Journal of Information and Communication Technology

Volume 17, Issue 66-67, Autumn and Winter 2025, pp. 34-53

An Architecture For Processes Analysis in Smart Factories Based on Big Data, Process Mining, and Machine Learning Techniques

Alireza Olyai[1], Shideh Saraeian[2*], Ali Nodehi[3]

[1] Department of Computer Engineering, Go.C., Islamic Azad University, Gorgan, Iran
[2] Department of Computer Engineering, Go.C., Islamic Azad University, Gorgan, Iran
[3] Department of Computer Engineering, Go.C., Islamic Azad University, Gorgan, Iran

Received: 08 January 2025, Revised: 28 May 2025, Accepted: 18 August 2025
Paper type: Research

Abstract

Due to the nature of smart factories and the use of new technologies such as cyber-physical systems, cloud computing, the Internet of Things, etc. in such environments, the volume of data generated has increased exponentially. Therefore, real-time processing of large volumes of high-speed data for process analysis is a difficult and challenging problem. In this case, big data analysis technologies, as a powerful tool, can play an important role in controlling processes. In this research, an architecture based on a combination of big data, process mining, and machine learning techniques for analyzing processes in smart factories is presented, which enables accurate and real-time analysis of processes in such environments. In fact, this architecture utilizes powerful big data analysis tools and new techniques such as process mining and by employing the logistic regression algorithm as a machine learning tool, it is able to extract valuable insights from the data generated in these environments. The results of the performance evaluation of the proposed architecture indicate that, based on the completeness and accuracy criteria in the context of process models discovery, it has achieved scores of 92.69% and 79.68%, respectively. Also, the best results were obtained for metrics such as prediction accuracy during the process model conformance checking phase using the logistic regression algorithm, reaching 99.5%. These results show the high capability of this architecture in real-time process analysis through the integration of advanced data analysis techniques.

Keywords: Smart Factory, Processes Analysis, Big Data, Process Mining, Machine Learning.

[*] Corresponding Author's email: shideh.saraeian@iau.ac.ir

دوفصلنامه
فناوری اطلاعات و ارتباطات ایران

سال هفدهم، شماره‌های ۶۶ و ۶۷، پاییز و زمستان ۱۴۰۴، صفحه ۳۴ الی ۵۳

یک معماری برای تحلیل فرآیندها در کارخانه‌های هوشمند بر اساس تکنیک‌های کلان‌داده، فرآیندکاوی و یادگیری ماشین

علیرضا اولیائی[۱]، شیده سرائیان[۲*]، علی نودهی[۳]

[۱] گروه کامپیوتر، واحد گرگان، دانشگاه آزاد اسلامی، گرگان، ایران

[۲] گروه کامپیوتر، واحد گرگان، دانشگاه آزاد اسلامی، گرگان، ایران

[۳] گروه کامپیوتر، واحد گرگان، دانشگاه آزاد اسلامی، گرگان، ایران

تاریخ دریافت: ۱۴۰۳/۱۰/۱۹ تاریخ بازبینی: ۱۴۰۴/۰۳/۰۷ تاریخ پذیرش: ۱۴۰۴/۰۵/۲۷

نوع مقاله: پژوهشی

چکیده

با توجه به ماهیت کارخانه‌های هوشمند و به‌کارگیری فناوری‌های نوینی مانند سیستم‌های سایبری فیزیکی، رایانش ابری، اینترنت اشیاء و غیره در این‌گونه محیط‌ها، حجم داده‌های تولید شده به‌صورت تصاعدی افزایش یافته است. بنابراین، پردازش بلادرنگ حجم زیادی از داده های با سرعت بالا به منظور تحلیل فرآیندها، مسئله‌ای سخت و چالش‌برانگیز است. در این شرایط، فناوری‌های تحلیل کلان‌داده‌ها به عنوان ابزاری قدرتمند می‌توانند نقش مهمی را در کنترل فرآیندها ایفا نمایند. در این پژوهش، یک معماری مبتنی بر ترکیب تکنیک‌های کلان‌داده، فرآیندکاوی و یادگیری ماشین برای تحلیل فرآیندها در کارخانه‌های هوشمند ارائه شده است که امکان تحلیل دقیق و بلادرنگ فرآیندها در این‌گونه محیط‌ها را فراهم می‌آورد. در حقیقت، این معماری از ابزارهای قدرتمند تحلیل کلان‌داده و تکنیک‌های نوینی مانند فرآیندکاوی بهره می‌برد و با به‌کارگیری الگوریتم رگرسیون لجستیک به عنوان یک ابزار یادگیری ماشین، قادر به استخراج بینش‌های ارزشمند از داده‌های تولید شده در این محیط‌ها است. نتایج ارزیابی معماری پیشنهادی نشان می‌دهد که این معماری بر اساس معیارهای کامل بودن و دقت در زمینه کاوش مدل‌های فرآیند به ترتیب به مقادیر ۹۲٫۶۹٪ و ۷۹٫۶۸٪دست یافته است. همچنین، بهترین نتایج در شاخص‌هایی نظیر دقت پیش‌بینی در مرحله بررسی انطباق مدل‌های فرآیند با استفاده از الگوریتم رگرسیون لجستیک معادل ۹۹٫۵٪ به‌دست آمده است. این نتایج نشان‌دهنده توانایی بالای این معماری در تحلیل‌های بلادرنگ فرآیندها از طریق ترکیب تکنیک‌های پیشرفته تحلیل داده است.

کلیدواژگان: کارخانه هوشمند، تحلیل فرآیندها، کلان‌داده‌ها، فرآیندکاوی، یادگیری ماشین.

* رایانامهٔ نویسنده مسؤول: shideh.saraeian@iau.ac.ir

۱- مقدمه

کارخانه هوشمند، نسل جدیدی از کارخانه‌ها است که با بهره‌گیری از فناوری‌های پیشرفته مانند سیستم‌های سایبری فیزیکی، اینترنت اشیاء، کلان‌داده‌ها، رایانش ابری، هوش مصنوعی، رباتیک و غیره باعث افزایش کیفیت و بهینگی تولید شده است. در کارخانه‌های هوشمند تمام عناصر (مانند ماشین‌ها، برنامه‌های کاربردی و غیره) در حال تولید کردن داده‌هایی می‌باشند که این داده‌ها عموما بزرگ و پیچیده هستند [۱]. بنابراین، علاوه بر داشتن کلان‌داده‌ها نیاز به ابزارهای قوی جهت دسترسی و تحلیل آنها می‌باشد که این مسئله می‌تواند به عنوان یک چالش مهم در نظر گرفته شود [۱]. اینترنت اشیاء یکی از منابع داده‌ای مهم تشکیل‌دهنده کلان‌داده‌ها است [۲]. این داده‌ها می‌توانند مربوط به کارخانه یا محیط کسب وکار باشند. از دید کسب وکار، تحلیل کلان‌داده‌ها در کارخانه‌های هوشمند می‌تواند اطلاعات مفیدی را برای سیستم‌های اطلاعاتی فراهم نماید [۳]. این کار توسط جمع‌آوری و تحلیل هوشمند مقدار گسترده‌ای از داده‌ها که از منابع مختلفی مانند گرایشات بازار، تقاضاهای جاری و آینده و غیره بدست می‌آیند، امکان‌پذیراست [۳]. با توجه به فناوری‌های مورد استفاده درکارخانه‌های هوشمند، داده‌های ساخت‌یافته و داده‌های جریانی دو نوع اصلی از داده‌هایی هستند که نقش مهمی را در این‌گونه محیط‌ها ایفا می‌نمایند. داده‌های ساخت‌یافته شامل داده‌هایی هستند که در یک قالب از پیش تعریف‌شده و سازمان‌یافته ذخیره می‌شوند (مانند داده‌های سیستم‌های اطلاعاتی). داده‌هایی که به صورت پیوسته و بلادرنگ تولید می‌شوند و نیاز به پردازش سریع دارند را داده‌های جریانی گویند. این داده‌ها معمولا توسط اینترنت اشیاء تولید می‌گردند.

یک فرآیند کسب وکار شامل مجموعه‌ای از فعالیت‌های مرتبط و پیوسته است که در یک سازمان برای رسیدن به یک هدف مشخص انجام می‌شود [۴]. کنترل عملکرد فرآیندهای کسب وکار، سازمان‌ها را قادر به بهبود فرآیندهای خود می‌سازد. عدم اندازه‌گیری کارایی فرآیندهای کسب و کار مانع کنترل و بهبود و در نتیجه مدیریت آنها می‌شود [۵]. بنابراین، تحلیل فرآیندها می‌تواند فوایدی مانند بهبود تصمیم‌گیری، افزایش بهره وری و کاهش هزینه‌ها را برای سازمان‌ها و صنایع فراهم آورد. فرآیندهای کسب و -کار در کارخانه‌های هوشمند بر اساس فناوری‌های جدید مانند اینترنت اشیاء و کلان‌داده‌ها اجرا می‌شوند. بنابراین، با توجه به اینکه حجم داده‌های فرآیندها به سرعت در حال رشد است، در کنار فراوانی رویدادهای مربوط به سیستم‌های سازمانی، کارخانه هوشمند و اینترنت اشیاء نیز موج جدیدی از داده‌های مربوط به فرآیندها را تولید می‌نمایند [۶] از این رو، اینترنت اشیاء نیز یکی از منابع داده‌ای است که می‌تواند جریانی از رویدادهای مورد استفاده در تحلیل فرآیندها را فراهم سازد. بنابراین، با توجه به ماهیت داده‌های مورد استفاده در کارخانه‌های هوشمند، کنترل اجرای فرآیندها در این‌گونه محیط‌ها باید توسط ابزاهای تحلیل کلان‌داده‌ها انجام گردد. این ابزارها می‌توانند در کنترل دقیق و بهینه فرآیندها نقش بسیار مهمی ایفا کنند.

از سوی دیگر، یکی از روش‌های قدرتمندی که می‌تواند برای کنترل فرآیندها به‌کار گرفته شود، تکنیک‌های فرآیندکاوی است. فرآیندکاوی به عنوان یکی از تکنیک‌های نوین مدیریت فرآیندهای کسب‌وکار[1]، ابزاری قدرتمند برای غنی‌سازی مدیریت کارخانه‌های هوشمند در جنبه‌های مختلف مانند کنترل و بهینه‌سازی فرآیندها محسوب می‌شود [۴]. فرآیندکاوی با تحلیل داده‌های ثبت شده از اجرای فرآیندها، امکان شناسایی انحرافات و بهبود کیفیت در فرآیندهای کسب‌وکار را فراهم می‌آورد. در حقیقت، کارخانه‌های هوشمند با تولید حجم عظیمی از داده‌ها، نیازمند ابزارهای قدرتمندی برای تحلیل و تصمیم‌گیری هستند. تحلیل کلان‌داده و فرآیندکاوی، دو فناوری مکمل هستند که می‌توانند به طور مشترک به این نیاز پاسخ دهند. تحلیل کلان‌داده با ارائه دید کلی از داده‌ها و فرآیندکاوی با تمرکز بر جزئیات فرآیندها، امکان بهبود مستمر، افزایش بهره‌وری و کاهش هزینه‌ها را در کارخانه‌های هوشمند فراهم می‌آورند.

همچنین، تکنیک‌های تحلیلی پیشرفته‌ای مانند الگوریتم‌های یادگیری ماشین نیز به عنوان ابزاری قدرتمند، می‌توانند نقش کلیدی را در تحلیل داده‌های تولید شده در کارخانه‌های هوشمند و بهینه‌سازی فرآیندها ایفا نمایند. درحقیقت، الگوریتم‌های یادگیری ماشین قادرند با یادگیری از داده‌های گذشته، مدل‌هایی ایجاد کنند که بتوانند رفتار سیستم را پیش‌بینی نموده و تصمیمات هوشمندانه‌ای اتخاذ کنند. بنابراین، برای فائق آمدن بر چالش تحلیل داده‌های بزرگ، متنوع و با سرعت زیاد جهت کنترل فرآیندها در کارخانه‌های هوشمند، این مقاله یک معماری مبتنی بر تکنیک‌های تحلیل‌کلان‌داده، فرآیندکاوی و یادگیری ماشین ارائه نموده است. در حقیقت، در این معماری از یکپارچه‌سازی تکنیک‌های پیشرفته تحلیل داده برای تحلیل فرآیندها استفاده شده است. استفاده از این تکنیک‌ها باعث می‌شود مدل‌های پیش‌بینی دقیق‌تری ایجاد شده و در نتیجه پشتیبانی بهتر و قوی‌تری در تحلیل فرآیندها فراهم گردد.

[1] Business Process Management (BPM)

بطورکلی، نوآوری‌های اصلی این پژوهش عبارت‌اند از:

- ارائه یک معماری کلان‌داده برای تحلیل داده‌ها در کارخانه‌های هوشمند؛
- یکپارچگی تکنیک‌های کلان‌داده، فرآیندکاوی و یادگیری ماشین برای کنترل دقیق فرآیندها در یک کارخانه هوشمند؛
- استفاده از مزایای الگوریتم‌های کاوشگر استقرایی و رگرسیون لجستیک برای استخراج و بررسی تطابق فرآیندها در معماری ارائه شده.

بقیه این مقاله به‌صورت زیر طبقه‌بندی شده است: بخش دوم به مرور ادبیات موضوع می‌پردازد. در بخش سوم معماری پیشنهادی معرفی می‌شود. بخش چهارم به ارزیابی کارایی و اعتبارسنجی معماری پیشنهادی اختصاص دارد. در بخش پنجم، نتایج حاصل از ارزیابی، بحث و بررسی می‌شود. سرانجام، بخش ششم به ارائه نتیجه‌گیری می‌پردازد.

۲- مرور ادبیات موضوع

۲-۱- ابزارهای تحلیل کلان‌داده‌ها

امروزه، حجم عظیمی از داده‌ها تولید می‌شوند که فرصت‌های بی‌نظیری را برای کسب وکارها فراهم می‌نماید. تحلیل این داده‌ها، به کسب وکارها کمک می‌کند تا تصمیمات بهتری بگیرند، روندها را پیش‌بینی کنند و از رقبا پیشی بگیرند. برای انجام این کار، به ابزارهای قدرتمندی نیاز است که بتوانند حجم عظیمی از داده‌ها را پردازش، تحلیل و بصری‌سازی کنند. ابزارهای تحلیل کلان‌داده را می‌توان به دسته‌های مختلفی شامل ابزارهای جمع‌آوری داده‌ها[1]، ابزارهای تحلیل داده‌ها[2]، ابزارهای ذخیره‌سازی داده‌ها[3] و ابزارهای ایجاد تقاضا[4] از داده‌ها تقسیم نمود. **دریافت داده،** اولین و یکی از مهم‌ترین مراحل در فرآیند تحلیل کلان‌داده است. این مرحله شامل جمع‌آوری داده‌ها از منابع مختلف، تبدیل آن‌ها به فرمتی قابل استفاده و بارگذاری در سیستم ذخیره‌سازی کلان‌داده است. برخی از ابزارهای محبوب مانند Apache Flume [۷]، Kafka [۷] و Apache Sqoop [۸] برای جمع‌آوری داده‌ها از منابع داده‌ای به‌کار می‌روند.

ابزارهای تحلیل داده، نقش بسیار مهمی در استخراج بینش‌های ارزشمند از حجم عظیمی از داده‌ها ایفا می‌کنند. این ابزارها باعث می‌شوند تا الگوها، روندها و ارتباطات پنهان در داده‌ها کشف شده و تصمیم‌گیری‌های مبتنی بر داده را بهبود می‌بخشند. در این رابطه، Hadoop یکی از محبوب‌ترین چارچوب‌های کلان‌داده است که از مدل برنامه‌نویسی MapReduce برای پردازش داده‌های دسته‌ای بزرگ استفاده می‌نماید. از سوی دیگر، ابزارهایی مانند Apache Storm [۷] و Apache Spark Streaming [۹-۱۰] نیز برای پردازش داده‌های جریانی در کلان‌داده‌ها استفاده می‌گردند.

بعضی از فناوری‌های ذخیره‌سازی داده‌ها شامل HDFS و HBase هستند که برای ذخیره‌سازی داده‌های بزرگ به‌کار می‌روند. HDFS[5] یکی از محبوب‌ترین سیستم‌های فایل توزیع‌شده است که برای ذخیره‌سازی داده‌های بزرگ در خوشه‌هایی طراحی شده است. HDFS داده‌ها را به بلوک‌های کوچک تقسیم نموده و آن‌ها را در چندین نود توزیع می‌کند تا در برابر خطا مقاوم باشند. HBase[6]، یک فناوری ذخیره‌سازی کلان‌داده است [۱۱]. به عبارت دیگر، HBase یک پایگاه داده توزیع شده NoSQL با قابلیت‌های مقیاس‌پذیری بسیار زیاد و تحمل‌پذیری عیب است که بر روی HDFS ساخته شده است [۱۲].

همچنین، ابزارهای ایجاد تقاضا از داده‌ها، این امکان را فراهم می‌سازند تا با استفاده از زبان‌های پرس‌وجو، داده‌های عظیم و پیچیده کاوش شوند. Apache Hive [۱۳]، Apache Pig [۱۴] و Apache Spark SQL [۹-۱۰] برخی از ابزارهای معروف ایجاد تقاضا در کلان‌داده‌ها می‌باشند که قادر به تحلیل کردن حجم بزرگی از داده‌ها هستند. Hive زبان SQL را پشتیبانی می‌کند. در حقیقت، Hive یک انبار داده مبتنی بر Hadoop است که به کاربران اجازه می‌دهد با استفاده از SQL به داده‌ها دسترسی پیدا کنند. Pig از یک زبان جریان داده‌ای به نام Pig Latin که مبتنی بر تقاضا است، پشتیبانی می‌نماید که نه تنها برای پردازش داده‌های ساخت‌یافته مانند Hive به‌کار می‌رود، بلکه قابلیت پردازش داده‌های غیرساخت‌یافته را نیز دارد. Apache Spark SQL نیز یک موتور پرس‌وجوی SQL بر روی Spark است که از پردازش داده‌های بزرگ به صورت موازی پشتیبانی می‌کند.

۲-۲- معماری‌های ارائه شده برای کلان‌داده‌ها

در ادبیات موضوع، طیف گسترده‌ای از معماری‌ها برای تحلیل کلان‌داده به منظور پاسخگویی به نیازهای خاص، طراحی شده‌اند.

[1] Data collection
[2] Data analyzing
[3] Data storing
[4] Data querying
[5] Hadoop Distributed File System
[6] Hadoop Database

در این رابطه، الکساکیس و همکاران[1] یک معماری کلان‌داده را برای سیستم‌های حمل و نقل هوشمند ارائه داده‌اند [۱۵]. این معماری می‌تواند بسیاری از موانع موجود در سیستم‌های قدیمی و کنونی را برطرف نماید. همچنین، در [۴] نویسندگان یک سیستم مدیریت فرآیند کسب وکار[2] را برای کارخانه‌های هوشمند توسعه داده‌اند که این سیستم از یک معماری تحلیل کلان‌داده نوآورانه به‌منظور نظارت بر اجرای فرآیندها بهره می‌برد. مانوگران و همکاران[3] یک معماری کلان‌داده مبتنی بر اینترنت اشیاء را به منظور نظارت و هشداردهی در زمینه مراقبت‌های پزشکی هوشمند پیشنهاد داده‌اند [۱۶]. در حقیقت، این معماری از دو زیرمعماری یکی برای جمع‌آوری و ذخیره‌سازی داده‌ها با استفاده از فناوری‌های کلان‌داده و دیگری برای یکپارچه‌سازی امن بین رایانش مه و رایانش ابری به‌کار می‌رود. با توجه به پیچیدگی روزافزون زنجیره‌های تامین، نیاز به ابزارهای قدرتمندی برای تحلیل داده‌ها و تصمیم‌گیری هوشمندانه بیش از پیش احساس می‌شود. در این رابطه، در [۱۷]، محققان یک معماری کلان‌داده را برای تحلیل زنجیره‌های تامین ارائه نمودند. آنها معتقدند که این معماری قابلیت‌هایی مانند کارایی، قابلیت اطمینان و بهینه‌سازی منابع را فراهم می‌آورد. المطیری و همکاران[4] یک رویکرد کلان‌داده را پیشنهاد داده‌اند که با بهره‌گیری از الگوریتم ژنتیک و تکنیک‌های مبتنی بر تقریب گرادیان، به طبقه‌بندی دقیق تصاویر پزشکی می‌پردازد [۱۸]. نویسندگان معتقدند که این روش باعث بهبود دقت تشخیص، ارائه هشدارهای بلادرنگ و کاهش زمان پردازش در حوزه سلامت می‌گردد. در [۱۹] یک معماری کلان‌داده به‌منظور شناسایی و تحلیل سایت‌های دارک وب[5] ارائه شده است. این معماری با استفاده از ترکیب فناوری‌های کلان‌داده، یادگیری ماشین و پردازش زبان طبیعی عمل می‌نماید. نتایج نشان می‌دهد، معماری پیشنهادی باعث شناسایی فعالیت‌های مشکوک مانند فیشینگ می‌شود.

سیریویرا و پایک[6] یک معماری مرجع چابک را برای تحلیل خودکار کلان‌داده‌ها برای کاربران لبه شبکه[7] ارائه نموده‌اند [۲۰]. آن‌ها معتقدند که این معماری می‌تواند به عنوان ابزاری مؤثر برای تحلیل داده‌ها در زیرساخت‌های مرتبط با صنعت و جامعه دیجیتال به‌کار رود

. در [۲۱] محققان یک معماری کلان‌داده مبتنی بر شبکه عصبی عمیق BilSTM[8] را پیشنهاد داده‌اند که برای پیش‌بینی و تصمیم‌گیری در حوزه‌های مالی و حسابداری به‌کار می‌رود. معماری پیشنهادی، قادر است با پردازش داده‌های حجیم، پیش‌بینی‌های دقیقی را ارئه دهد که می‌تواند باعث بهبود کارایی و دقت در تصمیم‌گیری‌های مالی شود. در [۲۲] نویسندگان یک معماری را با ترکیب رویکرد یادگیری ماشین و فناوری‌های تحلیل کلان‌داده برای شناسایی بیماران دیابتی در زمینه تصمیم‌گیری در حوزه مراقبت‌های پزشکی ارائه نمودند. آن‌ها با مقایسه الگوریتم‌های مختلف یادگیری ماشین در معماری پیشنهادی، نتیجه‌گیری کردند که الگوریتم رگرسیون لجستیک نتایج بهتری را نسبت به سایر الگوریتم‌ها داشته است.

مروری بر مشخصات برخی از معماری‌های ارائه شده برای کلان‌داده‌ها در جدول ۱ نشان داده شده است.

۲-۳- عملکردهای الگوریتم‌های یادگیری ماشین

یادگیری ماشین، شاخه‌ای از هوش مصنوعی است که به سیستم‌ها توانایی یادگیری و بهبود عملکرد بر اساس داده‌ها را می‌دهد. الگوریتم‌های یادگیری ماشین با شناسایی الگوها در داده‌ها، پیش‌بینی‌های دقیق و تصمیم‌گیری‌های هوشمندانه‌ای را ممکن می‌سازند. این الگوریتم‌ها در زمینه‌های مختلفی مانند کلان‌داده‌ها، رباتیک و غیره به‌کار می‌روند [۳۳]. برخی از الگوریتم‌های یادگیری ماشین شامل شبکه‌های عصبی مصنوعی، ماشین بردار پشتیان، آدابوست، رگرسیون لجستیک، درخت تصمیم و غیره هستند.

در سال‌های اخیر، این الگوریتم‌ها به دلیل توانایی‌های بی‌نظیرشان در حل مسائل پیچیده، توجه محققان و متخصصان از حوزه‌های گوناگونی مانند علوم کامپیوتر و مهندسی را به خود جلب کرده‌اند

[1] Alexakis et al.
[2] Business Process Management System (BPMS)
[3] Manogaran et al.
[4] Almutairi et al.
[5] Dark web
[6] Siriweera and Paik
[7] Edge computing
[8] Bidirectional Long Short-Term Memory

.

جدول ۱. مروری بر مشخصات بعضی از معماری‌های ارائه شده در زمینه کلان‌داده‌ها

سال	منبع	مشخصات معماری
۲۰۱۸	گوهر و همکاران[۱] [۲۳]	طراحی یک معماری کلان‌داده چندلایه‌ای برای تحلیل داده‌های اینترنت اشیای کوچک (IoST) با هدف استخراج اطلاعات ارزشمند و بهبود تصمیم‌گیری در سیستم‌های هوشمند.
۲۰۱۹	کنستانت نیکولاد و همکاران[۲] [۲۴]	تعریف یک معماری مبتنی بر کلان‌داده‌ها برای مدیریت و تحلیل داده‌های اینترنت اشیاء در زنجیره تأمین هوشمند به‌منظور بهبود کارایی و تصمیم‌گیری در زمینه لجستیک.
۲۰۲۰	سالیرنو و همکاران[۳] [۲۵]	یک معماری کلان‌داده چهار لایه‌ای برای نگهداری پیشگویانه خطوط راه‌آهن با هدف شناسایی خرابی‌ها در آن‌ها.
۲۰۲۱	سیماکوویچ و همکاران[۴] [۲۶]	ارائه یک معماری تحلیلی کلان‌داده برای اپراتورهای شبکه تلفن همراه به منظور افزایش سطح کیفیت خدمات و رضایت مشتریان.
۲۰۲۲	رایف وهمکاران[۵] [۲۷]	تعریف یک معماری کلان‌داده برای شهرهای هوشمند با هدف بهبود کیفیت زندگی شهروندان.
۲۰۲۳	آهیدوس و همکاران[۶] [۲۸]	ارائه یک معماری شش لایه‌ای برای کلان‌داده‌های مبتنی بر اینترنت اشیاء در زمینه آموزش.
۲۰۲۴	میلز و همکاران[۷] [۲۹]	یک معماری مبتنی بر رایانش ابری برای تحلیل کلان‌داده‌ها در محیط‌های 5G و رایانش لبه.
۲۰۲۴	ورنر و تای[۸] [۳۰]	توسعه‌ی یک معماری مرجع برای پردازش داده‌های کلان به‌صورت بدون سرور[۹] درمحیط‌های رایانش ابری به- منظور کاهش هزینه‌ها، افزایش کارایی و رفع چالش‌های مربوط به پردازش کلان‌داده‌ها.
۲۰۲۵	اسماعیل و همکاران[۱۰] [۳۱]	ارائه یک چارچوب برای پردازش داده‌های جریانی با استفاده از فناوری‌های کلان‌داده‌ها با هدف تحلیل بلادرنگ احساسات کاربران توییتر.
۲۰۲۵	ساراسوات و چوداری[۱۱] [۳۲]	ارائه یک رویکرد برای یکپارچه‌سازی تحلیل‌های کلان‌داده با رایانش ابری در بستر سیستم‌های ERP صنایع تولیدی.

در این رابطه، مایر و همکاران[۱۲] روشی مبتنی بر ترکیب تحلیل مولفه‌های اساسی[۱۳] و الگوریتم خوشه‌بندی DBSCAN[۱۴] را برای نظارت بر شرایط تولید و نگهداری پیشگویانه[۱۵] در کارخانه‌های هوشمند پیشنهاد داده‌اند [۳۴]. در [۲۴]، محققان برای کنترل زنجیره‌های تامین هوشمند یک سیستم تحلیل کلان‌داده مبتنی بر الگوریتم رگرسیون لجستیک را تعریف نموده‌اند. این معماری باعث انعطاف‌پذیری، کاهش هزینه‌ها و افزایش سرعت تصمیم‌گیری در زمینه لجستیک می‌گردد. همچنین، چو و همکاران[۱۶] یک روش جدید یادگیری ماشین که ترکیبی از روش‌های بدون‌نظارت و نیمه‌نظارت‌شده است را برای نگهداری پیشگویانه در کارخانه‌های هوشمند ارائه نموده‌اند [۳۵]. نویسندگان معتقدند رویکرد پیشنهادی، تسهیل یکپارچه‌سازی داده‌های ناهمگن تولید شده از دستگاه‌های مختلف را در راستای پشتیبانی از نگهداری پیشگویانه فراهم می‌آورد. در [۳۶]، محققان عملکرد الگوریتم‌های یادگیری ماشین مانند ماشین بردار پشتیبان و نایو بیز را در زمینه تشخیص نفوذ در شبکه‌های کامپیوتری ارزیابی نمودند. آن‌ها نتیجه گرفتند که الگوریتم ماشین بردار پشتیبان نتایج بهتری را در زمینه دقت و کاهش نرخ خطای طبقه‌بندی ارائه می‌دهد.

سرائیان و همکاران یک سیستم مدیریت فرآیند کسب وکار خودمختار را برای مدیریت فرآیندهای غیرقطعی با استفاده از عامل‌های[۱۷] هوشمند پیشنهاد داده‌اند [۳۷]. نویسندگان برای بهبود دقت تخمین پارامترها در این سیستم، از شبکه‌های عصبی پرسپترون چند لایه بهره برده‌اند. در [۲۵] یک معماری کلان داده مبتنی بر نوع خاصی از الگوریتم شبکه‌های عصبی بازگشتی یعنی شبکه LSTM[۱۸] ارائه شده است. در حقیقت، این رویکرد، یک مدل پیش‌بینی را معرفی می‌نماید که قادر به شناسایی خرابی‌های احتمالی در خطوط راه‌آهن می‌باشد. در [۳۸] محققان یک سیستم مدیریت فرآیند کسب وکار را برای تشخیص آناملی در فرآیندهای تولیدی افزایشی[۱۹] توسعه داده‌اند. در این سیستم از الگوریتم آدابوست به منظور شناسایی رفتارهای غیرعادی در مدل‌های فرآیند

[1] Gohar et al.
[2] Constante-Nicolalde et al.
[3] Salierno et al.
[4] Simaković et al.
[5] Raif et al.
[6] Ahaidous et al.
[7] Mills et al.
[8] Werener and Tai
[9] Serverless
[10] Ismail et al.
[11] Saraswat and Chouhari
[12] Maier et al.
[13] PCA (Principle Component Analysis)
[14] DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
[15] Predictive Maintenance
[16] Cho et al.
[17] Agents
[18] Long Short – Term Memory
[19] Additive Manufacturing Process

استفاده شده است. تیس و همکاران[1] یک معماری را برای بهبود پیش‌بینی مرگ‌ومیر در بیماران دیابتی با استفاده از روش‌های فرآیندکاوی و یادگیری عمیق ارائه داده‌اند[۳۹]. آن‌ها معتقدند که روش پیشنهادی در مقایسه با روش‌های مرسوم یادگیری ماشین، بر اساس مجموعه داده‌های مورد ارزیابی، عملکرد بسیار بهتری از خود نشان می‌دهد.

در [۲۸] یک معماری تحلیل کلان‌داده مبتنی بر الگوریتم درخت تصمیم در زمینه آموزش عالی پیشنهاد شده است. در این مقاله، محققان با تحلیل نمرات دانشجویان، روشی را برای نظارت و پیش‌بینی نمرات ارائه نمودند. احسانی و همکاران یک روش یادگیری ماشین را برای پیش‌بینی عمق کربوناسیون بتن تعریف نموده‌اند [۴۰]. در این پژوهش، از یک رویکرد نوآورانه به نام MOEA/D-ANN استفاده شده است. این رویکرد که ترکیبی از الگوریتم تکاملی چند هدفه مبتنی بر تجزیه و شبکه‌های عصبی مصنوعی است، به طور چشمگیری دقت پیش‌بینی را افزایش داده و زمان آموزش مدل را کاهش داده است. در [۴]، نویسندگان یک سیستم مدیریت فرآیند کسب‌وکار را برای کارخانه‌های هوشمند معرفی کرده‌اند. در این سیستم، از الگوریتم آدابوست برای نظارت بر وضعیت فرآیندها و تشخیص انحرافات استفاده شده است. با کمک این الگوریتم، امکان بهبود مستمر فرآیندها و افزایش بهره‌وری در کارخانه فراهم می‌گردد. در [۴۱] یک مدل طبقه‌بندی هوشمند بر اساس الگوریتم k نزدیکترین همسایه بهبودیافته به‌منظور افزایش دقت، سرعت و کیفیت طبقه‌بندی متون انگلیسی در کتابخانه‌هایی با حجم عظیم از منابع الکترونیکی ارائه شده است. همچنین، السیات و همکاران[2] یک رویکرد مبتنی بر الگوریتم یادگیری عمیق و ترکیب مدل‌ها را برای طبقه‌بندی تصاویر پزشکی با هدف شناسایی دقیق بیمارهای قلبی پیشنهاد داده‌اند [۴۲]. در روش پیشنهادی، محققان از مدل‌های از پیش آموزش‌دیده و انتخاب بهینه ترکیب آنها بهره گرفته‌اند تا دقت تشخیص را در زمینه‌های پزشکی افزایش دهند.

۲-۴- فرآیندکاوی

فرآیندکاوی یک حوزه تحقیقاتی جدید است که با استخراج دانش از گزارشات رویداد[3]تولیدشده توسط سیستم‌های اطلاعاتی، به تحلیل فرآیندهای یک سازمان می‌پردازد [۳۹]. در حقیقت، فرآیندکاوی شامل مجموعه‌ای از تکنیک‌ها است که با هدف کشف، نظارت و بهبود فرآیندها، از داده‌های موجود در سیستم‌های اطلاعاتی بهره می‌برد [۴۳]. تکنیک‌های فرآیندکاوی به سه دسته اصلی شامل کشف فرآیند[4]، بررسی مطابقت[5] و بهبود[6] تقسیم می‌شوند. الگوریتم‌های کشف فرآیند، با تحلیل داده‌های گزارشات رویداد، به‌صورت خودکار مدل‌های فرآیندها را تولید می‌نمایند.

در این رابطه، الگوریتم‌های متنوعی از جمله آلفا[7] [۴۴]، استخراج اکتشافی[8] [۴۵]، کاوشگر آی ال پی[9] [۴۶]، کاوشگر فازی[10] [۴۷] و کاوشگر استقرایی[11] [۴۸-۵۰]، برای استخراج مدل‌های فرآیند به‌کار گرفته می‌شوند. هر یک از این الگوریتم‌ها دارای مزایا و محدودیت‌های خاص خود بوده و انتخاب مناسب‌ترین الگوریتم به عوامل مختلفی از جمله نوع داده‌ها، پیچیدگی فرآیند و هدف تحلیل بستگی دارد. بررسی مطابقت دومین بخش از فرآیندکاوی است. تکنیک‌های بررسی مطابقت برای اهدافی مانند سنجش میزان انطباق بین مدل‌های فرآیند استخراج‌شده با واقعیت اجرایی (یعنی رفتار مشاهده شده) و همچنین ارزیابی تطابق بین مدل فرآیند با گزارش رویداد به‌کار می‌روند [۴۳]. برخی از الگوریتم‌های رایج بررسی مطابقت شامل فوت‌پرینت[12]، اجرای مجدد توکن[13] و همترازی[14] هستند. این روش‌ها در [۴۸] به‌صورت مفصل توضیح داده شده‌اند. سومین بخش از فرآیندکاوی، بهبود است. بهبود فرآیندها با استفاده از الگوریتم‌هایی صورت می‌گیرد که قادر به توسعه مدل فرآیند بر اساس داده‌های گزارش رویداد و همچنین ترمیم مدل‌های موجود می‌باشند.

همانطورکه پیش‌تر اشاره شد، با توجه به ماهیت کارخانه‌های هوشمند، پردازش بلادرنگ حجم عظیمی از داده‌های با سرعت زیاد به منظور تحلیل فرآیندها، مسئله‌ای سخت و چالش‌برانگیز است. همچنین، مطالعه معماری‌های ارائه شده در ادبیات موضوع نشان می‌دهد که بسیاری از طراحی‌های ارائه‌شده، فقط به معماری‌های کلان‌داده برای تحلیل داده‌ها بسنده نموده‌اند. نکته قابل توجه این است که ارائه معماری کلان‌داده به تنهایی کافی نیست؛ برای بهره‌برداری بهینه از آن، ترکیب با ابزارها و تکنیک‌های پیشرفته تحلیل داده ضروری است [۲۸]. در این زمینه، معماری‌های کلان‌داده از نظر ساختاری فقط به ارائه زیرساخت‌های ذخیره‌سازی

[1] Theis et al.
[2] Alsayat et al.
[3] Event logs
[4] Process discovery
[5] Conformance checking
[6] Enhancement
[7] Alpha
[8] Heuristic mining
[9] ILP miner
[10] Fuzzy miner
[11] Inductive miner
[12] Footprint
[13] Token Replay
[14] Alignment

و پردازش داده‌ها پرداخته‌اند و از نظر عملکردی فاقد پیش‌بینی تحلیلی بوده و تنها به جمع‌آوری و پردازش داده‌ها می‌پردازند.

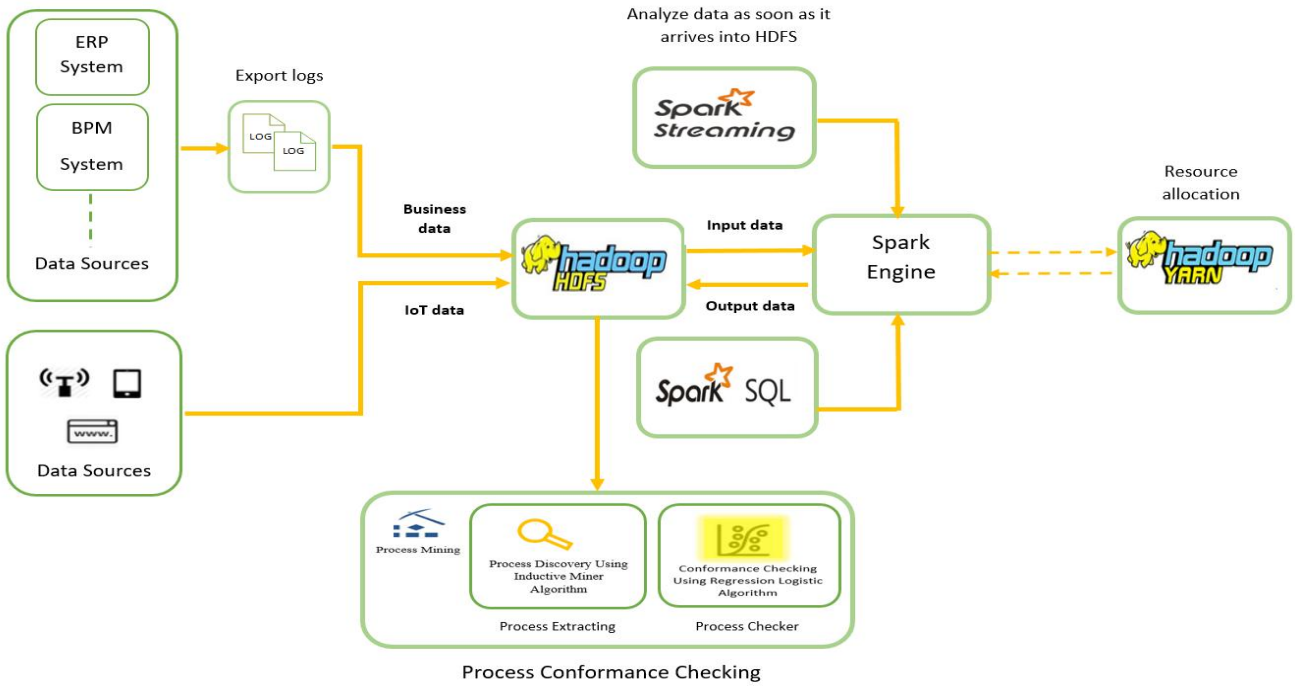
همچنین، برخی پژوهش‌های دیگر نیز صرفاً بر رویکردهای یادگیری ماشین متمرکز شده‌اند. با این حال، برای بهره‌برداری حداکثری از داده‌ها، ترکیب یادگیری ماشین با سایر روش‌های تحلیل داده، رویکردی کارآمدتر است. این‌گونه معماری‌ها، به لحاظ ساختاری نیازمند داده‌های آماده می‌باشند و عمدتا برای داده‌های حجیم و جریانی مناسب نیستند. از نظر عملکردی نیز، این معماری‌ها پیش-بینی‌هایی را ارائه می‌دهند بدون اینکه دید فرآیندی نسبت به تحلیل داشته باشند. در این زمینه، بعضی از طرح‌های پیشنهادی از ترکیب تکنیک‌های کلان‌داده‌ها و یادگیری ماشین به منظور تحلیل داده‌ها استفاده نموده‌اند. این معماری‌ها، اگرچه از نظر ساختاری کامل‌تر هستند، ولی قادر به شناسایی رفتار سیستم نمی‌باشند و به لحاظ عملکردی صرفا می‌توانند عملیات پیش‌بینی را انجام دهند و توانایی تحلیل گلوگاه‌ها، انحرافات یا بهینه‌سازی فرآیندها را ندارند. پژوهش حاضر، یک معماری را برای تحلیل دقیق و جامع فرآیندهای کارخانه‌های هوشمند ارائه می‌دهد. این معماری با تلفیق تکنیک‌های کلان- داده، فرآیندکاوی و یادگیری ماشین، امکان استخراج بینش‌های ارزشمند از داده‌ها را فراهم می‌نماید. یکی از مزایای بارز فرآیندکاوی در زمینه کشف مدل‌های فرآیند نسبت به روش‌های یادگیری ماشین، توانایی آن در ایجاد مدل‌های فرآیند قابل تفسیر است که به ما کمک می‌کند تا به طور مستقیم دلایل و علت‌های وقایع را درک کنیم [۳۹]. بنابراین، ترکیب این سه فناوری قدرتمند، دیدگاه کاملاً جدیدی را برای تحلیل فرآیندها فراهم می‌کند و به کسب وکارها اجازه می‌دهد تا به بینش‌های عمیق‌تری دست یابند و تصمیم‌گیری‌های بهتری اتخاذ نمایند. از این رو، معماری پیشنهادی دارای ویژگی‌های کارآمدی برای محیط‌هایی نظیر کارخانه‌های هوشمند است. این ویژگی‌ها عبارتند از:

- پردازش داده‌ها در مقیاس بالا به کمک فناوری‌های کلان‌داده؛
- پیش‌بینی و تحلیل‌های آماری از طریق روش‌های یادگیری ماشین؛
- بررسی و بهینه‌سازی رفتار سیستم با استفاده از تکنیک‌های فرآیندکاوی.

۳- معماری ارائه شده

معماری پیشنهادی، روشی برای تحلیل عمیق داده‌های کارخانه‌های هوشمند ارائه می‌دهد. این معماری با بهره‌گیری از تکنیک‌های پیشرفته کلان‌داده‌ها، فرآیندکاوی و یادگیری ماشین امکان کنترل دقیق فرآیندها را فراهم می‌آورد. همان‌طور که شکل ۱ نشان می‌دهد، این معماری شامل مولفه‌های مختلفی می‌باشد که می‌تواند قابلیت‌های تحلیلی قدرتمندی را برای تصمیم‌گیری بهینه در محیط‌های صنعتی هوشمند ارائه دهد. با گسترش کاربرد فناوری‌های نوین همچون اینترنت اشیاء و کلان‌داده در کارخانه‌های هوشمند، حجم عظیمی از داده‌های ساخت‌یافته تا داده‌های جریانی تولید می‌شوند. این حجم عظیم از داده‌ها، ضرورت استفاده از روش‌های نوین پردازش داده، به‌ویژه برای تحلیل داده‌های جریانی را دوچندان کرده است. به همین منظور، در معماری پیشنهادی، از تکنیک‌های پیشرفته تحلیل کلان‌داده برای پردازش بلادرنگ داده‌ها استفاده شده است. امروزه، Apache Hadoop و Apache Spark به عنوان ستون‌های اصلی پشته نرم‌افزار کلان‌داده شناخته می‌شوند [۵۱]. ابزارهای دیگر مانند Apache Hive و Apache HBase از قابلیت‌های آن‌ها برای اجرای عملیات خود بهره می‌برند [۵۱]. بنابراین، پایه و اساس معماری پیشنهادی بر روی دو چارچوب قدرتمند Hadoop و Spark استوار شده است. در این معماری، انواع منابع داده‌ای در نظر گرفته شده‌اند. ابتدا، از داده‌های ساخت‌یافته (مانند داده‌های سیستم‌های اطلاعاتی سازمانی) گزارشات رویداد استخراج شده و بر روی HDFS ذخیره می‌شوند تا در مراحل بعدی مورد تحلیل قرار گیرند. سپس، به‌صورت همزمان با ورود داده‌های ساخت‌یافته، داده‌های جریانی (مانند داده‌های اینترنت اشیاء) نیز به موتور پردازش Spark Engine ارسال می‌شوند تا از طریق مؤلفه Spark Streaming به‌صورت بلادرنگ و توزیع‌شده پردازش شوند. نتایج حاصل از این تحلیل نیز در HDFS ذخیره می‌گردد. پس از جمع‌آوری داده‌های ساخت‌یافته و داده‌های جریانی پردازش شده در HDFS، تمامی این داده‌ها به موتور پردازش Spark ارسال می‌شوند. در این مرحله، از مولفه SparkSQL برای اجرای پرس‌وجوهای تحلیلی و ترکیبی استفاده می‌شود.

این مرحله تحلیل، نقش اساسی در فراهم‌سازی داده‌های آماده برای فرآیندکاوی ایفا می‌کند. در این معماری، چارچوب YARN به عنوان یک مدیر منابع قدرتمند ایفای نقش می‌نماید. YARN با موتور Spark در تعامل مستقیم است و مسئولیت مدیریت منابع از جمله تخصیص، بهینه‌سازی و پشتیبانی از انواع مختلف پردازش را بر عهده دارد. پس از انجام تحلیل‌های کلان‌داده، نتایج به مولفه Process Conformance Checking به منظور انجام فرآیندکاوی ارسال می‌شوند. این عملیات شامل دو مرحله است؛ ابتدا مدل‌های فرآیند کشف شده و سپس تطابق آنها با داده‌های واقعی بررسی می‌شود. این مراحل در بخش‌های بعد بیشتر توضیح داده شده‌اند.



شکل۱. معماری پیشنهادی برای تحلیل فرآیندها در کارخانه‌های هوشمند

۳-۱- مولفه Process Conformance Checking

همانطور که شکل ۱ نشان می‌دهد، مولفه Process Conformance Checking در معماری پیشنهادی از دو ماژول تشکیل شده است. این دو ماژول دارای عملکردهای زیر می‌باشند:

- ماژول Process Extracting با تحلیل گزارشات رویداد، مدل‌های فرآیند را استخراج می‌نماید،
- ماژول Process Checker با بررسی تطابق مدل‌های فرآیند کشف‌شده با مدل‌های فرآیند در حال اجرا، انحرافات و گلوگاه‌های فرآیند را شناسایی می‌کند.

نحوه عملکرد این ماژول‌ها در مولفه Process Conformance Checking در شکل ۲ نشان داده شده است.

این رویکرد به کارخانه‌های هوشمند کمک می‌کند تا عملکرد فرآیندهای خود را بهینه‌سازی نمایند. این ماژول‌ها در بخش‌های بعدی با جزئیات بیشتر توضیح داده شده‌اند.

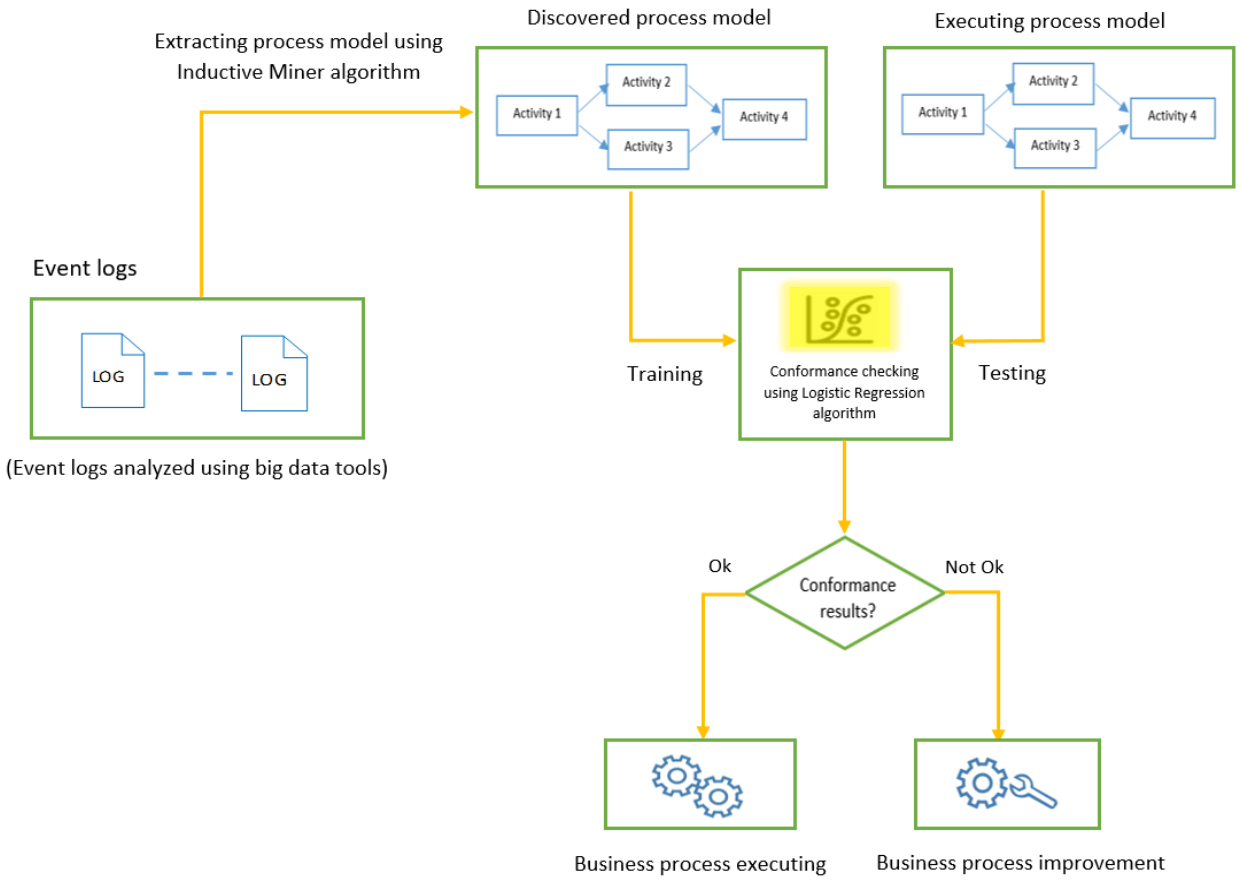
۳-۱-۱- ماژول Process Extracting

در ماژول Process Extracting، الگوریتم قدرتمند کاوشگر استقرایی با تحلیل گزارشات رویداد، مدل‌های فرآیند را به‌صورت خودکار استخراج می‌نماید. برخی از ویژگی‌ها و مزایای این الگوریتم عبارتند از [۴۸]:

- انعطاف‌پذیری؛
- استفاده از رویکرد درخت فرآیند[1] برای تولید مدل‌های فرآیند با کیفیت بالا نسبت به مدل‌های فرآیند تولید شده با زبان‌های مدل‌سازی دیگر مانند BPMN[2] و شبکه‌های پتری؛
- برخلاف الگوریتم‌های رایج کشف فرآیند مانند آلفا، استخراج اکتشافی و کاوشگر آی ال پی، تضمین تولید مدل‌های فرآیند Soundness (یعنی بدون بن‌بست[3] و سایر آناملی‌ها)؛
- برازندگی[4] بالای مدل فرآیند کشف‌شده نسبت به بازتولید رفتارهای گزارش رویداد.

الگوریتم کاوشگر استقرایی با استفاده از روش بازگشتی تقسیم و غلبه، مدل‌های فرآیند را کشف می‌کند. این الگوریتم، همانطور که در منابع [۴۸،۵۰] بیان شده است، یک گزارش رویداد را به چندین بخش کوچک‌تر تقسیم می‌کند و سپس به‌صورت بازگشتی بر روی هر بخش اجرا می‌شود تا زمانی که به ساده‌ترین شکل ممکن، یعنی یک فعالیت واحد، برسد. به این ترتیب، مسئله‌ی اصلی کشف مدل فرآیند، به چندین مسئله‌ی کوچک‌تر برای کشف زیرفرآیندها تقسیم می‌شود.

[1] Process Tree
[2] Business Process Model Notation
[3] Deadlock
[4] Fitness



شکل۲. توصیف عملکرد مولفه Process Conformance Checking

شبه‌کد الگوریتم کاوشگر استقرایی که برای کشف مدل‌های فرآیند به کار می‌رود، در الگوریتم ۱ ارائه شده است.

الگوریتم ۱. شبه‌کد الگوریتم کاوشگر استقرایی [۵۰]

```
function InductiveMiner(log)
// Base Case: If the log contains only one activity
b ← find BaseCase(log)
 if  InductiveMiner finds a basecase b in log then
    return b
 end if
 (⊕,∑1, …∑n) ← findCut(log)      // Find a suitable cut
// Split the log into partitions
 if InductiveMiner finds a cut c of operator ⊕ in log then
    log1…logn ← split log using c
    return⊕(InductiveMiner(log1),
    InductiveMiner(log2) , … InductiveMiner(logn))
 end if
// If no suitable cut is found, use a fallback strategy
    return fallThrought(log)
end function
```

نکته قابل توجه این است که مدل‌های فرآیند، علاوه بر نمایش گرافیکی رایج، می‌توانند به عنوان یک ماتریس علیتی[1] به‌صورت زیر نشان داده شوند [۳۸، ۴]:

$$CM = \{(In(act\_i), Out(act\_i)), \forall\, i = 1...n\}$$

به‌گونه‌ای که act_i نماینده یک فعالیت در گزارش رویداد است. In(act_i) و Out(act_i) به ترتیب مجموعه فعالیت‌هایی را نشان می‌دهند که باید پیش از act_i و پس از آن اجرا شوند. nتعداد کل فعالیت‌های موجود در گزارش رویداد است.

۳-۱-۲- ماژول Process Checker

پس از استخراج مدل‌های فرآیند، برای کنترل آن‌ها باید مدل‌های فرآیند کشف‌شده با مدل‌های فرآیند در حال اجرا مقایسه شوند. در حقیقت، با توجه به اینکه مدل‌های فرآیند استخراج‌شده، تصویری کلی از جریان کارهای کارخانه هوشمند را ارائه می‌دهند، مقایسه مداوم این مدل‌ها با وضعیت اجرایی واقعی فرآیندها، این

[1] Casual Matrix

امکان را فراهم می‌آورد تا از انطباق آن‌ها با یکدیگر اطمینان حاصل کنیم و در صورت لزوم، مدل‌ها را به‌روزرسانی نماییم. این امر به بهبود کنترل فرآیندها، افزایش بهره‌وری و کاهش خطاها کمک شایانی می‌کند. در ماژول Process Checker، عملیات کنترل فرآیندها از طریق الگوریتم رگرسیون لجستیک انجام می‌شود. این الگوریتم، یک ابزار قدرتمند و پرکاربرد در یادگیری ماشین است که برای حل طیف وسیعی از مسائل طبقه‌بندی استفاده می‌گردد. به طور کلی، رگرسیون لجستیک یک روش قدرتمند، قابل تفسیر، سریع و انعطاف‌پذیر برای پیش‌بینی متغیرهای وابسته دودویی است [۵۲]. برای مقایسه مدل‌های فرآیند، ابتدا الگوریتم با استفاده از مدل فرآیند کشف‌شده آموزش می‌بینند. سپس، الگوریتم آموزش‌دیده برای ارزیابی مدل فرآیند در حال اجرا به‌کار می‌رود. در نهایت، با مقایسه رفتار فرآیندهای در حال اجرا با مدل‌های کشف‌شده، در صورت وجود اختلاف، مدل‌های فرآیند بهبود داده می‌شوند. شبه‌کد الگوریتم رگرسیون لجستیک به منظور به‌کارگیری در ماژول Process Checker در الگوریتم ۲ نشان داده شده است.

۴- ارزیابی کارایی معماری پیشنهادی

در این پژوهش، برای ارزیابی معماری پیشنهادی از هدوپ نسخه ۳.۰.۰ و اسپارک نسخه ۳.۳.۱ استفاده شده است. آزمایشات بر روی یک سیستم با پردازنده‌ای دو هسته‌ای، ۱۲ گیگابایت حافظه‌ی رم و یک ترابایت فضای دیسک سخت انجام شده است. همچنین، با هدف ارزیابی کارایی مولفه Process Conformance Checking، الگوریتم‌های به‌کار رفته در آن به کمک ابزار ProM و زبان برنامه‌نویسی پایتون پیاده‌سازی شده‌اند. برای ارزیابی، مجموعه‌ای از گزارشات رویداد متنوع (جدول ۲) به عنوان داده‌های ورودی مورد استفاده قرار گرفته‌اند. با توجه به حجم بالای این گزارشات، فقط بخشی از یک نمونه از آن‌ها در شکل ۳ نشان داده شده است. این گزارشات رویداد توسط پیوند https://github.com/alireza178/event-log-evaluations/ قابل دسترس می‌باشند. همچنین، از معیارهای مختلفی نیز برای سنجش کارایی این مولفه استفاده شده است که در ادامه در زیربخش‌های بعدی به تفصیل شرح داده خواهند شد. شکل ۴ فرآیند کلی آزمایشات را برای ارزیابی کارایی مولفه Process Conformance Checking نشان می‌دهد.

الگوریتم۲. شبه‌کد الگوریتم رگرسیون لجستیک [۵۳]

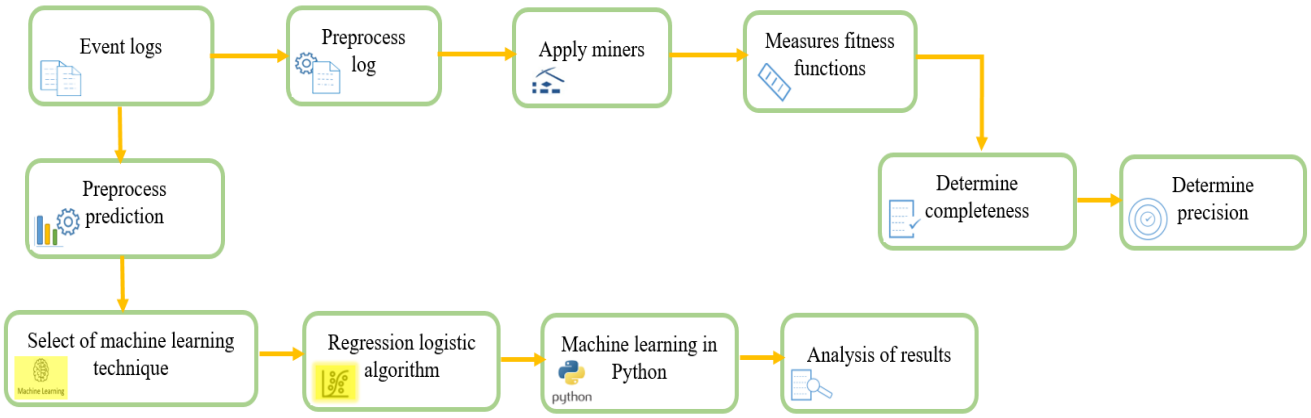
Input: Training data
(Train dataset← discovered process model,
Test dataset ← executing process model)
Begin
For i= 1 to k
For each training data instance d_i:
Set the target value for the regression to $z_i = \frac{y_i - P(1|d_j)}{[P(1|d_j) \times P(1-P(1|d_j))]}$
Initialize the weight of instance d_j to $[P(1|d_j) \times P(1 - P(1|d_j))]$
Finalize a f(j) to the data with class value (z_j) & weight(w_j)
Classification Label Decision
Assign (class label:1) if $P(1|d_j) > 0.5$, otherwise (class label:0)
End

جدول ۲. مشخصات فایل‌های گزارشات رویداد استفاده شده در آزمایشات

نام گزارش رویداد	توصیف مشخصات گزارش رویداد	ظرفیت	تعداد کیس‌ها	تعداد رویدادها
گزارش رویداد ۱	داده‌هایی از یک سیستم تولید مواد غذایی	۷۲٫۸ مگابایت	۲۰۱۳۵	۳۰۹۰۳۶
گزارش رویداد ۲	داده‌هایی درباره یک سیستم برنامه‌ریزی منابع سازمان	۴٫۸۴ مگابایت	۳۴۷۲۳	۱۰۳۴۶۹
گزارش رویداد ۳	داده‌هایی از تولیدات محصولات کشاورزی	۱٫۸۸ گیگابایت	۱۶۹	۱۰۴۸۵۷۵
گزارش رویداد ٤	داده‌هایی درباره زنجیره‌های تامین (داده‌های ساختیافته)	۹۱٫۱ مگابایت	۲۰۶۵۲	۱۸۰۵۱۹
گزارش رویداد ۵	داده‌هایی درباره زنجیره‌های تامین (داده‌های غیرساختیافته)	۹۱ مگابایت	۳۳۴۰	۴۶۹۹۷۷

Order Id	Product Name	order date (DateOrders)	Type	Days for sl	Days for sl	Benefit pe	Sales per customer	Delivery St	Late_deliv	Category I	Category I	Customer	Customer	Customer Email	Customer
1	Diamondback Women's Serene Classic Comfort Bi	1/1/2015 0:00	CASH	2	4	88.79	239.9799957	Advance sl	0	43	Camping &	Hickory	EE. UU.	XXXXXXXXX	Mary
2	Pelican Sunstream 100 Kayak	1/1/2015 0:21	PAYMENT	3	4	91.18	193.9900055	Advance sl	0	48	Water Spo	Chicago	EE. UU.	XXXXXXXXX	David
2	Nike Men's CJ Elite 2 TD Football Cleat	1/1/2015 0:21	PAYMENT	3	4	36.47	107.8899994	Advance sl	0	18	Men's Foo	Chicago	EE. UU.	XXXXXXXXX	David
2	Nike Men's Dri-FIT Victory Golf Polo	1/1/2015 0:21	PAYMENT	3	4	68.25	227.5	Advance sl	0	24	Women's /	Chicago	EE. UU.	XXXXXXXXX	David
4	Team Golf New England Patriots Putter Grip	1/1/2015 1:03	CASH	5	4	4.1	40.97999954	Late delive	1	40	Accessorie	San Anton	EE. UU.	XXXXXXXXX	Brian
4	Nike Men's Dri-FIT Victory Golf Polo	1/1/2015 1:03	CASH	5	4	60.27	123	Late delive	1	24	Women's /	San Anton	EE. UU.	XXXXXXXXX	Brian
4	Perfect Fitness Perfect Rip Deck	1/1/2015 1:03	CASH	5	4	26.13	296.9500122	Late delive	1	17	Cleats	San Anton	EE. UU.	XXXXXXXXX	Brian
4	O'Brien Men's Neoprene Life Vest	1/1/2015 1:03	CASH	5	4	33.59	159.9400024	Late delive	1	46	Indoor/Ou	San Anton	EE. UU.	XXXXXXXXX	Brian
5	Nike Men's CJ Elite 2 TD Football Cleat	1/1/2015 1:24	DEBIT	6	4	34.94	109.1900024	Late delive	1	18	Men's Foo	Caguas	Puerto Ric	XXXXXXXXX	Mary
5	Diamondback Women's Serene Classic Comfort Bi	1/1/2015 1:24	DEBIT	6	4	109.55	248.9799957	Late delive	1	43	Camping &	Caguas	Puerto Ric	XXXXXXXXX	Mary
5	Diamondback Women's Serene Classic Comfort Bi	1/1/2015 1:24	DEBIT	6	4	92.24	245.9799957	Late delive	1	43	Camping &	Caguas	Puerto Ric	XXXXXXXXX	Mary
5	Perfect Fitness Perfect Rip Deck	1/1/2015 1:24	DEBIT	6	4	143.98	299.9500122	Late delive	1	17	Cleats	Caguas	Puerto Ric	XXXXXXXXX	Mary
5	O'Brien Men's Neoprene Life Vest	1/1/2015 1:24	DEBIT	6	4	9.38	82.97000122	Late delive	1	46	Indoor/Ou	Caguas	Puerto Ric	XXXXXXXXX	Mary
7	Diamondback Women's Serene Classic Comfort Bi	1/1/2015 2:06	DEBIT	3	2	120.95	251.9799957	Late delive	1	43	Camping &	Miami	EE. UU.	XXXXXXXXX	Mary
7	Pelican Sunstream 100 Kayak	1/1/2015 2:06	DEBIT	3	2	78.4	195.9900055	Late delive	1	48	Water Spo	Miami	EE. UU.	XXXXXXXXX	Mary
7	Glove It Imperial Golf Towel	1/1/2015 2:06	DEBIT	3	2	4.58	77.55000305	Late delive	1	41	Trade-In	Miami	EE. UU.	XXXXXXXXX	Mary
8	O'Brien Men's Neoprene Life Vest	1/1/2015 2:27	TRANSFER	4	4	-131.15	163.9299927	Shipping o	0	46	Indoor/Ou	Caguas	Puerto Ric	XXXXXXXXX	Mary
8	Nike Men's Dri-FIT Victory Golf Polo	1/1/2015 2:27	TRANSFER	4	4	1.49	49.5	Shipping o	0	24	Women's /	Caguas	Puerto Ric	XXXXXXXXX	Mary
8	Perfect Fitness Perfect Rip Deck	1/1/2015 2:27	TRANSFER	4	4	16.88	149.3800049	Shipping o	0	17	Cleats	Caguas	Puerto Ric	XXXXXXXXX	Mary
8	Perfect Fitness Perfect Rip Deck	1/1/2015 2:27	TRANSFER	4	4	73.11	224.9600067	Shipping o	0	17	Cleats	Caguas	Puerto Ric	XXXXXXXXX	Mary
9	Pelican Sunstream 100 Kayak	1/1/2015 2:48	PAYMENT	5	4	96	199.9900055	Late delive	1	48	Water Spo	Lakewood	EE. UU.	XXXXXXXXX	Mary
9	Pelican Sunstream 100 Kayak	1/1/2015 2:48	PAYMENT	5	4	19.8	197.9900055	Late delive	1	48	Water Spo	Lakewood	EE. UU.	XXXXXXXXX	Mary
9	Nike Men's Free 5.0+ Running Shoe	1/1/2015 2:48	PAYMENT	5	4	3.8	189.9799957	Late delive	1	9	Cardio Eq	Lakewood	EE. UU.	XXXXXXXXX	Mary
10	Nike Men's CJ Elite 2 TD Football Cleat	1/1/2015 3:09	PAYMENT	6	4	28.73	110.4899979	Late delive	1	18	Men's Foo	Memphis	EE. UU.	XXXXXXXXX	Joshua
10	Glove It Women's Mod Oval 3-Zip Carry All Gol	1/1/2015 3:09	PAYMENT	6	4	-15.64	21.32999992	Late delive	1	41	Trade-In	Memphis	EE. UU.	XXXXXXXXX	Joshua
10	Pelican Sunstream 100 Kayak	1/1/2015 3:09	PAYMENT	6	4	-14.08	159.9900055	Late delive	1	48	Water Spo	Memphis	EE. UU.	XXXXXXXXX	Joshua
10	Pelican Sunstream 100 Kayak	1/1/2015 3:09	PAYMENT	6	4	72	149.9900055	Late delive	1	48	Water Spo	Memphis	EE. UU.	XXXXXXXXX	Joshua
10	O'Brien Men's Neoprene Life Vest	1/1/2015 3:09	PAYMENT	6	4	22.67	83.97000122	Late delive	1	46	Indoor/Ou	Memphis	EE. UU.	XXXXXXXXX	Joshua
11	Perfect Fitness Perfect Rip Deck	1/1/2015 3:30	PAYMENT	2	4	16.8	47.99000168	Advance sl	0	17	Cleats	Caguas	Puerto Ric	XXXXXXXXX	Nathan

شکل ۳. مشخصات بخشی از گزارش رویداد ۴



شکل ۴. مراحل انجام آزمایشات

۴-۱- معیارهای ارزیابی کارایی

در این بخش، برای ارزیابی عملکرد ماژول‌های Process Extracting و Process Checker در راستای کنترل فرآیندها، این ماژول‌ها در مولفه Process Conformance Checking توسط معیارهای مختلفی اعتبارسنجی شده‌اند. بر این اساس، توابع برازندگی برای ارزیابی الگوریتم‌های کشف مدل فرآیند به صورت زیر تعریف می‌شوند [۳۸]:

$$\text{Completeness} = \frac{\text{all considered activities of casual matrix} - \text{punishment}}{\text{total number of event log activities}} \quad (۱)$$

$$\text{Punishment} = \frac{\text{all incorrect relations of casual matrix}}{\text{number of log trace} - \text{number of traces incorrect relation} + 1} \quad (۲)$$

$$\text{Precision} = 1 - \max\{0, P_{dm}, P_{rm}\} \quad (۳)$$

$$P_{dm} = \frac{1}{\text{all enabled activites of the discoverd process model}} \quad (۴)$$

$$P_{rm} = \frac{1}{\textit{all enabled activites of the real process model}} \quad (۵)$$

همچنین، می‌توان برخی از معیارهای ارزیابی برای الگوریتم

بررسی انطباق فرآیند را به صورت زیر ارائه نمود [۳۸،۴۸]:

$$\text{True Positive Rate (TPR)} = \frac{\#\ \text{True Positive (TP)}}{\#\ \text{True Positive (TP)} + \#\ \text{False Negative(FN)}} \quad (۶)$$

$$\text{False Positive Rate (FPR)} = \frac{\#\ \text{False Positive (FP)}}{\#\ \text{False Positive (FP)} + \#\ \text{True Negative(TN)}} \quad (۷)$$

$$\text{Accuracy} = \frac{\#\ \text{True Positive (TP)} + \#\ \text{True Negative (TN)}}{N} \quad (۸)$$

$$N = \#\ \text{True Positive (TP)} + \#\ \text{True Negative(TN)} + \#\ \text{False Positive (FP)} + \#\ \text{False Negative (FN)} \quad (۹)$$

بنابراین، اگر نمونه‌های فرآیند درست و نادرست را به عنوان دو کلاس مجزا در نظر بگیریم، می‌توان پارامترهای استفاده شده در روابط ۶، ۷، ۸ و ۹ را به صورت زیر تعریف نمود:

- **مثبت صحیح**[1]: نمونه‌هایی از فرآیند که به درستی به عنوان نمونه‌های درست طبقه‌بندی شده‌اند.
- **منفی کاذب**[2]: نمونه‌هایی از فرآیند که به اشتباه به عنوان نادرست طبقه‌بندی شده‌اند، اما در واقع درست هستند.
- **منفی صحیح**[3]: نمونه‌هایی از فرآیند که به درستی، نمونه‌های نادرست طبقه‌بندی شده‌اند.
- **مثبت کاذب**[4]**:** نمونه‌هایی از فرآیند که به اشتباه به عنوان درست طبقه‌بندی شده‌اند، در حالی که در واقع نادرست هستند.

۴-۲- نتایج ارزیابی کارایی

برای ارزیابی کارایی ماژول Process Extracting در مولفه Process Conformance Checking، نیاز به پیاده‌سازی الگوریتم کاوشگر استقرایی است. بنابراین، این الگوریتم با استفاده از پلاگین مربوطه در نرم‌افزار ProM پیاده‌سازی می‌گردد. همچنین، برای سنجش عملکرد ماژول Process Extracting، مقایسه‌ای بین الگوریتم کاوشگر استقرایی با الگوریتم‌های آلفا، ژنتیک و استخراج اکتشافی انجام شده است. جدول‌های ۳ و ۴ پارامترهای استفاده شده در این الگوریتم‌ها را مشخص می‌نمایند.

جدول ۳. پارامترهای استفاده شده در الگوریتم ژنتیک

Population size	۱۰
Generation	۱۰۰
Extra behavior punishment	۰،۰۲۵
Mutation Probability	۰،۲
Crossover Probability	۰،۸
Elitism rate	۰،۲

جدول ۴. پارامترهای استفاده شده در الگوریتم استخراج اکتشافی

Long distance	۰،۹
Length two loops	۰،۹
Loops length one	۰،۹
Dependency	۰،۹

جدول ۵، نتایج نهایی ارزیابی عملکرد ماژول Process Extracting را بر اساس معیارهای انتخاب شده از جدول‌های ۳ و ۴ نشان می‌دهد. برای انجام این ارزیابی، گزارشات رویداد بیان شده در جدول ۲ به عنوان ورودی به ماژول وارد شده‌اند. پس از اجرای ماژول، نتایج به‌دست آمده در جدول ۵ ارائه گردیده است.

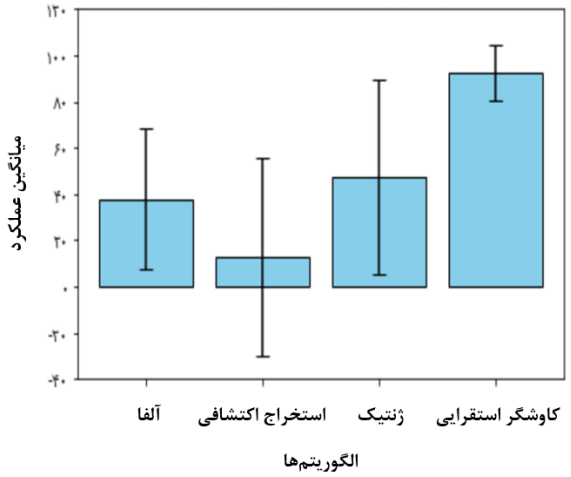
جدول ۵. نتایج مقایسه کارایی الگوریتم به‌کار رفته در ماژول Process Extracting با سایر الگوریتم‌ها

نام الگوریتم	کامل بودن (٪)	دقت (٪)
آلفا	۳۷،۹	۸۵،۱۸
	ارزیابی کلی:۶۱،۵۴	
استخراج اکتشافی	۱۲،۸	۷۶،۴۹
	ارزیابی کلی:۴۴،۶۴	
ژنتیک	۴۷،۳۳	۴۳،۵۸
	ارزیابی کلی:۴۵،۴۵	
کاوشگر استقرایی (این مقاله)	**۹۲،۶۹**	**۷۹،۶۸**
	ارزیابی کلی:۸۶،۱۸	

بر اساس نتایج جدول ۵، الگوریتم کاوشگر استقرایی با میزان دستیابی به نرخ کامل بودن[5]، ۹۲.۶۹ درصد و دقت[6]، ۷۹.۶۸ درصد عملکرد قابل قبولی از خود نشان می‌دهد. نکته قابل توجه این است که وقتی موازنه‌ای بین معیارهای کامل بودن و دقت مورد نیاز است و کیفیت کلی مدل کشف شده دارای اهمیت می‌باشد، الگوریتم کاوشگر استقرایی یک گزینه مناسب‌تری محسوب

[1] True Positive (TP)
[2] False Negative (FP)
[3] True Negative (TN)
[4] False Positive (FP)
[5] Completeness
[6] Precision

می‌شود. همچنین، برای بررسی تفاوت عملکرد الگوریتم‌های کاوش فرآیند از نظر معیار کامل بودن، تحلیل واریانس یک طرفه[1] انجام شده است. نتایج نشان می‌دهد که اختلاف معناداری بین الگوریتم‌ها وجود دارد. مقدار آماره‌ی آزمون F برابر با ۱۲,۱۲۷۸ با درجات آزادی۳ و ۱۶ محاسبه گردید و مقدار احتمال (p) برابر با ۰,۰۰۰۲ به‌دست آمده است. با توجه به اینکه مقدار p کمتر از سطح معنی‌دار ۰,۰۵ است، فرضیه صفر (که بیان می‌کند بین میانگین‌های گروه‌ها تفاوت معناداری وجود ندارد) رد می‌شود و این نتیجه نشان می‌دهد که حداقل، عملکرد یکی از الگوریتم‌ها به طور معنی‌داری با سایرین تفاوت دارد. همچنین، این یافته تاکید می‌نماید که نوع الگوریتم نیز تاثیر معناداری بر میزان کامل بودن فرآیند دارد. در این زمینه، شکل ۵ میانگین عملکرد چهار الگوریتم کاوش فرآیند را به همراه انحراف معیار آن‌ها نشان می‌دهد. همانطورکه در این نمودار مشخص شده است، الگوریتم کاوشگر استقرایی با میانگین عملکرد ۹۲,۶۹، بالاترین کارایی را در میان الگوریتم‌های دیگر داشته است؛ در مقابل، الگوریتم استخراج اکتشافی با میانگین ۱۲,۸ پایین‌ترین عملکرد را نشان داده است. با توجه به انحراف معیار پایین الگوریتم کاوشگر استقرایی، می‌توان نتیجه گرفت این الگوریتم علاوه بر عملکرد برتر، از پایداری بیشتری نسبت به الگوریتم‌های دیگر برخوردار بوده است.



شکل ۵. مقایسه میانگین و انحراف معیار عملکرد الگوریتم به‌کار رفته در ماژول Process Extracting با الگوریتم‌های دیگر

همچنین، برای تحلیل ترکیبی عملکرد الگوریتم‌ها بر اساس دو معیار کامل بودن و دقت، از تحلیل واریانس چندمتغیره[2] استفاده شده است. در این تحلیل، مقدار p به‌دست آمده برای تعامل بین نوع الگوریتم و معیارهای ارزیابی، ۰,۰۰۲۵۷۵ است. این مقدار نشان‌دهنده تاثیر معنی‌دار الگوریتم‌ها بر ترکیب این دو شاخص عملکردی می‌باشد. به‌طور مشخص، الگوریتم کاوشگر استقرایی بالاترین میزان کامل بودن را در بین الگوریتم‌های بررسی شده، ارائه داده است. اگرچه الگوریتم آلفا دقت بالاتری را نشان می‌دهد، اما میانگین کلی دو شاخص در الگوریتم کاوشگر استقرایی از سایر الگوریتم‌ها بالاتر بوده است. بنابراین، می‌توان نتیجه گرفت که الگوریتم کاوشگر استقرایی از نظر عملکرد کلی، برتری نسبی نسبت به سایر الگوریتم‌ها دارد.

همچنین، جدول ۶، مقایسه‌ای بین عملکرد الگوریتم‌های آلفا، استخراج اکتشافی، کاوشگر آی ال پی و کاوشگر استقرایی را بر اساس گزارشات رویداد استفاده شده در مطالعه‌ی بتاچی و همکاران[3] [۵۴] ارائه می‌دهد. نتایج این مطالعه در جدول ۶ نشان داده شده است.

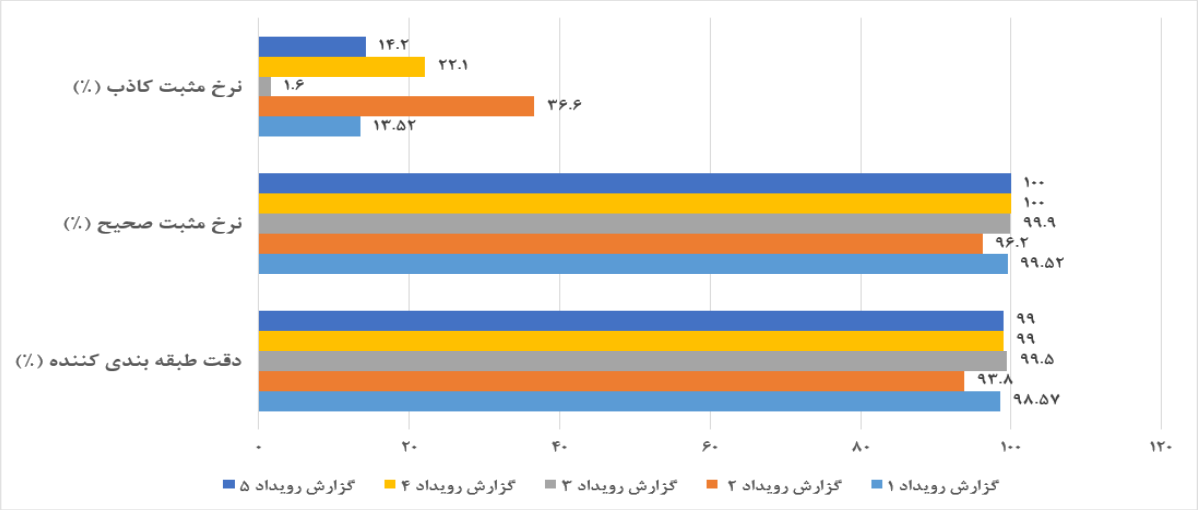
بر اساس مقایسه ارائه شده در جدول ۶، الگوریتم کاوشگر استقرایی دارای بالاترین میزان تناسب و دقت در مقایسه با سایر الگوریتم‌ها است.

به منظور ارزیابی دقیق کارایی ماژول Process Checker در مولفه Process Conformance Checking، الگوریتم رگرسیون لجستیک با استفاده از زبان برنامه‌نویسی پایتون پیاده‌سازی شده است. در این رابطه، گزارشات رویداد (جدول ۲) به نسبت ۷۰ درصد برای آموزش و ۳۰ درصد برای تست، تقسیم گردیده و به عنوان ورودی به مدل وارد شده‌اند. نتایج حاصل از اجرای این ماژول، به صورت نمودار در شکل ۶ ارائه شده است.

جدول ۶. مقایسه کارایی الگوریتم‌های آلفا، استخراج اکتشافی، کاوشگر آی ال پی و کاوشگر استقرایی [۵۴]

نام الگوریتم	تناسب (٪)	دقت (٪)
آلفا	۰,۸۳	۰,۷۷
استخراج اکتشافی	۰,۹۸	۰,۹۵
کاوشگر آی ال پی	۰,۹۰	۰,۷۷
کاوشگر استقرایی	**۰,۹۹**	**۰,۹۸**

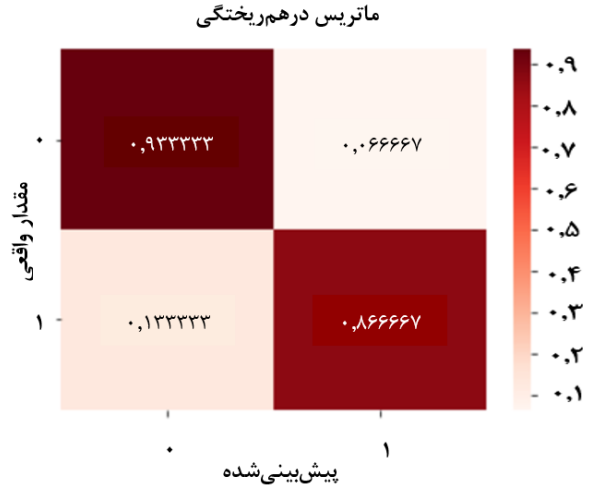
[1] One-Way ANOVA
[2] MANOVA
[3] Bettacchi et al.



شکل ۶. ارزیابی کارایی الگوریتم رگرسیون لجستیک

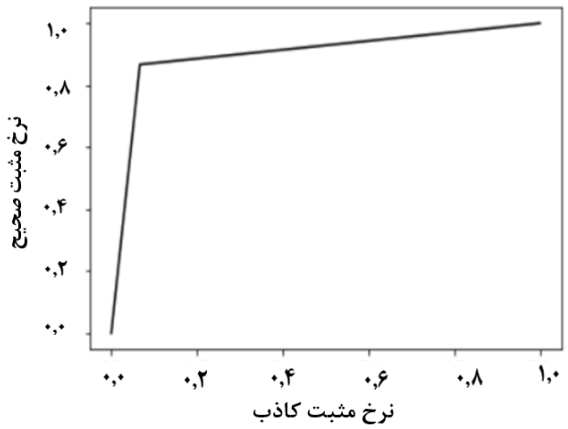
شکل ۶ نتایج ارزیابی عملکرد الگوریتم رگرسیون لجستیک به‌کار رفته در ماژول Process Checker را نشان می‌دهد. همان‌طور که مشاهده می‌شود، این الگوریتم عملکرد بسیار مناسبی در کنترل مدل‌های فرآیند بر اساس گزارشات رویداد مختلف از خود نشان داده است.

علاوه بر این، ماتریس درهم‌ریختگی[1] محاسبه شده برای بخشی از داده‌های گزارش رویداد ۱، به عنوان معیاری برای ارزیابی عملکرد مدل در تشخیص دقیق انواع مختلف داده‌ها، در شکل ۷ ارائه شده است. همانطور که مشاهده می‌شود، مدل در طبقه‌بندی کلاس‌های مختلف عملکرد بسیار خوبی دارد.



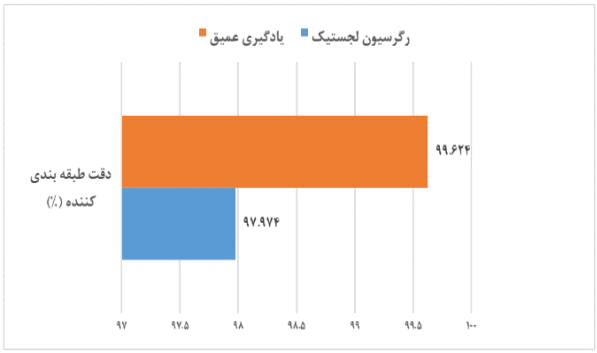
شکل ۷. یک مثال از ارزیابی مدل بر روی بخشی از داده‌های گزارش رویداد ۱

همچنین، شکل ۸ منحنی ROC[2] را برای داده‌های بررسی‌شده در شکل ۷ نمایش می‌دهد. مساحت زیر منحنی (AUC[3]) که معیاری برای سنجش عملکرد مدل‌های طبقه‌بندی است، برای این مدل برابر با ۰,۹۰ محاسبه شده است. هرچه مقدار AUC به ۱ نزدیک‌تر باشد، عملکرد مدل بهتر است. بنابراین، با توجه به مقدار AUCبه دست آمده، می‌توان گفت که مدل عملکرد قابل قبولی در تشخیص صحیح نمونه‌ها دارد.



شکل ۸. منحنی ROC برای ارزیابی طبقه‌بندی‌کننده

[1] Confusion Matrix
[2] Receiver Operating Characteristic
[3] Area Under the Curve



شکل ۹- مقایسه عملکرد الگوریتم‌های رگرسیون لجستیک و یادگیری عمیق

در این بخش، به منظور بررسی و مقایسه قابلیت‌های پیش‌بینی، عملکرد الگوریتم رگرسیون لجستیک در برابر الگوریتم یادگیری عمیق ارزیابی شده است. ارزیابی عملکرد الگوریتم یادگیری عمیق نیز با استفاده از همان گزارشات رویداد ارائه شده در جدول ۲ انجام گردیده است. در نهایت، نتایج دقت هر دو الگوریتم در قالب نمودار، در شکل ۹ ارائه شده است.

با توجه به مشاهدات آزمایشگاهی انجام شده و مقایسه آن‌ها با یافته‌های منابع مختلف، اگرچه الگوریتم یادگیری عمیق در پیش‌بینی‌های انجام گرفته دقت بالاتری ارائه نموده است (شکل ۹)، اما می‌توان نتیجه گرفت که الگوریتم رگرسیون لجستیک با توجه به ویژگی‌های خاص کارخانه‌های هوشمند نظیر تصمیم‌گیری بلادرنگ، مصرف بهینه منابع، تفسیرپذیری بالا و سادگی استقرار، نسبت به الگوریتم یادگیری عمیق گزینه‌ای کارامدتر محسوب می‌شود. بر این اساس، با مطالعه منابع گوناگون مانند [۵۵-۵۷] می‌توان این دو الگوریتم را بر پایه معیارهای مختلف مقایسه نمود.

الگوریتم یادگیری عمیق نیاز به زیرساخت محاسباتی سنگین (مانند GPU با توان محاسباتی بالا و حافظه زیاد) دارد. همچنین، برای پردازش داده‌های بزرگ و افزایش سرعت یادگیری، معمولاً لازم است تعداد لایه‌های شبکه افزایش یابد، که این موضوع به‌طور طبیعی باعث افزایش پیچیدگی محاسباتی، کندی اجرا و افزایش زمان پردازش می‌شود. در بسیاری از کاربردهای سازمانی، سرعت تصمیم‌گیری و سادگی مدل یک عامل مهم است. بنابراین رگرسیون لجستیک می‌تواند به‌سرعت و با تفسیرپذیری بالا، نتایج کاربردی‌تر ارائه دهد. در این زمینه، در تحلیل بلادرنگ فرآیندها، حتی چند میلی‌ثانیه تاخیر می‌تواند منجر به بروز مشکلاتی گردد.

الگوریتم رگرسیون لجستیک به دلیل ساختار ساده‌اش می‌تواند در بسترهای بلادرنگ و تحلیل داده‌های جریانی بدون تاخیر زیاد اجرا شود، در حالی‌که یادگیری عمیق زمان بیشتری برای پردازش نیاز دارد. همچنین، داده‌های جریانی در فرآیندهای کسب‌وکار، ممکن است شامل نویز، ناقص بودن و تنوع رفتاری باشند. الگوریتم یادگیری عمیق در چنین محیطی بدون تنظیم دقیق، ممکن است دچار بیش‌برازش گردد. در این شرایط، روش رگرسیون لجستیک مقاوم‌تر است و بهتر می‌تواند با داده‌های پرت یا ناقص مقابله نماید. علاوه بر این، یادگیری عمیق نیازمند تخصص بالا، زیرساخت مناسب و عملیات نگهداری مداوم است. مدل‌های ساده مانند رگرسیون لجستیک، پس از یک‌بار آموزش، می‌توانند به‌راحتی در سامانه‌های عملیاتی سازمانی مستقر شوند بدون آنکه بار مهندسی بالایی تحمیل کنند.

بنابراین، اگرچه الگوریتم یادگیری عمیق از نظر تئوری توانایی بیشتری دارد، نتایج آزمایشگاهی نشان می‌دهد که الگوریتم رگرسیون لجستیک تعادلی عملی‌تر بین عملکرد، تفسیرپذیری و کارایی استقرار ارائه می‌دهد. در محیط‌های کسب‌وکار با تصمیم‌گیری بلادرنگ، این روش، تصمیم‌گیری‌های سریع‌تر، یکپارچگی راحت‌تر و منطق شفاف‌تری فراهم می‌کند که اغلب از بهبودهای جزئی در دقت پیش‌بینی با ارزش‌تر هستند.

۴-۳- نتایج ارزیابی کارایی بر اساس داده‌های نویزی

برای ارزیابی میزان استحکام معماری پیشنهادی، الگوریتم‌های به‌کار رفته در آن شامل الگوریتم کاوشگر استقرایی و الگوریتم رگرسیون لجستیک بر اساس گزارشات رویداد ارائه شده در جدول ۲ و در سطوح مختلف نویز (۱۰٪، ۲۰٪ و ۳۰٪) مورد آزمایش قرار گرفته‌اند. در این پژوهش، نویزهای مورد مطالعه به دو دسته تقسیم شده‌اند. در بخش الگوریتم کاوش فرآیند، این نویزها شامل تغییر در نام فعالیت‌ها و ترتیب وقوع آن‌ها هستند. در بخش الگوریتم یادگیری ماشین، شامل اختلال در مقادیر برخی ویژگی‌ها و برچسب‌های کلاس می‌باشند. معیارهای ارزیابی شامل کامل بودن و دقت برای الگوریتم کاوشگر استقرایی و معیار دقت طبقه‌بندی‌کننده برای الگوریتم رگرسیون لجستیک در نظر گرفته شده‌اند. نتایج این ارزیابی‌ها در جداول ۷ و ۸ نشان داده شده‌اند.

جدول ۷. نتایج ارزیابی کارایی الگوریتم کاوشگر استقرایی بر اساس سطوح نویز ۱۰٪، ۲۰٪ و۳۰٪

نویز (٪)	کامل بودن (٪)	دقت (٪)
۰	۹۲,۶۹	**۷۹,۶۸**
۱۰	۸۹,۶۹	**۷۴,۸۶**
۲۰	۸۶,۳۲	**۷۲,۰۶**
۳۰	۸۳,۵۳	**۶۸,۶۱**

با توجه به اینکه الگوریتم کاوشگر استقرایی، به ویژه در معیار کامل بودن در برابر نویز مقاوم می‌باشد، همانطور که در جدول ۷ نشان داده شده است، با افزایش سطح نویز، مقادیر کامل بودن و دقت، کاهش یافته‌اند. با این حال، این کاهش تدریجی و یکنواخت بوده و نشان‌دهنده آن است که مدل فرآیند همچنان توانایی تطابق با داده‌های نویزی را دارد. در حقیقت، در این ارزیابی، حتی با ۳۰٪ نویز، مقادیر کامل بودن و دقت در محدوده قابل قبولی قرار گرفته‌اند که نشان‌دهنده این می‌باشد که هنوز معماری قادر است به‌طور نسبی فرآیندها را بازسازی نماید.

جدول ۸. عملکرد الگوریتم رگرسیون لجستیک در سطوح نویز ۱۰٪، ۲۰٪ و ۳۰٪

نویز (٪)	دقت طبقه‌بندی‌کننده
۰	۹۷,۹۷
۱۰	۸۹,۵۵
۲۰	۸۲,۴۹
۳۰	۶۷,۴۷

همان‌طور که در جدول ۸ نشان داده شده است، الگوریتم رگرسیون لجستیک نیز با افزایش نویز با کاهش دقت مواجه شده است. با این حال، کاهش دقت پیش‌بینی‌پذیر و منظم بوده و به‌صورت ناگهانی یا شدید رخ نداده است. دستیابی به دقت بالای ۸۹٪ با ۱۰٪ نویز و بیش از ۸۲٪ در ۲۰٪ نویز، نشان‌دهنده این است که مدل پیش‌بینی دارای ظرفیت تعمیم مناسبی بوده و دچار فروپاشی عملکردی نمی‌شود. حتی در شرایط ۳۰٪ نویز، مدل هنوز دقت بالاتر از حد تصادفی دارد که این امر نشان دهنده پایداری نسبی معماری است.

با توجه به نتایج ارائه‌شده، کاهش عملکرد هر دو الگوریتم به‌کار رفته در معماری پیشنهادی در برابر سطوح مختلف نویز، قابل انتظار و تدریجی بوده است. بنابراین، هیچ‌کدام از الگوریتم‌ها در مواجهه با نویز دچار فروپاشی شدید یا ناپایداری ناگهانی نشده‌اند و عملکرد آنها حتی در شرایط نویزی در محدوده قابل پذیرش قرار گرفته است. پس می‌توان نتیجه گرفت معماری طراحی‌شده از استحکام قابل قبولی در برابر نویز برخوردار است و برای کاربردهایی که داده‌ها ممکن است نویزی باشند، مناسب است.

۵- بحث

این پژوهش، یک معماری نوآورانه کلان‌داده را برای تحلیل و کنترل فرآیندها در یک کارخانه هوشمند ارائه می‌دهد. با ترکیب قدرتمند تکنیک‌های پیشرفته کلان‌داده، فرآیندکاوی و یادگیری ماشین، این معماری امکان نظارت دقیق، پیش‌بینی و بهینه‌سازی فرآیندها را فراهم می‌آورد. به عبارت دیگر، معماری پیشنهادی با استخراج بینش‌های عمیق از حجم عظیمی از داده‌ها، ابزاری قدرتمند برای تصمیم‌گیری داده‌محور و بهبود عملکرد کلی کارخانه هوشمند به شمار می‌رود.

همانطورکه پیش‌تر اشاره شد، معماری‌های کلان‌داده به تنهایی برای تحلیل داده‌ها کافی نیستند و برای بهره‌برداری بهینه از آن‌ها، ترکیب با ابزارها و تکنیک‌های پیشرفته تحلیل داده ضروری است [۲۸]. بنابراین، در معماری پیشنهادی برای دستیابی به کنترل دقیق فرآیندها، از ترکیب این تکنیک‌ها بهره گرفته شده است. این معماری با در نظر گرفتن ویژگی‌های کارخانه‌های هوشمند از جمله پویایی، توسعه‌پذیری و حجم بالای داده، طراحی شده است تا بتواند به طور مؤثری با این محیط سازگار شود. بدین منظور، مولفه‌های این معماری قابلیت‌های بالایی را درکنترل فرآیندها توسط استفاده از ابزارهای تحلیلی کلان‌داده، تکنیک‌های مدرنی مانند فرآیندکاوی و الگوریتم‌های یادگیری ماشین فراهم می‌آورند. در این رابطه، پردازش‌های کلان‌داده باعث می‌شود داده‌ها با سرعت بالایی پردازش شده و در نتیجه فایل‌های گزارش رویداد با بهینگی بیشتری برای مولفه‌های دیگر به‌کار گرفته شوند.

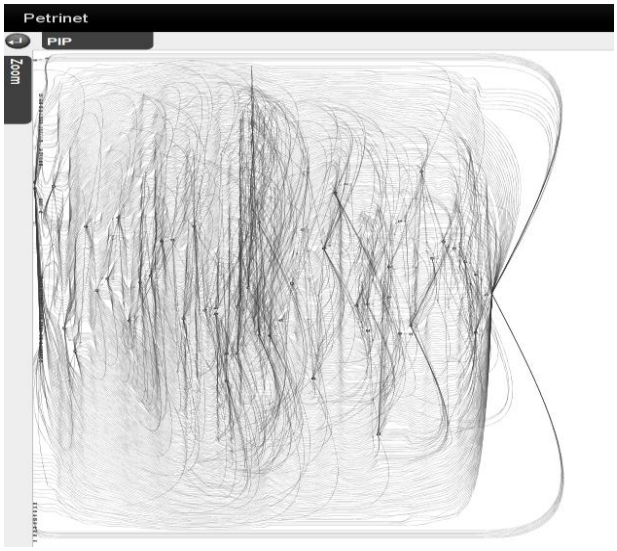
در این مقاله، الگوریتم کاوشگر استقرایی بر اساس معیارهای مختلفی ارزیابی شده است. با توجه به مزایای الگوریتم کاوشگر استقرایی در زمینه کشف مدل‌های فرآیند که در بخش ۱-۱-۳ تشریح شده است، نتایج ارزیابی‌های کمی ارائه شده در جدول‌های ۵ و ۶، برتری این الگوریتم را نسبت به الگوریتم‌های دیگر نشان می‌دهد. بنابراین، این الگوریتم، بر اساس معیارهای مختلف ارزیابی، عملکرد بهتری نسبت به سایر الگوریتم‌ها از خود نشان داده است.

علاوه بر این، در این پژوهش، الگوریتم رگرسیون لجستیک به عنوان ابزاری کارآمد برای ارزیابی مدل‌های فرآیند مورد استفاده قرار گرفته است. نتایج حاصل از ارزیابی کارایی این الگوریتم که در شکل ۶ به تصویر کشیده شده است، نشان دهنده دقت بالا و نرخ خطای پایین آن است.

برتری الگوریتم‌های یادگیری ماشین نسبت به الگوریتم‌های سنتی بررسی تطابق فرآیندها (مانند فوت‌پرینت، اجرای مجدد توکن و همترازی) در این است که آن‌ها توانایی بهتری در بهره-برداری از نتایج فرآیندکاوی برای بهبود فرآیندها دارند. درحقیقت، این الگوریتم‌ها مانند رگرسیون لجستیک با خودکارسازی فرآیندهای تحلیل و امکاناتی از قبیل طبقه‌بندی

مدل‌های فرآیند و تشخیص انحرافات، برای افزودن هوش مصنوعی به فرآیندکاوی به‌کار می‌روند و می‌توانند فرآیندکاوی سنتی را به فرآیندکاوی هوشمند تبدیل نمایند. علاوه بر این، روش‌های یادگیری ماشین مبتنی بر داده می‌باشند؛ در صورتی که اغلب روش‌های مرسوم بررسی مطابقت، مبتنی بر مدل هستند. بنابراین، در تکنیک‌های یادگیری ماشین نیاز به ایجاد مدل‌های فرآیند مانند مدل‌های پتری‌نت نمی‌باشد. از سوی دیگر، یکی از مشکلات تکنیک‌های مبتنی بر مدل نسبت به روش‌های یادگیری ماشین این است که این روش‌ها از مقیاس-پذیری ضعیف رنج می‌برند.

نکته قابل توجه در زمینه فرآیندکاوی، تعداد کیس‌ها و تعداد رویدادهای به‌کار رفته در یک کیس در یک گزارش رویداد است که می‌تواند گزارش رویداد را مقیاس‌پذیر نماید. به عبارت دیگر، از نظر ظرفیت یا مقیاس‌پذیری در کلان‌داده‌ها، منظور از کلان-داده در این پژوهش، گزارشات رویدادی هستند که تعداد کلانی از کیس و رویداد داشته باشند. پردازش مستقیم گزارشات رویداد بزرگ توسط الگوریتم‌های فرآیندکاوی، بدون بهره‌گیری از چارچوب‌های محاسباتی قدرتمند نظیر هدوپ، استخراج مدل-های فرآیند را غیرممکن می‌سازد یا منجر به استخراج مدل‌های فرآیند پیچیده می‌شود. این مدل‌ها در فرآیندکاوی به مدل‌های اسپاگتی موسوم‌اند. در مدل‌های اسپاگتی درک مدل فرآیند مشکل می‌شود و زمان زیادی برای تحلیل فرآیند نیاز است. برای مثال، شکل ۱۰ به وضوح نشان می‌دهد که عدم استفاده از چارچوب‌هایی مانند هدوپ در استخراج مدل فرآیند از گزارش رویداد ۴، منجر به تولید یک مدل اسپاگتی شده است.



شکل ۱۰. استخراج مدل فرآیند از گزارش رویداد ۴ در نرم افزار ProM

همچنین، با توجه به نقش محوری Apache Hadoop و Apache Spark در اکوسیستم کلان‌داده نسبت به برخی چارچوب‌های دیگر [۵۱]، معماری پیشنهادی بر پایه این دو چارچوب قدرتمند بنا شده است.

به‌طور کلی در سناریوهای بسیار بزرگ، مانند محیط‌های کارخانه‌های هوشمند که شامل تعداد زیادی رویداد و کیس هستند، مقیاس‌پذیری و کارایی معماری حائز اهمیت است. معماری پیشنهادی با تکیه برHDFS، Apache Spark طراحی شده است تا بتواند حجم عظیمی از گزارشات رویداد را پردازش نماید. بنابراین، داده‌ها ابتدا در بستر HDFS ذخیره شده و سپس با استفاده از Spark SQL و Spark Streaming تحلیل می‌شوند. HDFS، به‌عنوان سیستم ذخیره‌سازی توزیع‌شده، این گزارشات رویداد را به‌صورت کارآمد و مقیاس‌پذیر ذخیره نموده و دسترسی به داده‌ها را برای پردازش سریع فراهم می‌آورد. همچنین، ابزارهای Spark SQL و Spark Streaming امکان پردازش توزیع‌شده و کارآمد را هم در حالت دسته‌ای و هم در حالت جریانی فراهم می‌سازند. خروجی این مرحله به الگوریتم‌های کاوشگر استقرایی و رگرسیون لجستیک به منظور استخراج مدل فرآیند و ارزیابی میزان انطباق اجرای واقعی فرآیند با مدل استخراج‌شده منتقل می‌شود. بهره‌گیری از ویژگی‌های Apache Sparkباعث می‌شود این تحلیل‌ها به‌صورت توزیع‌شده و در مقیاس بالا انجام‌پذیر باشد. در نتیجه، این ترکیب از HDFS برای ذخیره‌سازی داده‌ها و Apache Spark برای پردازش آن‌ها، باعث می‌شود معماری ارائه‌شده توانایی پاسخ‌گویی به نیازهای پردازش و تحلیل در سناریوهای کلان‌داده‌ای را با حفظ دقت، سرعت و مقیاس‌پذیری داشته باشد.

بنابراین، با توجه به ماهیت پویا و مقیاس‌پذیر کارخانه‌های هوشمند و قابلیت داده‌های رویداد مورد استفاده در این محیط‌ها از نظر حجم، سرعت و تنوع و ویژگی‌های کلان‌داده‌ها (یعنی 4V[۱])، استفاده از ابزارهای تحلیلی کلان‌داده به همراه الگوریتم-های مورد استفاده در معماری پیشنهادی نه تنها مقیاس‌پذیری این معماری را به طور قابل توجهی بهبود می‌بخشد، بلکه امکان استخراج بینش‌های ارزشمند برای بهبود فرآیندها، افزایش بهره‌وری و کاهش هزینه‌ها را نیز فراهم می‌آورد. بدین ترتیب، با توجه به مزایای راهکارهای ارائه شده، معماری کلان‌داده‌ی پیشنهادی به عنوان یک رویکرد کارآمد برای پیاده‌سازی در کارخانه‌های هوشمند معرفی می‌گردد.

۶- نتیجه‌گیری

با توجه به ماهیت کارخانه‌های هوشمند و لزوم پردازش بلادرنگ حجم وسیع داده‌های تولید شده با سرعت زیاد، طراحی یک معماری کلان‌داده برای تحلیل فرآیندها ضروری است. بر این اساس، این مقاله یک معماری جدید را پیشنهاد می‌نماید که با ترکیب تکنیک‌های پیشرفته تحلیل داده‌ها مانند کلان‌داده‌ها، فرآیندکاوی و یادگیری ماشین به چالش تحلیل این نوع از داده‌ها پاسخ می‌دهد. نتایج ارزیابی نشان می‌دهد که معماری پیشنهادی به طور قابل توجهی در تحلیل و کنترل فرآیندها در محیط‌های صنعتی مانند کارخانه‌های هوشند به صورت کارآمد عمل می‌نماید.

در حقیقت، نوآوری اصلی این معماری، طراحی یک ساختار کلان‌داده برای پردازش بلادرنگ داده‌های حجیم و متنوع شامل داده‌های ساخت‌یافته و جریان داده‌ها، است. این ساختار با استفاده از چارچوب‌هایی مانند Hadoop و Spark و ترکیب تکنیک‌های پیشرفته‌ای مانند فرآیندکاوی و یادگیری ماشین، امکان تحلیل عمیق فرآیندها در کارخانه‌های هوشمند و تصمیم‌گیری‌های سریع و دقیق را فراهم می‌آورد.

علاوه بر این، این معماری با بهره‌گیری از الگوریتم‌های پیشرفته‌ای مانند کاوشگر استقرایی و رگرسیون لجستیک، در کنار تکنیک‌های کلان‌داده قادر به تحلیل فرآیندها و کشف روابط پنهان در داده‌های تولید شده در کارخانه‌های هوشمند است.

مراجع

[1] Mabkhot. M, Al-Ahmari. A, Salah. B, and Alkhalefah. H, "Requirements of the smart factory system: a survey and perspective," *Machines*, vol. 6, no. 2, pp. 23, 2018.

[2] Chen. M, Mao. S, and Liu. Y, "Big data: A survey," *Mobile networks and applications*, vol. 19, no. 2, pp. 171-209, 2014.

[3] Lee. J, Ardakani. H. D., Yang. S, and Bagheri. B, "Industrial big data analytics and cyber-physical systems for future maintenance & service innovation," *Procedia Cirp*, vol. 38, pp. 3-7, 2015.

[4] Olyai. A, Saraeian. S, and Nodehi. A, "Process mining-based business process management architecture: A case study in smart factories," *Scientia Iranica*, vol. 31, no. 14, 2024.

[5] Zur Muehlen. M, *Workflow-based Process Controlling: Foundation, Design and Application of workflow-driven Process Information Systems*, Logos Velrag Berlin, 2004.

[6] Polyvyanyy. A, Ouyang. C, Barros. A, and van der Aalst. W. M, "Process querying: Enabling business intelligence through query-based process analytics," *Decision Support Systems*, vol. 100, pp. 41-56, 2017.

[7] Liu. X, Iftikhar. N, and Xie. X, "Survey of real-time processing systems for big data," In *Proceedings of the 18th International Database Engineering & Applications Symposium, Porto, Portugal, July 7-9, 2014*, ACM, 2014, pp. 356-361.

[8] Nagdive. A S and Tugnayat. R M, "A review of Hadoop ecosystem for bigdata," *Int. J. Comput. Appl*, vol. 180, no.14, pp. 35-40, 2018.

[9] Shaikh. E, Mohiuddin. I, Alufaisan. Y, and Nahvi. I, "Apache spark: A big data processing engine," In *2019 2nd IEEE Middle East and North Africa COMMunications Conference (MENACOMM), Manama, Bahrain, November 19-21, 2019*, IEEE ,2019, pp. 1-6.

[10] Salloum. S, Dautov. R, Chen. X, Peng. P X, and Huang. J Z, "Big data analytics on Apache Spark," *International Journal of Data Science and Analytics*, vol. 1, pp.145-164 ,2016.

[11] Sahal. R, Breslin. J G, and Ali. M I, "Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case," *Journal of Manufacturing Systems*, vol. 54, pp. 138-151, 2020.

[12] Vora. M N, "Hadoop-HBase for large-scale data," In *Proceedings of 2011 International Conference on Computer Science and Network Technology, Harbin, China, December 24-26, 2011*, IEEE, 2011, pp. 601-605.

[13] Thusoo. A and et al., "Hive: a warehousing solution over a map-reduce framework," *Proceedings of the VLDB Endowment*, vol. *2*, no.2, pp.1626-1629, 2009.

[14] Bansal. K, Chawla. P, and Kurle. P, "Analyzing performance of apache pig and apache hive with Hadoop," In *Engineering Vibration, Communication and Information Processing: ICoEVCI, India*, Springer Singapore, 2019, pp. 41-51.

[15] Alexakis. T, Peppes. N, Demestichas. K, and Adamopoulou. E, " A distributed big data analytics architecture for vehicle sensor data," *Sensors*, vol. 23, no. 1, pp. 357, 2022.

[16] Manogaran. G and et al., "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system," *Future Generation Computer Systems*, vol. *82*, pp. 375-387, 2018.

[17] Biswas. S and Sen. J, "A proposed architecture for big data driven supply chain analytics," *arXiv preprint arXiv:1705.04958*, pp. 7-34, 2017.

[18] Almutairi. L, Abugabah, A, Alhumyani H, and Mohamed. A. A, "Intelligent biomedical image classification in a big data architecture using metaheuristic optimization and gradient approximation", *Wireless Networks*, vol.30, no. 8, pp. 7087-7108, 2024.

[19] Pastor-Galindo. J and et al., "A Big Data architecture for early identification and categorization of dark web sites", Future Generation Computer Systems, vol. 157, pp. 67-81, 2024.

[20] Siriweera. A and Paik. I, "AutoBDA: Model-driven Reference Architecture for Automated Big Data Analysis Framework", IEEE Transactions on Services Computing, 2025.

[21] Theodorakopoulos. L, Theodoropoulou. A, Kampiotis. G, and Kalliampakou. I, "NeuralACT: Accounting Analytics using Neural Network for Real-time Decision Making from Big Data", IEEE Access, 2025.

[22] Nauman. M and et al., "The Role of Big Data Analytics in Revolutionizing Diabetes Management and Healthcare Decision-Making", IEEE Access, 2025.

[23] Gohar. M and et al., "A big data analytics architecture for the internet of small things," *IEEE Communications Magazine*, vol. *56*, no. 2, pp.128-133, 2018.

[24] Constante-Nicolalde. F V, Pérez-Medina. J L, and Guerra-Terán. P, "A proposed architecture for iot big data analysis in smart supply chain fields," In *The international conference on advances in emerging trends and technologies*, Cham: Springer International Publishing, 2019, pp. 361-374.

[25] Salierno. G, Morvillo. S, Leonardi. L, and Cabri. G, "An architecture for predictive maintenance of railway points based on big data analytics," In *International Conference on Advanced*

Data in Libraries", Systems and Soft Computing, pp. 200186, 2025.

[42] Alsayat. A and et al., "Enhancing cardiac diagnostics: A deep learning ensemble approach for precise ECG image classification", Journal of Big Data, vol. 12, no., pp.7, 2025.

[43] Alizadeh. S, and Norani. A, "ICMA: a new efficient algorithm for process model discovery," *Applied Intelligence*, vol. *48*, no.11, pp. 4497-4514, 2018.

[44] Van der Aalst. W, Weijters. T, and Maruster. L, "Workflow mining: Discovering process models from event logs," *IEEE transactions on knowledge and data engineering*, vol.16, no. 9, pp. 1128-1142, 2004.

[45] Weijters. AJMM, Van der Aalst. WMP, Medeiros. AK, "Process mining with the heuristics miner algorithm," TU Eindhoven: BETA Working Paper Series, 2006.

[46] Van der Werf. J M E, van Dongen. B F, Hurkens. C A, and Serebrenik. A, "Process discovery using integer linear programming," In International conference on applications and theory of petri nets, Springer, Berlin, Heidelberg, 2008, pp. 368-387.

[47] Günther. C W, and Van Der Aalst. W M, "Fuzzy mining–adaptive process simplification based on multi-perspective metrics," In *International conference on business process management*, Springer Berlin Heidelberg, 2007, pp. 328-343.

[48] Van Der Aalst. W M P, *Process Mining-Data Science in Action*, 2rd ed., Springer, Berlin, Heidelberg, 2016.

[49] Leemans. S J, Fahland. D, and Van Der Aalst. W M, "Discovering block-structured process models from event logs containing infrequent behavior," In *Business Process Management Workshops: BPM 2013 International Workshops, Beijing, China, August 26, 2013, Revised Papers 11*, Springer international publishing, 2014, pp. 66-78.

[50] Leemans. S J, Fahland. D, and Van der Aalst. W M, "Scalable process discovery and conformance checking," *Software & Systems Modeling*, vol. 17, pp. 599-631, 2018.

[51] Genkin. M, Dehne. F, Shahmirza. A, Navarro. P, and Zhou. S, "Autonomic Architecture for Big Data Performance Optimization", In *Intelligent Systems Conference*, Cham: Springer Nature Switzerland, 2024, pp. 475-496.

[52] Pohar. M, Blas. M, and Turk. S, "Comparison of logistic regression and linear discriminant analysis: a simulation study," *Metodoloski zvezki*, vol.1, no. 1, pp.143, 2004.

[53] Kumar. V, "Evaluation of computationally intelligent techniques for breast cancer diagnosis," *Neural Computing and Applications*, vol. 33, no.8, pp. 3195-3208, 2021.

[54] Bettacchi. A, Polzonetti. A, and Re. B, "Understanding production chain business process using process mining: a case study in the manufacturing scenario", In *Advanced Information Systems Engineering Workshops: CAiSE 2016 International Workshops, Ljubljana, Slovenia, June 13-17, 2016, Proceedings 28*, Springer International Publishing, 2016, pp. 193-203.

[55] Ahmed. S F and et al. "Deep learning modelling techniques: current progress, applications, advantages, and challenges", Artificial Intelligence Review, vol.56, no. 11, pp. 13521-13617, 2023.

[56] Bailly. A and et al. "Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models", Computer Methods and Programs in Biomedicine, vol. 213, pp. 106504, 2022.

[57] Lu. Y and et al. "Comparison of machine learning and logistic regression models for predicting emergence delirium in elderly patients: A prospective study", International Journal of Medical Informatics, vol. 199, pp. 105888, 2025.

Information Systems Engineering, Cham: Springer International Publishing, 2020, pp. 29-40.

[26] Simaković. M N, Cica. Z G, and Masnikosa. I B. "Big Data architecture for mobile network operators," In *2021 15th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS), Nis, Serbia, October 20-22, 2021*, IEEE, 2021, pp. 283-286.

[27] Raif. M, Chehri. A, and Saadane. R, "Data architecture and big data analytics in smart cities," *Procedia Computer Science*, vol. 207, pp. 4123-4131.2022.

[28] Ahaidous. K, Tabaa. M, and Hachimi. H, "Towards IoT-Big Data architecture for future education," *Procedia Computer Science*, vol. 220, pp. 348-355.2023.

[29] Mills. N and et al., "A cloud-based architecture for explainable Big Data analytics using self-structuring Artificial Intelligence," *Discover Artificial Intelligence*, vol.*4*, no. 1, pp.33, 2024.

[30] Werner. S, and Tai. S," A reference architecture for serverless big data processing", Future Generation Computer Systems, vol. 155, pp. 179-192, 2024.

[31] Ismail. A, Sazali. F. H, Jawaddi. S. N. A, and Mutalib. S, " Stream ETL framework for twitter-based sentiment analysis: Leveraging big data technologies", Expert Systems with Applications, vol. 261, pp. 125523, 2025.

[32] Saraswat. J. K and Choudhari. S," Integrating big data and cloud computing into the existing system and performance impact: A case study in manufacturing", Technological Forecasting and Social Change, vol. 210, pp. 123883, 2025.

[33] Çınar. Z M and et al., "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," *Sustainability*, vol.*12*, no. 19, pp. 8211, 2020.

[34] Maier. A, Schriegel. S, and Niggemann. O, "Big data and machine learning for the smart factory—Solutions for condition monitoring, diagnosis and optimization," *Industrial Internet of Things: Cyber manufacturing Systems*, pp. 473-485, 2017.

[35] Cho. S and et al., "A hybrid machine learning approach for predictive maintenance in smart factories of the future," In *Advances in Production Management Systems. Smart Manufacturing for Industry 4.0: IFIP WG 5.7 International Conference, APMS 2018, Seoul, Korea, August 26-30, 2018, Proceedings, Part II*, Springer International Publishing, 2018, pp. 311-317.

[36] Halimaa. A and Sundarakantham. K, "Machine learning based intrusion detection system," In *2019 3rd International conference on trends in electronics and informatics (ICOEI), Tirunelveli, India, April 23-25, 2019*, IEEE, 2019, pp. 916-920.

[37] Saraeian. S, Shirazi. B, and Motameni. H, "Optimal autonomous architecture for uncertain processes management," *Information Sciences*, vol. 501, pp. 84-99, 2019.

[38] Saraeian. S, and Shirazi. B, "Process mining-based anomaly detection of additive manufacturing process activities using a game theory modeling approach," *Computers & Industrial Engineering*, vol. 146, pp. 106584, 2020.

[39] Theis. J, Galanter. W L, Boyd. A D, and Darabi. H, "Improving the in-hospital mortality prediction of diabetes ICU patients using a process mining/deep learning architecture," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no.1, pp. 388-399, 2021.

[40] Ehsani. M and et al., "Machine learning for predicting concrete carbonation depth: A comparative analysis and a novel feature selection," *Construction and Building Materials*, vol. 417, pp. 135331, 2024.

[41] Xu. Q, "Application of an Intelligent English Text Classification Model with Improved KNN Algorithm in the Context of Big