ارزیابی عملکرد زنجیره تأمین پایدار با استفاده از یادگیری ماشین و تحلیل پوششی داده شبکه‌ای

سینا سیاردوست تبریزی *
کیخسرو یاکیده **
محمود مرادی ***
مصطفی ابراهیم‌پور ازبری ****

تاریخ ارسال: ۱۴۰۲/۱۲/۰۴
تاریخ پذیرش: ۱۴۰۳/۰۶/۳۱

چکیده

صنعت پتروشیمی در ایران به‌عنوان صنعت مولد و یکی از پایه‌های اقتصاد کشور بوده و نظارت بر کارایی و عملکرد زنجیره تأمین پایدار در این صنعت یکی از فاکتورهای مهم برای مدیران جهت تصمیم‌گیری و تنظیم راهبردهای کلان توسعه پایدار است. باتوجه به اینکه تحلیل پوششی داده شبکه‌ای برای ارزیابی کارایی نسبی بین واحدهای تحت بررسی یک روش پذیرفته شده و معتبر در تحقیقات دانشگاهی بوده، یکی از چالش‌ها این حوزه محاسبه کارایی نسبی بین واحدهای همگن و مشابه است. در این پژوهش بر اساس مدلسازی ریاضی با استفاده از تحلیل پوششی داده شبکه‌ای (NDEA) ضمن بهره‌گیری از یادگیری ماشین بهترین الگورتیم برای خوشه‌بندی زنجیره تامین دوسطحی بین ۲۸ واحد پتروپالایشی فعال در ایران برای ۹۰ دوره زمانی با رویکرد پایداری انتخاب و نتایج با روش سنتی محاسبه کارایی بدون خوشه‌بندی مقایسه گردید. نتایج مقایسه سه الگورتیم مختلف یادگیری ماشین در خوشه بندی نشان داد که الگوریتم Deep Embedded Clustering بر اساس شاخص‌های سه گانه ارزیابی کیفیت خوشه‌بندی، به میزان ۱۰٪ از سایر الگورتیم‌ها کیفیت بهتری را بر روی مجموعه داده مورد مطالعه ارائه داده، ضمنا به صورت میانگین فاصله واحدهای ناکارا تا مرز کارایی خوشه خود به میزان ۱۰ تا ۲۰ درصد نسبت به محاسبه کارایی بدون خوشه‌بندی کاهش داشته است. این راهکار در تعیین برنامه بهبود عملیاتی‌تر برای واحدهای ناکارا بسیار مناسب است. همچنین مقایسه فاصله واحدهای ناکارا تا مرز کارایی در هر خوشه می‌تواند مبنای مناسب‌تری برای مقایسه کارایی واحدها در ارائه راهکار بهبود و سیاست گذاری‌های کلان مدیریتی در راستای توسعه محصولات در نظر گرفته شود. هدف این پژوهش نشان دادن تاثیر خوشه‌بندی در محاسبه کارایی نسبی است که از آن می‌توان برای ارزیابی کارایی سایر صنایع بهره جست.

واژگان کلیدی: یادگیری ماشین، زنجیره تأمین پایدار، تحلیل پوششی داده شبکه‌ای،

* دانشجوی دکترا، مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه گیلان، ایران، رشت
** استادیار گروه مدیریت، دانشکده مدیریت و اقتصاد، دانشگاه گیلان، ایران، رشت (yakideh@guilan.ac.ir)
*** دانشیار گروه مدیریت، دانشکده مدیریت و اقتصاد، دانشگاه گیلان، ایران، رشت
**** استاد گروه مدیریت، دانشکده مدیریت و اقتصاد، دانشگاه گیلان، ایران، رشت

مقدمه

مفهوم کارایی که عبارت است از نسبت حداکثر خروجی مطلوب به ورودی (منابع) مصرف شده، یکی از شاخص‌های ارزیابی کننده وضعیت واحدها است و کارایی نسبی معیاری جهت مقایسه بین واحدها معرفی شده‌است. حداکثر سازی خروجی مطلوب به ازای ورودی معین یا کاهش ورودی ها به ازای مقدار مشخصی خروجی از جمله راهکارهای افزایش کارایی که در روش تحلیل پوششی داده‌ها (DEA) با در نظر گرفتن مجموعه امکان تولید و تعیین مرز کارا به ارزیابی واحدهای تصمیم می‌پردازد. راهکارهای سنتی محاسبه کارایی بر مبنای DEA در حالت کلی به دو دسته پوششی و مضربی تقسیم می‌شوند. در فرم مضربی برای هر واحد تصمیم بهترین وزن‌های ممکن برای ورودی و خروجی‌های نظیرش محاسبه می‌شود و در نهایت کارایی نسبی بر اساس این وزن‌های بهینه، از تقسیم مجموع وزن‌دار خروجی‌ها بر مجموع وزن‌دار ورودی‌ها به‌دست می‌آید (Charnes et al., 1978). طراحی بهینه ساختار شبکه‌ای در زنجیره تأمین یکی از رویکردهای پژوهشی امروزی در حوزه مدیریت زنجیره تأمین (SCM) است. توجه به شاخص‌های کلیدی پایداری نیز از مهمترین بخش‌های مدیریت واحدهای صنعتی بر اساس مدل‌های پایه‌ای در طراحی ساختار شبکه زنجیره تأمین است (Hofmann, 2013). با ترکیب مفاهیم پایداری و ظهور زنجیره تأمین پایدار (SSCM) استفاده از حلقه بسته و جریان بازگشتی در طول زنجیره تأمین در نظر گرفته شد که هدف آن استفاده حداکثری از ضایعات و مواد بازیافتی و پس‌ماندها برای کاهش تأثیرات منفی در طول زنجیره تأمین است (Jabbarzadeh et al., 2018). یکی دیگر از وجه تمایز زنجیره تأمین پایدار با زنجیره تأمین‌های سنتی استفاده از شاخص‌های پایداری در محاسبه کارایی و عملکرد کل زنجیره تأمین است. در این پژوهش فرایندهای واحدهای تصمیم بر اساس سه حوزه اصلی رویکرد پایداری یعنی، مسئولیت اجتماعی، محیط زیست و اقتصاد طراحی و سنجیده می‌شود (Halog & Nguyen, 2017; Shokri Kahi et al., 2017).

همان طور که اشاره شد، از روش‌های تحلیل پوششی داده‌ها برای محاسبه کارایی در محیط‌های با پیچیدگی بالا که به‌دست آوردن تابع تولید امکان‌پذیر نیست استفاده می‌شود. لازم به ذکر است مدل کلاسیک DEA دو ایراد اصلی در محاسبه کارایی دارد، دیدگاه مدل‌سازی بر مبنای جعبه سیاه عیب اول است که در آن تنها ورودی‌ها و خروجی‌های کلی

واحد در نظر گرفته می‌شود پس فرایندهای درونی واحد در محاسبه کارایی بی‌تأثیر است. با در نظر گرفته شدن فرایندهای داخلی هر واحد در محاسبه کارایی محدودیت اول مدل سنتی DEA برداشته شد که از آن با تحلیل پوششی داده‌های شبکه‌ای (NDEA) یاد می‌شود. ایراد دیگر مدل سنتی تحلیل پوششی داده است است که فرض می‌کند ورودیهای یک دوره زمان به طور کامل برای تولید خروجی‌های همان دوره مورد استفاده است. با استناد به اینکه عملکرد یک سازمان در دوره‌های زمانی متوالی به یکدیگر مرتبط است، در عمل ارزیابی کارایی واحد صنعتی به صورت یک دوره زمانی مجزا و مستقل دقت پایینی خواهد داشت. لذا با توجه به اینکه تصمیمات گذشته و فعلی، عملکرد آینده را تحت تأثیر قرار خواهد داد ورودی و خروجی یک دوره بر دوره بعدی اثر گذار خواهد بود (Herrera-Restrepo et al., 2016). بسیاری از مقالات با در نظر گرفتن یک شبکه دو مرحله‌ای سعی در تشریح جزئیات فرایندهای داخلی و دقیق سازی محاسبات کارایی واحدهای تصمیم با این روش مدل‌سازی نمودند (Fang, 2020). پس برای پوشش عیب دوم تحلیل پوششی داده پویا (DDEA) توسعه یافت تا کارایی را بین چند دوره زمانی بررسی و محاسبه کند. با توجه به اینکه مورد مطالعه در این پژوهش در حوزه زنجیره تأمین پایدار است، با استفاده از نوعی از روش محاسبه کارایی بر اساس NDEA، محاسبه کارایی برای کل زنجیره تأمین صورت می‌گیرد و محاسبه کارایی و بسیاری از مسائل مربوط به تصمیم گیری چند معیاره در این مدل پیشنهادی لحاظ می‌گردد.

در کشورهایی که دارای منابع طبیعی نفت و گاز هستند فرایند استخراج نفت و انتقال آن به پالایشگاه در واحدهای پتروشیمی طی یک زنجیره تأمین دو یا چند سطحی در نظر گرفته می‌شود. در این پژوهش با درنظر گرفتن یک زنجیره تأمین دوسطحی متشکل از مرحله‌ی استخراج نفت (سطح۱) و پالایش نفت (سطح۲) با الگو برداری از مفهوم ارزیابی کارایی مبتنی بر تحلیل پوششی داده‌های شبکه‌ای (NDEA) مدل پیشنهادی ارائه شده‌است. در این مدل هر زنجیره تأمین مربوط به یک واحد پتروپالایشی به عنوان یک واحد تصمیم (DMU) در مدل شبکه‌ای حضور دارد در نتیجه علاوه بر ورودی و خروجی های ابتدا و انتهای شبکه، تبادلات بین دو سطح زنجیره نیز در محاسبات کارایی تأثیر دارد (Tone & Tsutsui, 2010).

یکی از چالش‌ها در مدیریت زنجیره تأمین ارائه راهکار بهبود عملیاتی و مقایسه

عادلانه بین واحدهای تصمیم و به تبع آن محاسبه کارایی رتبه‌بندی و ارائه راهکار بهبود است. در راستای پاسخدهی به این چالش می‌توان واحدهای مشابه و همگون را در یک دسته (یا خوشه) قرار داد، سپس محاسبات مبتنی بر تحلیل پوششی داده شبکه‌ای را داخل هر دسته به صورت مجزا انجام داد. در این روش کاراترین واحد در هر دسته به عنوان الگویی برای سایر واحدهای ناکارا در نظر گرفته می‌شود. از طرفی کاراترین واحد هر دسته کاندید جابه‌جایی به دسته بالاتر و آماده رقابت برای بهبود کارایی و پایداری خود با واحدهای تصمیم هم سطح خود در دوره‌های آتی خواهد شد. در همین راستا بهره‌گیری از روش‌های خوشه‌بندی در دستور کار قرار گرفت (Queiroz et al., 2022). در پژوهش‌های مشابه با بهره‌گیری از شبکه عصبی مصنوعی واحدهای تصمیم در خوشه‌های همگون قرار گرفتند و محاسبه کارایی با روش دو مرز کارا و ناکارا مورد ارزیابی و بررسی پژوهشگران بوده‌است (Yousefi et al., 2019). بهره‌گیری از کاربردهای شبکه عصبی مصنوعی صرفا منوط به خوشه‌بندی نیست همان طور که در مقاله (Yousefi et al., 2021) از آن در کنار تحلیل پوششی داده شبکه‌ای و پویا برای پیش‌بینی پایداری در دوره زمانی آتی نیز استفاده شده‌است. استفاده از هوش مصنوعی جهت پیش‌بینی کارایی جهت ارائه راهکار بهبود بیش از وقوع ناکارایی یکی از زمینه‌های پژوهشی شاخص در زنجیره تأمین با رویکرد پایداری است. یادگیری ماشین می‌تواند به عنوان یکی از زیرمجموعه‌های هوش مصنوعی در این بخش کاربرد داشته باشد. هدف این مقاله ارائه راهکاری ترکیبی با استفاده از یادگیری ماشین و NDEA است. نتایج نشان داد فاصله واحدهای ناکارا تا مرزکارای خوشه خود کمتر از فاصله آنها تا کاراترین واحد بین تمام واحدها است پس این روش می‌تواند مبنای مناسب‌تری جهت محاسبه کارایی نسبی و انتخاب الگو باشد که تاثیر آن در برنامه‌ریزی تولید و توسعه واحد نمایان می‌شود. بهره‌گیری از خوشه‌بندی با استفاده از یادگیری ماشین برای واحدهای پتروپالایشی در ایران برای اولین بار در این پژوهش استفاده شده‌است.

مرورادبیات و پیشینه پژوهش

مقالات مرتبط با این پژوهش به سه بخش اصلی تقسیم شده‌است. با توجه به عنوان مقاله اولویت اصلی در محاسبه کارایی خوشه‌بندی واحدهای تصمیم به دسته‌های همگون

بوده‌است که برای این قسمت تمرکز اصلی بر روی بررسی الگوریتم‌های زیر مجموعه یادگیری ماشین قرار گرفته است. در بخش بعدی هم با استفاده از تحلیل پوششی داده شبکه‌ای، ورودی و خروجی‌هایی با رویکرد پایداری در زنجیره تأمین دو سطحی صنعت پتروشیمی در ایران مورد بررسی قرار گرفت. به این ترتیب سه بخش اصلی در پیشینه پژوهش عبارت‌اند از: الف) بررسی کاربرد یادگیری ماشین در مدیریت زنجیره تأمین، ب) تحلیل پوششی داده شبکه‌ای و ج) طراحی شبکه زنجیره تأمین پایدار.

بررسی کاربرد یادگیری ماشین و هوش مصنوعی در مدیریت زنجیره تأمین

یکی از نکات قابل توجه بهره‌برداری از هوش مصنوعی، ایجاد دیدگاه توسعه در سازمان‌ها پیش از گسترش استفاده از این دانش است. اشاعه پذیرش تکنولوژی مبتنی بر هوش مصنوعی و یادگیری ماشین از ملزومات استقرار این تکنولوژی در فرایندهای مدیریت عملیات و مدیریت زنجیره تأمین به حساب می‌آید. در مقاله (Fosso Wamba et al., 2022) ضمن اشاره به کارکردهای عملی و اجرا شده در صنعت به ارتباط هوش مصنوعی با مدیریت عملیات و مدیریت زنجیره تأمین پرداخته شده است و هدف آن کمک به توسعه کاربردهای هوش مصنوعی در صنعت به ویژه کمک به تصمیم گیری، رتبه بندی، مدیریت عرضه و تقاضا و افزایش دانش و گسترش دیگاه مثبت برای بهره‌برداری هر چه بیشتر ذینفعان در این شاخه از علم تبیین شده‌است.

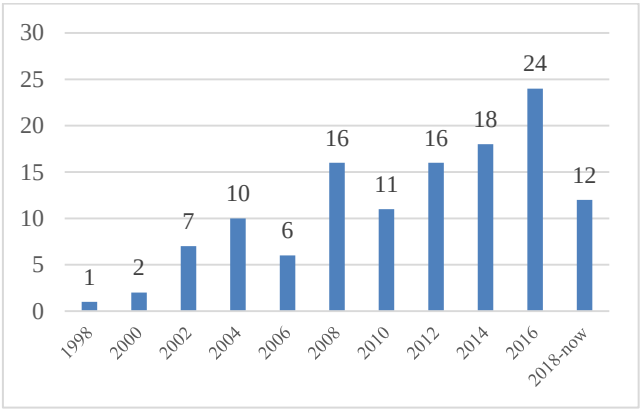
با ظهور انقلاب صنعتی چهارم ، تکنولوژی‌های نوظهور از قبیل داده‌های حجیم، هوش مصنوعی، زنجیره بلوکی، رباتیک، واقعیت افزوده، واقعیت مجازی، پرینتر سه بعدی، اینترنت اشیا و G۵ جایگاه و ارزش خود را در ارتباط بین زنجیره لجستیک و جامعه به وضوح نشان دادند. در این عصر اهمیت تطابق پذیری مدل کسب و کارها و فرایندهای لجستیک برای ایجاد محیط پویای تولید و مصرف بیش از پیش احساس می‌شود (Papadopoulos et al., 2022; Queiroz et al., 2020). در واقع نسل جدید مدیریت عملیات و زنجیره تأمین بدون حضور این تکنولوژی‌ها، غیرقابل تصور خواهد بود. نسل جدید این سامانه‌ها از سیستم پیشنهاد دهنده تولید هوشمند، قیمت گذاری پویا، ردیابی هوشمند محصول، سیستم خودکار جلوگیری از تاخیر در فرایندها، کاهش هزینه‌های انبارداری، سامانه هوشمند ارتباط لحظه‌ای با مشتری و توسعه محصول و

پایش لحظه‌ای منابع برای کاهش هزینه‌ها استفاده می‌نمایند (Dong et al., 2021). مدیران زنجیره تأمین و تصمیم گیرندگان اصلی این حوزه نیاز است که آگاهی کامل از جنبه‌های کارکرد هوش مصنوعی داشته باشند. البته توجه به این نکته حائز اهمیت است که صرف استفاده از هوش مصنوعی به تنهایی ایجاد بهره‌وری نخواهد کرد، بلکه نیاز است که تغییرات در مجموعه‌ای از استراتژی‌ها برای استقرار هوش مصنوعی در بدنه فرایندها صورت گیرد (Fosso Wamba et al., 2022). در صورتی که پیشنیازهای استقرار هوش مصنوعی در سازمان به درستی پیاده‌سازی شده باشد امکان بهره‌گیری از هوش مصنوعی و زیر مجموعه‌های آن در حوزه کنترل و برنامه ریزی تولید، پیش‌بینی تقاضا و کنترل کیفیت فراهم خواهد شد (Queiroz et al., 2022). در همین راستای از جمله کاربردهای یادگیری ماشین در پایش لحظه‌ای مشتری و بررسی فیدبک‌های آن است که با آنالیز دقیق این داده‌ها زنجیره تأمین می‌تواند به داده‌های واقعی دسترسی داشته باشد و خوشبینانه ترین پیش‌بینی را از وضعیت تولید برنامه ریزی نماید (Jacobsen & Tan, 2022). ترکیب یادگیری ماشین و داده‌های حجیم در صنعت فوتوولتائیک توانست که تأثیر بسیار زیادی در مدیریت کیفیت و افزایش کارایی سیستم تولید هوشمند با تولید دانش از داده‌های خام و با حجم بالا پیاده سازی کند (Xia et al., 2022).

رویکرد استفاده از یادگیری ماشین همگام با انقلاب صنعتی چهارم، منجر به افزایش دقت تصمیم گیری و جایگزین شدن ماشین به جای مغز انسان در تصمیم‌گیری‌های کلیدی و حساس شد که در آن دقت و سرعت عاملی تعیین کننده است. در واقع در این حالت کامپیوتر توانایی یادگیری را از داده‌های جمع آوری شده که به نوعی تجربیات پیشین است به‌دست می‌آورد (Mitchell, 1999). با تکیه بر توانایی یادگیری از داده‌ها و سرعت پردازش بالا، ماشین توانایی تشخیص الگوهای پیچیده و روابط بین متغیرهای مختلف و در نتیجه رسیدن به بهترین تصمیم را خواهد داشت (Marr, 2016). در حوزه زنجیره تأمین بهره گیری از یادگیری ماشین در تصمیم‌گیری چند معیاره و شرایط عدم قطعیت با تکیه بر توانایی تشخیص روابط غیرخطی منجر به پیش‌بینی دقیق و جلوگیری از وقوع تأثیرات اثر شلاقی خواهد شد که در نهایت نیز موجب افزایش کارایی کلی زنجیره تأمین می‌شود. یکی از نکات حائز اهمیت در استقرار یادگیری ماشین در فرایندها توجه به جایگاه آن در فرهنگ سازمان است. در پژوهشی در خصوص جایگاه

یادگیری ماشین در مدیریت زنجیره تأمین اشاره شد که اغلب پژوهش‌هایی که ازهوش مصنوعی بهره برده‌اند از شبکه عصبی مصنوعی(ANN) ، ماشین بردار پشتیبان(SVM) ، رگرسیون لجستیک (LR) و درخت تصمیم گیری (DT) استفاده کرده‌اند (Ni et al., 2020). در نتیجه در این پژوهش لزوم استفاده از یادگیری ماشین در حوزه زنجیره تأمین به عنوان زمینه پژوهشی مناسب تاکید شده‌است. توجه به قابلیت پیش‌بینی وضعیت در دوره‌های زمانی آتی همواره یکی از قابلیت‌های ارزشمند برای مدیران ناظر بر کارایی زنجیره تأمین است. پژوهشگران اغلب برای مدل‌سازی زنجیره از شبکه عصبی، یادگیری عمیق یا یادگیری ماشین، متناسب با نوع داده و حجم آن استفاده کرده‌اند. یکی از کارکردهای یادگیری ماشین در حوزه یادگیری بدون نظارت، خوشه‌بندی است که استفاده از این قابلیت در پژوهش‌های مدیریتی و به خصوص زنجیره تأمین کمتر مورد انتخاب پژوهشگران بوده‌است (Ni et al., 2020). در پژوهش حاضر با بررسی عمیق استفاده از یادگیری ماشین در خوشه‌بندی زنجیره تأمین با استناد به داده‌ها ۹۰ دوره زمانی کیفیت خوشه‌بندی بین روش‌های سه الگوریتم زیرمجموعه یادگیری ماشین بررسی و تحلیل شد تا به این سوال پاسخ داده شود که بهترین راه خوشه‌بندی با استفاده از یادگیری ماشین چیست و چگونه باید انتخاب شود؟

با مرور پژوهش‌های پیشین دیده شد، گرچه مسئله زنجیره تأمین پایدار برای صنعت پتروشیمی با روش‌های مختلف و رویکردهای متنوع بررسی شده‌است، اما به کارگیری از یادگیری ماشین در حوزه خوشه‌بندی واحدهای تصمیم در این حوزه مورد توجه پژوهشگران نبوده‌است. در ادامه با بررسی کیفیت خوشه‌بندی تعدادی از الگوریتم‌های زیر مجموعه یادگیری ماشین، بهترین روش خوشه‌بندی با استفاده از یادگیری ماشین تعیین شد. در این قسمت گزارشی از انواع الگوریتم‌های یادگیری ماشین که در حوزه زنجیره تأمین بین سال-های ۱۹۹۸ تا ۲۰۱۸ از ۱۲۳ مقاله مستخرج گردیده در شکل۱ قابل مشاهده‌است (Ni et al., 2020).



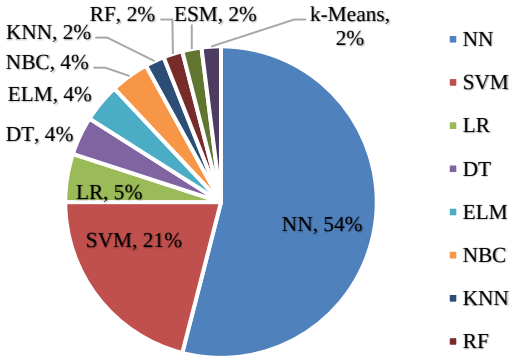
شکل۱: توزیع مقالات چاپ شده در حوزه یادگیری ماشین مرتبط با زنجیره تأمین

نکته جالب توجه رشد قابل توجه (پیک اول) استفاده از یادگیری ماشین در پژوهش‌ها حوزه زنجیره تأمین در حدود سال ۲۰۰۸ تا ۲۰۰۹ مصادف با بحران اقتصادی ثبت شده که از قابلیت یادگیری و پیش‌بینی کنندگی آن برای مدیریت عرضه و تقاضا بین بازیگران زنجیره استفاده شده‌است (Zhu et al., 2017). پیک دوم نیز در سال‌های ۲۰۱۶ تا ۲۰۱۷ مصادف با فراگیر شدن استفاده از Soft Computing در پژوهش‌های بین رشته‌ای بود. در شکل۲ به بررسی تنوع الگوریتم‌های یادگیری ماشین در حوزه مدیریت زنجیره تأمین طی ۲۰ سال اشاره شده‌است. همان طور که مشاهده می‌شود تمایل پژوهشگران برای استفاده از شبکه عصبی در بحث پیش‌بینی همواره مورد توجه بوده‌است در صورتی که در حال حاضر به دلیل به وجود آمدن انبار داده در سازمان‌ها پدیده داده حجیم که به تبع وجود پایگاه‌ها داده متعدد در صنایع و سازمان‌ها است بیش از پیش قابل مشاهده است. بهره‌گیری از داده کاوی و یادگیری ماشین به خصوص در بخش خوشه‌بندی و طبقه بندی می‌تواند بسیار کارا و موثر باشد. همان طور که در شکل ۲ قابل مشاهده است بیش از ۷۵٪ مقالات چاپ شده صرفا از شبکه عصبی (NN) و الگوریتم بردار ماشین پشتیبان (SVM) برای مقاصد مختلف که اغلب پیش‌بینی بوده‌است، استفاده کرده‌اند. در طبقه دیگر اشاره شد که از مجموعه مقالات چاپ شده در طی ۲۰ سال ۸۱٪ مقالات در دسته یادگیری نظارت شده، ۱۲٪ یادگیری بدون نظارت و تنها ۷٪ در دسته یادگیری تقویتی قرار گرفتند. از بررسی این آمار نتیجه

می‌شود که پرداختن به الگوریتم‌های زیرمجموعه یادگیری ماشین به خصوص یـادگیری بدون نظارت که خوشه‌بندی از مهمترین آن است در حوزه مدیریت زنجیره تأمین بسیار محدود بوده و زمینه بسیار مناسبی برای پژوهشگران معرفی می‌گردد.شکل۲ طبقه بندی الگوریتم‌های ML بر اساس میزان استفاده در مقالات را نشان می دهد.

تحلیل پوششی داده شبکه‌ای

بررسی فرایندهای چندگانه و داخلی در زنجیره تأمین یکی از عواملی است که منجر بـه ارائه راهکار تحلیل پوششی داده شبکه‌ای در ارزیابی کارایی واحدهای تصمیم شد. کائو و هوانگ [1](۲۰۰۸) به ارائه یک مدل مضربی و یک مدل جمعـی جهـت محاسـبه کـارایی زنجیره دو سطحی پرداختند (Chen et al., 2009; Kao & Hwang, 2008). آذر و همکاران در پژوهش با محوریت تحلیل پوششی داده شبکه‌ای در حـوزه محـیط زیسـت بـه ایـن راهکار پرداخته اند (Omid et al., 2021). در پژوهشی دیگر با ارائه یک ساختار دوسطحی به ارزیابی کارایی زنجیره تأمین در صـنعت هتلـداری پرداختـه شـد (Yin et al., 2020). پژوهشی دیگر با ارائه یک روش رتبه بندی فاصله‌ای برای ساختار دوسطحی بـا در نظـر گرفتن بهینه چندگانه اقدام به ارزیابی DMUها به روش تحلیل پوششـی داده شـبکه‌ای نمودد (Li et al., 2020). صنعت حمل و نقل هوایی و ارزیابی کارایی فرودگاه‌ها از جملـه حوزه‌هایی است که با در نظر گرفتن زنجیره تأمین چندسـطحی و رویکـرد NDEA بـه ارزیابی عملکرد واحدهای تصمیم پرداخته است (Zhang et al., 2021).



بهره‌گیری از DEA پوششی و مضربی برای ارزیابی عملکرد نسبی DMUها با ساختار

شبکه‌ای چند سطحی با ارائه مدلی مفهومی از دیگر کارهای مشابه با رویکرد شبکه‌ای است (Boloori et al., 2016). پاتریزی و همکاران مدل شبکه‌ای دوسطحی را در شرایط بازده به مقیاس متغیر با رویکردی متفاوت توسعه دادند (Patrizii, 2020). فخرموسوی و همکاران (۲۰۲۲) با در نظر گرفتن واحدهای تصمیم به فرم سیستم دو مرحله‌ای کارا در یک مجموعه مرجع مدلی بر مبنای اندازه دامنه تنظیم شده کاملا فازی با شرایط مکمل زائد قوی را مورد بررسی قرار دادند (FakhrMousavi et al., 2022). لوزانو و همکاران با بهره‌گیری از رویکرد کمترین بهبود در تحلیل پوششی داده شبکه‌ای با استفاده از دو رویکرد همکارانه و غیرهمکارانه به ارزیابی واحدهای تصمیم پرداختند (Lozano & Khezri, 2021).

طراحی شبکه زنجیره تامین پایدار

امروزه به دلیل تأثیر گازهای گلخانه‌ای و روند افزایشی گرم شدن زمین، تغییرات آب و هوایی و انتشار کربن دی اکسید، توجه به پایداری از جمله فاکتورهای اصلی در تحلیل بقای اقتصادی و اجتماعی در سازمان‌ها به شمار می‌رود (Acquaye et al., 2017). توجه به مسائل زیست محیطی و توجه به الگوهای جهانی برای جلوگیری از مشتقات کربن و حفاظت از محیط‌زیست سازمان‌ها را به سمت پیشبرد مفاهیم توسعه پایدار در صنایع سوق داده است. بر اساس این اصل مهم توجه به نیازهای فعلی در کنار حصول اطمینان از امکان توانایی بهره‌برداری نسل آتی از محیط زیست و منابع طبیعی تعریف می‌شود. پس جهت اعمال توسعه پایدار در سیاست‌های سازمان و صنایع، رویکرد اصلی اعمال محدودیت از طریق تأثیر فناوری بر منابع طبیعی در راستای حفظ محیط زیست و جلوگیری از افزایش اثرات مخرب تولید شده از صنایع تبیین شده‌است (Angelakoglou & Gaidajis, 2020). لذا محاسبه میزان پایداری در زنجیره تأمین صنایع به گونه‌ای به معنای محاسبه کارایی بر اساس شاخص‌های پایداری تعریف می‌شود. بدین ترتیب مسائل ارزیابی عملکرد محیط‌زیستی (پایداری) در سال‌های اخیر به عنوان شاخه مهمی از پژوهش در بین محققان همواره مورد توجه بوده‌است (Olafsson et al., 2014). یکی از تفاوت‌های بارز زنجیره تأمین پایدار با زنجیره تأمین سنتی توجه به جریان داده و تأثیر آن در بهبود فرایندهای زنجیره با رویکرد حفظ پایداری، کاهش ضایعات و رسیدن به

محصول با کیفیت‌تر است (Yu & Tseng, 2014). اگر بخواهیم برای زنجیره تأمین پایدار یک تعریف جامع ارائه دهیم عبارت است از:

- ایجاد زنجیره تأمین هماهنگ و تطبیق پذیر با اقتصاد که شاخص‌های کلیدی تجاری و سازمانی با رویکرد زیست محیطی و اجتماعی طراحی شود که ثمره آن مدیریت موثر و کارآمد جریان مواد مصرفی، اطلاعات و سرمایه، تولید و توزیع محصولات، بهبود سودآوری و تداوم قابلیت رقابت پذیری سازمان در برنامه‌های بلند مدت و کوتاه مدت می‌باشد (Ahi & Searcy, 2013).

در پژوهش دیگر با در نظر گرفتن نیروی کار شاغل و میزان سرمایه گذاری اولیه به عنوان ورودی و میزان تولید ناخالص و انتشار گاز دی‌اکسید کربن به عنوان خروجی عملکرد اقتصادی واحدهای صنعتی با محاسبه شاخص مالم کوئیست مورد ارزیابی قرار گرفته است (Zhou et al., 2010). سوزوکی و همکاران (۲۰۱۶) کارایی اقتصادی و زیست محیطی کشورهای حوزه اتحادیه اروپا، آسه آن و اپک را در قالب یک مدل DFM با عوامل ثابت مورد بررسی قرار داده اند که در آن متغیرهای کل مصرف انرژی و جمعیت به عنوان ورودی و تولید ناخالص داخلی و انتشار CO_2 به عنوان خروجی در نظر گرفته شد و سپس کارایی واحدها ارزیابی شد (Lenz et al., 2018). وانگ و همکاران با ارائه یک مدل SBM اقدام به بررسی بهبود کارایی ۲۵ کشور در بازه زمانی سال‌های ۲۰۱۰ تا ۲۰۱۷ نموده اند که در مدل خود نیروی کار، مصرف انرژی، سرمایه ثابت ناخالص به عنوان ورودی و میزان GDP و حجم گاز کربن دی‌اکسید به عنوان متغیرهای خروجی در نظر گرفته شده است (Wang et al., 2019).

روش تحقیق

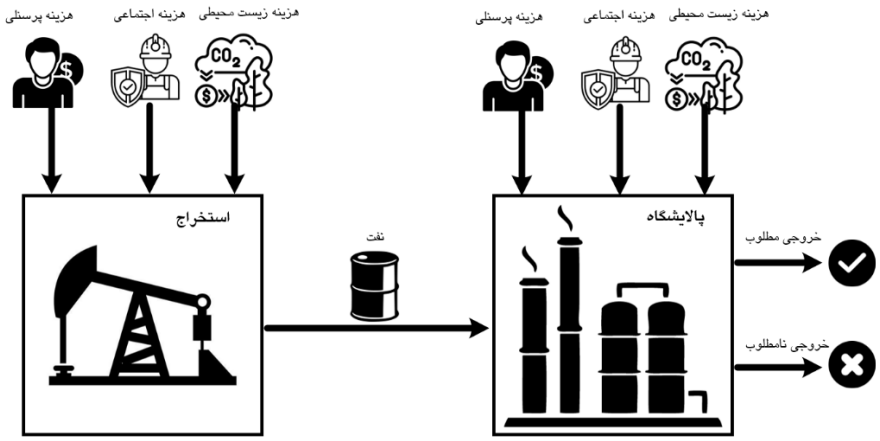
این پژوهش از بعد مخاطب، بنیادی (جامعه علمی)، از بعد هدف توصیفی و از بعد زمان مقطعی است. این پژوهش در نظر دارد تا با ارائه یک روش ترکیبی از تحلیل پوشش داده‌ها و یادگیری ماشین تاثیر خوشه‌بندی واحدهای همگون را در محاسبه کارایی نسبی بررسی نماید. مورد مطالعه ۲۸ واحد فعال پتروپالایش در ایران است که داده‌های آن در ۹۰ دوره زمانی جمع آوری شده است. با توجه به اینکه این روش در کلیه صنایع و برای سایر زنجیره تامین‌های فعال در صنعت کاربرد دارد، این پژوهش از بعد مخاطب

بنیادی است. در این مقاله ابتدا به کاربرد الگوریتم‌های زیر مجموعه یادگیری ماشین در خوشه‌بندی واحدهای همگن پرداخته شده و کیفیت خوشه‌بندی هر روش با استفاده از شاخص‌های استاندارد ارزیابی کیفیت خوشه‌بندی به تفکیک بررسی و مقایسه گردیده‌است. در مطالعات پیشین، کاربرد یادگیری ماشین در حوزه زنجیره تأمین بیشتر معطوف به پیش‌بینی عرضه و تقاضا بوده در صورتی که این پژوهش در خوشه خوشه‌بندی واحدها انجام شده است. در بخش بعدی ارزیابی عملکرد زنجیره تامین با رویکرد پایداری در نظر گرفته شد. نتایج در دو حالت مورد بررسی بدین شرح است: حالت اول استفاده از تحلیل پوششی داده بر روی کل داده‌ها و محاسبه میزان ناکارایی و حالت بعدی اجرای آن در هر خوشه به صورت مجزا و اندازه‌گیری مجدد میزان ناکارایی که در نهایت تمامی نتایج در بخش یافته‌ها بررسی شده‌است. داده‌های این پژوهش از بانک اطلاعاتی شرکت دانش بنیان سامانه اندیشمند پگاه استخراج گردید که یکی از فعالان نرم افزاری در حوزه زنجیره تأمین صنایع کلیدی در کشور است. در شکل۳ روند اجرای پژوهش قابل مشاهده است.

شکل۳ : روش شناسی پژوهش

طراحی شبکه زنجیره تأمین پایدار در صنعت پتروشیمی

در ابتدا به فرایند تولید محصولات پتروشیمی پرداخته می‌شود، در سطح اول که فاز استخراج در نظر گرفته شده‌است ابتدا نفت خام از میادین نفتی استخراج شده و در سطح دوم به همراه سایر ورودی‌هایی که از جنس پایداری در نظر گرفته شده‌اند به پالایشگاه وارد می‌شوند. در این بخش نیز خروجی سطح دو به خروجی‌های مطلوب و نامطلوب تقسیم می‌شود که در نهایت هر واحد تصمیم که یک زنجیره تأمین مجزا دوسطحی در این مقاله تعریف شده‌است، سعی در افزایش خروجی مطلوب و کاهش خروجی نامطلوب دارد.



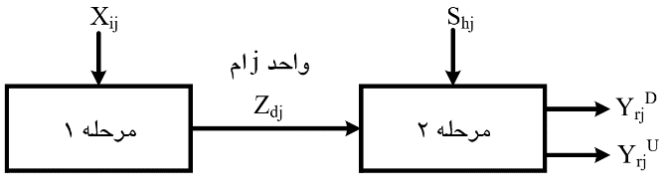
شکل۴ : زنجیره تأمین دو مرحله‌ای مدل پیشنهادی

لازم به ذکر است خروجی‌های مطلوب شامل کلیه مشتقات نفتی قابل مصرف در انتهای زنجیره تأمین در نظر گرفته شده‌است، برای مثال محصول‌های نهایی که بدون نیاز به فرایند مجدد قابل مصرف هستند عبارت‌اند از: کودهای شیمیایی، پلی‌اتیلن، اسیدها و حلال‌های شیمیایی و خوراک قابل مصرف در صنایع دارویی و سایر صنایع وابسته که بالاخره توسط مصرف کننده یا یک واحد صنعتی ثانویه قابل مصرف باشد. ***(لازم به ذکر است با توجه به تنوع محصولات خروجی واحدهای پتروشیمی جهت قابل قیاس بودن آنها، حجم معادلات آنها به عنوان درآمد حاصله از فروش محصولات با واحد دلار به عنوان خروجی مطلوب در نظر گرفته شده است)*** خروجی نامطلوب نیز عبارت است از ضایعات نفتی که قابل مصرف نبوده و نیاز است جهت پالایش مجدد به چرخه پالایش برگشت داده شود یا به عنوان زباله و دور زیر از چرخه خارج شود که در نهایت هدف کاهش این موارد خواهد بود ***(این خروجی نیز به واحد دلار در مجموعه داده در نظر گرفته شده‌است)***. در شکل ۴ شمایی از زنجیره تأمین دوسطحی پیشنهادی در این پژوهش قابل مشاهده است. ***(ورودی‌های سه گانه نیز از جنس هزینه و با واحد دلار در نظر گرفته شده‌اند(***

بررسی متغیرهای ورودی و خروجی

همان طور که در شکل ۴ قابل مشاهده است، ساختار دوسطحی مورد مطالعه در هر مرحله دارای ۳ ورودی است، این ورودی‌ها با ارجاع به ادبیات پژوهش در بخش زنجیره تأمین پایدار انتخاب شده‌اند و عبارت‌اند از:

- **هزینه‌های پرسنلی:** این ورودی به عنوان ورودی از جنس اقتصادی در هر مرحله در نظر گرفته شده و شامل کلیه هزینه پرسنلی اعم از کارمندی، مهندسی و سایر بخش‌های هر یک از واحدهای استخراج یا پالایشگاه می‌شود.
- **هزینه‌های ایمنی:** عبارت است کلیه هزینه‌هایی که صرف ایمنی کارگران و پرسنل اعم از هزینه HSE، بیمه و بهداشت که از بعد اجتماعی بوده و از جنس هزینه با رویکرد مسئولیت اجتماعی ماهیت ورودی دارد.
- **هزینه‌های زیست محیطی:** کلیه اقداماتی که در راستای کاهش CO2 و حفظ محیط زیست و جلوگیری از پخش و انتشار ضایعات و انواع آلودگی‌ها در محیط شده به صورت مجموعه‌ای از هزینه‌ها در نظر گرفته می‌شود. ضمنا این عامل در حوزه پایداری به عنوان عامل محیط زیستی در زنجیره تأمین ایفای نقش می‌کند.



شکل۵ : ساختار دوبخشی شبکه پیشنهادی

متغیرهای خروجی در شبکه به دو دسته خروجی مطلوب و نامطلوب تقسیم شده‌اند. خروجی مطلوب عبارت است از هر محصولی که پس از خروجی از برج‌های تقطیر قابل مصرف برای مصرف کننده باشد یا خود به عنوان خوراک برای واحد صنعتی دیگری قابل استفاده باشد که در ساختار دو سطحی به عنوان مجموع درآمد حاصل از این محصولات در نظر گرفته شده‌است. خروجی نامطلوب عبارت است از کلیه پسماندها و ضایعاتی که یاد قابل استفاده نبوده یا نیاز است برای پالایش مجدد به چرخه پالایشگاه

برگشت داده شود که این خروجی هم به صورت مجموع هزینه از دست رفته در زنجیره در نظر گرفته می‌شود.***(لازم به ذکر است با توجه به تفاوت تنوع محصولات واحدهای پتروشیمی جهت قیاس پذیر بودن حجم تولیدات به نسبت ورودی مصرفی هر دو خروجی مطلوب و نامطلوب با واحد دلار درنظر گرفته شده‌اند.)***

نفت خام حاصل از فرایند استخراج در مرحله اول نیز به عنوان خروجی مرحله (یا سطح) یک و ورودی مرحله (یا سطح) دو در نظر گرفته می‌شود که ماهیت دوگانه داشته و به عنوان متغیر داخلی شبکه حضور دارد. جهت معرفی متغیرها در معادلات تحلیل پوششی داده شبکه‌ای شکل ۵ خلاصه شده، ساختار دوسطحی را به صورت ساده تر نمایش داده است. در ساختار ارائه شده در شکل ۵، ورودی‌های مرحله اول با X نمایش داده شده‌اند که در صورت بررسی واحد j ام ورودی‌های مرحله اول به صورت X_{ij} در معادله در نظر گرفته خواهد شد که به معنای ورودی i ام مرحله ۱ از واحد j ام تعریف می‌شود. به همین ترتیب ورودی‌های مرحله ۲ با S نمایش داده شده‌است. خروجی مرحله ۱ که به نوعی یکی از ورودی‌های مرحله ۲ است با Z نمایش داده شده که Z_{ij} که در واقع خروجی مرحله ۱ واحد تحت بررسی j ام در نظر گرفته می‌شود. خروجی‌های مرحله ۲ که در واقع خروجی نهایی زنجیره هستند با Y^D به عنوان خروجی مطلوب و Y^U به عنوان خروجی نامطلوب نمایش داده شده‌اند. جدول ۱ متغیرهای مدل پیشنهادی را نمایش می‌دهد.

جدول۱: متغیرهای مدل پیشنهادی

متغیر	شرح
x_{ij}	ورودی i ام مرحله ۱ از واحد تحت بررسی j ام
z_{dj}	خروجی d ام مرحله ۱ از واحد تحت بررسی j ام
s_{hj}	ورودی h ام مرحله ۲ از واحد تحت بررسی j ام
y_{rj}^{U}	خروجی نامطلوب مرحله ۲ از واحد تحت بررسی j ام
y_{rj}^{D}	خروجی مطلوب مرحله ۲ از واحد تحت بررسی j ام
w_d	ضرایب خروجی مرحله اول (Zdj)
v_i	ضرایب ورودی مرحله اول (Xij)
q_h	ضرایب ورودی مرحله دوم (Shj)
u_r	ضرایب خروجی مطلوب مرحله دوم (yrjD)

برای محاسبه کارایی زنجیره تأمین از روش تحلیل پوششی داده شبکه‌ای چندین مدل ارائه شده‌است که دو مدل اصلی آن فرم جمعی و ضربی است. در محاسبه کارایی به فرم جمعی کارایی کل به صورت مجموع وزن‌دار کارایی دو مرحله محاسبه می‌شود به گونه‌ای که جمع ضرایب کارایی هر سطح برابر با یک در نظر گرفته شود ($e^* = w_1 e_1^{o*} + w_2 e_2^{o*} : w_1 + w_2 = 1$) که درواقع محاسبه اوزان کارایی هر مرحله به روش‌های مختلف در مدل‌های ریاضی تعریف شده‌است. این اوزان به گونه‌ای ضریب اهمیت و تأثیر کارایی هر مرحله در محاسبه کارایی کل زنجیره تأمین را معین می‌کند (Cook & Zhu, 2014). در محاسبه کارایی به فرم ضربی، کارایی کل برابر ضریب کارایی مرحله اول در مرحله دوم خواهد بود. بدین ترتیب ابتدا طبق معادله (۱) کارایی مرحله اول محاسبه می‌شود. سپس کارایی مرحله دوم با در نظر گرفتن ورودی و خروجی‌های وارد شده به مرحله دوم ضمن اعمال قید ثابت بودن مقادیر کارایی قبلی محاسبه شده است. پس از محاسبه کارایی هر مرحله برای به‌دست آوردن کارایی کل زنجیره تأمین کافی است دو عدد به‌دست آمده در هم ضرب شوند. نکته حائز اهمیت در این محاسبات توجه به اهمیت و اولویت بین مرحله ۱ و ۲ است. در صورتی که مرحله ۲ نسبت مرحله ۱ اولویت داشته باشد می‌توان ابتدا طبق معادله (۵) کارایی واحد دوم را محاسبه کرد و سپس قید ثابت بودن این کارایی مرحله دوم را در محاسبه کارایی کل دخیل کرد. بدین ترتیب با تقسیم کارایی کل بر روی کارایی مرحله دوم مقدار کارایی مرحله اول نیز قابل محاسبه خواهد بود. لازم به ذکر است این معادلات کارایی در ابتدا به فرم کسری نوشته شده بودند که با استفاده از تبدیل چارنز و کوپر به فرم معادله برنامه‌ریزی خطی به فرم (۱) تا (۵) تبدیل شدند.

بر اساس مدل پیشنهادی در شکل ۵ زنجیره مورد بررسی دارای خروجی مطلوب و نامطلوب است که نیاز است در محاسبات کارایی به روش NDEA لحاظ شود. برای به‌دست آوردن بیشینه کارایی نیاز است تا همزمان خروجی‌های مطلوب افزایش و خروجی‌های نا مطلوب کاهش یابند. به همین ترتیب جهت دخیل کردن خروجی‌ها نامطلوب در سنجش کارایی در انتهای معادل (۴) با استفاده از تبدیل ریاضی ماهیت خروجی نامطلوب به گونه‌ای تغییر پیدا کرد که میتوان آن را به عنوان یک خروجی مطلوب تحت عنوان $\hat{y}^D$ در محاسبات در نظر گرفت به این ترتیب $\hat{y}^D$ به عنوان یک y^D

در محاسبات مدل نهایی تأثیر گذار است. لازم به ذکر است مقدار b باید به اندازه‌ای مثبت و بزرگ در نظر گرفته شود که $\hat{y}^D = -y^U + b$ همواره مقدار مثبتی داشته باشد(Cook & Zhu, 2014). با استفاده از این تبدیل در صورتی که تلاش در افزایش $\hat{y}^D$ باشد در واقع مقدار اصلی خروجی نامطلوب در کمترین حالت خود قرار می‌گیرد، پس در مدل نهایی دو خروجی در نظر گرفته شد و حالت بهینه افزایش هر دو خروجی در نظر گرفته می‌شود. در بسیاری از مدل‌های پیشین کارایی کل به روش تحلیل پوششی داده شبکه‌ای بدون در نظر گرفتن عوامل نامطلوب پیاده‌سازی می‌شد در این مقاله جهت مقایسه نتایج به‌دست آمده و تأثیر خروجی نامطلوب در محاسبه کارایی زنجیره تأمین نتایج به سه روش پیاده‌سازی و مقایسه شده‌است (با فرض بازده به مقیاس ثابت):

- محاسبه کارایی زنجیره تأمین بدون در نظر گرفتن عوامل نامطلوب با استفاده از فرم ضربی
- محاسبه کارایی زنجیره تأمین با در نظر گرفتن همه عوامل با استفاده از فرم ضربی
- محاسبه میزان ناکارایی زنجیره تأمین با استفاده از مدل تابع فاصله جهت دار (DDF)

$$e^* = e_1^{o*} \times e_2^{o*} \qquad (1)$$

$$e_1^{o*}: \quad \max \sum_{d=1}^{D} w_d z_{do}$$
$$s.t. \quad \sum_{i=1}^{I} v_i x_{io} = 1$$
$$\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{I} v_i x_{ij} \le 0 \quad \forall j$$
$$v_i, w_d \ge 0$$

$$e_2^{o*}: \quad \max \sum_{r=1}^{R} u_r y_{ro}^D \qquad (2)$$
$$s.t. \quad \sum_{h=1}^{H} q_h s_{ho} + \sum_{d=1}^{D} w_d z_{do} = 1$$
$$\sum_{r=1}^{R} w_d z_{do} - e_1^{o*} \sum_{i=1}^{I} v_i x_{io} = 0$$
$$\sum_{r=1}^{R} u_r y_{rj}^D - \sum_{h=1}^{H} q_h s_{hj} - \sum_{d=1}^{D} w_d z_{dj} \le 0$$
$$\sum_{d=1}^{D} w_d z_{dj} - \sum_{i=1}^{I} v_i x_{ij} \le 0 \quad \forall j$$
$$u_r, q_h, w_d, v_i \ge 0 \quad \forall i,d,r,h$$

$$e^*: \quad \max \sum_{r=1}^{R} u_r y_{ro}^D \qquad (3)$$
$$s.t. \quad \sum_{i=1}^{I} v_i x_{io} + \sum_{h=1}^{H} q_h s_{ho} = 1$$
$$\sum_{r=1}^{R} u_r y_{rj}^D - \sum_{h=1}^{H} q_h s_{hj} - \sum_{i=1}^{I} v_i x_{ij} \le 0$$
$$u_r, q_h, w_d, v_i \ge 0 \quad \forall i,d,r,h$$

$$\hat{y}^D = -y^U + b \quad bias: b = positive\ value \qquad (4)$$

$$\pi_2^{o*} = \max \sum_{r=1}^{R} u_r y_{ro}^D \qquad (5)$$
$$s.t. \quad \sum_{r=1}^{R} u_r y_{rj}^D - \sum_{h=1}^{H} q_h s_{hj} - \sum_{d=1}^{D} w_d z_{dj} \le 0 \quad \forall j$$
$$\sum_{h=1}^{H} q_h s_{ho} + \sum_{d=1}^{D} w_d z_{do} = 1$$
$$w_d, q_h, u_r \ge 0, \forall d,h,r$$

یکی از روش‌های بررسی عوامل نامطلوب استفاده از مدل DFF یا تابع فاصله جهت

دار است، در این مدل ضمن فرض بازده به مقیاس ثابت (CRS)، استفاده از فرض دسترسی پذیری قوی می‌توان همزمان در جهت افزایش ورودی مطلوب و کاهش خروجی نامطلوب به صورت همزمان حرکت کرد. مقدار هدف در این مدل برنامه‌ریزی خطی برابر θ* (بهترین مقدار برای تتا) محاسبه می‌شود که میزان ناکارایی هر واحد را نشان می‌دهد. پس در این مدل درصورتی که مقدار نهایی θ = ۰ به‌دست آید، واحد کارا در نظر بوده و هر زنجیره تأمین به صورت یک بلک باکس در نظر گرفته می‌شود. لازم به ذکر است جهت مشاهده تأثیر استفاده از خوشه‌بندی در محاسبات مدل سه گانه مطرح شده هر یک به صورت مجزا در دوحالت پیاده‌سازی و مقایسه می‌شوند:

- **حالت اول)** مقایسه کارایی کل بین تمام واحدها (بدون خوشه‌بندی)
- **حالت دوم)** مقایسه کارایی واحدها در هر خوشه (به صورت مجزا)

در نهایت محاسبات برای ۲۸ واحد تصمیم (DMU) که در این پژوهش زنجیره تأمین صنایع پتروشیمی در ایران در نظر گرفته شده‌است. طبق سه مدل و هر یک در دو حالت بدون خوشه‌بندی و با خوشه‌بندی در نرم افزار گمز (GAMS) پیاده‌سازی شد و نتایج مورد تحلیل و بررسی قرار گرفت. ضمن توجه به ماهیت مدل سنتی تحلیل پوششی داده به فرم مدل CCR بهترین وزن‌ها برای تمامی ورودی و خروجی‌ها به گونه‌ای انتخاب می‌شود که هر واحد در بهینه‌ترین حالت خود در محاسبه کارایی درنظر گرفته شود، البته با قید اینکه هیچ کارایی بالاتر از ۱ در نظر گرفته نشود.

$$\max \quad \theta$$

$$s.t. \quad \sum_{j=1}^{n} \lambda_j Y_j^D \geq Y_o^D (1+\theta)$$

$$\sum_{j=1}^{n} \lambda_j Y_j^U \leq Y_o^U (1-\theta)$$

$$\sum_{j=1}^{n} \lambda_j X_j \leq X_o$$

$$\lambda_j \geq 0, \quad j = 1,...,n \quad , \theta \geq 0$$

پیش از محاسبه کارایی در نرم افزار گمز، ابتدا انواع روش‌های یادگیری ماشین در خوشه‌بندی واحدها به همراه فرایند ارزیابی کیفیت خوشه‌بندی پیاده‌سازی شد و بهترین الگوریتم خوشه‌بندی زیر مجموعه یادگیری ماشین برای مجموعه داده مورد

مطالعه انتخاب گردید. دلیل تمرکز بر روی یادگیری ماشین جهت خوشه‌بندی کمبود استفاده از یادگیری ماشین برای خوشه‌بندی زنجیره تأمین است که به عنوان یکی از شکاف‌های دانش مورد مطالعه این پژوهش در نظر گرفته شد. در قسمت بعد انواع روش های خوشه‌بندی و شاخص‌های ارزیابی کیفیت خوشه‌بندی معرفی شده‌است.

خوشه بندی با استفاده از یادگیری ماشین

تحلیل خوشه‌بندی و تحلیل طبقه بندی به عنوان زیر مجموعه‌ها، یادگیری ماشین به حساب می‌آیند. خوشه‌بندی در دسته یادگیری بدون نظارت و طبقه‌بندی در دسته یادگیری با نظارت تفکیک شده‌اند. خوشه‌بندی فرایندی است که در آن مجموعه اعضا به دسته‌هایی تقسیم می‌شوند که اعضای هر دسته بیشترین شباهت را به یکدیگر داشته باشند و از طرف دیگر اعضای هر دسته بیشترین تفاوت را نسبت به سایر دسته‌ها داشته باشند. با توجه به اینکه در خوشه‌بندی برچسبی برای توصیف دسته‌ها وجود ندارد، اعضا بر اساس ویژگی‌های مشترک و روش‌های اندازه‌گیری فاصل یا بیشترین شباهت بین اعضا به صورت خودکار در یک دسته (یا خوشه) قرار می‌گیرند. توصیف هر خوشه نیز پس از اتمام خوشه‌بندی بر اساس مشترکات اعضا و تفاوت بین خوشه‌ها تفسیر و تحلیل می‌شود. روش‌های متعددی در خوشه‌بندی وجود دارد که اجرا هر کدام از آنها بر روی مجموعه داده نتایج منحصر بفردی را حاصل می‌کند. در این مقاله سه الگوریتم از زیر مجموعه‌های یادگیری ماشین برای خوشه‌بندی استفاده شده و سپس نتایج و کیفیت خوشه‌بندی بین این سه روش بررسی گردیده است. هرچند اکثر الگوریتم‌های خوشه‌بندی مبنای یکسانی برای محاسبه بیشترین شباهت یا فاصله دارند و در جزئیات محاسبات ریاضی تفاوت‌هایی دیده می‌شود. استفاده از فاصله اقلیدسی، منهتن و مینکوفسکی از جمله تفاوت‌های محاسباتی بین انواع روش‌های خوشه‌بندی است. در ادامه به بررسی روش‌های K-Means، Hierarchial Clustering و Deep Embedded Clustering که زیر مجموعه یادگیری ماشین است، می‌پردازیم.

- **خوشه‌بندی K میانگین** (K-Means): یکی از محبوب‌ترین و پرکاربردترین روش‌های خوشه‌بندی در حوزه علوم داده و داده‌کاوی بر مبنای یادگیری بدون نظارت است. در این روش معمولا داده‌ها و متغیرهای چند بعدی با چند ویژگی

وجود دارند. در مورد مطالعه این پژوهش زنجیره تأمین صنایع پتروشیمی به فرم شبکه دوسطحی با چند ورودی و چند خروجی مفروض است که در نهایت آنچه اهمیت دارد، اندازه‌گیری بر مبنای میزان شباهت (که نوعی فاصله تعریف می‌شود) بین اعضای درون یک خوشه و بیشترین تفاوت (از نظر فاصله) بین دسته‌ها است. این فاصله می‌تواند با انواع محاسبات ریاضی تعریف شود یکی از پرکاربردترین فاصله‌ها در خوشه‌بندی به روش K-Means فاصله اقلیدسی است که در نهایت هدف کمینه بودن فاصله اعضای خوشه i ام از مرکز دسته و بیشینه بودن فاصله اقلیدسی مراکز دسته‌ها از یکدیگر می-باشد. به این روش خوشه‌بندی k-میانگین با الگوریتم لوید هم گفته می‌شود (Popat & Emmanuel, 2014; Rai & Singh, 2010).

- **خوشه‌بندی سلسله مراتبی** (Hierarchial Clustering): از جمله روش‌های یادگیری ماشین بدون نظارت است که رویکردی بالا به پایین دارد. در این روش هر خوشه ممکن است در بیش از یک خوشه قرار گیرد زیرا امکان دارد خود، زیرمجموعه خوشه بزرگتر در سطحی دیگر از فاصله قرار گرفته باشد. یکی از مهمترین نکات در پیاده‌سازی این روش انتخاب ویژگی‌های مناسب داده برای شروع محاسبات است که بحث استانداردسازی داده‌ها در این بخش برای جلوگیری از افزایش انحراف تابع فاصله مورد توجه پژوهشگران علوم داده قرار خواهد گرفت. درنهایت این روش با تمرکز بر ویژگی‌های تعیین شده توسط پژوهشگر بیشتر شباهت بین داده‌ها به صورت سلسله مراتبی مشخص می‌نماید که در صورت تعیین تعداد دسته‌ها نتیجه خوشه‌بندی در هر سطح قابل مشاهده خواهد بود (Murtagh & Contreras, 2012).
- **خوشه‌بندی بر مبنای یادگیری عمیق** (Deep Embedded Clustering): در حالتی که تعداد ویژگی‌ها و ابعاد داده زیاد باشد و حجم مناسبی از داده برای پیاده‌سازی یادگیری عمیق در دسترس باشد استفاده از این روش مناسب است. در این روش با بهره‌گیری از شبکه عصبی عمیق، DEC آموزش می‌بیند که بتواند داده‌ها با ابعاد بالا را به فرمی با ابعاد کاهش یافته تبدیل کند تا بتواند خوشه‌بندی را گونه‌ای ملموس‌تر پیاده‌سازی کند. استفاده از این روش بهبود

قابل توجهی در داده‌های با ابعاد زیاد داشت کـه در حـال حاضـر در بسـیاری از علوم که دارای داده‌های پیچیده هسـتند و نیـاز بـه خوشـه‌بنـدی دارنـد مـورد استفاده قرار می‌گیرد (Min et al., 2018; Xie et al., 2016). در قسمت بعد به دو بخش اصلی می‌پردازیم که در ارزیابی نتایج خوشـه‌بنـدی بسـیار حـائز اهمیـت هستند.

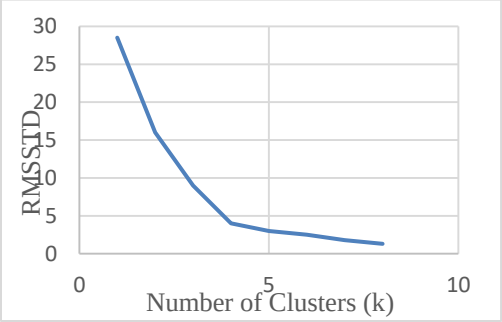
تعیین بهترین تعداد خوشه

برای تعیین بهترین تعداد خوشه (مقدار k) می‌توان از شاخص ریشـه میـانگین مربعـات انحراف استاندارد (RMSSTD) استفاده کرد. این شاخص میـزان شـباهت اعضـای درون یک خوشه را اندازه‌گیری می‌کند. برای به‌دست آوردن بهترین مقدار k، کافی است مقدار RMSSTD را به ازای مقادیر مختلف k محاسبه و در یک نمودار بـه فـرم خـط شکسـته رسم کرد. در شکل ۶ این نمـودار بـرای داده‌هـای مسـئله مـورد بحـث در مقالـه رسـم شده‌است، طبق این نمودار با افزایش مقدار k، میزان RMSSTD روند نزولی پیـدا کـرده است. نمودار در یک نقطه دچار شکستگی قابل توجه شده که آن را به فرمی شبیه آرنج دست شبیه کرده است، به این نقطه Elbow یا آرنج گفته می‌شود. مقادیر پـس از نقطـه elbow کاهش چندانی را در میزان RMSSTD ایجاد نمی‌کنـد. در نتیجـه مقـدار بهینـه برای تعداد خوشه‌ها در نقطه آرنج برآورد می‌گـردد. از ایـن روش بـرای تعیـین بهتـرین خوشه در مسائل مربوط به خوشه‌بندی استفاده می‌شود. با توجه به نمودار به‌دست آمده تعداد خوشه‌ها ۴ برآورد گردیده است.

برای محاسبه RMSSTD مقادیر زیر بدین ترتیب تعریف می‌شود:

- SS: مجموع مربعات درون گروه
- n: تعداد داده‌ها (یا همان نقاط)
- p: تعداد ابعاد (یا ویژگی‌ها)
- k: تعداد خوشه‌ها

$$RMSSTD = \sqrt{\frac{SS}{p(n-k)}}$$



شکل۶ : نمودار آرنج – تشخیص بهترین مقدار برای k

روش‌های ارزیابی نتایج خوشه‌بندی

ارزیابی نتیجه خوشه‌بندی (که یک فرایند بدون نظارت است) به عنوان گامی مهم در روند پژوهش تلقی می-شود. در حالت کلی دو معیار ارزیابی نتایج خوشه‌بندی وجود دارد. الف) معیارهای ارزیابی درونی (Internal Criteria Index) و ب) معیارهای ارزیابی بیرونی (External Criteria Index). معیارهای ارزیابی درونی سعی در اندازه‌گیری میزان شباهت اعضا خوشه و تفاوت آنها با سایر خوشه‌ها دارند. استفاده از روش تحلیل واریانس یکی از راهکارهای ارزیابی است، پس بر اساس این معیار آماری هدف به‌دست آوردن حالتی از خوشه‌بندی است که میزان پراکندگی درون خوشه‌ای کمینه (یعنی بیشترین شباهت بین اعضا) و میزان پراکندگی بین خوشه‌ای بیشینه (یعنی بیشترین تفاوت بین سایر خوشه‌ها) شود. در صورتی که تعداد ابعاد داده بیشتر از یک باشد نیاز است تا مجموع پراکندگی بین ابعاد تحت عنوان پراکندگی کل محاسبه گردد. در ادامه به بررسی سه شاخص ارزیابی استفاده شده در این پژوهش پرداخته خواهد شد:

الف) شاخص نیم‌رخ (Silhouette Score): این معیار هم به پیوستگی درون خوشه‌ای و هم به تفکیک پذیری بین خوشه‌ها بستگی دارد. در واقع مقدار شاخص سیلوت میزان تعلق به هر خوشه را در مقایسه با خوشه مجاور محاسبه می‌کند. برای فهم دقیق این شاخص نیاز به بررسی دو مفهوم کلیدی است:

- میانگین فاصله یک نقطه از خوشه با نقاط دیگر آن خوشه: که می‌تواند ملاکی برای ارزیابی تعلق نقطه x در خوشه‌اش در نظر گرفته شود. طبیعتا هر چه این مقدار

کمتر باشد نشان دهنده تعلق x به این خوشه است. این مقدار با a(i) به صورت زیر محاسبه می‌گردد:

$$a(i) = \frac{1}{n}\sum_{l=1}^{n} d(x,l)$$

- حداقل میانگین فاصله نقطه با خوشه‌های دیگر: فرض کنید نقطه x متعلق به خوشه ۱ باشد. حال میانگین فاصله این نقطه با نقاط خوشه‌های دیگر اندازه گرفته می شود، خوشه‌ای که دارای کمترین میانگین فاصله برای نقطه x باشد به عنوان خوشه مجاور در نظر گرفته میشود. مقدار میانگین فاصله برای نقطه x با سایر نقاط مجاور با b(i) نمایش داده می‌شود که Ck در آن خوشه k ام است.

$$b(i) = \min \frac{1}{n} \sum_{y_m \in C_l} \left(d(x, y_m)\right)$$

اگر مقدار a(i) کمتر از b(i) باشد مقدار Silhouette Score مثبت و اگر این شرط برقرار نباشد مقدار آن منفی خواهد شد و منفی شدن مقدار این شاخص نشان دهنده خوشه‌بندی ضعیف است زیرا در صورت منفی شدن نقطه x بیش از آنکه شبیه اعضای هم خوشه‌ای خود باشد شبیه خوشه مجاور شده که نشان دهنده کیفیت پایین نتیجه خوشه‌بندی است. مقدار این شاخص عددی بین ۱- و ۱+ است که در چه این مقدار با ۱ نزدیک تر باشد نشان دهنده کیفیت بالاتر خوشه‌بندی است. کم شدن این شاخص می تواند دلایل مختلفی داشته باشد، عدم انتخاب درست مقدار k، شباهت بسیار زیاد و تفکیک پذیر نبودن فضای داده یا عدم انتخاب بهترین الگوریتم برای خوشه‌بندی می تواند از جمله این عوامل تعیین کننده در مقدار معیار نیم‌رخ باشد. شاخص نیم‌رخ (سیلوت) برای نقطه x از رابطه زیر تعریف می‌شود که اگر میانگین این مقدار برای نقاط هر خوشه محاسبه گردد معیار ارزیابی هر خوشه به‌دست آمده همچنین میانگین کل مقادیر نیم‌رخ برای تمام خوشه‌ها نتیجه ارزیابی عملیات خوشه‌بندی را محاسبه می‌کند:

$$s(i) = \frac{b(i) - a(i)}{\max\left(b(i), a(i)\right)}$$

ب) شاخص دیویس – بولدین (Davies-Bouldin Score): برای محاسبه این شاخص نیاز است که دو معیار اندازه پراکندگی (که فرمول محاسبه آن شبیه فاصله مینکوفسکی نقاط هر خوشه از مرکز آن است و با Si نمایش داده شده) و عدم شباهت

بین خوشه‌ها (فاصله بین مرکز دو خوشه که با Dij نمایش داده شده) محاسبه گردد. سپس همان طور که در فرمول زیر دیده می‌شود مقدار Rij بدین صورت محاسبه می شود. طبق این فرمول در صورت میزان پراکندگی دو خوشه باهم جمع و در مخرج نیز عدم شباهت بین دو خوشه قرار گرفته است. هر چه خوشه‌ها پراکندگی بیشتری داشته باشند مقدار Rij بزرگتر و هرچه دو خوشه از هم فاصله کمتری داشته باشند هم مجدد بر روی افزایش Rij تأثیر مثبت می‌گذارد.

$$R_{ij} = \frac{S_i + S_j}{D_{ij}}$$

پس برای محاسبه شاخص دیویس-بولدین در یک روش خوشه‌بندی کافی است ابتدا بیشینه فاصله هر خوشه نسبت به سایر خوشه‌ها به‌دست آید که آن را با Ri نمایش می دهیم. سپس میانگین Ri ها برای همه خوشه‌های ایجاد شده توسط الگوریتم خوشه بندی محاسبه می‌شود که در حقیقت میانگین حداکثر نسبت پراکندگی درون به پراکندگی بین خوشه‌ها خواهد بود و هر مقدار آن کمتر باشد نشان دهنده کیفیت بالاتر عملیات خوشه‌بندی است.

$$DB\ score = \frac{\sum_{i=1}^{k} R_i}{k}$$

ج) شاخص کالینسکی-هاراباسز (Calinski-Harabasz Score): عبارت است از نسبت واریانس به نسبت میان پراکندگی درون خوشه‌ای و پراکندگی بین خوشه‌ای. به بیان دیگر به آن معیار نسبت واریانس (VRC) نیز گفته می‌شود که در این فرمول SSb مجموع واریانس بین خوشه‌ها و SSw مجموع واریانس درون خوشه‌ها در نظر گرفته شده‌است. مقادیر n و k هم به ترتیب تعداد اعضا نمونه و تعداد خوشه‌ها است:

$$VRC_k = \frac{SS_b}{SS_w} \times \frac{(n-k)}{k-1}$$

یکی دیگر از چالش‌های یادگیری ماشین در خوشه‌بندی حداکثر تعداد دفعات اجرای الگوریتم و تعداد دفعاتی است که الگوریتم به نقاط مختلفی به عنوان مرکز خوشه اجرا می‌شود که هدف آن رسیدن به بهترین نتیجه لازم در خوشه‌بندی است. لازم به ذکر است با توجه به اهمیت ارزیابی الگوریتم‌های خوشه‌بندی زیرمجموعه

یادگیری ماشین، کلیه توابع شاخص‌های معرفی شده در این بخش در کتابخانه scikit learn و Tensorflow موجود است که با استفاده از زبان برنامه نویسی پایتون پیاده‌سازی شده‌است. در بخش بعد هر کدام از الگوریتم اصلی معرفی شده با استفاده از زبان پایتون و در محیط VScode اجرا شده و مقایسه کیفیت عملکرد آن‌ها بر روی داده‌های مورد مطالعه مدنظر این پژوهش با استفاده از شاخص‌های سه گانه معرفی شده مورد ارزیابی قرار گرفته است.

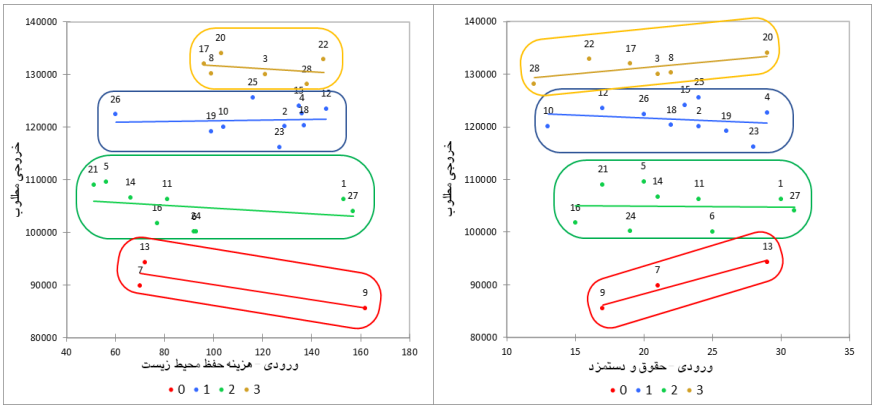
یافته‌های پژوهش

دراین قسمت ابتدا سه الگوریتم خوشه‌بندی زیر مجموعه یادگیری ماشین بر روی داده ها پیاده‌سازی شده‌است. سپس نتایج به‌دست آمده بر اساس سه شاخص ارزیابی کیفیت مورد بررسی قرار گرفت. در گام بعد پس از خوشه‌بندی واحدهای مشابه (که در واقع همان زنجیره تأمین صنایع پتروشیمی است) تحلیل پوششی داده شبکه‌ای طبق ساختار شبکه‌ای پیشنهادی و مدل معرفی شده در بخش ۳ به محاسبه کارایی نسبی بر مبنای معیارهای پایداری پرداخته است. همان طور که اشاره شد وضعیت ۲۸ زنجیره تأمین دوسطحی در صنایع پتروشیمی برای ۹۰ دوره زمانی جمع آوری شده‌است. بهترین کیفیت خوشه‌بندی در این ۹۰ دوره زمانی متعلق به دوره‌ای است که شاخص‌ها ارزیابی کیفیت در آن بهترین مقادیر را در بین تمام دوره‌ها زمانی داشته باشد. فضای داده در این پژوهش دارای ابعاد و پارامترهای متعددی است و امکان رسم نتایج به‌دست آمده به صورت دوبعدی به گونه‌ای که تمامی عوامل در شکلی واحد قابل رویت و مقایسه باشند وجود نداشت، لذا در شکل۷ نتایج خوشه‌بندی به روش K-Means به ازای سه ورودی مجزا و صرفا خروجی مطلوب در سه نمودار مجزا برای نشان دادن تفکیک پذیری مناسب داده‌ها رسم شده‌است. اصلی ترین موارد کاربرد یادگیری بدون نظارت می‌توان الف) تقسیم دیتاست بر مبنای ویژگی‌های مشترک، ب) تشخیص ناهنجاری‌ها یا داده‌ها پرت و ج) ساده‌سازی دیتاست و جمع آوری متغیرهایی با ویژگی مشابه. در صورت خلاصه در یک جمله هدف از یادگیری بدون نظارت مطالعه و بررسی ساختار ذاتی مجموعه داده مورد بررسی است که دو مسئله کاهش ابعاد و خوشه‌بندی از جمله کارکردهای اصلی آن است. هر کدام از این الگوریتم‌ها دارای محدودیت‌هایی هستند

برای مثال K-Means برای داده‌ها به فرم کروی پاسخ مناسبی نمی‌دهد.

یکی از مهمترین مراحلی که متخصصان علوم داده در انتخاب بهترین الگوریتم خوشه‌بندی آن را رعایت می‌کنند مشاهده تصویری فضای داده است. مشاهده پراکندگی و فرم داده گزینه‌های پیاده‌سازی نرم افزاری را برای رسیدن به بهترین جواب محدود می‌کند. البته امروزه با افزایش ابعاد داده‌ها در بسیاری از مواقع رسم نمودار بر اساس ویژگی‌های متعدد از فضای دو یا سه بعدی فراتر می‌رود و عملا رسم دیتاست میسر نخواهد بود. در این میان استفاده از روش‌های کاهش ابعاد (مانند PCA) پیشنهاد می شود که البته این راهکار هم در دیتاستی متناسب به متغیرها و ویژگی‌های منحصر بفرد آن برنامه ریزی می‌شود. شکل۷ صرفا عملکرد الگوریتم K-Means به ازای دو مورد از ورودی‌های زنجیره و خروجی مطلوب رسم کرده است. نتایج نشان دهنده عملکرد مناسب این الگوریتم در خوشه‌بندی است. در جدول۳ نتایج ارزیابی هر الگوریتم با بررسی سه شاخص معرفی شده در این مقاله ارائه شده‌است.

با توجه به نتایج مندرج در جدول الگوریتم Deep Embedded Clustering (DEC) در هر سه شاخص سیلوت، دیویس و کالینسکی بهترین عملکرد را نشان می‌دهد. در سطر انتهایی این جدول ضمن در نظر گرفتن هر زنجیره تأمین به صورت یک بلک باکس (درنظر نگرفتن فعالیت‌های بین دوسطح) DEA سنتی به روش CCR پیاده‌سازی شد که نتایج رتبه‌بندی واحدهای تصمیم بر اساس کارایی نسبی کل زنجیره را نشان می‌دهد.



شکل۷: نمودار خروجی الگوریتم K-Means به ازای سه ورودی مجزا و خروجی مطلوب

جهت ملموس شدن نتیجه خوشه‌بندی و عدم ارتباط آن با کارایی نسبی با استفاده از رنگ‌ها متفاوت و بر مبنای نتایج بهترین الگوریتم خوشه‌بندی که DEC معرفی شد نتایج بدین شرح تفسیر می‌گردد.

جدول۳: ارزیابی کیفیت خوشه‌بندی

DMU		۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲	۲۳	۲۴	۲۵	۲۶	۲۷	۲۸	Score
K-Memans	Silhouette	۲	۱	۳	۱	۲	۲	۰	۳	۰	۱	۲	۱	۰	۲	۱	۲	۳	۱	۱	۳	۲	۳	۱	۲	۱	۱	۲	۳	۰/۶۰۸
	Davies-Bouldin	۲	۱	۳	۱	۲	۲	۰	۳	۰	۱	۲	۱	۰	۲	۱	۲	۳	۱	۱	۳	۲	۳	۱	۲	۱	۱	۲	۳	۰/۴۷۲
	Calinski-Harabasz	۲	۰	۳	۰	۲	۲	۱	۳	۱	۰	۲	۰	۱	۲	۰	۲	۳	۰	۰	۳	۲	۳	۰	۲	۰	۰	۲	۳	۱۴۰/۱
Hierarchial	Silhouette	۰	۱	۳	۱	۰	۰	۲	۳	۲	۱	۰	۱	۲	۰	۱	۰	۳	۱	۱	۳	۰	۳	۱	۰	۱	۱	۰	۱	۰/۵۸۵
	Davies-Bouldin	۰	۱	۳	۱	۰	۰	۲	۳	۲	۱	۰	۱	۲	۰	۱	۰	۳	۱	۱	۳	۰	۳	۱	۰	۱	۱	۰	۱	۰/۴۶۹
	Calinski-Harabasz	۰	۱	۳	۱	۰	۰	۲	۳	۲	۱	۰	۱	۲	۰	۱	۰	۳	۱	۱	۳	۰	۳	۱	۰	۱	۱	۰	۱	۱۲۶
DEC	Silhouette	۲	۱	۳	۱	۲	۲	۰	۳	۰	۱	۲	۱	۰	۲	۱	۲	۳	۱	۱	۳	۲	۳	۱	۲	۱	۱	۲	۳	۰/۶۴۴
	Davies-Bouldin	۱	۲	۰	۲	۱	۱	۳	۰	۳	۲	۱	۲	۳	۱	۲	۱	۰	۲	۲	۰	۱	۰	۲	۱	۲	۲	۱	۰	۰/۳۵۸
	Calinski-Harabasz	۰	۱	۳	۱	۰	۰	۲	۳	۲	۱	۰	۱	۲	۰	۱	۰	۳	۱	۱	۳	۰	۳	۱	۰	۱	۱	۰	۳	۱۷۱
DEA-Efficiency Ranking		۲۵	۲۱	۹	۴	۲۷	۱۱	۶	۱۵	۱۲	۲۴	۲۲	۱۹	۲۰	۲۳	۲	۱۰	۱۴	۳	۱۸	۱۶	۸	۱۳	۲۸	۵	۱	۷	۱۷	۲۶	

شاید در گام اول در نظر گرفتن کارایی نسبی واحدها و دسته‌بندی آنها بر این اساس و قیاس برای ارائه راهکار بهبود روشی مناسب باشد ولی خوشه‌بندی وظیفه پیدا کردن و کشف مشابهت‌هایی را دارد که از نظر ظاهری برای کاربر ملموس نیست نتایج به‌دست آمده نشان می‌دهد که خوشه‌بندی واحدهای تصمیم به فرم زنجیره تأمین چند سطحی می‌تواند تا حد قابل توجهی ارائه راهکار بهبود عملیاتی برای واحدهای ناکارا در پی داشته باشد، ارائه راهکاری که قابل دسترس برای هر کدام از واحدها بوده و مدیران می توانند آن را به عنوان هدفی کوتاه مدت در اتخاذ استراتژی مناسب برای توسعه واحد صنعتی خود در نظر داشته باشند. و اما در خصوص واحدهای کارای در خوشه این نکته حائز اهمیت است که این واحدها در خوشه خود کارا هستند و اینها نیز نیاز است که برای خود راهکار بهبود داشته باشند. این واحدها می‌توانند با الگو قرار دادند کاراترین واحد کل برای توسعه خود برنامه ریزی نمایند تا بتوانند در دوره‌ها زمانی آتی در خوشه ای بالاتر برای رسین به نتایج بهتر رقابت نمایند.

پس از یافتن بهترین نتیجه برای خوشه‌بندی مبتنی بر یادگیری ماشین، طبق شکل۵، شبکه زنجیره تأمین پیشنهادی مدل‌سازی شد و معادلات برای محاسبه کارایی بر اساس تحلیل پوششی داده شبکه‌ای در نرم‌افزار GAMS پیاده‌سازدی گردید. در

جدول۴ ابتدا نتایج کارایی بدون در نظر گرفتن خوشه‌بندی برای کلیه واحدهای صنعتی پتروشیمی در ایران جمع آوری شده‌است. در این جدول با فرض بازده به مقیاس ثابت، معادلات در سه روش اشاره شده در بخش ۳ (روش تحقیق) پیاده‌سازی شده‌است. در این جدول کارایی کل زنجیره تأمین با استفاده از فرم ضربی، یعنی از حاصلضرب کارایی مرحله اول و دوم در دو حالت بدون در نظر گرفتن خروجی نامطلوب و با درنظر گفتن خروجی نامطلوب به روش تحلیل پوشش داده شبکه‌ای (NDEA) محاسبه گردیده است. به علاوه مقدار ناکارایی هر واحد نیز با استفاده از مدل DDF (تابع فاصله جهت دار) با مقدار تتا در جدول قابل مشاهده است. جدول ۵ مقادیر کارایی کل و ناکارایی هر واحد در خوشه محاسبه کرده است.

در جدول۵ همان طور که مشاهده می‌شود کارایی کل از ضرب کارایی مرحله اول (استخراج) در مرحله دوم (پالایشگاه) و با اولویت مرحله دوم یعنی پالایشگاه مورد محاسبه قرار گرفته است. تعداد واحدهایی که دارای کارایی ۱ شده‌اند (واحدهای کارا) در روش خوشه‌بندی مورد بیشتر است. پس می‌توان با استفاده از رویکرد خوشه‌بندی کارایی واحدهای پتروشیمی مورد مطالعه را با رویکرد خوش‌بینانه ارزیابی نمود و واحدهای ناکارا می‌توانند جهت برنامه ریزی برای توسعه واحد خود، شاخص‌های کاراترین واحد خوشه خود را برای برنامه-ریزی کوتاه مدت در نظر داشته باشند.

همان طور که در در روش تحقیق اشاره شد میزان k (تعداد خوشه) به روش elbow برای داده‌ها مورد مطالعه مقداری برابر ۴ به‌دست آمد، البته محاسبات برای k=3 هم انجام شد که نتایج خوشه‌بندی مورد تأیید نبود. لازم به ذکر است با توجه به اینکه خوشه ۳ و ۴ از نظر تعداد واحد از خوشه ۱ و ۲ کمتر بودند در بخشی جداگانه تر کیب شده و به عنوان یک خوشه در نظر گرفته شدند تا تفاوت نتایج به‌دست آمده مورد ارزیابی قرار گیرد. نتایج نشان داد که درصورتی که تعداد اعضای یک خوشه از ۵ کمتر باشد محاسبات کارایی نتایج مناسبی را نمی‌دهد به همین دلیل نیاز است که پس از محاسبه k و نتایج به‌دست آمده در صورت برای جمع بندی نهایی و تحلیل دو دسته‌ای که کمترین اعضا را دارند در یک دسته بزرگتر قرار بگیرند تا نتایج برای اعضای خوشه قابل تفسیر باشد. در بخش بعد به تحلیل نتایج به‌دست آمده پرداخته شده‌است.

بحث و نتیجه‌گیری

در مهمترین پدیده قرن حاضر ظهور حجم عظیمی از داده در سازمان‌ها و صنایع است، بهره‌گیری از علوم داده در پالایش و تفسیر این داده‌ها یکی از مهمترین فاکتورهای ایجاد مزیت رقابتی در صنعت است، درواقع این داده‌ها به عنوان دارایی ضمنی در سازمان‌ها تلقی می‌شوند. پژوهش حاضر توجه ویژه به جایگاه علوم داده در مدیریت زنجیره تامین را در روش تحقیق در نظر گرفت و به طور خاص کاربرد الگوریتم‌های یادگیری ماشین را در حوزه خوشه‌بندی مورد بررسی قرارداده است. در همین راستا ضمن بررسی الگوریتم‌های به روز و مورد استفاده در مقالات و مجلات بین‌المللی، جایگاه این قبیل بررسی‌ها در پژوهش‌های پیشین مرتبط با محاسبه کارایی نسبی و پایداری در زنجیره تأمین بررسی شد تا یک خط سیر پژوهشی در حوزه کارایی زنجیره تأمین با استفاده از NDEA با بهره‌گیری از علوم داده ترسیم گردد. این پژوهش نشان داد که تأثیر دسته بندی بهینه واحدهای تصمیم به خوشه‌هایی با بیشترین شباهت می تواند راهکار عملی بهتری را نسبت به روش‌های سنتی محاسبه کارایی نسبی ارائه دهد. خوشه‌بندی DMUها باعث شد تا واحدهای ناکارا هر خوشه فاصله کمتری را برای بهبود با کاراترین واحد دسته خود داشته باشند که این امر می‌تواند معیاری مناسب برای هدف گذاری کوتاه مدت واحدهای صنعتی برای رسیدن به کاراترین واحد خوشه خود در نظر گرفته شود. از طرفی با توجه به اینکه کارایی نسبی برای تمامی DMUها بدون در نظر گرفتن خوشه‌بندی هم اجرا شده‌است، واحدهای کارای هر خوشه نیز می‌توانند هدف خود را رسیدن شاخص‌های کاراترین واحد در بین ۲۸ واحد صنعتی قرار دهند تا محاسبات به روش پیشنهاد شده در این مقاله بتواند سیر توسعه عملیاتی برای تمامی واحدها را در نظر گرفته باشد. نتایج مشاهده شده در جدول ۴ و ۵ محاسبه کارایی مرحله‌های ۱ و ۲ زنجیره تأمین را در دو حالت بدون خوشه‌بندی و با خوشه‌بندی نشان داد. همان طور که اشاره شد کارایی به فرم ضربی و بر اساس مدل ارائه شده در شکل ۵ در نرم افزار گمز پیاده‌سازی شد.

جدول۲: داده یک دوره زمانی و نتیجه خوشه‌بندی به سه روش

DMU	استخراج نفت				پلایشگاه							
	هزینه زیست محیطی	هزینه مسئولیت اجتماعی	هزینه پرسنلی	نفت سبک تولیدی اوپک (میلیون دلار)	هزینه زیست محیطی	هزینه مسئولیت اجتماعی	هزینه پرسنلی	فروش محصولات نفتی اوپک (میلیون دلار)	ضایعات نفتی (میلیون دلار)	K-Means	Hierarchianl	DEC
۱	۶۵/۵۳	۲۴/۲۴	۱۷/۰۶	۳۸۲۳/۹۳	۸۷/۹۷	۲۱/۵۴	۱۲/۵۷	۴۶۲۶/۹۵	۲۶۷/۶۷	۲	۰	۲
۲	۵۵/۶۵	۳۵/۰۱	۸/۹۸	۲۸۷۶/۰۲	۷۳/۶۱	۱۲/۵۷	۱۴/۳۶	۳۴۷۹/۹۹	۲۰۱/۳۲	۱	۱	۱
۳	۸۱/۶۸	۲۶/۹۳	۸/۹۸	۳۵۶۹/۹۰	۳۹/۵۰	۲۵/۱۳	۱۱/۶۷	۴۳۱۹/۵۷	۲۴۹/۸۹	۳	۳	۳
۴	۸۲/۵۸	۴۰/۳۹	۱۳/۴۶	۴۴۷۳/۸۱	۵۲/۹۶	۱۱/۶۷	۱۵/۲۶	۵۴۱۳/۳۱	۳۱۳/۱۷	۱	۱	۱
۵	۲۳/۳۴	۳۶/۸۰	۵/۳۹	۵۱۱۴/۷۳	۳۲/۳۱	۲۳/۳۴	۱۴/۳۶	۶۱۸۸/۸۲	۳۵۸/۰۳	۲	۰	۲
۶	۲۹/۶۲	۳۶/۸۰	۱۴/۳۶	۳۳۴۱/۹۰	۶۱/۹۴	۳۵/۰۱	۹/۸۷	۴۰۴۳/۶۹	۲۳۳/۹۳	۲	۰	۲
۷	۳۴/۱۱	۳۲/۳۱	۷/۱۸	۶۱۲۲/۷۷	۳۵/۹۱	۳۷/۷۰	۱۳/۴۶	۷۴۰۸/۵۵	۴۲۸/۵۹	۰	۲	۰
۸	۳۹/۵۰	۱۹/۷۵	۱۵/۲۶	۲۴۰۷/۴۶	۵۹/۲۴	۳۳/۲۱	۶/۲۸	۲۹۱۳/۰۲	۱۶۸/۵۲	۳	۳	۳
۹	۸۰/۷۹	۱۷/۹۵	۶/۲۸	۲۷۳۲/۴۰	۸۱/۶۸	۱۲/۵۷	۱۰/۷۷	۳۳۰۶/۲۱	۱۹۱/۲۷	۰	۲	۰
۱۰	۳۵/۹۱	۳۸/۶۰	۴/۴۹	۵۲۸۲/۵۸	۶۸/۲۲	۱۰/۷۷	۸/۰۸	۶۳۹۱/۹۳	۳۶۹/۷۸	۱	۱	۱
۱۱	۵۳/۸۶	۱۷/۹۵	۶/۲۸	۳۵۳۷/۵۸	۲۶/۹۳	۳۴/۱۱	۱۷/۰۶	۴۲۸۰/۴۷	۲۴۷/۶۳	۲	۰	۲
۱۲	۸۲/۵۸	۱۶/۱۶	۷/۱۸	۴۰۰۳/۴۵	۶۳/۷۳	۱۲/۵۷	۸/۹۸	۴۸۴۴/۱۸	۲۸۰/۲۴	۱	۱	۱
۱۳	۱۷/۹۵	۱۸/۸۵	۱۳/۴۶	۵۱۶۲/۳۰	۵۳/۸۶	۱۸/۸۵	۱۵/۲۶	۶۲۴۶/۳۸	۳۶۱/۳۶	۰	۲	۰
۱۴	۳۵/۹۱	۳۹/۵۰	۱۵/۲۶	۲۴۹۹/۹۱	۲۹/۶۲	۱۴/۳۶	۵/۳۹	۳۰۲۴/۹۰	۱۷۴/۹۹	۲	۰	۲
۱۵	۵۹/۲۴	۳۹/۵۰	۵/۳۹	۴۵۱۶/۰۰	۷۵/۴۰	۴۰/۳۹	۱۷/۰۶	۵۴۶۴/۳۶	۳۱۶/۱۲	۱	۱	۱
۱۶	۲۶/۹۳	۳۴/۱۱	۸/۰۸	۴۱۶۱/۴۴	۵۰/۲۷	۲۶/۰۳	۶/۲۸	۵۰۳۵/۳۴	۲۹۱/۳۰	۲	۰	۲
۱۷	۳۴/۱۱	۱۷/۰۶	۱۴/۳۶	۲۶۴۳/۵۴	۶۱/۹۴	۳۲/۳۱	۴/۴۹	۳۱۹۸/۶۸	۱۸۵/۰۵	۳	۳	۳
۱۸	۸۶/۱۷	۳۷/۷۰	۸/۰۸	۲۲۵۱/۲۷	۵۱/۱۷	۱۰/۷۷	۱۳/۴۶	۲۷۲۴/۰۴	۱۵۷/۵۹	۱	۱	۱
۱۹	۳۲/۳۱	۱۸/۸۵	۱۴/۳۶	۳۱۸۳/۰۱	۶۶/۴۳	۱۹/۷۵	۱۱/۶۷	۳۸۵۱/۴۵	۲۲۲/۸۱	۱	۱	۱
۲۰	۷۸/۰۹	۲۵/۱۳	۱۵/۲۶	۴۴۹۱/۷۷	۲۵/۱۳	۲۹/۶۲	۱۳/۴۶	۵۴۳۵/۰۴	۳۱۴/۴۲	۳	۳	۳
۲۱	۳۵/۹۱	۲۶/۹۳	۱۱/۶۷	۲۴۵۸/۶۲	۱۵/۲۶	۲۰/۶۵	۵/۳۹	۲۹۷۴/۹۳	۱۷۲/۱۰	۲	۰	۲
۲۲	۷۰/۹۱	۸/۰۸	۸/۹۸	۴۷۴۴/۰۰	۷۴/۵۰	۱۶/۱۶	۶/۲۸	۵۷۴۰/۲۴	۳۳۲/۰۸	۳	۳	۳
۲۳	۵۲/۹۶	۴۰/۳۹	۱۵/۲۶	۵۵۸۹/۵۷	۷۳/۶۱	۱۴/۳۶	۱۲/۵۷	۶۷۶۳/۳۹	۳۹۱/۲۷	۱	۱	۱
۲۴	۱۷/۹۵	۹/۸۷	۵/۳۹	۳۰۰۲/۵۹	۷۵/۴۰	۲۲/۴۴	۱۳/۴۶	۳۶۳۳/۱۳	۲۱۰/۱۸	۲	۰	۲
۲۵	۵۳/۸۶	۳۵/۹۱	۱۵/۲۶	۴۵۷۶/۱۴	۶۱/۹۴	۱۲/۵۷	۸/۰۸	۵۵۳۷/۱۳	۳۲۰/۳۳	۱	۱	۱
۲۶	۱۸/۸۵	۱۶/۱۶	۸/۹۸	۴۵۵۸/۱۹	۴۰/۳۹	۱۲/۵۷	۱۰/۷۷	۵۵۱۵/۴۱	۳۱۹/۰۷	۱	۱	۱
۲۷	۸۱/۶۸	۱۴/۳۶	۱۷/۰۶	۲۸۴۶/۴۰	۷۵/۴۰	۴۰/۳۹	۱۳/۴۶	۳۴۴۴/۱۵	۱۹۹/۲۵	۲	۰	۲
۲۸	۷۳/۶۱	۲۹/۶۲	۷/۱۸	۴۸۸۴/۰۳	۶۴/۶۳	۳۷/۷۰	۴/۴۹	۵۹۰۹/۶۸	۳۴۱/۸۸	۳	۱	۳

جدول۴ : نتایج محاسبه کارایی و ناکارایی برای کل واحدهای تصمیم

کارایی کل زنجیره	کارایی مرحله دوم	کارایی مرحله اول	کارایی کل زنجیره (بدون خروجی نامطلوب)	کارایی مرحله دوم (بدون خروجی نامطلوب)	کارایی مرحله اول (بدون خروجی نامطلوب)	ناکارایی (تتا)	DMU
۰/۴۶	۱	۰/۴۶	۰/۴۶	۱	۰/۴۶	۰/۲۳	۱
۰/۴۰	۱	۰/۴	۰/۴۰	۱	۰/۴	۰/۲۲	۲
۰/۵۵	۱	۰/۵۵	۰/۵۵	۱	۰/۵۵	۰/۱۲	۳
۰/۴۶	۱	۰/۴۶	۰/۴۶	۱	۰/۴۶	۰/۲۴	۴
۱/۰۰	۱	۱	۱/۰۰	۱	۱	۰/۰۷	۵
۰/۲۶	۰/۵۸	۰/۴۵	۰/۲۴	۰/۵۴	۰/۴۵	۰/۲۷	۶
۱/۰۰	۱	۱	۱/۰۰	۱	۱	۰	۷
۰/۳۲	۰/۸۱	۰/۳۹	۰/۱۹	۰/۴۸	۰/۳۹	۰/۳۱	۸
۰/۵۹	۰/۹۵	۰/۶۲	۰/۲۷	۰/۴۴	۰/۶۲	۰/۱۳	۹
۱/۰۰	۱	۱	۱/۰۰	۱	۱	۰	۱۰
۰/۶۱	۰/۷۶	۰/۸	۰/۵۸	۰/۷۳	۰/۸	۰/۲۲	۱۱
۰/۸۰	۰/۹۳	۰/۸۶	۰/۶۵	۰/۷۵	۰/۸۶	۰/۱۲	۱۲
۱/۰۰	۱	۱	۱/۰۰	۱	۱	۰/۰۳	۱۳
۰/۲۹	۱	۰/۲۹	۰/۲۹	۱	۰/۲۹	۰/۰۵	۱۴
۰/۴۱	۰/۵۲	۰/۷۸	۰/۴۰	۰/۵۱	۰/۷۸	۰/۲۹	۱۵
۰/۷۰	۱	۰/۷	۰/۷۰	۱	۰/۷	۰/۱	۱۶
۰/۴۹	۱	۰/۴۹	۰/۴۹	۱	۰/۴۹	۰/۱۷	۱۷
۰/۳۲	۱	۰/۳۲	۰/۳۲	۱	۰/۳۲	۰/۱۳	۱۸
۰/۳۶	۰/۶۵	۰/۵۶	۰/۲۸	۰/۵	۰/۵۶	۰/۰۳	۱۹
۰/۵۲	۱	۰/۵۲	۰/۵۲	۱	۰/۵۲	۰/۱۷	۲۰
۰/۳۶	۱	۰/۳۶	۰/۳۵	۰/۹۸	۰/۳۶	۰/۰۷	۲۱
۱/۰۰	۱	۱	۱/۰۰	۱	۱	۰	۲۲
۰/۵۱	۰/۸۸	۰/۵۸	۰/۵۱	۰/۸۸	۰/۵۸	۰/۳۶	۲۳
۱/۰۰	۱	۱	۱/۰۰	۱	۱	۰/۰۷	۲۴
۰/۴۹	۱	۰/۴۹	۰/۴۹	۱	۰/۴۹	۰/۱۹	۲۵
۱/۰۰	۱	۱	۱/۰۰	۱	۱	۰	۲۶
۰/۱۸	۰/۴۱	۰/۴۳	۰/۱۵	۰/۳۶	۰/۴۳	۰/۱۹	۲۷
۰/۸۳	۱	۰/۸۳	۰/۸۳	۱	۰/۸۳	۰	۲۸

جدول ۵: نتایج محاسبات کارایی به تفکیک هر خوشه

کارایی کل زنجیره	کارایی مرحله دوم	کارایی مرحله اول	کارایی کل زنجیره (بدون خروجی نامطلوب)	کارایی مرحله دوم (بدون خروجی نامطلوب)	کارایی مرحله اول (بدون خروجی نامطلوب)	ناکارایی (نتا)	DMU	ردیف
خوشه ۱								
۰/۵۲	۱/۰۰	۰/۵۲	۰/۵۲	۱/۰۰	۰/۵۲	۰/۱۵	۱	۱
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۵	۲
۰/۳۴	۰/۶۰	۰/۵۶	۰/۳۳	۰/۵۹	۰/۵۶	۰/۱۹	۶	۳
۰/۸۸	۱/۰۰	۰/۸۸	۰/۸۸	۱/۰۰	۰/۸۸	۰/۱۷	۱۱	۴
۰/۳۶	۱/۰۰	۰/۳۶	۰/۳۵	۰/۹۶	۰/۳۶	۰/۱۲	۱۴	۵
۰/۷۶	۱/۰۰	۰/۷۶	۰/۷۶	۱/۰۰	۰/۷۶	۰/۰۰	۱۶	۶
۰/۳۹	۱/۰۰	۰/۳۹	۰/۳۹	۱/۰۰	۰/۳۹	۰/۰۷	۲۱	۷
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۲۴	۸
۰/۶۵	۱/۰۰	۰/۶۵	۰/۶۵	۱/۰۰	۰/۶۵	۰/۱۰	۲۷	۹
خوشه ۲								
۰/۳۹	۰/۸۶	۰/۴۵	۰/۲۲	۰/۴۹	۰/۴۵	۰/۳۲	۲	۱
۰/۵۲	۰/۹۹	۰/۵۳	۰/۴۹	۰/۹۲	۰/۵۳	۰/۲۰	۴	۲
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۱۰	۳
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۱۲	۴
۰/۴۸	۰/۶۰	۰/۸۰	۰/۴۶	۰/۵۸	۰/۸۰	۰/۱۴	۱۵	۵
۰/۳۵	۱/۰۰	۰/۳۵	۰/۳۵	۱/۰۰	۰/۳۵	۰/۱۰	۱۸	۶
۰/۵۷	۰/۹۵	۰/۶۰	۰/۳۲	۰/۵۳	۰/۶۰	۰/۱۴	۱۹	۷
۰/۵۵	۰/۸۸	۰/۶۲	۰/۵۵	۰/۸۸	۰/۶۲	۰/۲۸	۲۳	۸
۰/۵۳	۱/۰۰	۰/۵۳	۰/۵۳	۱/۰۰	۰/۵۳	۰/۱۲	۲۵	۹
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۲۶	۱۰
خوشه ۳								
۰/۶۶	۱/۰۰	۰/۶۶	۰/۵۲	۰/۷۹	۰/۶۶	۰/۱۲	۳	۱
۰/۸۰	۱/۰۰	۰/۸۰	۰/۴۰	۰/۵۰	۰/۸۰	۰/۱۶	۸	۲
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۱۷	۳
۰/۸۳	۱/۰۰	۰/۸۳	۰/۸۳	۱/۰۰	۰/۸۳	۰/۰۸	۲۰	۴
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۲۲	۵
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۲۸	۶
خوشه ۴								
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۷	۱
۰/۷۲	۱/۰۰	۰/۷۲	۰/۵۷	۰/۷۹	۰/۷۲	۰/۰۴	۹	۲
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۱۳	۳
ترکیب خوشه ۳ و ۴								
۰/۵۵	۱/۰۰	۰/۵۵	۰/۵۵	۱/۰۰	۰/۵۵	۰/۱۲	۳	۱
۰/۳۹	۱/۰۰	۰/۳۹	۰/۱۹	۰/۴۹	۰/۳۹	۰/۱۸	۸	۲
۰/۴۹	۱/۰۰	۰/۴۹	۰/۴۹	۱/۰۰	۰/۴۹	۰/۰۷	۱۷	۳
۰/۵۶	۱/۰۰	۰/۵۶	۰/۵۶	۱/۰۰	۰/۵۶	۰/۰۶	۲۰	۴
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۰	۲۲	۵
۰/۸۳	۱/۰۰	۰/۸۳	۰/۸۳	۱/۰۰	۰/۸۳	۰/۰۸	۲۸	۶
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۷	۷	۷
۰/۶۲	۱/۰۰	۰/۶۲	۰/۴۶	۰/۷۴	۰/۶۲	۰/۱۳	۹	۸
۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۰/۰۴	۱۳	۹

نتایج نشان داد که در صورت خوشه‌بندی میزان ناکارایی برای هر واحد تصمیم کاهش می‌یابد، یعنی واحدهای هر خوشه برای رسیدن به کارایی نسبی کافی است که کاراترین واحد خوشه خود را الگو قراردهند. لازم به ذکر است با توجه به پیاده‌سازی تحلیل پوششی داده به فرم ضربی و الویت قرار دادن سطح ۲، منجر شد که اکثر وزن‌ها در معادله به گونه‌ای در راستای حداکثرسازی کارایی مرحله ۲ در نظر گرفته شوند به همین دلیل در اکثر موارد تأثیر مرحله پالایشگاه در محاسبه کارایی نسبی و محاسبه ناکارایی به روش DDF مشهود است. خوشه‌بندی و ارائه راهکار بهبود سلسله مراتبی می تواند واحدهای صنعتی را در بهبود و توسعه مستمر ترغیب کند که نتایج شبیه‌سازی هم این ادعا را ثابت می‌نماید. در بررسی دیگر بین محاسبه کارایی بدون خروجی نامطلوب و با خروجی نامطلوب نتایج نشان داد که با توجه به تأثیر خروجی نامطلوب که با یک تبدیل خطی به عنوان خروجی مطلوب تحت عنوان $\hat{y}$ در نظر گرفته شد میزان کارایی کل و کارایی مرحله دوم در تمامی موارد نسبت به حالت بدون خروجی نامطلوب افزایش پیدا کرد که خود نشان دهنده تأثیر پذیری مثبت دخیل شدن این عامل در افزایش دقت محاسبات کارایی نسبی است. این راهکار منجر به ارائه بهبود مستمر و توسعه پلکانی در صنایع می‌شود که خود یکی از شاخص‌ها توسعه پایدار در اصول مدیریتی است. بهره‌گیری از تکنیک‌های علوم داده در مباحث مدیریت زنجیره تأمین زمینه ظهور پژوهش‌های میان رشته‌ای خواهد شد که می‌تواند تأثیر به سزایی در بهبود و توسعه واحدهای صنعتی داشته باشد. برای ادامه روند این پژوهش نیز می‌توان روش های خوشه‌بندی بیشتری از قبیل خوشه‌بندی با استفاده از DEA را بر روی داده‌های مورد مطالعه در نظر گرفت. روش خوشه‌بندی مبتنی بر DEA از توابع تولید تکه‌ای به‌دست‌آمده از روش DEA را برای خوشه‌بندی داده‌ها با انواع ورودی و خروجی در نظر می‌گیرد. بنابراین، هر واحد تصمیم‌گیری ارزیابی‌شده نه تنها خوشه‌ای که به آن تعلق دارد را می‌شناسد، بلکه نوع تابع تولید دسته خود را در نظر می‌گیرد. این روش برای تصمیم‌گیری‌های مدیریتی مهم است که در آن مدیران علاقه دارند که تغییرات مورد نیاز در ترکیب منابع ورودی و خروجی را شناسایی کنند تا بر این اساس بتوانند خود را در یک خوشه مطلوب قرار دهند. از طرفی با توجه به تأثیر کارایی دوره‌های زمانی گذشته بر کارایی دوره زمانی آتی، می‌توان از مدل تحلیل پوششی داده شبکه‌ای پویا

برای تحلیل زنجیره تأمین دوسطحی با در نظر گرفتن حلقه بازگشتی برگشت ضایعات به چرخه پالایش و اضافه شدن عوامل بیشتر در محاسبه کارایی نسبی استفاده کرد.

پی‌نوشت‌ها

1. Kao and Hwang

منابع

Acquaye, A., Feng, K., Oppon, E., Salhi, S., Ibn-Mohammed, T., Genovese, A., & Hubacek, K. (2017). Measuring the environmental sustainability performance of global supply chains: A multi-regional input-output analysis for carbon, sulphur oxide and water footprints. *Journal of environmental management*, *187*, 571-585.

Ahi, P., & Searcy, C. (2013). A comparative literature analysis of definitions for green and sustainable supply chain management. *Journal of cleaner production*, *52*, 329-341.

Angelakoglou, K., & Gaidajis, G. (2020). A conceptual framework to evaluate the environmental sustainability performance of mining industrial facilities. *Sustainability*, *12*(5), 2135.

Boloori, F., Afsharian, M., & Pourmahmoud, J. (2016). Equivalent multiplier and envelopment DEA models for measuring efficiency under general network structures. *Measurement*, *80*, 259-269.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429-444.

Chen, Y., Cook, W. D., Li, N., & Zhu, J. (2009). Additive efficiency decomposition in two-stage DEA. *European journal of operational research*, *196*(3), 1170-1176.

Cook, W. D., & Zhu, J. (2014). Data envelopment analysis: A handbook of modeling internal structure and network.

Dong, C., Akram, A., Andersson, D., Arnäs, P.-O., & Stefansson, G. (2021). The impact of emerging and disruptive technologies on freight transportation in the digital era: current state and future trends. *The International Journal of Logistics Management*, *32*(2), 386-412.

FakhrMousavi, S. M., Amirteimoori, A., Kordrostami, S., & Vaez-Ghasemi, M. (2022). Estimation of efficiency of two-stage processes using a fully fuzzy range-adjusted measure approach and strong complementary slackness conditions. *Modern Research in Decision Making*, *7*(2), 29-51.

Fang, L. (2020). Stage efficiency evaluation in a two-stage network data envelopment analysis model with weight priority. *Omega*, *97*, 102081.

Fosso Wamba, S., Queiroz, M. M., Guthrie, C., & Braganza, A. (2022). Industry

experiences of artificial intelligence (AI): Benefits and challenges in operations and supply chain management. In (Vol. 33, pp. 1493-1497): Taylor & Francis.

Halog, A., & Nguyen, N. H. (2017). Evaluating supply chain networks by incorporating the triple dimensions of sustainability paradigm. *Implementing Triple Bottom Line Sustainability into Global Supply Chains*, 165-186.

Herrera-Restrepo, O., Triantis, K., Trainor, J., Murray-Tuite, P., & Edara, P. (2016). A multi-perspective dynamic network performance efficiency measurement of an evacuation: A dynamic network-DEA approach. *Omega*, *60*, 45-59.

Hofmann, E. (2013). Supply Chain Management: Strategy, Planning and Operation, S. Chopra, P. Meindl.

Jabbarzadeh, A., Haughton, M., & Khosrojerdi, A. (2018). Closed-loop supply chain network design under disruption risks: A robust approach with real world application. *Computers & industrial engineering*, *116*, 178-191.

Jacobsen, H., & Tan, K. H. (2022). Improving food safety through data pattern discovery in a sensor-based monitoring system. *Production Planning & Control*, *33*(16), 1548-1558.

Kao, C., & Hwang, S.-N. (2008). Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European journal of operational research*, *185*(1), 418-429.

Lenz, N. V., Šegota, A., & Maradin, D. (2018). Total-factor energy efficiency in EU: Do environmental impacts matter? *International Journal of Energy Economics and Policy*, *8*(3), 92-96.

Li, Y., Shi, X., Emrouznejad, A., & Liang, L. (2020). Ranking intervals for two-stage production systems. *Journal of the Operational Research Society*, *71*(2), 209-224.

Lozano, S., & Khezri, S. (2021). Network DEA smallest improvement approach. *Omega*, *98*, 102140.

Marr, B. (2016). Why everyone must get ready for the 4th industrial revolution. *The Forbes*.

Min, E., Guo, X., Liu, Q., Zhang, G., Cui, J., & Long, J. (2018). A survey of clustering with deep learning: From the perspective of network architecture. *IEEE Access*, *6*, 39501-39514.

Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*, *42*(11), 30-36.

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(1), 86-97.

Ni, D., Xiao, Z., & Lim, M. K. (2020). A systematic review of the research trends of machine learning in supply chain management. *International Journal of Machine Learning and Cybernetics*, *11*, 1463-1482.

Olafsson, S., Cook, D., Davidsdottir, B., & Johannsdottir, L. (2014). Measuring countries ′environmental sustainability performance–A review and case study of Iceland. *Renewable and sustainable energy reviews*, *39*, 934-948.

Omid, A., Azar, A., Dehghan Nayeri, M., & Moghbel, A. (2021). Developing a network Data Envelopment Analysis approach to compare the environmental efficiency of active industries in Tehran. *Management Research in Iran*, *25*(3), 193-216.

Papadopoulos, T., Singh, S. P., Spanaki, K., Gunasekaran, A., & Dubey, R. (2022). Towards the next generation of manufacturing: implications of big data and digitalization in the context of industry 4.0. In (Vol. 33, pp. 101-104): Taylor & Francis.

Patrizii, V. (2020). On network two stages variable returns to scale Dea models. *Omega*, *97*, 102084.

Popat, S. K., & Emmanuel, M. (2014). Review and comparative study of clustering techniques. *International journal of computer science and information technologies*, *5*(1), 805-812.

Queiroz, M. M., Fosso Wamba, S., Machado, M. C., & Telles, R. (2020). Smart production systems drivers for business process management improvement: An integrative framework. *Business Process Management Journal*, *26*(5), 1075-1092.

Queiroz, M. M., Ivanov, D., Dolgui, A., & Fosso Wamba, S. (2022). Impacts of epidemic outbreaks on supply chains: mapping a research agenda amid the COVID-19 pandemic through a structured literature review. *Annals of operations research*, *319*(1), 1159-1196.

Rai, P., & Singh, S. (2010). A survey of clustering techniques. *International Journal of Computer Applications*, *7*(12), 1-5.

Shokri Kahi, V., Yousefi, S., Shabanpour, H., & Farzipoor Saen, R. (2017). How to evaluate sustainability of supply chains? A dynamic network DEA approach. *Industrial Management & Data Systems*, *117*(9), 1866-1889.

Tone, K., & Tsutsui, M. (2010). Dynamic DEA: A slacks-based measure approach. *Omega*, *38*(3-4), 145-156.

Wang, L.-W., Le, K.-D., & Nguyen, T.-D. (2019). Assessment of the energy efficiency improvement of twenty-five countries: a DEA approach. *Energies*, *12*(8), 1535.

Xia, H., An, W., Zhang, Z., & Liu, G. (2022). Managing production systems with machine learning: a case analysis of Suzhou GCL photovoltaic technology. *Production Planning & Control*, *33*(16), 1559-1572.

Xie, J., Girshick, R., & Farhadi, A. (2016). Unsupervised deep embedding for clustering analysis. International conference on machine learning,

Yin, P., Chu, J., Wu, J., Ding, J., Yang, M., & Wang, Y. (2020). A DEA-based two-stage network approach for hotel performance analysis: An internal cooperation

perspective. *Omega*, *93*, 102035.

Yousefi, S., Shabanpour, H., & Saen, R. F. (2021). Sustainable clustering of customers using capacitive artificial neural networks: a case study in Pegah Distribution Company. *RAIRO-operations Research*, *55*(1), 51-60.

Yousefi, S., Soltani, R., Bonyadi Naeini, A., & Farzipoor Saen, R. (2019). A robust hybrid artificial neural network double frontier data envelopment analysis approach for assessing sustainability of power plants under uncertainty. *Expert Systems*, *36*(5), e12435.

Yu, V. F., & Tseng, L.-C. (2014). Measuring social compliance performance in the global sustainable supply chain: an AHP approach. *Journal of Information and Optimization Sciences*, *35*(1), 47-72.

Zhang, Q., Koutmos, D., Chen, K., & Zhu, J. (2021). Using operational and stock analytics to measure airline performance: A network DEA approach. *Decision Sciences*, *52*(3), 720-748.

Zhou, P., Ang, B., & Han, J. (2010). Total factor carbon emission performance: a Malmquist index analysis. *Energy Economics*, *32*(1), 194-201.

Zhu, Y., Xie, C., Wang, G.-J., & Yan, X.-G. (2017). Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. *Neural Computing and Applications*, *28*, 41-50.

Assessing Sustainability of Supply Chain Performance using Machine Learning and Network Data Envelopment Analysis

Sina Sayardoost Tabrizi [1]
Keikhosro Yakideh[*2]
Mahmoud Moradi [3]
Mostafa Ebrahimpour [4]

Abstract
Using network data envelopment analysis (NDEA) models to assess the efficiency of Decision Making Units (DMUs) is a widely accepted method in academic research. An ongoing challenge in this field involves the computation and implementation of enhancement solutions within homogeneous clusters utilizing Machine Learning techniques. The primary aim of this paper is to identify the optimal clustering algorithm for a two-stage sustainable supply chain within the petrochemical industry in Iran. Subsequently, the application of NDEA within each cluster aims to ascertain efficiency levels and devise improvement strategies to facilitate a more targeted development approach for inefficient units. This paper investigates the best clustering algorithms in the area of Machine Learning by using quality measurement indicators and using Network Data Envelopment Analysis (NDEA) for measuring the efficiency of DMUs with sustainability approach. Upon examination, it has been determined that the Deep Embedded Clustering algorithm yields the most favorable results when applied to the data set. Furthermore, the comparison of the clustering result with the standard NDEA model has demonstrated the utility of clustering and comparing units in homogeneous categories for the purpose of efficiency calculation and determining the distance to the efficient frontier. This article, showed that how to find the best algorithm for two-stage supply chain clustering. Also, by comparing the effect of clustering on measuring the distance of inefficient units to the efficiency frontier, it was shown that clustering of units can play a significant role in planning to reach a practical development plan in each cluster.

Machine learning, Sustainable supply chain, Network data envelopment analysis, Efficiency, Clustering

1 PhD Candidate of Industrial Management, Faculty of Management and Economics, University of Guilan, Rasht
2 Assistant Professor, Faculty of Management and Economics, University of Guilan, Rasht, Iran (yakideh@guilan.ac.ir)
3 Associate Professor, Faculty of Management and Economics, University of Guilan, Rasht, Iran
4 Professor of the Management and Economics Faculty, University of Guilan, Rasht, Iran