگوریتم‌های یادگیری عمیق به‌عنوان یک روند نوین در نگهداری پیش‌بینانه رشد چشمگیری داشته است. مدل‌هایی مانند شبکه‌های عصبی کانولوشنی[۴] (CNN) و شبکه‌های حافظه طولانی‌مدت[۵] (LSTM)

[1] Corrective Maintenance
[2] Preventive Maintenance
[3] Condition-Based Maintenance
[4] Convolutional Neural Network
[5] Long Short-Term Memory

توانایی بالایی در استخراج الگوهای پیچیده از داده‌های سنسوری چندبعدی دارند و برای برآورد عمر باقی‌مانده تجهیزات مورد استفاده قرار گرفته‌اند [۱۰]. همچنین رویکردهای بیزین داده محور، با بهره‌گیری از داده‌های به دست‌آمده از سیستم‌های متصل، امکان پیش‌بینی مستقیم و دقیق‌تر را فراهم کرده‌اند [۱۱]. این تحولات فناورانه مسیر نگهداری پیش‌بینانه را از یک ابزار صرفاً تشخیصی به یک سیستم هوشمند و پیش‌بینی‌کننده ارتقا داده‌اند. نگهداری پیش‌بینانه یکی از رویکردهای نوین در مدیریت سلامت ماشین‌آلات و سیستم‌های مکانیکی است که با استفاده از تحلیل داده‌های واقعی عملکرد، به پیش‌بینی احتمال وقوع خرابی‌های آتی می‌پردازد. این روش در تضاد با نگهداری اصلاحی است که تنها پس از بروز خرابی اقدام به تعمیر می‌کند و نیز نسبت به نگهداری پیشگیرانه که به‌صورت زمان‌بندی شده و بدون توجه به وضعیت واقعی تجهیزات انجام می‌شود، عملکرد به‌مراتب هوشمندانه‌تری دارد [۱۲]. در حوزه حمل‌ونقل و ناوگان خودرو، نگهداری پیش‌بینانه به‌طور فزاینده‌ای در حال گسترش است؛ چرا که با فراهم آوردن امکان تشخیص به‌موقع مشکلات، می‌تواند هزینه‌های تعمیرات، زمان‌های توقف و مخاطرات ایمنی را به‌شدت کاهش دهد. این نوع نگهداری معمولاً متکی بر داده‌هایی نظیر کارکرد وسایل نقلیه، سنسورهای وضعیت فنی، سوابق تعمیرات و رفتار رانندگی است و با تحلیل آن‌ها به کمک مدل‌های هوش مصنوعی، پیش‌بینی دقیقی از احتمال خرابی ارائه می‌دهد [۱۳]. مطالعات متعددی نشان داده‌اند که استفاده از نگهداری پیش‌بینانه در مدیریت ناوگان‌های حمل‌ونقل می‌تواند منجر به کاهش بیش از ۳۰ درصدی هزینه‌های نگهداری، افزایش عمر مفید قطعات و بهبود تصمیم‌گیری‌های مدیریتی شود [۱۴]. همچنین، با رشد فناوری‌هایی مانند اینترنت اشیا و کلان‌داده[1]، قابلیت‌های نگهداری پیش‌بینانه در تحلیل‌های لحظه‌ای و تصمیم‌گیری بلادرنگ به‌شدت گسترش یافته است [۱۵].

در دهه‌ی اخیر، استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی خرابی‌ها و نیاز به تعمیرات در وسایل نقلیه به‌عنوان یکی از محورهای اصلی تحقیقات حوزه مهندسی نگهداری و هوش مصنوعی مطرح شده است. هدف اصلی این مطالعات، تشخیص زودهنگام خطاها و پیش‌بینی خرابی قبل از وقوع آن‌ها از طریق تحلیل داده‌های حسگرها و پارامترهای عملکردی خودرو است. پژوهش‌های اخیر نشان می‌دهند که ترکیب مدل‌های مبتنی بر داده با روش‌های تحلیلی کلاسیک، موجب افزایش دقت و کاهش هزینه‌های عملیاتی در سیستم‌های نگهداری پیشگویانه شده است [۱۶]. در مطالعه‌ای دیگر مدل‌های یادگیری ماشین بر اساس داده‌های تله‌متری در سیستم‌های انرژی و خودرو به منظور شناسایی الگوهای خطا طراحی شدند. نتایج نشان دادند که استفاده از الگوریتم‌هایی نظیر RF و XGBoost توانایی بالایی در شناسایی ناهنجاری‌ها دارد. همچنین ردی و رائو در کتاب خود به بررسی کاربردهای یادگیری عمیق و ویژگی‌های استخراج شده از داده‌های حسگری خودرو پرداختند و بر اهمیت ترکیب داده‌های زمان واقعی و روش‌های یادگیری ترکیبی تأکید کردند [۱۷].

مطالعات پیشین مانند [۱۸, ۱۹] نیز از الگوریتم‌های SVM و NB برای تشخیص نقص در سیستم‌های ترمز و موتور استفاده کرده‌اند و گزارش داده‌اند که انتخاب ویژگی‌های مناسب نقش کلیدی در بهبود عملکرد مدل دارد. از سوی دیگر، تحقیقات [۴, ۵] به بررسی استفاده از مدل‌های ترکیبی[2] مانند Stacking و Bagging پرداخته‌اند که موجب افزایش پایداری و کاهش حساسیت مدل نسبت به نویز داده‌ها شده است.

در مطالعات جدیدتر، کاربرد یادگیری عمیق و مدل‌های ترکیبی در تعمیر و نگهداری خودروها نیز رو به گسترش است. به عنوان مثال، [۲۰] با استفاده از شبکه‌های عصبی کانولوشنی و داده‌های لرزشی توانستند خرابی‌های سیستم انتقال نیرو را با دقت بیش از ۹۵ درصد پیش‌بینی کنند. همچنین [۲۱] نشان دادند که ترکیب روش‌های یادگیری ماشین با تحلیل مبتنی بر اینترنت اشیاء[3] (IoT) می‌تواند عملکرد سیستم‌های پایش خودرو را در زمان واقعی به‌طور قابل توجهی بهبود بخشد. در جدول ۱ مرور ادبیات به صورت خلاصه و مقایسه‌ای بین پژوهش‌های انجام شده در این حوزه انجام شده است.

[1] Big Data
[2] Ensemble Learning
[3] Internet of Things

جدول ۱ پیشینه پژوهش

مرجع	نوع داده	سطح تحلیل	الگوریتم	مسئله	محدودیت
[۱۶]	داده‌های تله‌متری سیستم‌های انرژی و خودرو	تشخیص خطا	RF, XGBoost	شناسایی ناهنجاری و الگوهای خطا	تمرکز بر تشخیص خطا، نه تصمیم نیاز به تعمیر در سطح ناوگان
[۱۷]	داده‌های حسگری خودرو	استخراج ویژگی و پیش‌بینی خرابی	RF,DT	تحلیل داده‌های زمان واقعی	تمرکز بیشتر بر استخراج ویژگی؛ فاقد مقایسه جامع مدل‌های کلاسیک و ترکیبی
[۹]	داده‌های سیستم ترمز	خرابی قطعه خاص	SVM	تشخیص نقص سیستم ترمز	محدود به یک زیرسیستم خاص
[۱۰]	داده‌های موتور خودرو	خرابی قطعه خاص	Naïve Bayes	تشخیص نقص موتور	عدم بررسی ویژگی‌های عملیاتی و رفتاری
[۳]	داده‌های صنعتی	طبقه‌بندی خرابی	Stacking	بهبود پایداری مدل	استفاده محدود از داده‌های واقعی ناوگان چندویژگی
[۲۲]	داده‌های نویزی صنعتی	طبقه‌بندی	Bagging	کاهش حساسیت به نویز	تمرکز بر عملکرد الگوریتم، نه چارچوب جامع داده‌ای
[۲۳]	داده‌های لرزشی سیستم انتقال نیرو	پیش‌بینی خرابی قطعه	CNN	پیش‌بینی خرابی با دقت بالا	داده تک‌منبعی (لرزش)، عدم ترکیب سوابق تعمیراتی
[۲۴]	داده‌های IoT خودرو	پایش بلادرنگ	مدل‌های ترکیبی ML, IoT	بهبود مانیتورینگ زمان واقعی	تمرکز بر زیرساختIoT ، نه مقایسه نظام‌مند مدل‌ها
[۱۵]	داده‌های سنسوری و کلان‌داده	نگهداری پیش‌بینانه عمومی	KNN, DT	کاهش هزینه و توقف	اغلب فاقد تحلیل مقایسه‌ای گسترده بین الگوریتم‌ها

با وجود گسترش پژوهش‌ها در حوزه نگهداری پیش‌بینانه خودرو، بررسی نظام‌مند ادبیات نشان می‌دهد که چندین خلأ اساسی همچنان پابرجاست. نخست آن‌که بخش عمده‌ای از مطالعات پیشین بر پیش‌بینی خرابی یک قطعه یا زیرسیستم خاص (مانند موتور، سیستم ترمز یا انتقال نیرو) متمرکز بوده‌اند و کمتر به مسئله تصمیم‌گیری درباره «نیاز به تعمیر» در سطح کلان ناوگان پرداخته‌اند؛ در حالی‌که در مدیریت ناوگان، تصمیم‌گیری عملیاتی معمولاً در سطح خودرو و نه قطعه منفرد انجام می‌شود. دوم آن‌که بسیاری از پژوهش‌ها از مجموعه‌داده‌های محدود، آزمایشگاهی یا شبیه‌سازی‌شده استفاده کرده‌اند که تعمیم‌پذیری نتایج آن‌ها به شرایط واقعی عملیاتی ناوگان‌های شهری و خدماتی با تردید همراه است. سوم، در اغلب مطالعات تمرکز بر تعداد محدودی از ویژگی‌های فنی مانند کارکرد یا دمای موتور بوده و ترکیب همزمان ویژگی‌های عملیاتی (مانند الگوهای رانندگی و مصرف سوخت)، ویژگی‌های فنی (وضعیت ترمز و باتری) و سوابق تعمیراتی و گزارش‌های خطا به‌صورت یک چارچوب یکپارچه کمتر مورد بررسی قرار گرفته است. افزون بر این، مقایسه جامع و نظام‌مند میان الگوریتم‌های کلاسیک طبقه‌بندی و روش‌های پیشرفته مبتنی بر بوستینگ و مدل‌های ترکیبی بر روی داده‌های واقعی ناوگان در ادبیات محدود گزارش شده است. این شکاف‌ها نشان می‌دهد که هنوز نیاز به توسعه چارچوبی داده‌محور و مقایسه‌ای برای پیش‌بینی نیاز به تعمیر در سطح ناوگان خودرو وجود دارد.

با توجه به خلأهای شناسایی‌شده در ادبیات، مسئله اصلی این پژوهش به‌صورت مشخص و آزمون‌پذیر چنین تعریف می‌شود: آیا می‌توان با بهره‌گیری از یک چارچوب داده‌محور مبتنی بر ترکیب همزمان ویژگی‌های فنی، عملیاتی و سوابق تعمیراتی خودرو، نیاز به انجام تعمیرات را در سطح ناوگان با دقت و قابلیت اطمینان بالاتر نسبت به مدل‌های کلاسیک پیش‌بینی کرد؟ به بیان دقیق‌تر، این پژوهش در پی آن است که مشخص کند آیا استفاده از الگوریتم‌های پیشرفته مبتنی بر درخت و روش‌های ترکیبی نظیر Boosting و Bagging می‌تواند عملکرد پیش‌بینی را به‌طور معناداری نسبت به الگوریتم‌های سنتی طبقه‌بندی مانند رگرسیون لجستیک، KNN یا SVM بهبود دهد یا خیر. پاسخ به این پرسش، نه‌تنها از منظر فنی اهمیت دارد، بلکه می‌تواند مبنایی برای توسعه سیستم‌های تصمیم‌یار در مدیریت هوشمند نگهداری ناوگان فراهم سازد.

اگرچه مطالعات متعددی پیرامون نگهداری پیش‌بینانه در حوزه حمل‌ونقل صورت گرفته است، اما پژوهش حاضر با بهره‌گیری همزمان از ویژگی‌های عملیاتی و فنی خودرو در قالب یک مجموعه داده ترکیبی و پیاده‌سازی مدل‌های متنوع یادگیری ماشین و الگوریتم‌های ترکیبی[۱]، گامی نوآورانه در افزایش دقت پیش‌بینی نیاز به تعمیرات در ناوگان‌های خودرویی برداشته است. این ترکیب داده محور و مدل محور در ادبیات پیشین به‌صورت محدود یا ناقص مورد بررسی قرار گرفته است.

۳- روش تحقیق

با توجه به شکاف‌های شناسایی‌شده در ادبیات، این پژوهش بر این فرض استوار است که ترکیب همزمان ویژگی‌های فنی، عملیاتی و سابقه‌ای خودرو می‌تواند الگوهای پیچیده‌تری از نیاز به تعمیر را آشکار سازد که توسط مدل‌های کلاسیک به‌طور کامل قابل استخراج نیستند. بر این اساس، یک چارچوب مفهومی سه‌مرحله‌ای طراحی شد: (۱) یکپارچه‌سازی ویژگی‌های چندبعدی خودرو، (۲) ارزیابی نظام‌مند الگوریتم‌های نماینده‌ی رویکردهای خطی، غیرخطی و تجمیعی و (۳) مقایسه عملکرد آن‌ها بر اساس شاخص‌های چندمعیاره. این چارچوب مفهومی در بخش روش تحقیق از طریق پیاده‌سازی ۱۲ الگوریتم یادگیری ماشین و ارزیابی کمی آن‌ها عملیاتی شده است.

داده‌های مورد استفاده در این پژوهش شامل مجموعه‌ای از اطلاعات مربوط به وضعیت فنی و عملکردی خودروها است که با هدف پیش‌بینی نیاز به تعمیر و نگهداری جمع‌آوری شده‌اند. مجموعه داده مورد استفاده ۳۳۴۱۸ رکورد است که شامل متغیرهایی مانند مسافت طی‌شده[۲]، سابقه نگهداری[۳]، سن خودرو[۴]، اندازه موتور[۵]، وضعیت ترمز و تایر، نوع سوخت و مدل خودرو است. در مجموع، داده‌ها دارای ۲۰ ویژگی[۶] مستقل و یک متغیر وابسته‌ی دودویی نیاز به نگهداری و تعمیرات[۷] هستند که بیانگر نیاز یا عدم نیاز خودرو به سرویس است. قبل از آموزش مدل‌ها، داده‌ها از نظر وجود مقادیر گمشده، نویز و مقادیر پرت سبب شد که دقت و کارایی الگوریتم‌ها در فرآیند پیش‌بینی بهبود یابد و مقایسه‌ی مدل‌ها در شرایطی منصفانه انجام شود. بررسی و پاک‌سازی شدند. برای اطمینان از پایداری مدل‌ها، داده‌ها به دو بخش آموزش[۸] (۸۰ درصد داده‌ها) و آزمون[۹] (۲۰ درصد داده‌ها) تقسیم شدند. همچنین با نرمال‌سازی ویژگی‌های عددی و کدگذاری ویژگی‌های کیفی با استفاده از One-Hot Encoding و متوازن‌سازی داده‌ها با روش SMOTE، داده‌ها برای ورود به الگوریتم‌های یادگیری ماشین آماده‌سازی شده است.

پس از طراحی و پیاده‌سازی مدل‌های یادگیری ماشین، لازم است میزان دقت و اعتبار عملکرد آن‌ها مورد ارزیابی قرار گیرد. هدف اصلی این پژوهش، پیش‌بینی نیاز به انجام تعمیرات با بهره‌گیری از الگوریتم‌های مختلف یادگیری ماشین است؛ ازاین‌رو، یکی از پرسش‌های اساسی تحقیق این است که کدام یک از الگوریتم‌ها در شناسایی خودروهای نیازمند تعمیر عملکرد بهتری دارد. برای پاسخ به این سؤال، از مجموعه‌ای از شاخص‌های ارزیابی مدل که مبتنی بر ماتریس درهم‌ریختگی[۱۰] هستند، استفاده شده است. شاخص‌های مورد بررسی شامل دقت[۱۱]، صحت[۱۲]، بازخوانی[۱۳]، F1-Score و ROC-AUC است. در میان این شاخص‌ها، دقت و صحت از اهمیت ویژه‌ای برخوردارند، زیرا میزان توانایی مدل در شناسایی درست نمونه‌ها را نشان می‌دهند [۲۵]. شاخص‌های صحت و بازخوانی میزان توانایی الگوریتم در تشخیص صحیح نمونه‌های مثبت را بیان می‌کنند [۲۶, ۲۷]. علاوه بر این، معیار F1-Score به عنوان میانگین هارمونیک دقت و بازخوانی، توازن میان این دو شاخص را منعکس می‌کند [۲۸]. در این

[1] Ensemble
[2] Mileage
[3] Maintenance History
[4] Vehicle Age
[5] Engine Size
[6] Feature
[7] Need_Maintenance
[8] Train
[9] Test
[10] Confusion Matrix
[11] Accuracy
[12] Precision
[13] Recall

میان، حساسیت[1] که با عنوان نرخ مثبت درست[2] نیز شناخته می‌شود، درصد نمونه‌های مثبت به‌درستی شناسایی‌شده را نمایش می‌دهد، درحالی‌که ویژگی[3] نشان‌دهنده‌ی نسبت نمونه‌های منفی است که مدل به‌درستی آن‌ها را پیش‌بینی کرده است. شرح دقیق فرمول‌ها و نحوه محاسبه‌ی هر یک از این معیارها در جدول ۲ آورده شده است.

جدول ۲ ماتریس درهم‌ریختگی

Confusion Matrix		Predicted	
		positive	Negative
Actual	positive	TP	FP
	Negative	FN	TN

در فرآیند ارزیابی عملکرد مدل‌های یادگیری ماشین، چهار شاخص اصلی از ماتریس درهم‌ریختگی استخراج می‌شوند که شامل موارد زیر است:

- مثبت درست[4] (TP): تعداد نمونه‌هایی از کلاس مثبت که مدل به درستی آن‌ها را به‌عنوان مثبت تشخیص داده است.
- منفی درست[5] (TN): تعداد نمونه‌های منفی که مدل به درستی به‌عنوان منفی شناسایی کرده است.
- مثبت نادرست[6] (FP): تعداد نمونه‌های منفی که مدل به اشتباه به‌عنوان مثبت پیش‌بینی کرده است.
- منفی نادرست[7] (FN): تعداد نمونه‌های مثبت که مدل به اشتباه در گروه منفی قرار داده است.

این مقادیر مبنای محاسبه شاخص‌های مهمی چون دقت، صحت، بازخوانی و F1-Score هستند و امکان ارزیابی جامع عملکرد مدل در شناسایی صحیح کلاس‌ها را فراهم می‌سازند. نحوه محاسبه این معیارها در جدول ۳ قابل مشاده است.

جدول ۳ معیارهای ماتریس در هم‌ریختگی

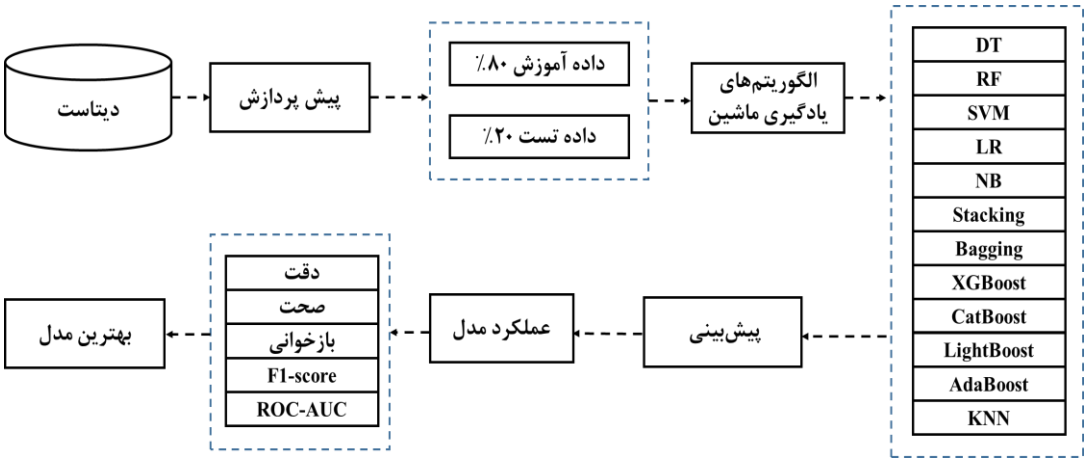
معیار	فرمول
Accuracy	(TP + TN) / (TP + FP + FN + TN)
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)
F1-score	2*Precision*Recall / (Precision + Recall)
Sensitivity	TP / (TP + FN)
Specificity	TN / (TN + FP)

فلوچارت شکل ۱، به‌صورت گام‌به‌گام فرآیند توسعه مدل‌های پیش‌بینی در این پژوهش را نشان می‌دهد. این فرآیند با گردآوری و پیش‌پردازش داده‌ها آغاز می‌شود؛ در این مرحله داده‌ها پاک‌سازی، نرمال‌سازی و آماده‌سازی می‌شوند تا برای مدل‌سازی مناسب باشند. در گام بعدی، داده‌ها به‌منظور ارزیابی منصفانه مدل، به دو بخش آموزش و آزمون تقسیم می‌شوند، به‌طوری‌که ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای آزمون مورد استفاده قرار می‌گیرد. در این پژوهش به‌منظور ارزیابی عملکرد روش‌های مختلف یادگیری ماشین در پیش‌بینی نیاز به تعمیر و نگهداری خودرو، از دوازده الگوریتم پرکاربرد و معتبر استفاده شده است. این الگوریتم‌ها نماینده‌ی گروه‌های اصلی مدل‌های یادگیری ماشین هستند و شامل موارد زیر هستند:

[1] Sensitivity
[2] True Positive Rate
[3] Specificity
[4] True Positive
[5] True Negative
[6] False Positive
[7] False Negative

- الگوریتم‌های پایه[۱]: رگرسیون لجستیک[۲] (LR) [۲۹]، ماشین بردار پشتیبان[۳] (SVM) [۳۰]، نزدیک‌ترین همسایه‌ها[۴] (KNN) [۳۱] و طبقه‌بند بیز ساده[۵] (NB) [۳۲].
- الگوریتم‌های درختی[۶]: درخت تصمیم[۷] (DT) و جنگل تصادفی[۸] (RF) [۳۳].
- الگوریتم‌های ترکیبی[۹]: Bagging [۳۴]، AdaBoost [۳۵]، XGBoost [۳۶]، LightGBM [۳۷]، CatBoost [۳۸] و Stacking [۳۹].

این الگوریتم‌ها از آن جهت انتخاب شده‌اند که پوشش کاملی از روش‌های خطی، غیرخطی و ترکیبی را ارائه می‌دهند و به‌صورت گسترده در مسائل طبقه‌بندی مشابه مورد استفاده قرار گرفته‌اند. هدف از به‌کارگیری هم‌زمان این دوازده مدل، مقایسه‌ی میزان دقت، بازخوانی، F1-Score و شاخص ROC-AUC و تعیین بهترین الگوریتم در تشخیص وضعیت نگهداری خودرو بوده است.



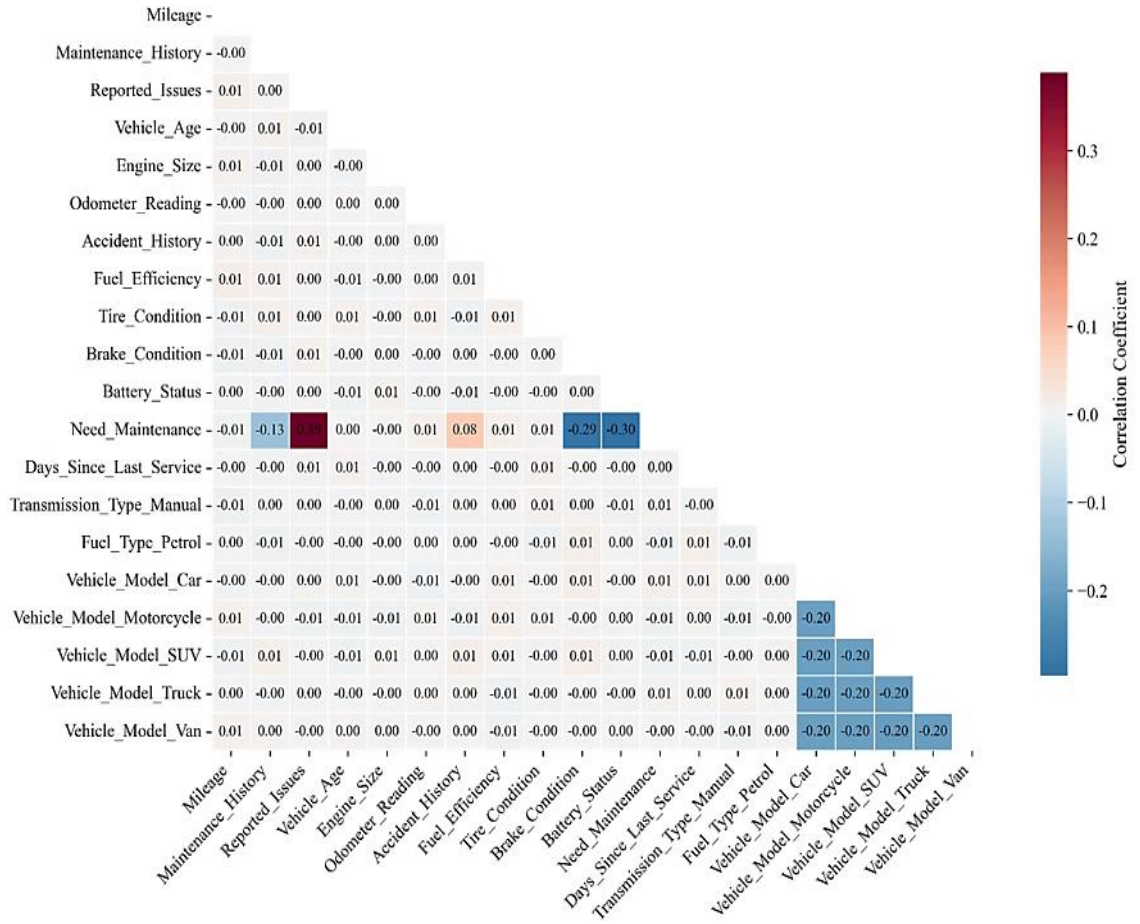
شکل ۱ فلوچارت تحقیق

۴- بحث و نتایج

در فرایندهای یادگیری ماشین، شناسایی میزان ارتباط میان ویژگی‌های ورودی از اهمیت بالایی برخوردار است. به‌منظور بررسی این ارتباط، از ماتریس همبستگی[۱۰] استفاده شده است که در آن هر سطر و ستون نشان‌دهنده یکی از ویژگی‌های مجموعه داده است. این ماتریس دارای ابعاد ۲۰×۲۰ و کاملاً متقارن است. برای محاسبه میزان همبستگی بین ویژگی‌ها دو رویکرد آماری متفاوت به‌کار گرفته شد: روش پیرسون[۱۱] به‌عنوان یک روش پارامتری و روش اسپیرمن[۱۲] به‌عنوان یک روش ناپارامتری. در شکل۲ ماتریس همبستگی اسپیرمن و در شکل ۳ ماتریس همبستگی پیرسون ارائه شده‌اند. نتایج ماتریس همبستگی نشان داد که بیشتر ویژگی‌های مجموعه داده از استقلال نسبی برخوردارند و همبستگی بالایی بین آن‌ها مشاهده نمی‌شود. این موضوع حاکی از آن است که داده‌ها فاقد تکرار اطلاعات بوده و هر ویژگی نقش منحصربه‌فردی در مدل‌سازی دارد. در میان ویژگی‌ها، بیشترین همبستگی مثبت بین سوابق نگهداری و نیاز به سرویس با مقدار تقریبی ۰/۳۹ مشاهده شد که بیانگر آن است خودروهایی

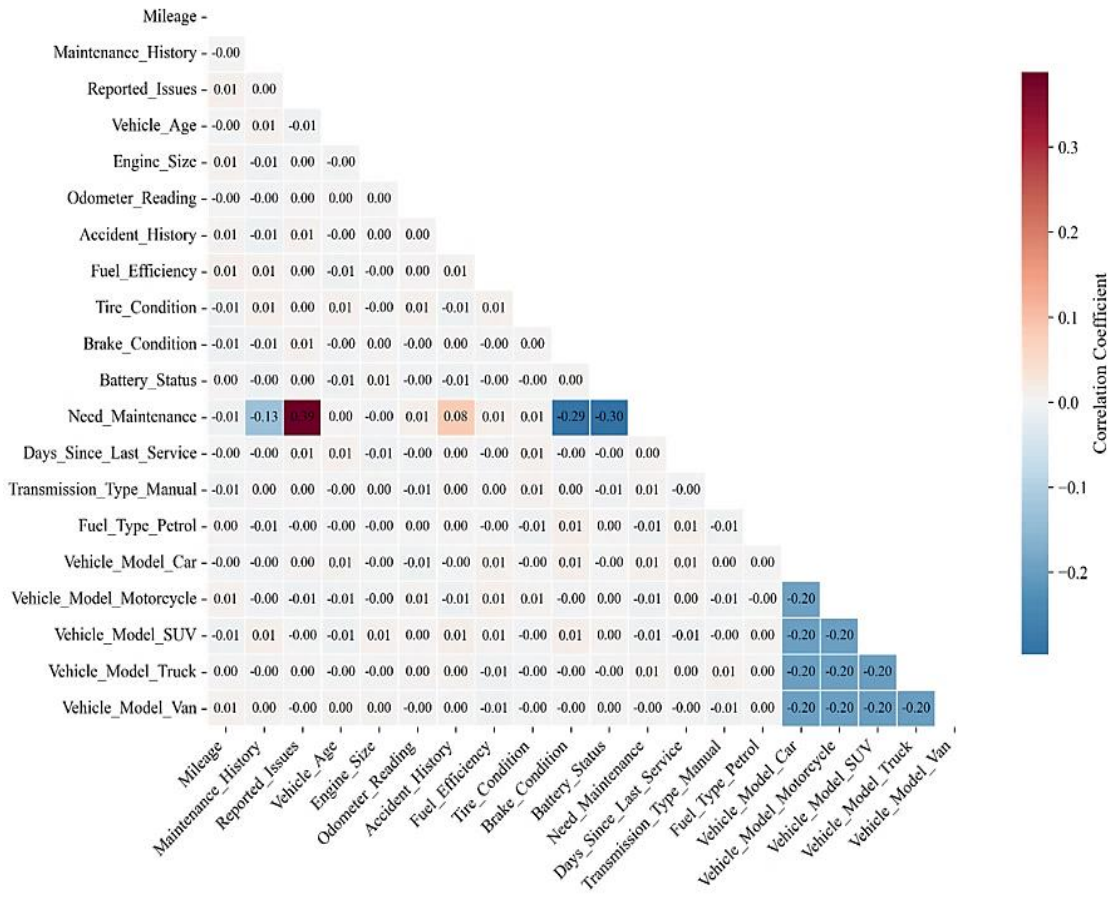
[1] Base Learners
[2] Logistic Regression
[3] Support Vector Machine
[4] K-Nearest Neighbors
[5] Naïve Bayes
[6] Tree-based Models
[7] Decision Tree
[8] Random Forest
[9] Ensemble Methods
[10] Correlation Matrix
[11] Pearson Correlation Coefficient
[12] Spearman Correlation Coefficient

که سابقه سرویس بیشتری دارند، احتمال نیاز به تعمیر مجدد در آن‌ها بالاتر است. همچنین، همبستگی منفی بین وضعیت ترمز[1] و وضعیت باتری[2] با متغیر هدف نشان می‌دهد که هرچه این اجزا در وضعیت بهتری قرار داشته باشند، احتمال نیاز به سرویس کاهش می‌یابد. به‌طورکلی، نتایج این تحلیل بیانگر آن است که ویژگی‌های فنی و سابقه نگهداری خودرو تأثیر مستقیم‌تری بر نیاز به سرویس دارند، در حالی‌که متغیرهایی مانند نوع سوخت یا مدل خودرو نقش ثانویه‌تری ایفا می‌کنند. پایین بودن ضرایب همبستگی در اغلب ویژگی‌ها نیز تأیید می‌کند که داده‌ها چندبعدی و بدون هم‌خطی شدید هستند؛ بنابراین، این ساختار می‌تواند عملکرد مدل‌های یادگیری ماشین را در مرحله پیش‌بینی بهبود بخشد.



شکل ۲ ماتریس همبستگی پیرسون

[1] Brake_Condition
[2] Battery_Status



شکل ۳ ماتریس همبستگی اسپیرمن

نتایج ارائه‌شده در این بخش، به‌طور مستقیم در راستای آزمون چارچوب مفهومی و پرسش اصلی پژوهش مبنی بر برتری مدل‌های ترکیبی نسبت به مدل‌های کلاسیک تحلیل می‌شوند. به‌منظور ارزیابی و مقایسه‌ی کارایی الگوریتم‌های مختلف یادگیری ماشین در تشخیص وضعیت سیستم مورد بررسی، مجموعه‌ای از معیارهای ارزیابی شامل دقت، بازخوانی، صحت، امتیاز و مساحت زیر منحنی ROC محاسبه شد. هر یک از این معیارها جنبه متفاوتی از عملکرد مدل را نشان می‌دهند. به‌طور خاص، دقت نشان می‌دهد چه نسبتی از پیش‌بینی‌های مثبت مدل صحیح بوده است، درحالی‌که بازخوانی بیانگر توانایی مدل در شناسایی نمونه‌های مثبت واقعی است. F1-Score به‌عنوان میانگین هارمونیک دقت و بازخوانی، تعادلی میان این دو ویژگی برقرار می‌کند. از سوی دیگر، شاخص ROC-AUC نمایانگر توان کلی مدل در تمایز میان دو کلاس است؛ هرچه مقدار آن به ۱ نزدیک‌تر باشد، مدل قدرت تفکیک بهتری دارد.

بررسی جدول ۴ نشان می‌دهد که الگوریتم‌های LightGBM، Bagging و Stacking عملکرد بسیار بالایی از خود نشان داده‌اند. به‌ویژه الگوریتم LightGBM با ROC-AUC برابر ۰/۹۴۷۵، F1-score معادل ۰/۹۶۱۳ و صحت کلی ۰/۹۷۰۶ بهترین نتایج را در میان همه‌ی مدل‌ها ارائه داده است. این موضوع بیانگر آن است که LightGBM به‌خوبی توانسته تعادل بین کاهش خطاهای مثبت و منفی را برقرار کند و در عین حال با سرعت و پایداری بالا به طبقه‌بندی دقیق نمونه‌ها بپردازد. از آنجا که LightGBM مبتنی‌بر درخت‌های گرادیان بوستینگ با روش تقسیم داده کارآمد است، معمولاً در مسائل با ویژگی‌های متعدد عملکردی برتر از سایر روش‌ها دارد.

الگوریتم Bagging نیز با F1-score (۰/۹۵۶۲) و دقت ۰/۹۷۸۵ عملکردی بسیار نزدیک به LightGBM داشته است. این شباهت عملکرد بیانگر آن است که رویکردهای مبتنی بر تجمیع مدل‌ها در داده‌های این پژوهش اثربخشی بالایی دارند. در واقع،

ترکیب چندین مدل تصمیم‌گیری در قالب Bagging باعث افزایش پایداری و کاهش واریانس مدل شده است. همچنین مقدار بالای ROC-AUC (۰/۹۴۳۱) در این روش تأییدی بر قدرت بالای آن در تشخیص درست نمونه‌هاست.

در بین مدل‌های ترکیبی، الگوریتم Stacking نیز با مقدار بالای ROC-AUC (۰/۹۴۵۱) و F1-score (۰/۹۶) توانسته است دقت و بازخوانی بالایی به‌دست آورد. علت این عملکرد قوی آن است که Stacking با بهره‌گیری از چندین مدل پایه مانند SVM، RF و LR و ترکیب خروجی آن‌ها توسط یک مدل متا، تعادلی میان خطاهای طبقه‌بندی ایجاد کرده و به یک تصمیم نهایی بهینه رسیده است. این رفتار نشان‌دهنده‌ی توانایی بالای مدل در استخراج روابط پیچیده بین متغیرها است.

از میان مدل‌های کلاسیک، پایه عملکرد چشمگیری داشته است؛ به‌طوری که با ROC-AUC (۰/۹۴۶۴) و F1-score (۰/۹۳۳۹) از بسیاری از مدل‌های پیشرفته‌تر مانند AdaBoost و XGBoost نیز عملکرد بهتری داشته است. این نتیجه می‌تواند به توزیع نسبتاً مستقل ویژگی‌های داده اشاره داشته باشد، چراکه فرض استقلال ویژگی‌ها در NB، در صورت صدق نسبی در داده، منجر به عملکردی بسیار کارآمد و سریع می‌شود.

در مقابل، مدل‌های SVM و KNN عملکرد نسبتاً ضعیف‌تری از خود نشان داده‌اند. برای مثال، الگوریتم SVM با ROC-AUC (۰/۸۶۹۲) و F1-score (۰/۸۸۹۰) در مقایسه با سایر روش‌ها دقت کمتری دارد. همچنین الگوریتم KNN با F1-score (۰/۸۹۵۷) و بازخوانی ۰/۸۳۵۱ احتمالاً به‌دلیل وابستگی زیاد به فاصله‌ی نمونه‌ها و توزیع غیرخطی داده، نتوانسته است مرز تصمیم مناسبی ایجاد کند.

الگوریتم‌های درختی مانند RF و DT عملکردی نسبتاً مناسب ولی نه در حد مدل‌های ترکیبی داشته‌اند. به‌عنوان مثال، RF با ROC-AUC (۰/۹۵۰) و F1-score (۰/۹۳۷۹) از پایداری خوبی برخوردار است و توانسته است میان دقت و بازخوانی تعادل برقرار کند. در مقابل، DT با ROC-AUC (۰/۹۲۰۰) و F1-score (۰/۹۲۶۹) کمی ضعیف‌تر عمل کرده است که این امر به‌دلیل حساسیت ذاتی درخت تصمیم به بیش‌برازش[1] است.

در میان مدل‌های بوستینگ، CatBoost با ROC-AUC (۰/۹۳۲۰) و F1-score (۰/۹۳۹۲) عملکردی نسبتاً پایدار داشته و توانسته است میان دقت (۰/۹۰۴۸) و بازخوانی (۰/۹۲۰۵) توازن برقرار کند. این عملکرد مناسب، ناشی از توان CatBoost در مدیریت داده‌های نامتوازن و دسته‌بندی ویژگی‌های متنوع است. در مقابل، XGBoost با وجود شهرت بالا، در این مجموعه داده نتوانسته عملکرد مطلوبی ارائه دهد؛ زیرا با ROC-AUC (۰/۹۱۷۸) و دقت ۰/۸۱۰۹ در رتبه‌های پایین‌تر قرار گرفته است.

در مجموع، نتایج نشان می‌دهد که مدل‌های مبتنی بر تجمیع از جمله Bagging، Stacking وLightGBM در مقایسه با مدل‌های کلاسیک و پایه، از پایداری و دقت بالاتری برخوردارند. این امر بیانگر آن است که ترکیب خروجی چندین مدل پایه می‌تواند با کاهش واریانس و افزایش تعمیم‌پذیری، کارایی نهایی سیستم را بهبود بخشد. می‌توان نتیجه گرفت که در مسئله حاضر، داده‌ها ساختاری غیرخطی و نسبتاً پیچیده دارند و بنابراین، مدل‌هایی که از رویکردهای درختی یا ترکیبی استفاده می‌کنند، عملکرد بهتری نسبت به مدل‌های ساده‌تری مانند SVM یا KNN از خود نشان داده‌اند. الگوریتم LightGBM در نهایت به‌عنوان بهترین مدل انتخاب شد، چراکه نه‌تنها بالاترین مقدار ROC-AUC و F1-score را داراست، بلکه تعادل مطلوبی میان دقت، بازخوانی و صحت کلی برقرار کرده است. این نتیجه مؤید آن است که روش‌های بوستینگ مبتنی بر گرادیان می‌توانند در تحلیل داده‌های مشابه این پژوهش، گزینه‌ای بسیار مناسب و قابل اعتماد باشند.

نتایج این پژوهش نشان می‌دهد که مدل‌های مبتنی بر تجمیع و گرادیان بوستینگ عملکردی برتر نسبت به مدل‌های کلاسیک دارند. این یافته با نتایج مطالعات پیشین در حوزه نگهداری پیش‌بینانه همسو است، جایی که روش‌های Ensemble به دلیل توانایی در کاهش واریانس و مدل‌سازی روابط غیرخطی پیچیده، عملکرد دقیق‌تری نسبت به مدل‌های منفرد ارائه کرده‌اند. با این حال، در برخی مطالعات پیشین گزارش شده بود که XGBoost عملکرد برتری نسبت به سایر روش‌ها دارد، در حالی که در مجموعه داده حاضر، LightGBMعملکرد بهتری نشان داد. این تفاوت می‌تواند ناشی از ساختار ویژگی‌ها، توزیع کلاس‌ها و

[1] Overfitting

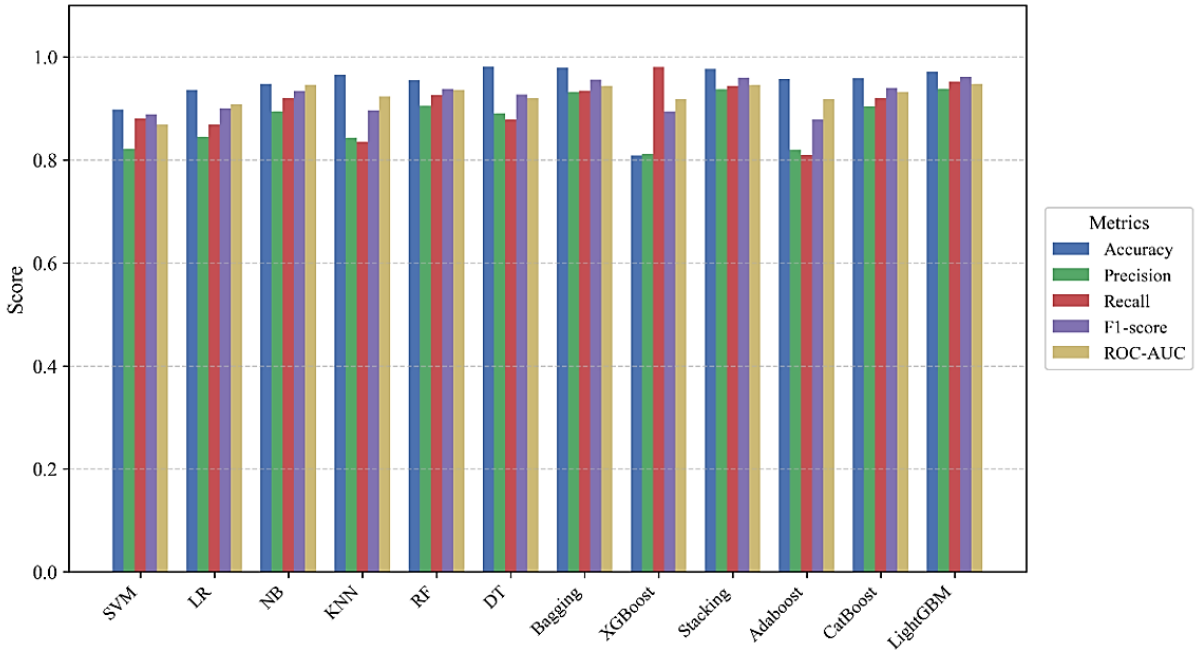
ابعاد داده باشد.

جدول ۴ نتایج ارزیابی الگوریتم‌ها

الگوریتم	دقت	صحت	بازخوانی	F1-score	ROC-AUC
SVM	۰/۸۲۲۱	۰/۸۹۷۸	۰/۸۸۰۴	۰/۸۸۹۰	۰/۸۶۹۲
LR	۰/۸۴۵۲	۰/۹۳۵۰	۰/۸۶۹۱	۰/۹۰۰۹	۰/۹۰۷۸
NB	۰/۸۹۴۵	۰/۹۴۷۵	۰/۹۲۰۷	۰/۹۳۳۹	۰/۹۴۶۴
KNN	۰/۸۴۲۶	۰/۹۶۵۸	۰/۸۳۵۱	۰/۸۹۵۷	۰/۹۲۲۷
RF	۰/۹۰۵۳	۰/۹۵۴۵	۰/۹۲۵۴	۰/۹۳۷۹	۰/۹۳۵۰
DT	۰/۸۸۹۴	۰/۹۸۰۹	۰/۸۷۸۵	۰/۹۲۶۹	۰/۹۲۰۰
Bagging	۰/۹۳۱۶	۰/۹۷۸۶	۰/۹۳۴۷	۰/۹۵۶۲	۰/۹۴۳۱
XGBoost	۰/۸۱۰۹	۰/۸۰۸۹	۰/۹۹۸۹	۰/۸۹۳۹	۰/۹۱۷۸
Stacking	۰/۹۳۷۳	۰/۹۷۶۷	۰/۹۴۳۹	۰/۹۶۰۰	۰/۹۴۵۱
Adaboost	۰/۸۲۰۳	۰/۹۵۸۱	۰/۸۱۰۲	۰/۸۷۸۰	۰/۹۱۷۹
CatBoost	۰/۹۰۴۸	۰/۹۵۸۶	۰/۹۲۰۵	۰/۹۳۹۲	۰/۹۳۲۰
Lightgbm	۰/۹۳۸۸	۰/۹۷۰۶	۰/۹۵۲۲	۰/۹۶۱۳	۰/۹۴۷۵

همان‌گونه که در شکل ۴ مشاهده می‌شود، مقادیر شاخص‌های ارزیابی برای الگوریتم‌های مختلف محاسبه و مقایسه شده‌اند. نتایج نشان می‌دهد که مدل‌های ترکیبی مبتنی بر روش‌های تجمیع از جمله LightGBM، Bagging و Stacking عملکردی برتر نسبت به سایر مدل‌ها دارند. الگوریتم LightGBM با داشتن مقادیر بالاتر در تمام معیارها، به‌ویژه F1-score و ROC-AUC، بهترین تعادل را میان دقت و بازخوانی ایجاد کرده است. در مقابل، مدل‌های پایه مانند SVM و KNN عملکرد ضعیف‌تری داشته‌اند. این یافته‌ها تأکید می‌کند که بهره‌گیری از مدل‌های ترکیبی به‌ویژه در داده‌های پیچیده، می‌تواند منجر به بهبود معنی‌دار دقت و پایداری سیستم‌های پیش‌بینی شود. مقادیر بالای ROC-AUC در اکثر مدل‌های Ensemble (بیش از ۰/۹۴) نشان‌دهنده ثبات عملکرد آن‌ها در برابر تغییرات داده آموزشی است. نزدیکی مقادیر F1-score و Accuracy در مدل‌های برتر نیز بیانگر عدم سوگیری شدید نسبت به یک کلاس خاص است. این موضوع اهمیت دارد زیرا در مسائل تعمیرات، خطای نوع دوم (عدم شناسایی خودروی نیازمند تعمیر) می‌تواند پیامدهای عملیاتی جدی داشته باشد.

برتری مدل‌های درختی و ترکیبی نشان می‌دهد که روابط بین متغیرهای مستقل و متغیر هدف احتمالاً غیرخطی و دارای تعاملات پیچیده هستند. عملکرد نسبتاً ضعیف SVM و KNN نیز مؤید آن است که داده‌ها فاقد مرز تصمیم خطی ساده بوده و ساختار آن‌ها نیازمند مدل‌هایی با ظرفیت بالاتر برای استخراج تعاملات چندمتغیره است. همچنین عملکرد مناسب Naïve Bayes می‌تواند نشان دهد که برخی ویژگی‌ها دارای توزیع‌های نسبتاً مستقل یا شبه‌مستقل بوده‌اند.



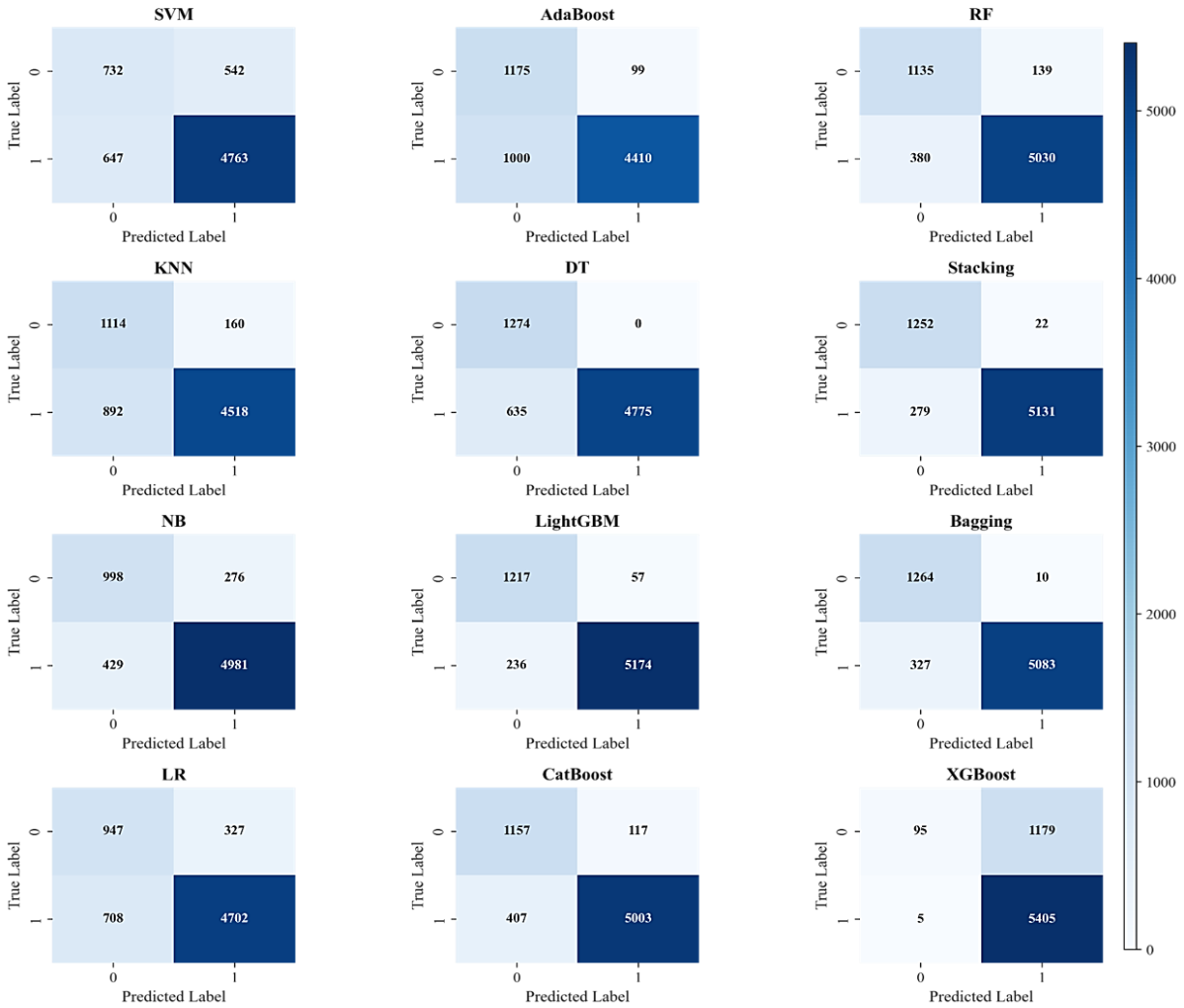
شکل ۴ مقایسه نتایج ارزیابی الگوریتم‌ها

به‌منظور بررسی دقیق عملکرد مدل‌های مختلف، از ماتریس درهم‌ریختگی برای ارزیابی نحوه‌ی پیش‌بینی کلاس‌ها استفاده شد. همان‌گونه که در شکل ۵ نشان داده شده است، این ماتریس‌ها توزیع نمونه‌های درست و نادرست را در دو کلاس (۰ و ۱) برای هر الگوریتم به‌صورت مجزا نمایش می‌دهند. در تمامی مدل‌ها مشاهده می‌شود که مقدار پیش‌بینی صحیح برای کلاس غالب (برچسب ۱) به‌طور چشمگیری بیشتر از مقادیر نادرست است، که نشان‌دهنده‌ی توان بالای مدل‌ها در شناسایی الگوهای مثبت داده‌ها است. نتایج نشان می‌دهد که الگوریتم‌های ترکیبی و مبتنی بر گرادیان بوستینگ (LightGBM، Stacking و Bagging) بهترین عملکرد را از نظر F1-score، ROC-AUC، صحت و دقت ارائه می‌دهند. LightGBM با F1-score (0.9613) و ROC-AUC (0.9475) تعادل مطلوب میان شناسایی نمونه‌های مثبت و منفی برقرار کرده و توانایی تفکیک کلاس‌ها را به‌طور چشمگیری نشان می‌دهد. Stacking و Bagging نیز عملکرد نزدیک به LightGBM دارند و نشان‌دهنده قدرت ترکیب مدل‌های پایه در کاهش واریانس و افزایش پایداری پیش‌بینی هستند. در مقابل، مدل‌های کلاسیک مانند SVM و KNN به دلیل محدودیت در مدل‌سازی مرزهای غیرخطی، عملکرد ضعیف‌تری دارند. مدل NB، علی‌رغم سادگی، توانسته با F1-score (0.9339) و ROC-AUC (0.9464) نتایج قابل قبولی ارائه دهد، که ناشی از استقلال نسبی ویژگی‌ها در داده است. در بین مدل‌های درختی و بوستینگ، CatBoost و RF عملکرد نسبتاً پایدار و مناسبی دارند، درحالی‌که XGBoost در این مجموعه داده عملکردی پایین‌تر نشان داده است. به‌طورکلی، داده‌های مورد مطالعه دارای ساختار پیچیده و غیرخطی هستند و مدل‌های ترکیبی و مبتنی بر گرادیان بوستینگ در چنین شرایطی بهترین کارایی را از خود نشان داده است.

در مسئله حاضر، خطای منفی کاذب (FN) به معنای پیش‌بینی «عدم نیاز به تعمیر» برای خودرویی است که در واقع نیازمند سرویس است. این نوع خطا پرریسک‌ترین حالت در مدیریت ناوگان محسوب می‌شود، زیرا می‌تواند منجر به خرابی ناگهانی در حین عملیات، افزایش تعمیرات اضطراری، توقف‌های پیش‌بینی‌نشده، کاهش ایمنی و حتی بروز حوادث شود. نتایج نشان داد که مدل LightGBM با بازخوانی ۰/۹۵۲۲، بیش از ۹۵ درصد خودروهای واقعاً نیازمند تعمیر را به‌درستی شناسایی می‌کند. این میزان دقت در شناسایی موارد بحرانی، به معنای کاهش قابل توجه احتمال خرابی‌های ناگهانی در مقیاس عملیاتی است. در ناوگان‌های بزرگ، حتی کاهش چند درصدی در نرخ FN می‌تواند از ده‌ها توقف اضطراری در سال جلوگیری کرده و پایداری عملیاتی سیستم را به شکل محسوسی افزایش دهد.

از سوی دیگر، خطای مثبت کاذب (FP) بیانگر پیش‌بینی نیاز به تعمیر برای خودرویی است که در واقع سالم است. این نوع خطا اگرچه از منظر ایمنی کم‌خطرتر است، اما پیامدهای اقتصادی مستقیم دارد؛ از جمله انجام سرویس‌های غیرضروری، افزایش

مصرف قطعات یدکی، اتلاف نیروی انسانی و کاهش بهره‌وری ناوگان. مدل پیشنهادی با صحت ۰/۹۷۰۶ نشان می‌دهد که اکثریت قریب به اتفاق پیش‌بینی‌های مثبت آن صحیح هستند. این موضوع به معنای کاهش اعزام غیرضروری خودروهای سالم به تعمیرگاه و در نتیجه بهینه‌سازی تخصیص منابع نگهداری است. تعادل هم‌زمان میان کاهش FN و FP که در مقدار بالای F1-score منعکس شده، بیانگر آن است که مدل نه‌تنها ریسک عملیاتی را کاهش می‌دهد، بلکه هزینه‌های نگهداری را نیز کنترل می‌کند. علاوه بر این، مقدار بالای ROC-AUC (۰/۹۴۷۵) نشان‌دهنده توان تفکیک بالای مدل در تمایز میان خودروهای سالم و معیوب است. این ویژگی از منظر مدیریتی اهمیت راهبردی دارد، زیرا امکان تنظیم آستانه تصمیم‌گیری بر اساس سیاست‌های سازمان را فراهم می‌کند. برای مثال، در شرایطی که ایمنی اولویت بالاتری دارد، می‌توان آستانه تصمیم را به‌گونه‌ای تنظیم کرد که بازخوانی افزایش یابد و احتمال از دست رفتن موارد بحرانی به حداقل برسد. در مقابل، در شرایط محدودیت منابع، می‌توان تمرکز را بر افزایش صحت قرار داد تا از انجام تعمیرات غیرضروری جلوگیری شود. بنابراین، مدل پیشنهادی تنها یک ابزار پیش‌بینی نیست، بلکه یک ابزار پشتیبان تصمیم منعطف برای مدیریت ریسک و هزینه است. در مقیاس کلان سازمانی، بهبود عملکرد پیش‌بینی در حد چند درصد می‌تواند آثار اقتصادی قابل توجهی به همراه داشته باشد. کاهش خرابی‌های ناگهانی منجر به افزایش دسترس‌پذیری ناوگان، بهبود سطح خدمات، افزایش رضایت کاربران و کاهش هزینه‌های تعمیرات اضطراری می‌شود. هم‌زمان، کاهش سرویس‌های غیرضروری موجب افزایش بهره‌وری عملیاتی و استفاده بهینه از منابع انسانی و مالی خواهد شد. بر این اساس، استقرار مدل LightGBM در چارچوب یک سامانه نگهداری پیش‌بینانه می‌تواند به‌عنوان گامی مؤثر در جهت پیاده‌سازی مدیریت هوشمند ناوگان و گذار از نگهداری واکنشی به نگهداری داده‌محور تلقی شود.



شکل ۵ ماتریس درهم‌ریختگی

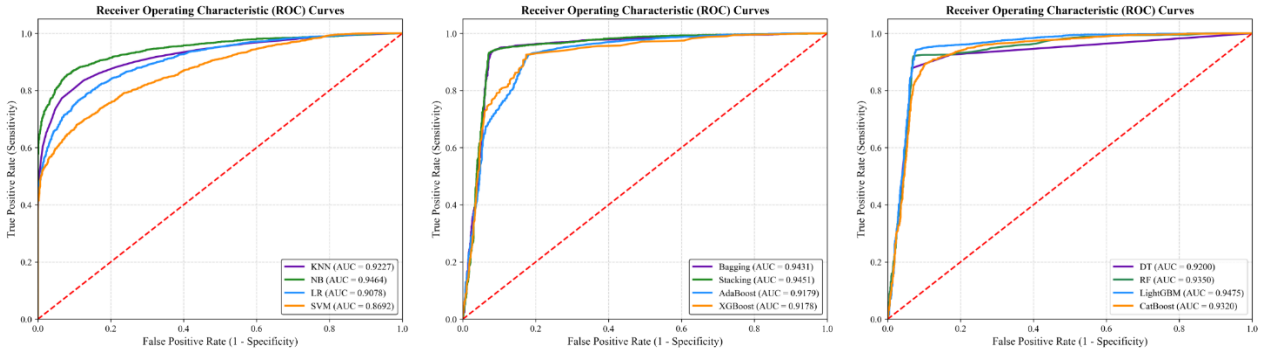
برای ارزیابی دقیق‌تر توانایی مدل‌ها در تشخیص صحیح کلاس‌ها، از منحنی ROC Curve و شاخص مساحت زیر منحنی ROC استفاده شد. در این نمودارها، محور افقی نرخ مثبت کاذب[1] و محور عمودی نرخ مثبت واقعی[2] را نشان می‌دهد. هرچه منحنی به سمت گوشه بالا-چپ نزدیک‌تر باشد، عملکرد مدل از نظر تفکیک‌پذیری میان کلاس‌ها بهتر است که در شکل ۶ نمایش داده شده است.

در نمودار اول که شامل مدل‌های SVM، KNN، NB و LR است، مشاهده می‌شود که NB و LR بالاترین مقادیر AUC را دارند (به‌ترتیب حدود ۰/۹۴ و ۰/۹۰) که نشان‌دهنده دقت بالا در تفکیک کلاس‌هاست. در مقابل، مدل‌های KNN و SVM عملکرد ضعیف‌تری نسبت به سایرین داشته‌اند و منحنی آن‌ها فاصله بیشتری از ناحیه ایده‌آل دارد. این موضوع نشان می‌دهد که روش‌های خطی و احتمالاتی (مانند LR و NB) در این داده‌ها کارآمدتر از روش‌های مبتنی بر همسایگی بوده‌اند.

در نمودار دوم که شامل مدل‌های DT، RF، LightGBM و CatBoost است، منحنی‌ها در نزدیکی یکدیگر و بسیار نزدیک به محور بالا-چپ قرار گرفته‌اند. این موضوع نشان می‌دهد که تمام مدل‌های مبتنی بر درخت در تشخیص صحیح کلاس‌ها عملکرد بسیار مطلوبی دارند. در میان آن‌ها، مدل‌های LightGBM و CatBoost با مقادیر AUC حدود ۰/۹۵ بهترین عملکرد را نشان داده‌اند و مدل DT نیز علی‌رغم سادگی ساختار، عملکرد قابل قبولی داشته است.

در نمودار سوم که مدل‌های Bagging، Stacking، XGBoost و AdaBoost را شامل می‌شود، منحنی‌ها تقریباً روی هم منطبق شده‌اند و همگی مقادیر AUC بالاتر از ۰/۹۴ را نشان می‌دهند. در میان آن‌ها، مدل Stacking بالاترین مقدار AUC را دارد و نشان می‌دهد که ترکیب چندین الگوریتم در یک چارچوب ترکیبی می‌تواند دقت نهایی را به شکل معناداری افزایش دهد. در مقابل، مدل AdaBoost نسبت به سایر روش‌های ترکیبی کمی عملکرد ضعیف‌تری داشته است، هرچند همچنان در محدوده‌ی عملکرد بالا قرار دارد.

با بررسی سه نمودار ROC مشخص می‌شود که مدل‌های ترکیبی مبتنی بر تجمیع مدل‌ها از جمله Stacking، Bagging، XGBoost، LightGBM و CatBoost بهترین عملکرد را در تفکیک داده‌ها داشته‌اند و بیشترین مقدار AUC بیش از ۹۴/ را کسب کرده‌اند. این نتایج نشان می‌دهد که بهره‌گیری از روش‌های تجمیع مدل‌ها نسبت به مدل‌های مستقل و کلاسیک توانایی بیشتری در یادگیری الگوهای پیچیده داده‌ها دارد و موجب افزایش پایداری و دقت نهایی سیستم تصمیم‌گیری می‌شود.



شکل ۶ نمودار ROC-AUC

از منظر نظری، نتایج این پژوهش تأیید می‌کند که در مسائل پیش‌بینی نگهداری خودرو که شامل ویژگی‌های ناهمگون فنی و عملیاتی هستند، مدل‌های مبتنی بر گرادیان بوستینگ می‌توانند چارچوبی مناسب برای مدل‌سازی وابستگی‌های پیچیده فراهم کنند. از منظر کاربردی، انتخاب LightGBM به‌عنوان مدل نهایی می‌تواند در توسعه سیستم‌های تصمیم‌یار مدیریت ناوگان مورد استفاده قرار گیرد و به کاهش هزینه‌های توقف اضطراری و افزایش قابلیت اطمینان عملیاتی منجر شود. توان بالای این مدل در

[1] False Positive Rate
[2] True Positive Rate

حفظ تعادل میان دقت و بازخوانی، اهمیت آن را در محیط‌های واقعی که هزینه خطا نامتقارن است، دوچندان می‌کند.

۵- نتیجه‌گیری

در این پژوهش، با هدف پیش‌بینی نیاز به تعمیرات و بهبود تصمیم‌گیری در حوزه نگهداری و تعمیرات پیش‌بینانه، از مجموعه‌ای متنوع از الگوریتم‌های یادگیری ماشین استفاده شد. داده‌های جمع‌آوری شده پس از انجام مراحل پیش‌پردازش شامل پاک‌سازی، نرمال‌سازی و متوازن‌سازی، برای آموزش و ارزیابی مدل‌ها مورد استفاده قرار گرفتند. به منظور اطمینان از صحت نتایج، فرآیند اعتبارسنجی متقابل به‌کار گرفته شد و عملکرد هر یک از مدل‌ها با استفاده از شاخص‌های آماری مختلف نظیر دقت، صحت، بازخوانی، F1-Score و منحنی ROC-AUC مورد ارزیابی قرار گرفت. نتایج حاصل از مقایسه مدل‌ها نشان داد که الگوریتم‌های LightGBM، Bagging و NB عملکرد بهتری نسبت به سایر مدل‌ها داشته‌اند. مدل LightGBM با مقدار AUC برابر با ۰/۹۴۷۵ و F1-score معادل ۰/۹۶۱۳، بالاترین کارایی را در میان مدل‌های بررسی‌شده نشان داد. این امر بیانگر توانایی بالای این الگوریتم در تشخیص صحیح نمونه‌های مثبت و منفی است. به‌طورکلی، مشاهده شد که روش‌های ترکیبی مانند Bagging و Stacking نیز در افزایش پایداری و کاهش خطای پیش‌بینی تأثیر قابل‌توجهی دارند. از سوی دیگر، عملکرد الگوریتم‌های ساده‌تر مانند Adaboost و KNN در مقایسه با مدل‌های مبتنی بر بوستینگ و درخت تصمیم، کمتر بود. با توجه به نتایج به‌دست‌آمده، می‌توان نتیجه گرفت که استفاده از الگوریتم‌های مبتنی بر درخت و روش‌های ترکیبی در پیش‌بینی تعمیرات تجهیزات عملکرد بسیار مؤثری دارند و می‌توانند به‌عنوان ابزار تصمیم‌یار در سیستم‌های مدیریت نگهداری مورد استفاده قرار گیرند. این مدل‌ها با ارائه پیش‌بینی‌های دقیق‌تر، موجب کاهش توقف‌های ناگهانی و بهبود بهره‌وری عملیاتی خواهند شد.

پیشنهاد می‌شود در مطالعات آتی، از شبکه‌های عصبی عمیق[1] به‌ویژه مدل‌های مبتنی بر LSTM و GRU برای پیش‌بینی‌های زمانی استفاده شود تا الگوهای پیچیده‌تر در داده‌ها بهتر شناسایی شوند. به‌کارگیری روش‌های انتخاب ویژگی[2] مانند الگوریتم ژنتیک یا بهینه‌سازی ازدحام ذرات می‌تواند به بهبود دقت مدل‌ها و کاهش پیچیدگی محاسباتی کمک کند. ترکیب داده‌های سنسوری لحظه‌ای[3] با مدل‌های یادگیری ماشین می‌تواند منجر به ایجاد سیستم‌های هوشمند نگهداری و تعمیرات پیش‌گویانه شود که در زمان واقعی تصمیم‌گیری می‌کنند.

۶- منابع

[1] Laxmi MM, Varshini KVS, Sandhya B, Prasmitha B, Sruthi B, Tanisha A. Hybrid Ensemble Learning for Real-Time Predictive Maintenance in Vehicular Engine Health Monitoring Systems. Frontiers in Collaborative Research. 2025;3(2):1-11.

[2] Hossain M, Rahman M, Ramasamy D. Artificial intelligence-driven vehicle fault diagnosis to revolutionize automotive maintenance: A review. Computer Modeling in Engineering & Sciences. 2024;141(2):951.

[3] Samrat B, Thomas P, Kumar S, Benila A, Bhardwaj R, Vigenesh M, editors. Industrial informatics in optimizing software-defined vehicles for logistics. 2024 IEEE 2nd International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP); 2024: IEEE.

[4] Li H, Shaukat H, Zhu R, Bin Kaleem M, Wu Y. Fault Detection of Li–Ion Batteries in Electric Vehicles: A Comprehensive Review. Sustainability. 2025;17(14):6322.

[5] Jegadeeshwaran R, Sugumaran V. Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. Mechanical systems and signal processing. 2015;52:436-46.

[6] Dunn BD. Materials and processes: for spacecraft and high reliability applications: Springer; 2016.

[7] Moubray J. Reliability-centered maintenance: Industrial Press Inc.; 2001.

[8] Zhang W, Yang D, Wang H. Data-driven methods for predictive maintenance of industrial equipment: A survey. IEEE systems journal. 2019;13(3):2213-27.

[1] Deep Learning
[2] Feature Selection
[3] Real-time

[9] Lee J, Bagheri B, Kao H-A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing letters. 2015;3:18-23.

[10] Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring. Mechanical systems and signal processing. 2019;115:213-37.

[11] Mosallam A, Medjaher K, Zerhouni N. Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. Journal of Intelligent Manufacturing. 2016;27(5):1037-48.

[12] Mobley RK. An introduction to predictive maintenance: Elsevier; 2002.

[13] Selcuk S. Predictive maintenance, its implementation and latest trends. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture. 2017;231(9):1670-9.

[14] Lee J, Bagheri B, Kao H-A, editors. Recent advances and trends of cyber-physical systems and big data analytics in industrial informatics. International proceeding of int conference on industrial informatics (INDIN); 2014.

[15] Tao F, Qi Q, Liu A, Kusiak A. Data-driven smart manufacturing. Journal of manufacturing systems. 2018;48:157-69.

[16] Yurdagül HH, Zaim U, Seller A, Özdemir H, Uyğur G, Akay MF, editors. Fault Prediction in Energy Systems: Telemetric Data and Machine Learning Approaches. 2025 9th International Symposium on Innovative Approaches in Smart Technologies (ISAS); 2025: IEEE.

[17] Reddy TV, Rao KM. Recent Trends in VLSI and Semiconductor Packaging: CRC Press; 2025.

[18] Baviskar[1] PV, Nayak C, editors. A Review on Equipment Health Monitoring Using Machine Learning Techniques. Proceedings of the International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022); 2023: Springer Nature.

[19] Radhakrishnan P, Ramaiyan K, Vinayagam A, Veerasamy V. A stacking ensemble classification model for detection and classification of power quality disturbances in PV integrated power network. Measurement. 2021;175:109025.

[20] Soresini F, Barri D, Cazzaniga I, Ballo F, Mastinu G, Gobbi M. Artificial Intelligence for Fault Detection of Automotive Electric Motors. Machines 2025; 13: 457.

[21] Patel K, Kolla D, Yadav G, Waghmode U, Mannava SK, Narayanan A, editors. IoT-Enabled Predictive Maintenance System for Smart Manufacturing Plants. 2025 10th International Conference on Communication and Electronics Systems (ICCES); 2025: IEEE.

[22] Benkaihoul S, Khadar S, Özüpak Y, Aslan E, Almalki MM, Mossa MA. Advanced fault classification in induction motors for electric vehicles using a stacking ensemble learning approach. World Electric Vehicle Journal. 2025;16(11):614.

[23] Dettinger F, Jazdi N, Weyrich M, Brandl L, Reuss H-C, Pecha U, et al., editors. Machine-Learning-Based Fault Detection in Electric Vehicle Powertrains Using a Digital Twin. 23rd Stuttgart International Symposium; 2023: SAE Technical Paper.

[24] Dui H, Dong X, Chen L, Wang Y. IoT-enabled fault prediction and maintenance for smart charging piles. IEEE Internet of Things Journal. 2023;10(23):21061-75.

[25] Guo B, Ding Y, Yao L, Liang Y, Yu Z. The future of false information detection on social media: New perspectives and trends. ACM Computing Surveys (CSUR). 2020;53(4):1-36.

[26] Afroz S, Brennan M, Greenstadt R, editors. Detecting hoaxes, frauds, and deception in writing style online. 2012 IEEE symposium on security and privacy; 2012: IEEE.

[27] Cantarella M, Fraccaroli N, Volpe R. Does fake news affect voting behaviour? Research Policy. 2023;52(1):104628.

[28] Goutte C, Gaussier E, editors. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. European conference on information retrieval; 2005: Springer.

[29] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons; 2013.

[30] Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20(3):273-97.

[31] Cover T, Hart P. Nearest neighbor pattern classification. IEEE transactions on information theory. 1967;13(1):21-7.

[32] John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. arXiv preprint arXiv:13024964. 2013.

[33] Breiman L. Random forests. Machine learning. 2001;45(1):5-32.

[34] Breiman L. Bagging predictors. Machine learning. 1996;24(2):123-40.

[35] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences. 1997;55(1):119-39.

[36] Chen T, Guestrin C, editors. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016.

[37] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems. 2017;30.

[38] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems. 2018;31.

[39] Wolpert DH. Stacked generalization. Neural networks. 1992;5(2):241-59.