Journal of Industrial Innovations 3(3) (2025) 39 – 60

Industrial Innovations

https://jii.araku.ac.ir

ISSN: 3092-7811

# Development of a Data Mining-Based Predictive Model for Petrochemical Compressor Failures: A Case Study of Mahabad Petrochemical Complex

**Farhad Chavoshini [a], Mehdi Yousefinejad Attari [a*], Fahime Lotfian Delouyi [b]**

[a] Department of Industrial Engineering, Bon.C., Islamic Azad University, Bonab, Iran.
[b] Department of Mechanical Engineering, Facuity of Engineering, Zabol University, Zabol, Iran.

Original Article

Use your device to scan and read the article online

Citation: Chavoshini F, Yousefinejad Attari M, Lotfian Delouyi F. Development of a Data Mining-Based Predictive Model for Petrochemical Compressor Failures: A Case Study of Mahabad Petrochemical Complex. Industrial Innovations. 2025;3 (3):39-60.

https://doi.org/10.61882/jii.3.3.39

KEYWORDS

Data Mining;
Petrochemical Industry;
Clustering;
Decision Tree;
Critical Root Cause.

ABSTRACT

A majority of production equipment worldwide today experiences failures in its working life cycle and must be repaired and maintained. Since production downtime incurs enormous financial implications, it is extremely crucial to find solutions for taking preventive action. Moreover, the use of sophisticated equipment—whose spare parts are expensive and sometimes unavailable due to political or technical limitations—has elevated the status of maintenance systems within companies, with it becoming a key determinant in productivity. A maintenance system not only reduces the final cost of production but also impacts the overall organizational value chain. Elements such as production speed and delivery time, product quality, production line reliability, and organizational flexibility all come under the influence of maintenance decisions. Therefore, in industries such as petrochemicals, where high-level machinery is utilized in the production systems, equipment reliability and maintenance gain greater significance. In this study, to analyze and predict equipment failures by using data mining methods, a 1304640 record dataset measured between December 20, 2021, and May 20, 2022, from 17 measurement sensors on one of the 3000 -series compressors of the Mahabad Petrochemical Complex were used. These data were examined, analyzed, and mined with Weka software. The statistical summaries and correlation relationships among the data were studied initially. Then, the reasons for equipment failure and the critical path leading to compressor shutdown were evaluated using the application of K-Means clustering, decision trees, and feature selection methods. Finally, based on the findings, recommendations were provided to focus maintenance efforts on critical sensors.

## Extended Abstract

### 1. Introduction

Maintenance is regarded as a crucial industrial process that significantly affects both costs and reliability and plays a direct role in enabling companies to compete on the basis of low cost, high quality, and superior performance. Consequently, its importance within organizations has been steadily increasing. In recent years, maintenance management policies have evolved from simple systems into complex and proactive approaches. Under such conditions, corrective and preventive activities are implemented based on inspections, monitoring, and forecasting in order to ensure continuous operation, improved productivity, and enhanced reliability.

* Corresponding author.
E-mail address: mahdi108108@iau.ir
DOI: https://doi.org/10.61882/jii.3.3.39
Received: November 19, 2025; Received in revised form: Junuary 27, 2026; Accepted: Junuary 27, 2026.
Article type: Research Paper

With the continuous advancement of information technology in the petrochemical industry, the volume of operational data has been growing exponentially. A critical challenge is how to optimally manage and efficiently analyze these data in order to remain competitive in the global market while achieving high levels of productivity and reliability.

## 2. Statement of the Problem

Fault detection and prediction in industrial systems have become essential due to their direct impact on system performance and efficiency. Process gas compressors play a critical role in the oil, gas, and petrochemical industries. Instrumentation systems monitor control parameters such as temperature, pressure, flow rate, and vibration, and acceptable operating ranges are defined for each parameter. Any deviation beyond these limits may result in equipment or system shutdown. Therefore, the availability of an intelligent system capable of analyzing recorded data and detecting abnormal changes in vibration, bearing temperature, oil condition, unusual noises, cooling system performance, leakage or tube blockage, pressure variations, and seal failures prior to an incident is extremely valuable.

The aim of this study is to apply data mining techniques to predict critical operating states of compressors in the petrochemical industry. Factors influencing compressor failure include compressor gas parameters such as inlet and outlet temperatures, suction and discharge pressures, as well as environmental conditions. To achieve this objective, the required data were collected from 17 sensors installed on a compressor at the Mahabad Petrochemical Complex.

## 3. Methodology

This study employed K-means clustering, decision tree analysis, and feature selection techniques to identify critical features. All computations were performed using the Weka software. Based on the type of equipment and the extracted time-series data, various numbers of clusters were tested to determine the optimal clustering configuration. The sum of squared errors was calculated for each case, and the number of clusters with the minimum error was selected as the optimal solution. Subsequently, critical clusters were identified. These clusters require further investigation to identify failure risks and implement preventive measures. For decision tree construction, a random tree algorithm was used. This algorithm is a supervised classification method and belongs to the category of ensemble learning techniques.

## 4. Results

The performance evaluation results of the decision tree indicate only a slight discrepancy between predicted and actual values, demonstrating the high accuracy and efficiency of the proposed model. Based on the combined results of the three methods, critical sensors were identified, the reasons for their importance were analyzed, the causes of their malfunctions were determined, and recommendations were proposed to prevent compressor shutdowns.

## 5. Conclusion

In this study, data mining techniques were applied to analyze and predict critical operating states of compressors in the petrochemical industry. Three methods—K-means clustering, decision tree analysis, and feature selection—were used to identify critical features. Based on the obtained results, critical sensors were identified and practical recommendations were proposed to prevent compressor shutdowns

نوآوری‌های صنعتی

شاپا: ۷۸۱۱-۳۰۹۲

# توسعه مدل پیش‌بینی خرابی کمپرسورهای پتروشیمی مبتنی بر داده‌کاوی: مطالعه موردی مجتمع پتروشیمی مهاباد

فرهاد چاوشینی [الف]، مهدی یوسفی‌نژاد عطاری [ب*]، فهیمه لطفیان دلوئی [ج]

[الف] گروه مهندسی صنایع، واحد بناب، دانشگاه آزاد اسلامی، بناب، ایران. Farhad_m670@yahoo.com

[ب] گروه مهندسی صنایع، واحد بناب، دانشگاه آزاد اسلامی، بناب، ایران. mahdi108108@iau.ir

[ج] گروه مهندسی مکانیک، دانشکده فنی و مهندسی، دانشگاه زابل، زابل، ایران. fahime.lotfian@uoz.ac.ir

**واژگان کلیدی**

داده‌کاوی؛
صنعت پتروشیمی؛
خوشه‌بندی؛
درخت تصمیم؛
مسیر بحرانی.

تاریخ دریافت: ۱۴۰۴/۰۸/۲۸

تاریخ بازنگری: ۱۴۰۴/۱۱/۰۷

تاریخ پذیرش: ۱۴۰۴/۱۱/۰۷

**چکیده**

تجهیزات مورد استفاده در فرآیند تولید، در طول چرخه عمر عملیاتی خود نیازمند نگهداری و تعمیرات هستند. توقف خطوط تولید، اغلب هزینه‌های بالایی دارد. یافتن راهکارهایی برای اقدامات پیش‌گیرانه، اهمیت بسزایی دارد. از سوی دیگر، استفاده از تجهیزات پیشرفته با قطعات یدکی گران‌قیمت، و دارای محدودیت‌های سیاسی یا فنی برای خرید، جایگاه سیستم نگهداری و تعمیرات را در سازمان ارتقاء داده و آن را به یکی از مسائل کلیدی در حوزه بهره‌وری تبدیل کرده است. سامانه نگهداری و تعمیرات، نه‌تنها در کاهش هزینه نهایی تولید نقش دارد، بلکه بر کل زنجیره ارزش سازمان نیز اثرگذار است. عواملی مانند سرعت تحویل محصول به مشتری، کیفیت محصول، قابلیت اطمینان تولید، تحت‌تأثیر تصمیمات مرتبط با نگهداری و تعمیرات قرار دارند. در صنایعی مانند پتروشیمی که از ماشین‌آلات پیشرفته در سیستم تولید استفاده می‌شود، بر اهمیت قابلیت اطمینان تجهیزات افزوده می‌شود. در این پژوهش، به‌منظور مطالعه و پیش‌بینی خرابی تجهیزات، از روش‌های مبتنی بر داده‌کاوی شامل ۱۳۰۴۶۴۰ داده در بازه زمانی ۲۰ دسامبر ۲۰۲۱ تا ۲۰ می ۲۰۲۲ استفاده شده که توسط ۱۷ حسگر اندازه‌گیری مربوط به یکی از کمپرسورهای ۳۰۰۰ پتروشیمی مهاباد ثبت شده‌اند. این داده‌ها با استفاده از نرم‌افزار Weka مورد تحلیل و داده‌کاوی قرار گرفته‌اند. در ابتدا، توصیف‌های آماری و روابط همبستگی میان داده‌ها تحلیل شده است. سپس با استفاده از روش‌های خوشه‌بندی K-Means، درخت تصمیم، و انتخاب ویژگی‌ها، علل خرابی تجهیزات و مسیرهای بحرانی منجر به توقف کمپرسور مورد ارزیابی قرار گرفته است. در پایان، با توجه به نتایج ، پیشنهاداتی جهت نگهداری و تعمیرات حسگرهای بحرانی ارائه شده است.

## ۱- مقدمه

تعمیر و نگهداری به‌عنوان یک فرآیند حیاتی در صنعت، تأثیر قابل‌توجهی بر هزینه‌ها و قابلیت اطمینان دارد و بر توانایی شرکت برای رقابت در قیمت پایین، کیفیت بالا و عملکرد تأثیرگذار است [۱]. ازاین‌رو جایگاه آن در سازمان‌ها رو به رشد است [۲]. در طی سال‌های اخیر، خطی مشی مدیران صنعتی در زمینه نگهداری و تعمیرات، از یک سیستم ساده به سیستمی پیچیده

* نویسنده مسئول؛
آدرس پست الکترونیک: mahdi108108@iau.ir

و فعّال تبدیل شده است. در چنین شرایطی، بر اساس بازرسی‌ها، پایش‌ها و پیش‌بینی‌ها، فعالیت‌های اصلاحی و پیشگیرانه انجام می‌شود تا بهره‌برداری مستمر، افزایش بهره‌وری و قابلیت اطمینان تضمین شود [۳].

با بهبود مستمر فناوری اطلاعات در صنعت پتروشیمی، داده‌های پایه به‌صورت نمایی در حال افزایش است. مسئله مهم این است که چگونه می‌توان این داده‌ها را به‌طور بهینه مدیریت و به‌صورت کارآمد تحلیل کرد تا امکان رقابت در عرصه جهانی را داشته و به بهره‌وری و قابلیت اطمینان بالایی دست یابد [۴].

با پیشرفت صنایع، تشخیص و پیش‌بینی خطا در سیستم‌های صنعتی امری حیاتی است و بازدهی و عملکرد این سیستم‌ها را تحت تأثیر قرار می‌دهد. بدین منظور، استفاده از اینترنت اشیاء، هوش مصنوعی و یادگیری عمیق، به‌عنوان روش‌هایی کارا برای تشخیص و پیش‌بینی خطا مورد بررسی و تحقیق قرار گرفته است. می‌توان با تشخیص زودهنگام خطا، از وقوع خطا جلوگیری کرد. از مزایای این روش‌ها می‌توان به افزایش دقت و سرعت، قابلیت پیش‌بینی خطاهای احتمالی قبل از وقوع آن‌ها و کاهش هزینه و زمان نگهداری و تعمیرات با جلوگیری از ایجاد خطا اشاره کرد.

در صنعت نفت و گاز و پتروشیمی، کمپرسورهای گاز فرآیندی نقش بسیار مهمی ایفا می‌کنند. از کمپرسور فرآیندی در سیکل‌های بسته برای ثابت نگه‌داشتن مقدار جریان سیال استفاده می‌شود. هر رآکتور که محصول در آن تولید می‌شود، دارای یک سیستم بازیابی گاز است که برای کمک به حفظ ترکیبات گاز راکتور استفاده می‌شود.

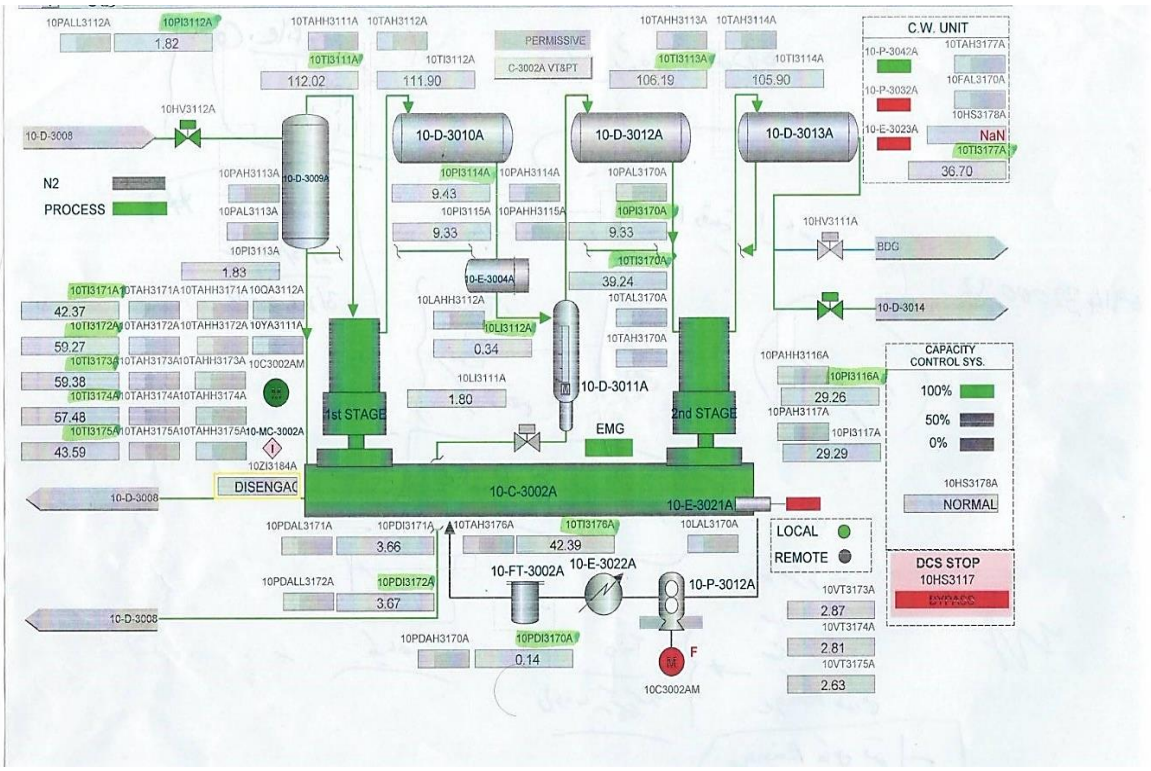
ترکیبات گازی که در واکنش تولید باقی‌مانده و با کاتالیست داخل آن به محصول پودر تبدیل نشده است، مجدد با کمک این کمپرسور به چرخه تولید و تزریق داخل رآکتورها برمی‌گردد. خرابی این تجهیزات منجر به توقف تولید در سیستم فرآیندی مجموعه خواهد شد.

کمپرسورهای رفت و برگشتی فشار گازها را از ۱.۵ بار تا ۲۹ بار افزایش می‌دهند. این کمپرسورها می‌تواند روغنی یا بدون روغن باشد. امّا در صنایع نفت، گاز و پتروشیمی، به دلیل حساسیت بالا به آلودگی، معمولاً از نوع بدون روغن استفاده می‌شود. حتی ورود حدود یک درصد روغن به جریان خروجی می‌تواند باعث آلودگی محصول نهایی شود. تنها در مواردی که آلودگی اهمیتی نداشته باشد، از کمپرسورهای روغنی استفاده می‌شود. در کمپرسورهای پیستونی، قطعات فرسایشی متعددی مانند سوپاپ‌ها، رینگ‌های تراکم و روغنی و یاتاقان‌ها وجود دارد که موجب افزایش هزینه نگهداری و تعمیرات می‌شوند. علاوه بر این، به دلیل ضرورت استفاده از سوپاپ‌های مکش و تراکم، احتمال خرابی و کاهش کارایی وجود دارد.

پارامترهای کنترلی مانند دما، فشار، دبی و لرزش توسط تجهیزات ابزار دقیق پایش می‌شوند و برای هر یک محدوده مجاز تعریف شده است. هر مقدار خارج از این محدوده می‌تواند موجب توقف تجهیز یا سیستم شود. بنابراین، وجود یک سیستم هوشمند با قابلیت تحلیل داده‌های ثبت‌شده، که بتواند تغییرات غیرعادی در لرزش، دمای بیرینگ‌ها، وضعیت روغن، صداهای غیرعادی، عملکرد سیستم خنک‌کننده، نشتی یا انسداد تیوب‌ها، تغییرات فشار و خرابی در آب‌بندی‌ها را پیش از بروز حادثه تشخیص دهد، بسیار ارزشمند است. با استفاده از الگوریتم‌های پیش‌بینی خطا، می‌توان بهبود چشمگیری در عملکرد سیستم‌ها داشت. برای استفاده از هوش مصنوعی در تشخیص خطا، نیازمند داده‌های با کیفیت بالا و همگن در بازه زمانی مورد نظر هستیم. همچنین، باید از روش‌های پردازش داده بهره برد تا بتوانیم دقیق و سریع خطاهای سیستم را تشخیص داده و پیش‌بینی کنیم. بنابراین با انجام تعمیرات دوره‌ای و تشخیص به‌موقع تعویض قطعات مصرفی، از توقف سیستم، خرابی‌های جبران‌ناپذیری که منجر به تعمیرات قطعات اصلی و یا تعویض کامل تجهیزات می‌شود، جلوگیری کرده و هزینه‌ها و نفر ساعت کار تعمیرات انجام‌شده را کاهش دهیم.

در این پژوهش، هدف پیش‌بینی حالات بحرانی کمپرسورها در صنعت پتروشیمی با استفاده از روش‌های داده‌کاوی است. معیارهای مؤثر بر خرابی کمپرسور شامل پارامترهای گازهای کمپرسور از جمله دمای ورودی و خروجی و فشار ساکشن و تخلیه کمپرسور و همچنین متغیرهای محیطی مورد بررسی قرار گرفته است. بدین منظور داده‌های مورد نیاز از ۱۷ حسگر نصب شده بر روی یک کمپرسور در مجتمع پتروشیمی مهاباد جمع‌آوری شده‌اند. در بخش فرآیندی تولید محصول پلی‌اتیلن پتروشیمی

مهاباد، ۴ عدد کمپرسور ۳۰۰۰ وجود دارد که وظیفه ترکیب و بازگرداندن گازها به رآکتورها را بر عهده دارند. برای تولید مستمر، ضروری است که این کمپرسورها در وضعیت بهره‌برداری باشند. شکل ۱، نمایی شماتیک از حسگرهای نصب‌شده بر روی کمپرسور را نشان می‌دهد.




**شکل ۱ حسگرهای نصب‌شده بر روی کمپرسور**

سپس تحلیل آماری داده‌ها شامل شاخص‌های مرکزی، پراکندگی و هیستوگرام ویژگی‌ها انجام شده است. از روش خوشه‌بندی K-Means، درخت تصمیم و انتخاب ویژگی برای شناسایی ویژگی‌های بحرانی بهره برده شده است. محاسبات با نرم‌افزار Weka انجام شده است. با توجه به نوع تجهیزات و داده‌های زمانی استخراج‌شده، برای تعیین تعداد بهینه خوشه‌ها، تعداد خوشه‌های مختلف امتحان، معیار مربع خطا محاسبه‌شده و در نهایت تعداد خوشه‌ای که کمترین مربع خطا را دارد به‌عنوان خوشه‌های بهینه انتخاب شده است. خوشه‌های بحرانی شناسایی خواهد شد. این خوشه نیاز به بررسی بیشتر جهت شناسایی ریسک‌های خرابی و پیش‌گیری از آن را دارد. برای ترسیم درخت تصمیم، از درخت تصادفی که یک طبقه‌بندی نظارت‌شده است و در دسته الگوریتم‌های یادگیری تجمعی قرار می‌گیرد بهره برده شده است. الگوریتم درخت تصادفی در مقایسه با سایر الگوریتم‌های تجمعی، از مزایایی همچون توان تعمیم‌پذیری بالا، سازگاری و دقت بیشتر برخوردار است. علاوه بر این، این الگوریتم قادر است مسائل غیرخطی و بیش‌برازش را با دقت بالا حل کند. الگوریتم درخت تصادفی همچنین از سطح بالایی از انعطاف‌پذیری و سرعت بالای آموزش برخوردار است [۵]. نتایج ارزیابی عملکرد درخت تصمیم نشان‌گر اختلاف اندک بین مقادیر پیش‌بینی‌شده و مقادیر واقعی و در نتیجه کارایی و دقت بالای مدل در پیش‌بینی است. با توجه به نتایج مجموعه سه روش، حسگرهای بحرانی شناسایی، دلایل اهمیت آن‌ها تحلیل، علل اختلال در آن‌ها شناسایی و پیشنهادات جهت پیشگیری از توقف کمپرسور ارائه شده است.

بازرسی چندحسگری در مقایسه با رویکرد تک‌حسگری، اطلاعات جامع‌تری درباره وضعیت سلامت ماشین ارائه می‌دهد. مطالعات پیشین عمدتاً بر داده‌های تک‌حسگری بوده‌اند که از نظر هزینه و میزان اطلاعات ارائه‌شده بهینه نبوده‌اند. در این پژوهش، پارامترهای فشار و دما اندازه‌گیری و پایش شده‌اند [۶،۷].

علاوه بر این، کمپرسورها از اجزای اصلی صنعت محسوب می‌شوند و نادیده گرفتن وضعیت سلامت آن‌ها امکان‌پذیر نیست. بااین‌حال، بازرسی و پایش کمپرسورها در مطالعات اخیر کمتر مورد توجه قرار گرفته است. نوآوری این مقاله در تلفیق گردآوری داده‌های چندحسگری با حجم بالا، داده‌کاوی با استفاده از روش‌های خوشه‌بندی K-Means، درخت تصادفی و انتخاب ویژگی برای شناسایی ویژگی‌های بحرانی است. این رویکردها با هدف پوشش خلأهای علمی و کاربردی موجود در این حوزه ارائه شده است.

این مقاله به این صورت سازمان‌دهی شده است: در بخش ۲، مرور ادبیات مرتبط ارائه شده است. در بخش ۳، روش‌های داده‌کاوی مورد استفاده برای شناسایی حسگرهای بحرانی شرح داده شده است. در بخش ۴، نتایج تحلیل داده‌ها و شناسایی حسگرهای بحرانی مورد بحث قرار گرفته است. در پایان در بخش ۵، نتیجه‌گیری و پیشنهادات برای تحقیقات آتی مطرح شده است.

## ۲- پیشینه تحقیق

کشف دانش درون داده‌ها در عصر اطلاعات یکی از مفاهیمی است که روزبه‌روز بر اهمیت آن افزوده می‌شود. امروزه دسترسی به داده‌های حاصل از محیط‌های تولیدی، همچون داده‌های مربوط به فرآیند و عملیات یا سوابق فعالیت‌های نگهداری و تعمیرات گذشته، امکان‌های جدیدی را برای تحلیل‌های پیشرفته داده از جمله پیش‌بینی رفتار خرابی فراهم ساخته است. با استفاده از روش‌های تحلیل داده، برآورد ریسک و احتمال وقوع خرابی دیگر ماهیتی صرفاً ذهنی نخواهد داشت و صرفاً وابسته به تجربه و دانش کارکنان نخواهد بود. در این راستا پژوهش‌های متعددی در زمینه تلفیق روش‌های نگهداری و تعمیرات با تکنیک‌های داده‌کاوی و هوش مصنوعی انجام شده است که سیر تکامل آن‌ها نشان‌دهنده حرکت از مدل‌های توصیفی به سمت مدل‌های پیش‌بینی و تصمیم‌یار پیشرفته است.

صالحیان و جهان (۲۰۲۲) در یک مدل توصیفی، به بیان تجربیات استقرار سیستم نگهداری تعمیرات، در شرکت توزیع گاز گلستان پرداختند. در گام‌های اولیه مدل‌های پیش‌بینی، برخی مطالعات به شناسایی ریسک‌ها در نگهداری پرداختند [۸]. به‌عنوان نمونه، فیلز و همکاران[1] (۲۰۲۱) یک روش تحلیل حالات و آثار خرابی (FMEA) مبتنی بر داده را با استفاده از مدل‌های یادگیری عمیق بر روی داده‌های تاریخی و عملیاتی مرحله بهره‌برداری از کالاهای سرمایه‌ای صنعتی ارائه کردند. چارچوب پیشنهادی با استفاده از یک مطالعه موردی در بخش هوانوردی پیاده‌سازی و اعتبارسنجی شد و نتایج نشان داد که دقت پیش‌بینی خرابی حدود ۹۵ درصد است [۹]. همچنین، پنگ و همکاران[2] (۲۰۱۷) با تمرکز بر صنعت ساختمان، رویکردهای داده‌کاوی مبتنی بر مدل‌سازی اطلاعات ساختمان (BIM) ارائه دادند تا الگوها و قوانین معنادار استخراج شود. تکنیک‌هایی چون خوشه‌بندی، تشخیص داده‌های ناهنجار و استخراج الگو را برای پیش‌بینی، هشدار زودهنگام و تصمیم‌گیری بکار بردند که منجر به بهبود استفاده از منابع در مرحله نگهداری و بهره‌برداری شد [۱۰]. آتشگر و سراوانی (۲۰۱۶) با ترکیب شبکه عصبی و بلوک‌های تصمیم‌گیری، مدلی برای تشخیص هم‌زمان خرابی و علت آن در ماشین‌های تراش ارائه کردند که با استفاده از توزیع وایبل آموزش دیده و دقت ۹۹.۵ درصدی در پیش‌بینی داشت [۱۱]. کریمی و افشار کاظمی[3] (۲۰۱۶) نیز با بهره‌گیری از الگوریتم درخت C&R، فواصل میان خرابی دستگاه‌های خودپرداز را پیش‌بینی کرده و گروه‌بندی آن‌ها را بر اساس زمان خرابی انجام دادند [۱۲]. باقری‌قادیکلایی[4] (۲۰۲۰) با به‌کارگیری روش‌های داده‌کاوی شامل خوشه‌بندی، الگوریتم GRI و طبقه‌بندی بر روی داده‌های یک شرکت تولیدکننده دستگاه‌های ATM، الگوها و قوانین معنادار را استخراج کردند [۱۳].

تحقیقات دیگری مانند رودریگرز و همکاران[5] (۲۰۲۳) با هدف بهبود نگهداری پیش‌گویانه توربین‌های بادی، به تحلیل داده‌های SCADA با استفاده از الگوریتم خوشه‌بندی K-Means پرداختند. نتایج نشان داد که این روش‌ها قادر به شناسایی

[1] Filz et al.
[2] Peng et al.
[3] Karimi and Afshr Kazemi
[4] Bagherighadikolaeia
[5] Rodriguez et al.

داده‌های پرت و رفتارهای غیرعادی توربین‌ها بوده و می‌توانند زمینه‌ساز تصمیم‌گیری دقیق‌تر متخصصان و بهینه‌سازی راهبردهای نگهداری باشند [۱۴]. چنگ و همکاران[1] (۲۰۱۳) با استفاده از رویکرد داده‌کاوی و بهره‌گیری از الگوریتم CART، به تحلیل ۳۴۹ مورد از حوادث شغلی عمده در صنایع پتروشیمی تایوان طی سال‌های ۲۰۰۰ تا ۲۰۱۰ پرداخته شد. یافته‌های آن‌ها نشان داد که طراحی و نگهداشت اصولی تجهیزات همراه با اجرای نظام‌مند آموزش‌های ایمنی و استانداردهای مدیریتی، می‌تواند در پیشگیری از وقوع حوادثی نظیر آتش‌سوزی، انفجار و نشت مواد و در نتیجه کاهش پیامدهای اجتماعی ناشی از آن‌ها، نقش بسزایی ایفا کند [۱۵].

در سال‌های اخیر، درخت تصمیم به‌عنوان یکی از ابزارهای پرکاربرد در این حوزه، در مطالعات متعددی به‌کاررفته است. ارنا و همکاران[2] (۲۰۲۲) با استفاده از داده‌های یک شرکت صنایع غذایی، به بررسی کاربرد درخت تصمیم برای مدیریت تعمیر و نگهداری پیش‌گویانه (PdM) در دو نوع خرابی گیربکس خط بو دادن دانه‌های روغنی پرداختند. هدف مطالعه، توسعه سیستمی برای پشتیبانی از تصمیم‌گیری در پیاده‌سازی PdM بود که با توجه به کیفیت داده‌ها، شدت و قابلیت تشخیص خرابی‌های بالقوه (شناسایی‌شده از طریق FMECA)، و هزینه‌های مستقیم و غیرمستقیم نگهداری، بهترین استراتژی را پیشنهاد دهد. درخت تصمیم امکان تحلیل سناریوهای مختلف را فراهم کرده و نشان داد در چه شرایطی PdM نسبت به تعمیرات اصلاحی فعلی، صرفه‌جویی اقتصادی دارد. نتایج حاکی از سادگی و اثربخشی روش در پشتیبانی تصمیمات نگهداری و ارزیابی هزینه‌ها در سطوح مختلف خرابی است [۱۶]. سنتوس و کاستا[3] (۲۰۰۸) با بهره‌گیری از مطالعه موردی و رویکرد داده‌کاوی مبتنی بر درخت‌های تصمیم مختلف، به پایش وضعیت برخط موتورهای القایی پرداختند [۱۷]. همچنین تحقیقات مشابهی مانند ترن و همکاران[4] (۲۰۰۹) نیز مدل‌های ترکیبی درخت تصمیم و سیستم‌های فازی را برای عیب‌یابی موتور القایی ارائه دادند [۱۸].

کاربردهای داده‌کاوی به حوزه‌های دیگر نگهداری و تعمیرات نیز گسترش یافته است. الله بوخش و همکاران[5] (۲۰۱۹) باهدف مطالعه توسعه مدل‌های پیش‌بینی از داده‌های موجود از یک آژانس راه‌آهن از تکنیک‌های طبقه‌بندی مبتنی بر درخت یادگیری ماشین برای پیش‌بینی نیاز تعمیر و نگهداری، نوع فعالیت و وضعیت ماشه سوئیچ‌های راه‌آهن استفاده کردند. با استفاده از داده‌های یک فرآیند کاری در حال استفاده، مدل‌های پیش‌بینی بر اساس درخت تصمیم، جنگل تصادفی و درخت‌های تقویت‌شده گرادیان توسعه دادند [۱۹]. در همین راستا، برمیستر و همکاران[6] (۲۰۲۳) قابلیت یک رویکرد فعال را با استفاده از داده‌های تولید برای جلوگیری از تعمیر و نگهداری آینده ارائه دادند. همچنین کاربرد مدل‌های یادگیری ماشینی قابل تفسیر را برای درک رابطه بین ویژگی‌های خطاهای تولید و خطاهای بازرسی بعدی نشان دادند [۲۰].

کامل[7] (۲۰۲۲) با استفاده از هوش مصنوعی مدلی برای پیش‌بینی وضعیت سلامت ماشین‌ها ارائه دادند. در این داده‌ها اطلاعاتی مانند دمای هوا و فرآیند، سرعت چرخش، گشتاور، میزان سایش دستگاه و نوع محصول تولیدی ثبت شد. خروجی مدل، وضعیت خرابی ماشین به‌صورت باینری (درست یا نادرست) بود. یک شبکه عصبی مصنوعی توسعه یافت که توانست وضعیت ماشین را با دقت پیش‌بینی کند. نتایج نشان داد که به‌کارگیری هوش مصنوعی در تعمیر و نگهداری، به بهبود بهره‌وری و صرفه‌جویی مالی منجر می‌شود [۲۱].

روش‌های متعارف هوش مصنوعی نقش قابل‌توجهی در پیشرفت پایش وضعیت کمپرسورها، به‌ویژه در حوزه‌های تشخیص خطا، عیب‌یابی و نگهداری و تعمیرات پیش‌بینانه ایفا کرده‌اند. این روش‌ها با بهره‌گیری از الگوریتم‌های یادگیری ماشین، حجم عظیمی از داده‌ها را پردازش کرده، الگوهای پنهان را شناسایی نموده و پیش‌بینی‌های آگاهانه‌ای درباره سلامت و عملکرد کمپرسورها ارائه می‌دهند.

[1] Cheng et al.
[2] Arena et al.
[3] Santos and Costa
[4] Tran et al.
[5] Allah Bukhsh et al.
[6] Burmeister et al.
[7] Kamel

ماشین‌های بردار پشتیبان (SVM) در پایش وضعیت کمپرسورها، به‌ویژه در تشخیص و طبقه‌بندی خطا، کارایی بالایی از خود نشان داده‌اند. توانایی این روش در مدیریت داده‌های با ابعاد بالا و انجام طبقه‌بندی غیرخطی، آن را به گزینه‌ای مناسب برای تحلیل الگوهای پیچیده عملکرد کمپرسورها تبدیل کرده است. در پژوهش فالزون و کلدزیج[1] (۲۰۱۷)، از SVM برای تحلیل داده‌های ارتعاشی پردازش‌شده با تبدیل موجک گسسته استفاده شد. این رویکرد امکان تشخیص خرابی در فنرهای شیر تخلیه سمت میل‌لنگ یک کمپرسور رفت‌وبرگشتی را فراهم کرد و کارایی SVM را در شناسایی خرابی اجزای خاص نشان داد [۲۲]. در مطالعه‌ای دیگر، گلمرادی و همکاران (۲۰۱۷a) روشی مبتنی بر SVM برای پایش وضعیت کمپرسور ارائه کردند که دقت ۱۰۰ درصدی در تشخیص خطا را گزارش نمودند [۲۳].

درخت‌های تصمیم نیز از جمله روش‌های پرکاربرد هوش مصنوعی متعارف هستند که با تقسیم داده‌ها به شاخه‌های مختلف بر اساس مقادیر ویژگی‌ها، فرآیند پیش‌بینی را انجام می‌دهند. این روش به دلیل ساختار شفاف و قابل تفسیر، از شهود بالایی برخوردار است و مدل تصمیم‌گیری را به‌صورت یک ساختار درختی نمایش می‌دهد.

در پایش وضعیت کمپرسورها، درخت‌های تصمیم قادرند حالت‌های عملکردی مانند سالم یا معیوب را با تحلیل ویژگی‌های ورودی طبقه‌بندی کنند. سادگی در پیاده‌سازی، توانایی پردازش داده‌های عددی و کیفی، و ارائه قواعد تصمیم‌گیری شفاف از مزایای اصلی این روش به‌شمار می‌روند. بااین‌حال، درخت‌های تصمیم در مواجهه با داده‌های پیچیده یا نویزی مستعد بیش‌برازش هستند که می‌تواند توان تعمیم‌پذیری مدل را کاهش دهد. باوجوداین محدودیت، ترکیب درخت‌های تصمیم با روش‌های تجمعی مانند جنگل تصادفی می‌تواند دقت و پایداری مدل را افزایش دهد [۲۴]. در پژوهش گلمرادی و همکاران (۲۰۱۷b) یک سیستم استنتاج فازی را با ترکیب تبدیل موجک و درخت تصمیم برای عیب‌یابی کمپرسورها ارائه کردند. در این مطالعه از الگوریتم J48 برای طراحی مدل طبقه‌بندی و انتخاب ویژگی‌های مؤثر استفاده شد که دقت کلی ۹۳.۳۳ درصدی در تشخیص عیوب کمپرسور به‌دست آمد [۲۵].

از جمله روش‌های یادگیری تجمعی، می‌توان به جنگل‌های تصادفی اشاره کرد، که با ایجاد تعداد زیادی درخت تصمیم و تجمیع خروجی آن‌ها، عملکرد پیش‌بینی را بهبود می‌بخشند. این رویکرد با کاهش بیش‌برازش درخت‌های منفرد، موجب افزایش تعمیم‌پذیری مدل می‌شود. جنگل تصادفی در پایش وضعیت کمپرسورها، برای تحلیل داده‌های پیچیده حاصل از حسگرهایی مانند ارتعاش، دما و فشار به‌طور گسترده استفاده شده است. النگوان و همکاران[2] (۲۰۱۵) از جنگل تصادفی برای تشخیص خرابی شیرهای کمپرسور رفت‌وبرگشتی استفاده کردند و نشان دادند که این روش نسبت به SVM و رگرسیون لجستیک دقت بالاتری دارد [۲۶].

الگوریتم k-همسایه نزدیک یک روش یادگیری غیرپارامتریک و مبتنی بر نمونه است که به‌طور گسترده در پایش وضعیت کمپرسورها به‌کار می‌رود. سادگی و کارایی این روش، آن را برای تشخیص و عیب‌یابی خطا مناسب ساخته است. در مطالعه آراوینث و ساگوماران[3] (۲۰۲۳)، الگوریتم k-همسایه نزدیک در تحلیل ویژگی‌های استخراج‌شده از سیگنال‌های ارتعاشی عملکرد قابل قبولی در طبقه‌بندی خطا نشان داد [۲۷].

شبکه‌های بیزی مدل‌های گرافی احتمالاتی هستند که روابط شرطی بین متغیرها را از طریق یک گراف جهت‌دار بدون دور نمایش می‌دهند. این شبکه‌ها در پایش وضعیت کمپرسورها امکان مدل‌سازی روابط میان پارامترهای عملیاتی و شرایط خرابی را فراهم کرده و در تشخیص و نگهداری پیش‌بینانه نقش مؤثری دارند. نسایب و لکحال[4] (۲۰۲۲) مدلی مبتنی بر شبکه بیزی ارائه کردند که با تلفیق داده‌های ارتعاشی و پارامترهای عملیاتی، دقت پیش‌بینی خرابی را افزایش داد [۲۸].

شبکه‌های عصبی مصنوعی به یکی از ابزارها در پایش وضعیت کمپرسورها تبدیل شده‌اند و چارچوبی برای تشخیص خطا، ارزیابی عملکرد و نگهداری پیش‌بینانه فراهم می‌کنند. توانایی این شبکه‌ها در مدل‌سازی روابط غیرخطی پیچیده، آن‌ها را برای

[1] Falzone and Kolodziej
[2] Elangovan et al.
[3] Aravinth and Sugumaran
[4] Nessaib and Lakehal

تحلیل داده‌های حسگری و شناسایی ناهنجاری‌های عملیاتی کارآمد ساخته است. مطالعات مختلفی کارایی شبکه‌های عصبی مصنوعی را در تشخیص شرایط سالم و معیوب کمپرسورها نشان داده‌اند [۲۹].

در مجموع، مرور این مطالعات نشان می‌دهد حوزه نگهداری و تعمیرات از تحلیل‌های سنتی و توصیفی به سمت سیستم‌های هوشمند و پیش‌بینی‌گر مبتنی بر داده‌کاوی و یادگیری ماشین حرکت کرده است. جایی که تصمیم‌گیری نه صرفاً بر اساس تجربه، بلکه بر پایه تحلیل دقیق داده‌ها و مدل‌های پیشرفته صورت می‌گیرد.

در این تحقیق به مطالعه و پیش‌بینی حالات بحرانی کمپرسور در صنعت پتروشیمی با بهره بردن از روش‌های داده‌کاوی و حالت چند حسگری پرداخته شده است. بدین منظور از سه روش خوشه‌بندی K-Means، درخت تصادفی که در مقایسه با سایر الگوریتم‌های تجمعی، از مزایایی همچون توان تعمیم‌پذیری بالا، سازگاری و دقت بیشتر برخوردار است، و انتخاب ویژگی برای شناسایی ویژگی‌های بحرانی بهره برده شده است. با توجه به نتایج حاصل، حسگرهای بحرانی شناسایی شده و پیشنهادات جهت پیشگیری از توقف کمپرسور ارائه شده است.

## ۳- رویکرد داده‌کاوی در پیش‌بینی خرابی کمپرسور

در این بخش، پس از شناسایی ویژگی‌های مؤثر بر خرابی کمپرسور، داده‌های مورد نیاز این تحقیق گردآوری شده است. پس از بررسی و تحلیل توصیفات آماری داده‌ها شامل شاخص‌های مرکزی و شاخص‌های پراکندگی، هیستوگرام تغییرات ویژگی‌های اندازه‌گیری شده در طی زمان ترسیم شده است. در ادامه با استفاده از روش خوشه‌بندی K-Means، درخت تصمیم و تکنیک انتخاب ویژگی، حسگرهای بحرانی شناسایی شده‌اند. در این تحقیق برای محاسبات داده‌کاوی، از نرم‌افزار Weka استفاده شده است.

### ۳-۱- تشکیل پایگاه داده (گردآوری اولیه داده‌ها)

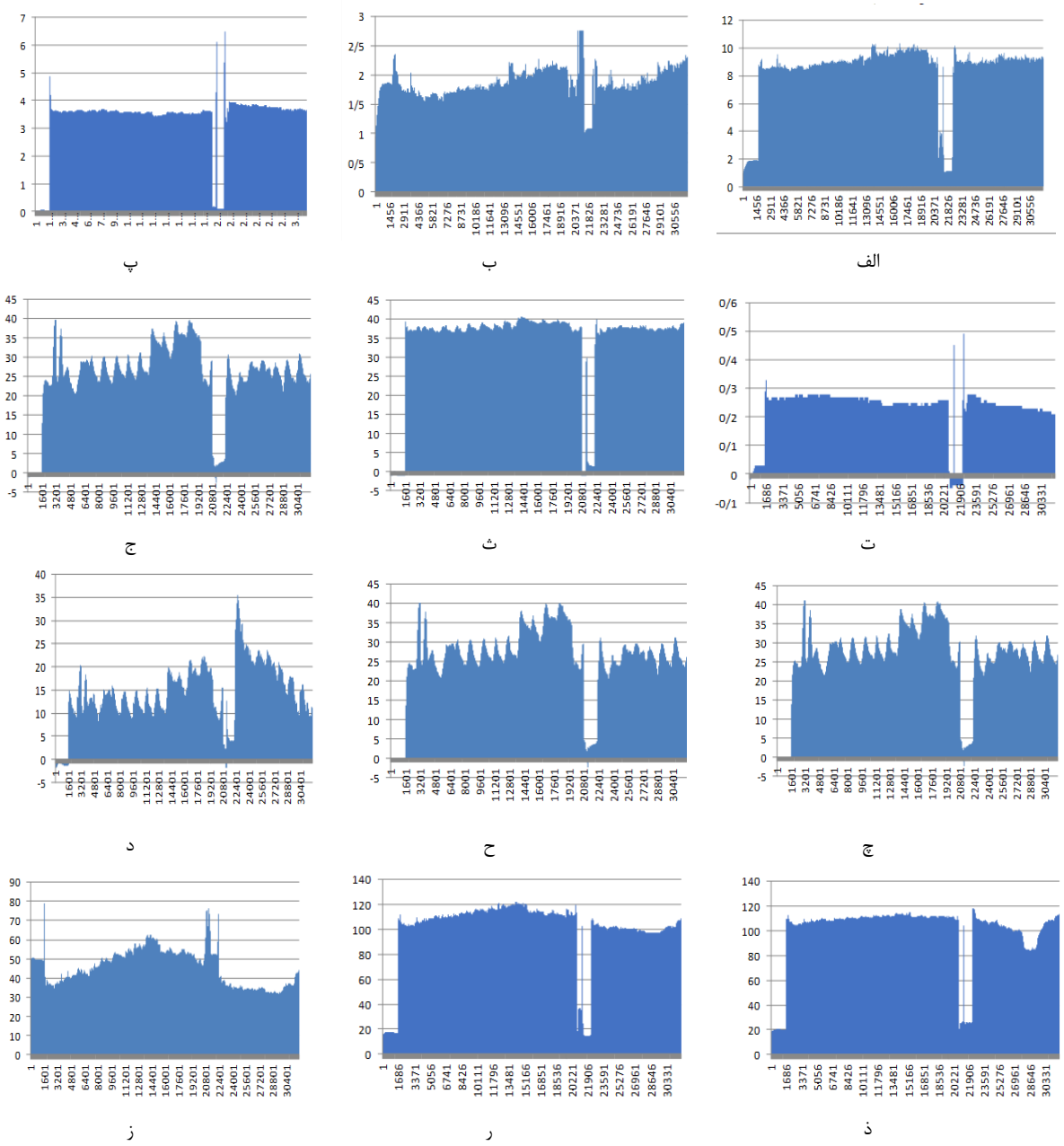
در این مرحله، تمامی داده‌های مورد نیاز این تحقیق شناسایی و گردآوری شد. در نتیجه معیارهای مؤثر بر خرابی کمپرسور شامل پارامترهای فرایندی گاز کمپرسور (مانند دمای ورودی و خروجی، فشار مکش و فشار تخلیه) و همچنین متغیرهای محیطی مورد بررسی قرار گرفتند. داده‌های این پژوهش شامل ۱۳۰۴۶۴۰ نمونه است که با هدف مطالعه و پیش‌بینی خرابی کمپرسور، از ۱۷ حسگر نصب‌شده بر روی یک کمپرسور در مجتمع پتروشیمی مهاباد گردآوری شده‌اند.

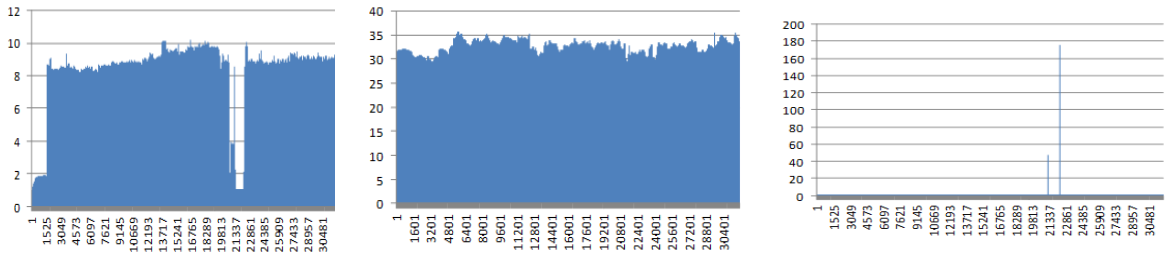
داده‌ها در بازه زمانی ۲۰ دسامبر ۲۰۲۱ تا ۲۰ می ۲۰۲۲، با فاصله زمانی شش بار در هر دقیقه (میانگین‌گیری شده) جمع‌آوری‌شده است. در جدول ۱، نام ۱۷ ویژگی اثرگذار بر خرابی کمپرسور به همراه خلاصه توصیفات آماری داده‌ها شامل شاخص‌های مرکزی و شاخص‌های پراکندگی گزارش شده است. شایان‌ذکر است که تمامی ویژگی‌های انتخاب‌شده از نوع پیوسته هستند. واحد اندازه‌گیری فشار، بار و واحد اندازه‌گیری دما سانتی‌گراد است. همچنین، برای دسته‌بندی، داده‌ها به ترتیب زمانی مرتب شده‌اند. سپس تحلیل هیستوگرام ویژگی‌های اندازه‌گیری شده در طول زمان در شکل ۲ ارائه‌شده است.

**جدول ۱ ویژگی‌های اثرگذار بر خرابی کمپرسور و شاخص‌های مرکزیت و پراکندگی آن‌ها**

| شماره | کد ویژگی | نام ویژگی | میانگین | میانگین هرس شده | میانه | مد | چارک Q1 | چارک Q2 | چارک Q3 | واریانس |
|---|---|---|---|---|---|---|---|---|---|---|
| ۱ | PI3114 | فشار خروجی مرحله دوم | ۸.۰۱۸ | ۸.۶۸۱ | ۸.۸۴ | ۸.۹۱ | ۸.۵۵ | ۸.۸۴ | ۹.۰۴ | ۶.۲۴۱ |
| ۲ | PI3112 | فشار مکش | ۱.۷۳۰ | ۱.۷۳۸ | ۱.۷ | ۱.۶۳ | ۱.۵۸ | ۱.۷ | ۱.۹۱ | ۰.۰۸۶ |
| ۳ | PDI3172 | اختلاف فشار مسیر روغن | ۳.۲۳۰ | ۳.۵۴۹ | ۳.۶۲ | ۳.۵۸ | ۳.۵۴ | ۳.۶۲ | ۳.۶۸ | ۱.۲۷۶ |
| ۴ | PDI3170 | اختلاف فشار در فیلتر | ۰.۲۰۲ | ۰.۲۱۸ | ۰.۲۲ | ۰.۲ | ۰.۲ | ۰.۲۲ | ۰.۲۴ | ۰.۰۰۶ |
| ۵ | TI3176 | دمای روغن | ۳۵.۲۴۴ | ۳۸.۳۱۴ | ۳۸.۴ | -۱.۱۳ | ۳۷.۲۸ | ۳۸.۴ | ۴۰.۴۲ | ۱۱۵.۳۵۳ |
| ۶ | TI3171 | دمای بیرینگ الکتروموتور۱ | ۲۸.۵۳۹ | ۲۹.۵۹۷ | ۳۰.۳۸ | ۳۳.۸ | ۲۳.۸ | ۳۰.۳۸ | ۳۵.۰۶ | ۹۶.۸۱۸ |
| ۷ | TI3172 | دمای سیم‌پیچ الکتروموتور۱ | ۳۰.۷۲۱ | ۳۱.۶۰۴ | ۳۰.۳ | ۲۳.۴۹ | ۲۴.۵۹ | ۳۰.۳ | ۳۹.۴۷ | ۱۳۵.۶۶۶ |
| ۸ | TI3173 | دمای سیم‌پیچ الکتروموتور۲ | ۳۱.۸۶۱ | ۳۲.۸۰۱ | ۳۱.۵۱ | ۲۴.۵۹ | ۲۵.۷۴ | ۳۱.۵۱ | ۴۰.۶۹ | ۱۴۰.۲۳۵ |

| شماره | کد ویژگی | نام ویژگی | میانگین | میانگین هرس شده | میانه | مد | چارک Q1 | چارک Q2 | چارک Q3 | واریانس |
|---|---|---|---|---|---|---|---|---|---|---|
| ۹ | TI3174 | دمای سیم‌پیچ الکتروموتور۳ | ۳۱.۱۴۴ | ۳۲.۰۲۱ | ۳۰.۷ | -۰.۴۶ | ۲۵.۰۴ | ۳۰.۷ | ۳۹.۸۲ | ۱۳۴.۶۰۶ |
| ۱۰ | TI3175 | دمای بیرینگ الکتروموتور۲ | ۱۹.۲۰۸ | ۱۸.۸۷۶ | ۱۸.۶۵ | ۹.۳۱ | ۱۲.۱۲ | ۱۸.۶۵ | ۲۵.۵۳ | ۸۲.۳۴۹ |
| ۱۱ | TI3111 | دمای خروجی مرحله اول | ۱۰۱.۲۰۶ | ۱۰۸.۷۵۶ | ۱۱۱.۷۳ | ۱۱۳.۰۵ | ۱۰۷.۲۸ | ۱۱۱.۷۳ | ۱۱۳.۸ | ۷۶۴.۷۴۵ |
| ۱۲ | TI3113 | دمای خروجی مرحله دوم | ۹۷.۵۰۵ | ۱۰۵.۰۷۸ | ۱۰۶.۵۹ | ۱۰۶.۱۹ | ۱۰۲.۴۳ | ۱۰۶.۵۹ | ۱۱۰.۱۷ | ۸۰۱.۵۹۴ |
| ۱۳ | TI3170 | دمای ورودی مرحله اول | ۴۳.۵۵۹ | ۴۳.۰۵۲ | ۴۱.۷۵ | ۳۹.۳۶ | ۳۸.۸۴ | ۴۱.۷۵ | ۴۹.۰۵ | ۴۹.۹۶۸ |
| ۱۴ | LI3112 | سطح روغن مخزن ۳۰۱۱ | ۱.۵۱۲ | ۱.۴۶۰ | ۱.۴۶ | ۱.۴۶ | ۱.۴۶ | ۱.۴۶ | ۱.۴۶ | ۷.۴۸۹ |
| ۱۵ | TI3177 | دمای خنک کاری بدنه | ۳۱.۹۶۰ | ۳۱.۹۶۰ | ۳۲.۰۳ | ۳۱.۸۹ | ۳۰.۸۸ | ۳۲.۰۳ | ۳۲.۹۵ | ۱.۹۳۸ |
| ۱۶ | PI3170 | فشار ورودی گاز مرحله دوم | ۷.۹۱۳۲ | ۸.۵۴۴ | ۸.۷ | ۸.۸ | ۸.۴۱ | ۸.۷ | ۸.۹ | ۵.۷۸۱ |
| ۱۷ | PI3116 | فشار خروجی کمپرسور | ۲۵.۱۳۹ | ۲۷.۶۹۶ | ۲۸.۲۲ | ۲۸.۸۳ | ۲۷.۷۷ | ۲۸.۲۲ | ۲۸.۷۷ | ۷۵.۲۶۳ |





ک گ و

**شکل ۲: تحلیل هیستوگرام ویژگی‌های اندازه‌گیری شده: الف: فشار خروجی مرحله دوم برحسب زمان، ب: فشار مکش برحسب زمان، پ: فشار مسیر روغن برحسب زمان، ت: اختلاف فشار در فیلتر برحسب زمان، ث: دمای روغن برحسب زمان، ج: دمای سیم‌پیچ الکتروموتور۱ برحسب زمان، چ: دمای سیم‌پیچ الکتروموتور۲ برحسب زمان، ح: دمای سیم‌پیچ الکتروموتور۳ برحسب زمان، د: دمای بیرینگ الکتروموتور۲ برحسب زمان، ذ: دمای خروجی مرحله اول برحسب زمان، ر: دمای خروجی مرحله دوم برحسب زمان، ز: دمای ورودی مرحله اول برحسب زمان، ک: سطح روغن مخزن ۳۰۱۱ برحسب زمان، گ: دمای خنک کاری بدنه برحسب زمان، و: فشار ورودی گاز مرحله دوم برحسب زمان، ی: فشار خروجی کمپرسور برحسب زمان، ه: دمای بیرینگ الکتروموتور۱ برحسب زمان.**

جدول ۲، نتایج تست همبستگی پیرسون بین ۱۷ ویژگی شناسایی‌شده نشان می‌دهد. ضریب همبستگی پیرسون بزرگ‌تر از ۰.۹، بیان‌گر رابطه همبستگی خطی به نسبت قوی بین ویژگی‌ها است. همان‌طور که در جدول ۲ نشان داده شده است، مقدار این ضریب برای ویژگی‌های TI3174، TI3173، TI3172، به‌صورت دوبه‌دو برابر ۱ است. بنابراین بین این ویژگی‌ها، رابطه قوی، مثبت و معناداری برقرار است. همچنین نتایج حاکی از رابطه قوی و مثبت بین فشار ورودی گاز مرحله دوم و فشار خروجی مرحله دوم با ضریب همبستگی ۰.۹۹۸ است. بین داده‌های دمای خروجی گاز مرحله دوم و دمای خروجی مرحله اول نیز رابطه قوی و مثبت با ضریب همبستگی ۰.۹۸۵ برقرار است.

### ۳-۲- خوشه‌بندی K-Means

خوشه‌بندی K-Means یکی از مهم‌ترین الگوریتم‌های داده‌کاوی است که داده‌ها را به تعدادی خوشه تقسیم می‌کند. هر نقطه داده به گروهی تخصیص داده می‌شود که به سایر نقاط داده آن گروه شباهت بیشتری داشته و با نقاط داده در گروه‌های دیگر متفاوت باشد. بنیان این گروه‌بندی بر مبنای اندازه‌گیری شباهت و عدم شباهت بین داده‌ها است. این روش، داده‌ها را بر اساس شباهت صفات از هم تفکیک می‌کند.

هدف اصلی روش خوشه‌بندی K-Means یافتن تعداد خوشه‌ها به‌گونه‌ای است که مجموع مربعات فواصل درون خوشه‌ای، حداقل شود. اضافه‌کردن خوشه‌ها تا جایی ادامه می‌یابد که افزودن یک خوشه دیگر، منجر به کاهش مجموع مربعات فواصل نگردد.

مقادیر گم‌شده با استفاده از میانه و مد جایگزین شده است. با تعریف ۱۷ حسگر انتخاب‌شده به‌عنوان ویژگی و انتخاب ۵ خوشه ( K=5 )، ۹۹٪ از داده‌ها در محاسبات استفاده شد و مجموع مربعات فواصل درون خوشه‌ای، حداقل شد. اضافه‌کردن خوشه دیگر، منجر به کاهش مجموع مربعات فواصل نشد. نتایج حاصل در جدول ۳ برای نشان دادن شباهت هر ویژگی با هر خوشه ارائه شده است. در محاسبات انجام شده از روش خوشه‌بندی K-Means ساده استفاده شده و ۷۶۸۸۲ سطر داده مورد بررسی قرار داده شده است. مراکز شروع اولیه خوشه‌ها به‌صورت تصادفی تعیین شدند و در ۱۹ تکرار، مجموع مربع خطای درون خوشه‌ای برابر با ۳۶۹۲.۸۸ به دست آمد.

از جمله الگوهای کشف‌شده می‌توان به این اشاره کرد که خوشه‌ها می‌توانند نشان‌دهنده رفتارهای مشابه در حسگرها باشند. به‌عنوان نمونه، خوشه‌های با دما و فشار بالا ممکن است نشان‌دهنده شرایط غیرعادی باشند، درحالی‌که خوشه‌های مربوط به شرایط نرمال می‌توانند نشان‌دهنده وضعیت عملیاتی عادی باشند. همچنین، اگر خوشه‌ای با داده‌های بسیار متفاوت وجود داشته

باشد می‌توان آن را به‌عنوان انحراف از عملکرد عادی تجهیز در نظر گرفت.

**جدول ۲ ضریب همبستگی پیرسون بین ویژگی‌های شناسایی‌شده**

| | PI3114 | PI3112 | PDI3172 | PDI3170 | TI3176 | TI3171 | TI3172 | TI3173 | TI3174 | TI3175 | TI3111 | TI3113 | TI3170 | LI3112 | TI3177 | PI3170 | PI3116 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PI3114 | 1 | 0.572 | 0.967 | 0.914 | 0.949 | 0.728 | 0.696 | 0.703 | 0.696 | 0.435 | 0.955 | 0.969 | -0.449 | 0.005 | 0.059 | 0.998 | 0.983 |
| PI3112 | 0.57 | 1 | 0.46 | 0.481 | 0.42 | 0.195 | 0.114 | 0.121 | 0.113 | -0.118 | 0.436 | 0.462 | -0.118 | -0.010 | 0.144 | 0.574 | 0.474 |
| PDI3172 | 0.967 | 0.46 | 1 | 0.933 | 0.958 | 0.748 | 0.688 | 0.695 | 0.688 | 0.448 | 0.960 | 0.963 | -0.527 | -0.001 | 0.027 | 0.968 | 0.985 |
| PDI3170 | 0.914 | 0.481 | 0.933 | 1 | 0.876 | 0.573 | 0.500 | 0.508 | 0.500 | 0.247 | 0.881 | 0.922 | -0.393 | -0.003 | 0.136 | 0.918 | 0.914 |
| TI3176 | 0.949 | 0.42 | 0.958 | 0.876 | 1 | 0.823 | 0.798 | 0.804 | 0.798 | 0.569 | 0.971 | 0.966 | -0.452 | 0.001 | 0.039 | 0.950 | 0.964 |
| TI3171 | 0.728 | 0.195 | 0.748 | 0.573 | 0.823 | 1 | 0.885 | 0.887 | 0.885 | 0.825 | 0.776 | 0.731 | -0.529 | -0.007 | -0.105 | 0.727 | 0.769 |
| TI3172 | 0.696 | 0.114 | 0.688 | 0.500 | 0.798 | 0.885 | 1 | 1.000 | 1.000 | 0.870 | 0.769 | 0.724 | -0.310 | -0.027 | -0.061 | 0.695 | 0.724 |
| TI3173 | 0.703 | 0.121 | 0.695 | 0.508 | 0.804 | 0.887 | 1.000 | 1 | 1.000 | 0.868 | 0.775 | 0.731 | -0.313 | -0.027 | -0.059 | 0.702 | 0.731 |
| TI3174 | 0.696 | 0.113 | 0.688 | 0.500 | 0.798 | 0.885 | 1.000 | 1.000 | 1 | 0.871 | 0.769 | 0.723 | -0.309 | -0.026 | -0.061 | 0.694 | 0.723 |
| TI3175 | 0.435 | -0.118 | 0.448 | 0.247 | 0.569 | 0.825 | 0.870 | 0.868 | 0.871 | 1 | 0.518 | 0.452 | -0.345 | 0.008 | -0.121 | 0.433 | 0.483 |
| TI3111 | 0.955 | 0.436 | 0.960 | 0.881 | 0.971 | 0.776 | 0.769 | 0.775 | 0.769 | 0.518 | 1 | 0.985 | -0.427 | 0.010 | 0.048 | 0.956 | 0.976 |
| TI3113 | 0.969 | 0.462 | 0.963 | 0.922 | 0.966 | 0.731 | 0.724 | 0.731 | 0.723 | 0.452 | 0.985 | 1 | -0.368 | 0.006 | 0.095 | 0.970 | 0.978 |
| TI3170 | -0.449 | -0.181 | -0.527 | -0.393 | -0.452 | -0.529 | -0.310 | -0.313 | -0.309 | -0.345 | -0.427 | -0.368 | 1 | -0.016 | 0.368 | -0.444 | -0.523 |
| LI3112 | 0.005 | -0.010 | -0.001 | -0.003 | 0.001 | -0.007 | -0.027 | -0.027 | -0.026 | 0.008 | 0.010 | 0.006 | -0.016 | 1 | -0.005 | 0.006 | 0.008 |
| TI3177 | 0.059 | 0.144 | 0.027 | 0.136 | 0.039 | -0.105 | -0.061 | -0.059 | -0.061 | -0.121 | 0.048 | 0.095 | 0.368 | -0.005 | 1 | 0.065 | 0.025 |
| PI3170 | 0.998 | 0.574 | 0.968 | 0.918 | 0.950 | 0.727 | 0.695 | 0.702 | 0.694 | 0.433 | 0.956 | 0.970 | -0.444 | 0.006 | 0.065 | 1 | 0.983 |
| PI3116 | 0.983 | 0.474 | 0.985 | 0.914 | 0.964 | 0.769 | 0.724 | 0.731 | 0.723 | 0.483 | 0.976 | 0.978 | -0.523 | 0.008 | 0.025 | 0.983 | 1 |

در جدول ۳، فاصله بین هر ویژگی با مرکز خوشه به‌دست آمده است. هرچه فاصله کمتر باشد، تأثیر ویژگی بر آن خوشه بیشتر است و هرچه فاصله بیشتر باشد، تأثیر آن کمتر خواهد بود. بنابراین، با توجه به اثر هر ویژگی بر هر خوشه، خوشه‌بندی ۱۷ ویژگی انتخاب شده در جدول ۴ نمایش داده شده است.

**جدول ۳ محاسبات شباهت خوشه‌بندی داده‌ها**

| کل | 5 | 4 | 3 | 2 | 1 | ویژگی‌ها |
|---|---|---|---|---|---|---|
| -0.0023 | 0.2987 | -2.6063 | 0.3306 | 0.4419 | -2.8092 | PI3114 |
| -0.0044 | -0.4685 | -0.5655 | 0.4368 | 0.4057 | -1.8447 | PI3112 |
| -0.0023 | 0.2987 | -2.6063 | 0.3306 | 0.4419 | -2.8092 | PI3114 |
| -0.0044 | -0.4685 | -0.5655 | 0.4368 | 0.4057 | -1.8447 | PI3112 |
| -0.0022 | 0.3629 | -2.789 | 0.3719 | 0.3234 | -2.6947 | PDI3172 |
| -0.002 | 0.0032 | -2.7541 | 0.602 | 0.2713 | -2.4634 | PDI3170 |
| -0.0025 | 0.5008 | -3.2326 | 0.212 | 0.3952 | -2.3405 | TI3176 |
| -0.0022 | 0.9566 | -2.7673 | -0.2223 | 0.3184 | -1.6747 | TI3171 |
| -0.002 | 1.1826 | -2.5547 | -0.4408 | 0.3546 | -1.5895 | TI3172 |
| -0.0021 | 1.1723 | -2.5742 | -0.431 | 0.3591 | -1.6101 | TI3173 |
| -0.0021 | 1.1796 | -2.5647 | -0.4415 | 0.3575 | -1.5814 | TI3174 |
| -0.0016 | 1.311 | -1.976 | -0.5973 | 0.151 | -0.8253 | TI3175 |
| -0.0028 | 0.5022 | -2.8685 | 0.2336 | 0.3776 | -2.6222 | TI3111 |
| -0.0027 | 0.3845 | -2.8261 | 0.3223 | 0.3707 | -2.6708 | TI3113 |
| -0.0047 | -0.3424 | 1.1039 | -0.2882 | 0.093 | 1.595 | TI3170 |
| -0.0001 | -0.019 | -0.0283 | 0.0283 | -0.0133 | -0.016 | LI3112 |
| -0.0176 | -0.1994 | -0.6506 | 0.4146 | -0.3977 | 0.1669 | TI3177 |
| -0.0025 | 0.2945 | -2.6407 | 0.329 | 0.4486 | -2.7924 | PI3170 |
| -0.0023 | 0.4197 | -2.721 | 0.3286 | 0.3454 | -2.7918 | PI3116 |
| 76882 | 19143 | 3392 | 28533 | 20309 | 5505 | تعداد داده |
| 99% | 25% | 4% | 37% | 26% | 7% | درصد داده |

تحلیل خوشه‌ها نشان می‌دهد که خوشه ۳، خوشه ۲ و خوشه ۵، به ترتیب با شامل بودن ٪۳۷، ٪۲۶ و ٪۲۵ از داده‌ها، حاوی اکثریت داده‌ها هستند و الگوهای غالب، عملکرد نرمال یا شرایط معمولی کمپرسور را نشان می‌دهد. خوشه ۴ با داشتن ٪۴ داده‌ها، کوچک‌ترین خوشه است که می‌تواند نشان‌دهنده رفتارهای غیرمعمول یا نقاط انحرافی، یا داده‌های خاص (مانند ناهنجاری‌ها یا داده‌های خارج از محدوده) باشد. بنابراین خوشه‌بندی به خوبی داده‌ها را در ۵ خوشه تقسیم کرده است، که خوشه‌های بزرگ نشان دهنده شرایط عادی و خوشه‌های کوچک نمایانگر نقاط انحرافی یا شرایط خاص تجهیزات هستند.

**جدول ۴ خوشه‌بندی ویژگی‌های انتخاب شده**

| شماره خوشه | C1 | C2 | C3 | C4 | C5 |
|---|---|---|---|---|---|
| ویژگی‌ها | TI3177 | PI3112<br>PDI3172<br>TI3173<br>TI3174<br>TI3175<br>TI3170<br>LI3112 | TI3176<br>TI3171<br>TI3172<br>TI3111<br>TI3113<br>PI3116 | | PI3114<br>PDI3170<br>PI3170 |

### ۳-۳- درخت تصمیم

امروزه کارخانه‌ها و تاسیسات با چالش‌ها و مشکلات زیادی برای افزایش بهره‌وری و کاهش هزینه‌ها مواجه می‌شوند و استراتژی‌های نگهداری و تعمیرات نقش بسزایی در رسیدن به این اهداف دارند. در سال‌های اخیر تلاش‌های فراوانی شده است تا با بهره‌گیری از سیستم‌های فناوری اطلاعات، روش‌های متنوع نگهداری و تعمیرات به‌صورت مؤثر پیاده‌سازی و اجرا شوند. نگهداری و تعمیرات پیش‌گویانه یکی از استراتژی‌های نگهداری و تعمیرات است که بر مبنای آن، در بازه‌های زمانی معیّن، تعدادی از پارامترهای تجهیزات اندازه‌گیری می‌شود و بر اساس این داده‌ها، زمان رخداد خرابی احتمالی آینده پیش‌بینی شده و برای تعمیر و یا تعویض قطعات و تجهیزات تصمیم‌گیری می‌شود. درخت تصمیم از جمله ابزارهای کارآمد برای عملیاتی‌سازی نگهداری و تعمیرات پیش‌گویانه است.

درخت تصمیم ابزاری برای تصمیم‌گیری است که اهداف، معیارها، گزینه‌ها و میزان سود و سرمایه‌گذاری را به تصمیم‌گیرنده نشان دهد. درخت تصمیم درختی است که در آن نمونه‌ها را به نحوی دسته‌بندی می‌کند که از ریشه به سمت پایین رشد می‌کنند و در نهایت، به گره‌های برگ می‌رسد [۳۰]. اولین گره آن نشان‌دهنده هدف تصمیم‌گیری است. یک درخت تصمیم، دربردارنده، انتخاب‌ها و پیشامدهای محتمل یا نتیجه این انتخاب‌ها است، که در نهایت نتایج، زنجیره تصمیمات و پیشامدهای احتمالی را نشان می‌دهد و تصمیم‌گیرندگان قادر هستند بر مبنای این نتایج، تصمیمات خود را بهینه کنند و از این نتایج برای پیش‌بینی وقایع آینده بهره ببرند. برای ایجاد یک درخت تصمیم باید موارد زیر مد نظر قرار گیرد:

شناسایی نقاط تصمیم و انتخاب‌های ممکن در هر سطح

شناسایی احتمالات و بازه پیشامدها در هر سطح

تخمین مقادیر عددی برای تحلیل احتمال نتایج عملکرد، هزینه‌ها و سود حاصل

تحلیل ارزش انتخاب‌ها برای انتخاب یک مسیر

شایان‌ذکر است که این درخت، نتیجه قطعی و نهایی را به مدیران ارائه نمی‌دهد. بلکه مسیرهای بهینه را به آن‌ها پیشنهاد می‌دهد تا آن‌ها متناسب با سیاست‌های مورد نظر خود بهترین مسیر را انتخاب کنند و تحلیل دقیق‌تری از شرایط و پیش‌بینی بهتری از آینده داشته باشند. درخت تصمیم یکی از ابزارهای رایج در داده‌کاوی می‌باشد که برای دسته‌بندی و پیش‌بینی و تحلیل سیستماتیک در شرایط عدم قطعیت استفاده می‌شود. درخت‌های تصمیم می‌توانند با داده‌های پیچیده سروکار داشته باشند. امّا این موضوع به معنای دشوار بودن درک عملکرد الگوریتم نیست. یک درخت تصمیم از چهار بخش کلیدی تشکیل می‌شود:

گره ریشه که نقطه شروع فرآیند تصمیم‌گیری را نشان می‌دهد.

گره‌های داخلی که حاوی گره فرزند هستند و در مراحل میانی فرآیند تصمیم‌گیری قرار دارند. گره‌های داخلی شامل گره‌های تصمیم، گره‌های شانس، و گره‌های پایانی هستند.

شاخه‌ها که خطوط اتصال گره‌ها هستند و نتایج احتمالی یک تصمیم را نشان می‌دهند.

گره‌های برگ که فاقد گره فرزند بوده و نتیجه نهایی فرآیند تصمیم‌گیری را نشان می‌دهد.

از جمله درخت‌های تصمیم، درخت تصادفی است. درخت تصادفی یک طبقه‌بندی نظارت‌شده است که در دسته الگوریتم‌های یادگیری تجمعی قرار می‌گیرد و از ترکیب چندین یادگیرنده منفرد تشکیل می‌شود. درخت‌های تصادفی در واقع ترکیبی از دو الگوریتم شناخته‌شده در یادگیری ماشین هستند: درخت‌های مدل منفرد و ایده جنگل تصادفی. جنگل تصادفی یک روش یادگیری ترکیبی برای طبقه‌بندی داده‌ها است که از تعداد زیادی درخت تصمیم در زمان آموزش تشکیل شده‌است و خروجی این الگوریتم برای طبقه‌بندی مجموعه‌ای از داده‌ها مورداستفاده قرار می‌گیرد. درخت‌های مدل، درخت‌های تصمیمی هستند که در هر برگ آن‌ها، یک مدل خطی قرار دارد که برای زیرفضای محلی مربوط به همان برگ بهینه‌شده است. الگوریتم درخت تصادفی قابلیت انجام وظایف طبقه‌بندی و رگرسیون را دارد و یک درخت تصمیم است که از طریق یک چرخه تصادفی توسعه می‌یابد.

در یک درخت تصمیم استاندارد، هر گره با استفاده از بهترین تقسیم‌بندی در میان کل مجموعه داده‌ها به زیرگره‌ها تفکیک می‌شود. برای پیاده‌سازی الگوریتم RT، داده‌های ورودی به دسته‌بندی درختی تخصیص داده می‌شوند، به‌طوری‌که هر درخت به‌صورت مستقل عمل طبقه‌بندی را بر روی داده‌های ورودی انجام می‌دهد. در نهایت، کلاسی که بیشترین فراوانی را در میان خروجی درخت‌ها داشته باشد، به‌عنوان خروجی نهایی سیستم انتخاب می‌شود. الگوریتم یادگیری تجمعی RT در مقایسه با سایر الگوریتم‌های تجمعی، از مزایایی همچون توان تعمیم‌پذیری بالا، سازگاری و دقت بیشتر برخوردار است. علاوه بر این، این الگوریتم قادر است مسائل غیرخطی و بیش‌برازش را با دقت بالا حل کند. الگوریتم درخت تصادفی همچنین از سطح بالایی از انعطاف‌پذیری و سرعت بالای آموزش برخوردار است [۵]. جنگل‌های تصادفی عملکرد درخت‌های تصمیم منفرد را به‌طور چشمگیری بهبود می‌بخشند. این بهبود از طریق دو روش تصادفی‌سازی حاصل می‌شود:

در مرحله آموزش، برای هر درخت به‌صورت با جایگزینی نمونه‌گیری می‌شود.

در زمان رشد درخت، به‌جای محاسبه بهترین تقسیم برای هر گره میان همه ویژگی‌ها، تنها یک زیرمجموعه تصادفی از ویژگی‌ها در هر گره در نظر گرفته می‌شود و بهترین تقسیم در همان زیرمجموعه انتخاب می‌شود [۳۱].

برای اجرای درخت تصمیم در این پژوهش، از درخت تصادفی استفاده شده است. ٪۷۰ از داده‌ها برای آموزش و باقیمانده داده‌ها برای آزمون استفاده شده است. درخت تصمیم حاصل در شکل ۳ نمایش داده شده است.

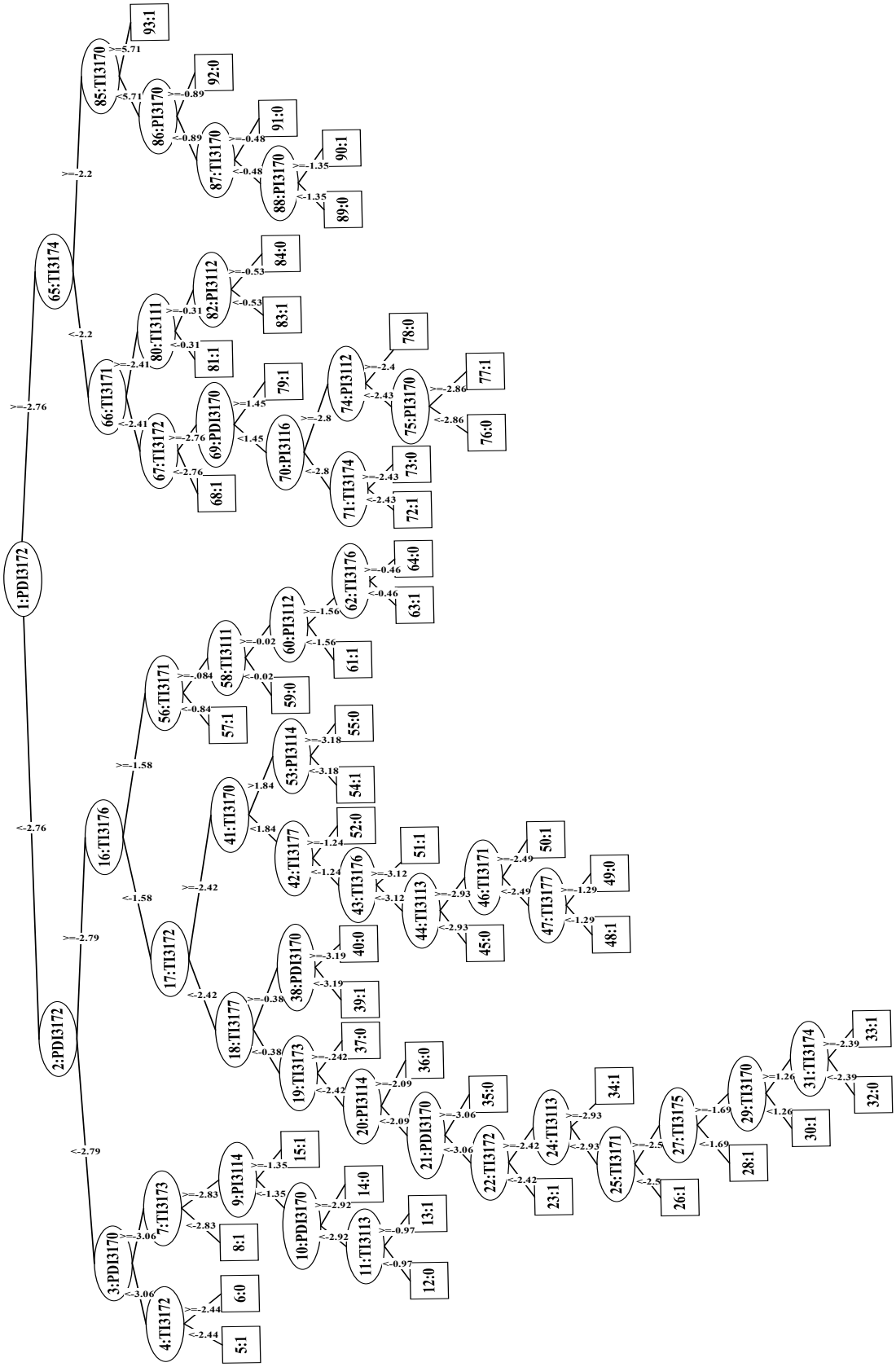
نتایج ارزیابی عملکرد درخت تصمیم با توجه به داده‌های آزمون در جدول ۵ گزارش‌شده است. کل سطرهای داده موردبررسی برای ترسیم درخت تصمیم، برابر با ۷۶۸۸۲ بوده که ۲۳۰۶۵ مورد آن به آزمون اختصاص داده‌شده است. مدل با داده‌های آموزش، آموزش داده شده و سپس با داده‌های آزمون اعتبارسنجی شد. مقدار ضریب همبستگی نزدیک به ۱ نشان‌گر ارتباط بسیار قوی بین مقادیر واقعی و پیش‌بینی‌شده توسط مدل است. همچنین مقادیر اندک MAE، RMSE، RAE، RRSE نشان‌گر اختلاف ناچیز بین مقادیر پیش‌بینی‌شده و مقادیر واقعی و در نتیجه کارایی و دقت بالای مدل در پیش‌بینی است.

**جدول ۵ ارزیابی عملکرد درخت تصمیم**

| | |
|---|---|
| تعداد کل داده‌ها | 76882 |
| تعداد داده‌های آزمون | 23065 |
| ضریب همبستگی[1] (C.O) | 0.9758 |
| میانگین خطای مطلق[2] (MAE) | 0.0006 |
| ریشه میانگین مربعات خطا[3] (RMSE) | 0.0246 |
| خطای مطلق نسبی[4] (RAE) | 2.4327% |
| ریشه خطای مطلق نسبی[5] (RRSE) | 21.9622 % |

[1] Correlation Coefficient
[2] Mean Absolute Error
[3] Root Mean Squared Error
[4] Relative Absolute Error
[5] Root Relative Squared Error

در درخت تصمیم ارائه شده، شاخه‌ای که به برگ «صفر» ختم می‌شود، نشان‌دهنده در حال بهره‌برداری بودن کمپرسور است. ویژگی‌ها و آستانه‌های اندازه‌گیری که باعث حفظ وضعیت بهره‌برداری می‌شوند، شرایطی را مشخص می‌کنند که در آن کمپرسور پایدار باقی می‌ماند. از سوی دیگر شاخه‌ای که به برگ «یک» ختم می‌شود، نشان‌گر توقف یا از کار افتادگی کمپرسور است. مسیر منتهی به برگ «یک» را مسیر بحرانی می‌نامیم. تحلیل طولانی‌ترین مسیر منتهی به «یک» نشان می‌دهد که چه زنجیره‌ای از عوامل و مقادیر باعث ایجاد خطا یا توقف کمپرسور می‌شوند.




**شکل ۳ درخت تصمیم**

برای شناسایی طولانی‌ترین شاخه منتهی به «صفر»، از گره اولیه شروع کرده و به سمت برگ‌های منتهی به «صفر» حرکت می‌کنیم. در این حالت آستانه ویژگی‌های گره‌های مسیر که منجر به حفظ وضعیت بهره‌برداری می‌شوند، به‌صورت زیر تحلیل می‌شوند:

گره ۱: مقدار ویژگی مربوط به حسگر PDI3172 کمتر از ۲.۷۶ باشد (PDI3172 < 2.76)

گره ۲: مقدار ویژگی مربوط به حسگرPDI3172 کمتر از ۲.۷۹ باشد (PDI3172 < 2.79).

گره ۳: مقدار ویژگی مربوط به حسگر TI3170 بیشتر مساوی ۳.۰۶ باشد (TI3170 ≥ 3.06)

گره ۴: مقدار ویژگی مربوط به حسگر TI3172 کمتر از ۲.۴۲ باشد (TI3172 < 2.42).

گره ۵: مقدار ویژگی مربوط به حسگر TI3113 کمتر از ۲.۳۹ باشد (TI3113 < 2.39)

برگ «صفر»: کمپرسور در وضعیت بهره‌برداری قرار دارد.

برای شناسایی طولانی‌ترین شاخه منتهی به «یک»، نیز از گره اولیه شروع کرده و به سمت برگ‌های منتهی به «یک» حرکت می‌کنیم. در این حالت آستانه ویژگی‌های گره‌های مسیر که منجر به توقف یا از کار افتادگی کمپرسور می‌شوند، تحلیل می‌شوند:

گره ۱: مقدار ویژگی مربوط به حسگر PDI3172 بیشتر یا مساوی ۲.۷۶ باشد (PDI3172 ≥2.76).

گره ۶۵: مقدار ویژگی مربوط به حسگر TI3174 بیشتر یا مساوی ۲.۲۲ باشد (TI3174 ≥ 2.22).

گره ۸۵: مقدار ویژگی مربوط به حسگر TI3170 بیشتر یا مساوی ۵.۷۱ باشد (TI3170 ≥ 5.71).

گره ۸۶: مقدار ویژگی مربوط به حسگر TI3170 کمتر از ۰.۸۹- باشد (TI3170 < -0.89).

برگ «یک»: کمپرسور در وضعیت توقف یا ازکارافتادگی قرار دارد.

بنابراین برای جلوگیری از توقف باید مقادیر حسگرها به‌دقت پایش شوند تا از محدوده ایمن خارج نشوند. همچنین، مسیرهای منتهی به توقف، مسیرهای بحرانی بوده که شرایط غیرمعمول در حسگرها، به‌ویژه تغییرات ناگهانی، باید سریع شناسایی و رفع شوند.

تحلیل‌ها نشان می‌دهد که شرایط حسگرها در مسیر بالا که منتهی به برگ «یک» می‌شود، بحرانی‌ترین مسیر بوده و بیشترین تأثیر بر توقف کمپرسور را دارد. ویژگی‌های PDI3172 و TI3170 نماینده فشار و دمای اصلی سیستم هستند و تغییرات شدید در آن‌ها، عملکرد کمپرسور را با اختلال مواجه می‌کند.

درخت تصمیم با ۹۳ گره ایجادشده که نشان می‌دهد مدل توانسته است با پیچیدگی متوسط، داده‌ها را به‌خوبی دسته‌بندی کند. معیارهای متنوعی از داده‌های حسگرها مورد استفاده قرار گرفته‌اند. ویژگی‌های TI3170 و PDI3172 و PI3114 و TI3111 نقش مهمی در تعیین مسیر درخت دارند. برای اطمینان از تعمیم‌پذیری مدل، می‌توان از داده‌های جدید برای ارزیابی استفاده کرد یا اعتبارسنجی متقابل انجام داد.

تجزیه‌وتحلیل درخت تصمیم شامل سنجش جوانب مثبت و منفی مسیرها و انتخاب بهترین گزینه از ساختار درخت است.

### ۴-۳- انتخاب ویژگی‌های مهم

در این مرحله به انتخاب ویژگی‌های مهم از مجموعه داده‌ها پرداخته شده است. هدف، کاهش تعداد ویژگی‌ها، بهبود کارایی الگوریتم‌های یادگیری ماشین و کاهش پیچیدگی مدل است. این روش ویژگی‌هایی را که ارتباط بیشتری با متغیر هدف دارند را شناسایی می‌کند و ویژگی‌هایی که اطلاعات کمی ارائه می‌دهند یا وابستگی زیادی به سایر ویژگی‌ها دارند، حذف می‌کند. با کاهش تعداد ویژگی‌ها، داده‌ها ساده‌تر و پردازش سریع‌تر می‌شود، چراکه حذف نویز و داده‌های غیرضروری می‌تواند دقت مدل را

افزایش دهد.

برای انتخاب ویژگی‌ها از روش cfsSubsetEval در نرم‌افزار Weka استفاده شد. این روش ویژگی‌هایی را انتخاب می‌کند که ارتباط بالایی با متغیر هدف داشته و وابستگی کمی به یکدیگر دارند. نتایج نشان داد که سه حسگر PI3114، PDI3172، TI3111 به‌عنوان مهم‌ترین ویژگی‌ها انتخاب شده‌اند، که بیشترین تأثیر را بر متغیر هدف دارند و می‌توانند در بهبود دقت مدل و کاهش پیچیدگی مفید باشند.

## ۴- تحلیل نتایج

با تحلیل مسیرهای بحرانی در درخت تصادفی شکل ۳ و خوشه‌های ایجاد شده ، حسگرهای بحرانی واقع‌شده در مسیرهای منجر به توقف، که حائز اهمیت و توجه بیشتر هستند مشخص شدند. این موارد به شرح زیر است:

### الف. حسگر بحرانی اصلی TI3170

دلایل اهمیت این حسگر عبارت است از:

- این حسگر در بحرانی‌ترین شاخه قرار دارد و مقدار آن در تصمیم‌گیری تأثیر بسزایی دارد.
- در مسیرهای منتهی به توقف، مقدار آن در چندین مرحله به آستانه‌های بحرانی ($5.71 \geq$ و $-0.89 >$) رسیده است.
- تغییرات ناگهانی و شدید در مقدار این حسگر (از مقدار مثبت بزرگ به منفی) نشان‌دهنده شرایط غیرعادی است که مستقیماً به توقف منجر می‌شود.

در رابطه با ارتباط این حسگر با عملکرد کمپرسور، این حسگر دمایی را اندازه‌گیری می‌کند که تغییرات آن با عملکرد کمپرسور ارتباط مستقیم دارد. همچنین، افزایش بیش‌ازحد یا کاهش ناگهانی در مقدار آن می‌تواند نشان‌دهنده مشکلاتی مانند خرابی مکانیکی، افزایش دما، یا تغییرات فشار باشد.

پس از فشرده‌سازی در مرحله اول، گاز داغ به یک خنک‌کننده بین مرحله‌ای فرستاده می‌شود تا دمای آن کاهش یابد. سپس وارد مرحله دوم می‌شود. اگر این خنک‌کننده عملکرد مناسبی نداشته باشد یا دچار گرفتگی و نشت شود، گاز به‌طور کامل خنک نمی‌شود و دمای گاز ورودی به مرحله دوم افزایش می‌یابد. این افزایش دما باعث افزایش کار مورد نیاز برای فشرده‌سازی در مرحله دوم و کاهش بازده کلی می‌شود که ممکن است به دلیل اختلال در قطعات زیر باشد:

- مبدل حرارتی بین مرحله‌ای[1] که رسوب و گرفتگی لوله‌ها و نشتی داخلی آن (اختلال سیال خنک‌کننده و گاز) منجر به کاهش دبی آن می‌شود.
- شیرها و کنترل‌کننده‌های جریان که عملکرد ناقص آن‌ها باعث عدم تنظیم صحیح دبی خنک‌کننده است.
- مشکل در مرحله اول کمپرسور که ناشی از شیرها و پیستون‌ها است. اگر پیستون یا رینگ آسیب‌دیده باشند، گاز فشرده‌شده در مرحله اول بیشتر از حد معمول داغ شده و تلفات مکانیکی بالا و اصطکاک بیش‌ازحد، دمای گاز را افزایش می‌دهد.

### ب. حسگر بحرانی دوم PDI3172

دلایل اهمیت این حسگر عبارت است از:

- این حسگر در چندین مسیر مهم، از جمله مسیرهای بحرانی منتهی به توقف، ظاهر شده است و درصورتی‌که مقدار آن بیشتر از ۲.۷۶ باشد، سیستم به سمت وضعیت بحرانی حرکت می‌کند.
- این حسگر نماینده فشار یا جریانی است که شرایط عملکردی کمپرسور را پایش می‌کند.

[1] Intercooler

در رابطه با ارتباط این حسگر با عملکرد کمپرسور، فشار غیرعادی می‌تواند به مشکلاتی مانند انسداد یا خرابی سیستم منجر شود. پایش دقیق این حسگر برای پیشگیری از افزایش فشار ضروری است. اختلاف فشار روغن معمولاً نشان‌دهنده سلامت سیستم روغن‌کاری است. این اختلاف فشار بین فشار خروجی پمپ روغن و فشار میل لنگ یا کارتر اندازه‌گیری می‌شود و تغییرات زیاد آن می‌تواند نشانه‌ای از اختلالات زیر باشد:

- خرابی یا ضعف پمپ روغن (خرابی یا ورود هوا یا گاز به داخل روغن)
- انسداد در مسیر روغن‌کاری و گرفتگی فیلتر روغن
- نشتی داخلی در سیستم روغن‌کاری
- تغییرات دمایی زیاد

**ج. حسگر بحرانی سوم TI3174**

دلایل اهمیت این حسگر عبارت است از:

- این حسگر در مسیرهای بحرانی نقش دارد و درصورتی‌که مقدار آن بیشتر از ۲.۲۲ شود، شرایط سیستم به سمت توقف پیش می‌رود.
- در تصمیم‌گیری‌های چندمرحله‌ای مربوط به توقف، تغییرات این حسگر نیز تأثیرگذار است.

در رابطه با ارتباط این حسگر با عملکرد کمپرسور، این حسگر دمایی را اندازه‌گیری می‌کند که افزایش آن می‌تواند نشان‌دهنده بار اضافی یا نقص در سیستم خنک‌کننده باشد. از جمله دلایل نقص این سیستم‌ها عبارتند از:

- دلایل مکانیکی که ممکن است از بار بیش‌ازحد و یا روغن‌کاری نامناسب باشد.
- دلایل الکتریکی که عمدتاً به افت ولتاژ، نوسان برق، اختلاف زیاد بین ولتاژ یا جریان فازها، ضعف عایق‌بندی سیم‌پیچ‌ها، وضعیت کابل‌ها و خرابی بلبرینگ‌های موتور مربوط می‌شود.

با توجه به شناسایی حسگرهای بحرانی، موارد زیر جهت پیشگیری از توقف کمپرسور پیشنهاد می‌گردد:

- ✓ پایش مداوم: استفاده از سیستم‌های پایش پیشرفته به‌منظور پایش ویژگی TI3170 و شناسایی تغییرات ناگهانی.
- ✓ تنظیم آستانه‌های هشدار: ایجاد سیستم هشدار سریع برای شناسایی تغییرات ناگهانی، به‌ویژه برای حسگرهای PDI3172 و TI3174، تا در صورت نزدیک شدن به آستانه‌های بحرانی، اقدامات اصلاحی انجام شود.

تعمیر و نگهداری پیش‌گیرانه: بررسی دوره‌ای شرایط محیطی و فنی سیستم خنک‌کننده و کنترل فشار برای جلوگیری از بروز شرایط بحرانی.

## ۵- نتیجه‌گیری و پیشنهادهای تحقیقات آتی

در این پژوهش، سعی شده است تا با کمک روش‌های داده‌کاوی و نرم‌افزار Weka، یک مدل پیش‌بینی رخداد خرابی برای تجهیزات کمپرسورهای ۳۰۰۰ مجتمع پتروشیمی مهاباد طراحی و تحلیل گردد. در حال حاضر، اتکا به روش‌های تجربی، دستی و بصری به دلیل ناتوانی در ارائه هشدارها و پیش‌بینی‌های به‌موقع، نمی‌تواند پشتیبان مؤثری برای واحدهای بهره‌برداری نگهداری و تعمیرات باشند. بنابراین بکارگیری روش‌های نوین داده‌کاوی و هوش مصنوعی می‌تواند کمک قابل‌توجهی در پیشگیری از خرابی تجهیزات، برنامه‌ریزی برای رفع خرابی، و تدارک به‌موقع قطعات مورد نیاز فراهم آورد. این امر منجر به تولید مستمر در مجموعه، جلوگیری از توقف تولید و خرابی‌های ناگهانی و هزینه‌بر خواهد شد. استفاده از این تکنیک‌ها، همچنین زمان تعمیرات را کاهش داده، فرآیند نگهداری را بهبود بخشیده، سودآوری را افزایش داده و دقت پیش‌بینی را ارتقاء می‌دهد.

به همین منظور ابتدا مجموعه‌ای از داده‌های حسگرها مشخص، در یک بازه زمانی معین جمع‌آوری گردید. سپس داده‌ها با بهره‌گیری از الگوریتم‌های ضریب همبستگی، خوشه‌بندی، درخت تصمیم، و انتخاب ویژگی در محیط Weka تجزیه‌وتحلیل شدند. نتایج این تحلیل‌ها از جمله مسیرهای بحرانی در درخت تصمیم، خوشه‌های شناسایی‌شده و ویژگی‌های کلیدی، منجر به تعیین حسگرهای بحرانی گردید. پس‌ازآن دلایل اهمیت این حسگرهای بحرانی و علل اختلال در آن‌ها تحلیل شد و پیشنهاداتی جهت پیشگیری از توقف کمپرسور ارائه شد.

با وجود دستاوردهای این مطالعه، همچنان زمینه برای بهبود و توسعه وجود دارد. تحقیقات آتی می‌تواند شامل به‌کارگیری الگوریتم‌های پیشرفته‌تر یادگیری ماشین، ارزیابی دقت مدل‌ها در شرایط واقعی بهره‌برداری، اولویت‌بندی نقاط بحرانی و تعمیراتی و بررسی تأثیر استفاده از سیستم‌های پیش‌بینی مبتنی بر داده بر کاهش هزینه‌های نگهداری و توقفات غیرمترقبه باشد. همچنین، توسعه یک سامانه هوشمند یکپارچه که قابلیت پایش لحظه‌ای، هشداردهی خودکار و پیشنهاد برنامه تعمیرات بهینه را فراهم آورد، می‌تواند گام مؤثری در ارتقاء مدیریت نگهداری و تعمیرات در صنایع پتروشیمی باشد.

## ۶- تضاد منافع

نویسندگان اعلام می‌کنند که هیچ‌گونه تضاد منافع مرتبط با تحقیق حاضر ندارند و نتایج به‌صورت بی‌طرفانه و بدون دخالت منافع شخصی یا حرفه‌ای به‌دست‌آمده است.

## ۷- منابع


[1] Shams H, Khorasgani GRH, Esfadan GA, Farsijani H, Shahmansouri A. Presentation of a Mathematical Model to Examine the Economic Advantages of Maintenance Strategies. System Engineering and Productivity. 2025;5(2). doi: https://doi.org/10.22034/sep.2025.20 51634.1264. [in Persian]

[2] Gorji MA, Jamali MB. Comprehensive Model for Evaluating Maintenance and Repair Policies Based on Interval fuzzy Numbers. System Engineering and Productivity. 2022;2(1). doi: https://doi.org/10.22034/sep.2022.243404. [in Persian]

[3] Izadikhah M, Garshasbi D. Using Data Mining and Three Decision Tree Algorithms to Optimize the Repair and Maintenance Process. Journal of New Researches In Mathematics. 2019;5(17):167–78. doi: https://sid.ir/paper/370289. [in Persian]

[4] Wang L. Data Mining Technology Integrates Analysis in Petrochemical Testing Data. Procedia Computer Science. 2025;261:946–53. doi: https://doi.org/10.1016/j.procs.2025.04.482.

[5] Giwa SO, Taziwa RT, Sharifpur M. Dependence of composition-based approaches on hybrid biodiesel fuel properties prediction using artificial neural network and random tree algorithms. Renewable Energy. 2023;128. doi: https://doi.org/10.1016/j.renene.2023.119324.

[6] Mallia J, Francalanza E, Xuereb P, Borg M, Refalo P. Implementation of an intelligence-based framework for anomaly Implementation of an intelligence-based framework for anomaly detection on the demand-side of sustainable compressed air systems. Procedia Computer Science. 2024;232:1554–63. doi: https://doi.org/10.1016/j.procs.2024.01.153.

[7] Namuduri S, Narayanan BN, Davuluru VSP, Burton L, Bhansali S. Deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors Journal of The Electrochemical Society. 2020;167(3). doi: https://doi.org/10.1149/1945-7111/ab67a8.

[8] Salehian Z, Jahan A. Establishing a Reliability-based Maintenance methodology in a Gas Pressure Reduction System. System Engineering and Productivity. 2022;1(1). doi: 10.22034/sep.2022.243401. [in Persian]

[9] Filz M-A, Langner JEB, Herrmann C, Thiede S. Data-driven failure mode and effect analysis (FMEA) to enhance maintenance planning. Computers in Industry. 2021;129:103451. doi: https://doi.org/10.1016/j.compind.2021.103451.

[10] Peng Y, Lin J-R, Zhang J-P, Hu Z-Z. A hybrid data mining approach on BIM-based building operation and maintenance. Building and Environment. 2017;126:483–95. doi: https://doi.org/10.1016/j.buildenv.2017.09.030.

[11] Atashhgar K, Saravany A. Failure prediction and diagnostic analysis for lathes using a hybrid approach including artificial neural netswork and decision making block. Internnational Journal of Indusrial Engineering & Prodction Management. 2016;27(2):143–56. [in Persian]

[12] Karimi M, Afshr Kazemi mohammad ali. Predicting failures and planning ATM maintenance by using data mining technique. Modern Research in Decision Making. 2016;1(3):113–30. [in Persian]

[13] Bagherighadikolaeia SP, Ghousi R, Haeri A. A Data Mining Approach for Forecasting Failure Root Causes: A Case Study in an Automated Teller Machine (ATM) Manufacturing Company. Journal of Optimization in Industrial Engineering. 2020;13(2). doi: 10.22094/JOIE.2020.1863364.1630.

[14] Rodriguez PC, Marti-Puig P, Caiafa CF, Serra-Serra M, Cusidó J, Solé-Casals J. Exploratory Analysis of SCADA Data from Wind Turbines Using the K-Means Clustering Algorithm for Predictive Maintenance Purposes. Machines Article. 2023;11(270). doi: https://doi.org/10.3390/machines11020270.

[15] Cheng C-W, Yao H-Q, Wu T-C. Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. Journal of Loss Prevention in the Process Industries. 2013;26(6):1269–78. doi: https://doi.org/10.1016/j.jlp.2013.07.002.

[16] Arena S, Florian E, Zennaro I, Orrù PF, Sgarbossa F. A novel decision support system for managing predictive maintenance strategies based on machine learning approaches. Safety Science. 2022;146. doi: https://doi.org/10.1016/j.ssci.2021.105529.

[17] Santos SP, Costa JAF. Application of multiple decision trees for condition monitoring in induction motors. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). 2008. p. 3736–41. doi: 10.1109/IJCNN.2008.4634334.

[18] Tran VT, Yang B-S, Oh M-S, Tan ACC. Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. Expert Systems with Applications. 2009;36(2, Part 1):1840–9. doi: https://doi.org/10.1016/j.eswa.2007.12.010.

[19] Allah Bukhsh Z, Saeed A, Stipanovic I, Doree AG. Predictive maintenance using tree-based classification techniques: A case of railway switches. Transportation Research Part C: Emerging Technologies. 2019;101:35–54. doi: https://doi.org/10.1016/j.trc.2019.02.001.

[20] Burmeister N, Frederiksen RD, Hog E, Nielsen P. Exploration of Production Data for Predictive Maintenance of Industrial Equipment: A Case Study. IEEE Access. 2023;11:102025–37. doi: https://doi.org/10.1109/ACCESS.2023.3315842.

[21] Kamel H. Artificial intelligence for predictive maintenance. In: Journal of Physics: Conference Series. 2022; 2299(1). doi: https://doi.org/10.1088/1742-6596/2299/1/012001

[22] Falzone S, Kolodziej JR. Condition monitoring of a reciprocating compressor using wavelet transformation and support vector machines. Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM. 2017;39–45. doi: https://doi.org/10.36001/phmconf.2017.v9i1.2191.

[23] Golmoradi M, Ebrahimi E, Javidan M. Compressor fault diagnosis based on SVM and GA. Vibroengineering Procedia. 2017;12:49–53. doi: https://doi.org/10.21595/vp.2017.18392.

[24] QUINLAN J.R. Induction of Decision Trees. Machine Learning. 1986;1:81–106. doi: https://doi.org/10.1007/BF00116251.

[25] Golmoradi M, Ebrahimi E, Javidan M. Fault diagnosis of compressor based on decision tree and fuzzy inference system. Vibroengineering Procedia. 2017;12:54–60. doi: https://doi.org/10.21595/vp.2017.18398.

[26] Elangovan M, Kumar SS, Bharathi Ganesh HB. Condition monitoring of a valve in a reciprocating compressor using machine learning approach. International Journal of Applied Engineering Research. 2015;10(13):33078–81.

[27] Aravinth S, Sugumaran V. Prediction of air compressor condition using vibration signals and machine learning algorithms. JVC/Journal of Vibration and Control. 2023;29(5–6):1342–51. doi: https://doi.org/10.1177/10775463211062330.

[28] Nessaib K, Lakehal A. Multi Sources Information Fusion Based on Bayesian Network Method to Improve the Fault Prediction of Centrifugal Compressor. Strojnicky Casopis. 2022;72(1):109–24. doi: https://doi.org/10.2478/scjme-2022-0011.

[29] Salehi S, Sadedel M. Condition Monitoring of Reciprocating Compressors using Probabilistic Neural Network and Optimization with Genetic Algorithm. Iranian Journal of Mechanical Engineering Transactions of ISME. 2023;25(3). [in Persian]

[30] Jamali M, Fargi Z, Rabiee M. Data mining of depressed patients to improve and evaluate its relationship with music. System Engineering and Productivity. 2021;1(4):49–73. doi: https://doi.org/10.22034/sep.2022.243416.

[31] Kalmegh SR. Comparative Analysis of WEKA Data Mining Algorithm RandomForest, RandomTree and LADTree for Classification of Indigenous News Data. International Journal of Emerging Technology and Advanced Engineering. 2008;9001(1):507–17.