



شرکت بین المللی مهندسی سیستم ها و اتوماسیون-ایریسا



شرکت سامانه های پشتیبان تصمیم-تاکبن

## مستند فاز طراحی و پیاده سازی مدل پیش بینی نرخ مصرف قطعات یدکی فولاد مبارکه

بهمن ماه ۱۴۰۰

## تاریخچه مدرک

بازنگری	تاریخ	بخش ایجاد کننده	توضیحات
۱	۱۴۰۰/۱۰/۰۴	شرکت سامانه های پشتیبان تصمیم_ تاکبن	
۲	۱۴۰۰/۱۱/۰۹	شرکت سامانه های پشتیبان تصمیم_ تاکبن	اعمال اصلاحات مورد نیاز کاربر

## فهرست مطالب

۱- مقدمه	۴
۲- متدولوژی انجام پروژه	۵
۳- درک و شناخت کسب و کار	۷
۳-۱ چارت سازمانی	۷
۳-۲ دسته بندی موجود قطعات یدکی	۷
۳-۳ فرایند کنترل موجودی قطعات یدکی	۹
۴- درک و شناخت و آماده سازی داده ها	۱۰
۵- طراحی و مدل سازی	۱۲
۵-۱ روش های داده کاوی مورد استفاده	۱۳
۵-۲ رویکرد مدل سازی قطعات یدکی	۲۶
۵-۳ مدل سازی دسته پرمصرف	۲۷
۵-۴ مدل سازی دسته کم مصرف	۲۸
۵-۵ مدل سازی دسته تعمیرپذیر	۳۰
۶- ارزیابی و استقرار	۳۳
۶-۱ ارزیابی نتایج دسته پرمصرف	۳۶
۶-۲ ارزیابی نتایج دسته کم مصرف	۴۲
۶-۳ ارزیابی نتایج دسته تعمیرپذیر	۴۸
۶-۴ استقرار در سیستم مدیریت کالا	۵۳
۷- جمع بندی و پیشنهادات آتی	۵۵

## ۱- مقدمه

افزایش نقش مدیریت خرید و تدارکات طی پنجاه سال اخیر منجر به تغییر نگرش سازمان ها نسبت به واحد سازمانی خرید از وظایف پشتیبانی و ستادی به فعالیت تخصصی و حرفه ای متناسب با سطح بالای تخصص شده است. به طوری که امروزه واحد خرید در سازمان ها، با مدیریت دارایی های ثابت به عنوان دست اندرکاران اصلی سازمان محسوب می شوند.

در سازمان های بزرگ تولیدی، عدم تامین به موقع کالا، موجب بروز مشکلات بحرانی در تحویل به موقع محصولات و خدمات به مشتریان می شود. در برنامه ریزی خرید به همان میزان که در دسترس بودن کالا، در زمان مورد نیاز حائز اهمیت بوده، متعادل بودن و تحت کنترل بودن هزینه ها نیز بسیار مهم است؛ یعنی از طرفی خرید بیش از نیاز در برنامه ریزی خرید، سود حاصله را محدود کرده و از طرف دیگر سرمایه گذاری کم باعث توقف در فرایندهای سازمان شده و در نتیجه شکست در کسب و کار مربوطه را سبب می شود.

یکی از موضوعات مهم در زنجیره تامین، موضوع کنترل موجودی مواد، قطعات و محصولات است. در این بین به دلیل حساسیت بالایی که قطعات یدکی در صنایعی همچون فولاد سازی دارند، این قطعات می بایست مورد توجه ویژه قرار گیرد. در واقع می توان گفت پیچیده ترین نوع برنامه ریزی خرید در صنایع فولادی، به تامین قطعات یدکی برمی گردد و تصمیم گیری دقیق باعث اثربخشی بسیار بالایی در این دسته خواهد شد. از آن جایی که عوامل هزینه ای تاثیرگذار بر سیاست های موجودی در تامین قطعات یدکی در این صنایع بسیار بالا است، لذا کوچکترین اشتباهی منجر به تحمیل هزینه های سنگین به سازمان خواهد شد.

در تعیین مقدار و نقطه سفارش قطعات یدکی آیتم های زیادی دخالت دارد. یکی از مهم ترین این پارامترها میزان نرخ مصرف قطعه یدکی است که تاثیر مستقیم در مقدار سفارش دهی خواهد داشت. همان گونه که بیان گردید محاسبه دقیق این پارامتر تاثیر به سزایی در کاهش هزینه های نگهداری و کمبود و کاهش حجم قطعات راکد و مازاد در انبار می گردد. لذا شناخت دقیق داده ها و در نظرگیری عوامل تاثیرگذار بر مصرف قطعات نقش بسزایی در پیش بینی دقیق تر مصرف خواهد داشت.

در پروژه پیش رو با عنوان "پیش بینی نرخ مصرف قطعات یدکی فولاد مبارکه" که با هدف تحلیل داده و ارائه مدل پیش بینی نرخ مصرف قطعات یدکی انجام شده است، از روش های داده کاوی به منظور دسته بندی قطعات یدکی فعال در فولاد مبارکه و ارائه الگوی پیش بینی نرخ مصرف استفاده گردیده است. در این مستند به ارائه روش های مدل سازی، ارزیابی و پیاده سازی مدل پیش بینی نرخ مصرف در هر دسته از قطعات پرداخته شده و نتایج حاصل از ارزیابی مدل به تفکیک روش های پیشنهادی آورده شده است.

## ۲- متدولوژی انجام پروژه

مهم ترین فعالیت در انجام هر پروژه تعیین روش و شناسایی گام های اصلی برای پیشبرد آن است که با توجه به ماهیت و روش انجام پروژه پیش بینی نرخ مصرف قطعات یدکی با روش داده کاوی، متدولوژی شناخته شده <sup>۱</sup>CRISP-DM انتخاب گردیده است. این فرآیند دارای شش مرحله اصلی است. این شش مرحله از درک نیازهای اصلی کسب و کار شروع می شود و در نهایت به ارائه راهکاری برای آن ختم می گردد. علاوه بر وجود

توالی در انجام این گام ها، در عمل رفت و

برگشت های زیادی بین مراحل وجود دارند.

مراحل انجام این متدولوژی به شرح زیر است:

➤ درک و شناخت کسب و کار

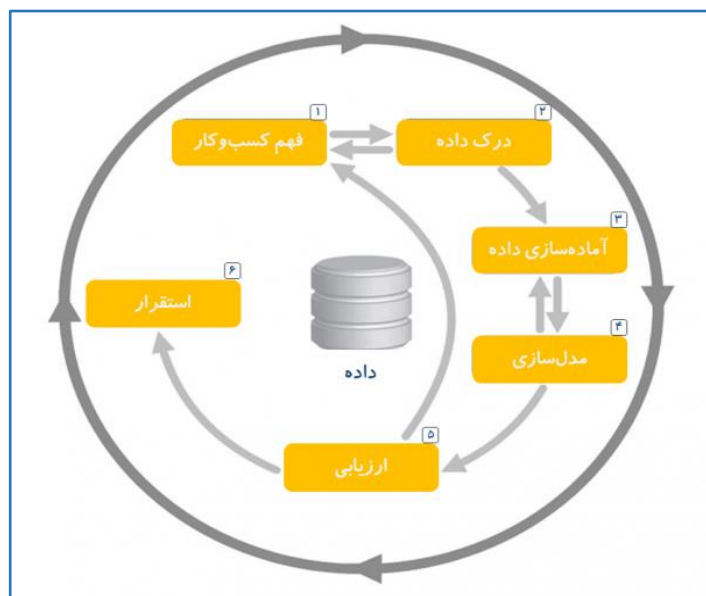
➤ درک و شناخت داده ها

➤ آماده سازی داده ها

➤ مدل سازی

➤ ارزیابی

➤ استقرار



<sup>۱</sup> CRoss Industry Standard Process for Data Mining

در راستای انجام این پروژه جلسات متعدد برای شناخت، تحلیل داده بررسی مدل سازی و ارزیابی خروجی و نتایج با واحدهای مختلف سازمان به شرح جدول زیر برگزار گردید:

ردیف	عنوان جلسه	تعداد جلسات
۱	شناخت فرایند سفارش دهی واحد SMO و MPC نواحی	۶
۲	شناخت داده ها و فرایند موجود در سیستم کنترل موجودی قطعات یدکی	۲
۳	بررسی، تحلیل و انطباق داده های سیستم های مدیریت کالا و نت	۳
۴	بررسی، تحلیل و اصلاح داده های نامنطبق مربوط به گردش کالا	۳
۵	بررسی اشکالات داده ای و فرایندی موجود بین واحد SMO و CEW	۲
۶	بررسی فرایند قطعات برگشتی و نواقص داده ای موجود	۲
۷	برگزاری جلسات بین واحدی نت و SMO و تعمیرگاه مرکزی برای اصلاح فرایند بین واحدی	۳
۸	ارائه و بررسی نتایج مدل پیاده سازی شده بر روی قطعات پرمصرف و کم مصرف	۳
۹	ارائه و بررسی نتایج مدل پیاده سازی شده بر روی قطعات تعمیرپذیر	۴
۱۰	ارائه گزارشات مدیریتی شناخت و تحلیل داده های موجود و اعلام نواقص داده ای و فرایندی موجود و روند پیشرفت پروژه	۵
۱۱	بررسی نسخه اولیه مستند فاز طراحی و نحوه مدل سازی و بحث های مربوط به قابلیت های کانتور داده	۱
۱۲	جلسه بررسی الزامات و محورهای لازم برای کانتور داده	۲
۱۳	جلسه بررسی و تصمیم گیری در خصوص بحث داده های کانتور داده	۲
۱۴	جلسات بررسی و ارزیابی نتایج مدل داده کاوی پیاده سازی شده با کاربر	۴
۱۵	جلسات نهایی تست پذیرش و تحویل مدل پیاده سازی شده در سیستم مدیریت کالا	۲
	جمع جلسات	۴۴

در ادامه این مستند بر اساس گام های تعریف شده در متدولوژی به شناخت داده و کسب و کار مربوط به نحوه مصرف قطعات یدکی فولاد مبارکه و نحوه مدل سازی و ارزیابی نتایج پرداخته شده است. لازم به ذکر است جزییات کامل مربوط به شناخت مساله و فرایند کنترل موجودی قطعات یدکی در مستند دیگری تحت عنوان " فاز شناخت پروژه پیش بینی نرخ مصرف قطعات یدکی شرکت فولاد مبارکه " ارائه گردیده است.

### ۳- درک و شناخت کسب و کار

#### ۳-۱ چارت سازمانی

در شکل زیر بخشی از چارت سازمانی معاونت بهره برداری مرتبط با پروژه پیش بینی نرخ مصرف آورده شده است.



کاربران فعلی سیستم کنترل موجودی قطعات یدکی، پرسنل دفتر فنی قطعات یدکی SMO و برنامه ریزی و کنترل نگهداری و تعمیرات MPC نواحی می باشند که به صورت روزانه از سیستم گزارش می گیرند و در صورت مشاهده قطعه یدکی Under Stock، در صورت لزوم پیشنهاد خرید صادر می کنند.

#### ۳-۲ دسته بندی موجود قطعات یدکی

در سیستم کنونی مدیریت قطعات یدکی، قطعات یدکی بر اساس میزان مصرف به سه دسته کلی زیر تحت عنوان مدل مدیریتی دسته بندی می شوند:

- قطعات پرمصرف (RE): قطعاتی که میزان مصرف در ۳ سال اخیر بیش از یک عدد باشد.

- قطعات کم مصرف (SA): قطعاتی که میزان مصرف در ۳ سال اخیر کمتر یا مساوی با یک عدد باشد.
- قطعات تعمیرپذیر (RR): قطعات با قابلیت تعمیر که در شناسنامه کالا به عنوان قطعه تعمیر پذیر مشخص شده اند.

با توجه به این که هدف از انجام این پروژه، پیش بینی نرخ مصرف برای قطعات یدکی فعال و غیرراکد با قابلیت صدور پیشنهاد خرید است، تعداد قطعات یدکی در هر مدل مدیریتی در زمان تهیه گزارش به شرح زیر است:

مدل مدیریتی	تعداد کل قطعات یدکی	تعداد قطعات یدکی فعال غیر راکد
RE	۵۰,۹۸۸	۳۴,۹۱۷
RR	۲۳,۹۸۶	۱۰,۵۴۴
SA	۱۷۰,۲۱۶	۳۰,۴۹۵

در این روش دسته بندی در زمان تعریف کد قطعه یدکی در سیستم برای اولین بار، بر اساس شناخت کارشناس یا نظر تامین کننده، مدل مدیریتی در شناسنامه کالا وارد شده و تا ۵ سال بعد از تاریخ تعریف قطعه مشروط به نداشتن هرگونه خرید یا خروج از انبار به صورت اتوماتیک به روز نمی شود، مگر در مواردی که با درخواست کارشناسی و پس از دریافت تاییدات لازم مدل مدیریتی قبل از ۵ سال به روز گردد.

برای قطعاتی که بیشتر از ۵ سال از تعریف قطعه در سیستم گذشته است، مدل مدیریتی قطعات پرمصرف و کم مصرف که بر اساس میزان مصرف دسته بندی شده اند، با استفاده از فانکشنی که به صورت هفتگی در سیستم اجرا می شود به روز می گردند. در تبدیل مدل مدیریتی برای قطعات کم مصرف و پرمصرف به تعمیرپذیر و بالعکس با درخواست کارشناس و ارائه سوابق تعمیرات، مدل مدیریتی تغییر داده می شود. فانکشن به روز رسانی مدل مدیریتی بر اساس سوابق خروج از انبار، برگشت قطعات به انبار، خریدهای مستقیم و اضطراری و با در نظرگیری ملاحظات مربوط به تغییر ماهیت قطعات، مدل مدیریتی قطعات را با در نظرگیری نرخ مصرف در ۵ سال اخیر به روز می نماید.



### ۳-۳ فرایند کنترل موجودی قطعات یدکی

سیستم انبارداری قطعات یدکی فولاد مبارکه به شکل متمرکز بوده و عمده قطعات یدکی در انبار مرکزی نگهداری می شوند. خروج قطعه از انبار مرکزی به معنی مصرف قطعه بوده و در هر یک از نواحی انبارهای پای کاری وجود داشته که موجودی آن ها در سیستم لحاظ نمی شود. برای قطعات با مدل مدیریتی RE و SA زمانی قطعه مصرف شده است که از انبار خارج شود و نیز برای قطعات RR، قراضه یا اسقاط شدن قطعه نشان دهنده مصرف آن قطعه است.

نقطه سفارش (Optimal Stock) قطعات یدکی توسط فرمول زیر محاسبه می شود.

$$(Order\ Point)\ Optimal\ Stock = \frac{X.DT}{۳۶۰} * SF$$

- X: میانگین مصرف قطعه در ۵ سال گذشته به اضافه مقدار گذشته در سال جاری
- DT: میانگین زمان چهار خرید آخر
- S.F: Safety Factor یک ضریب اطمینان است که بر حسب اهمیت قطعه

هر زمان که مجموع موجودی انبار و سفارش در راه از نقطه سفارش کمتر شد، قطعه به زیر سطح موجودی رسیده است و با در نظرگیری موجودی در راه و برگشتی عمدتاً پیشنهاد خرید صادر می شود.

مقدار X در حال حاضر از طریق روش های میانگین گیری متحرک بر اساس تاریخچه ۵ سال گذشته انجام می گیرد. هدف از انجام این پروژه پیش بینی نرخ مصرف قطعات یدکی بر اساس تاریخچه داده های انبار، نگهداری و تعمیرات و تعمیرگاه مرکزی با در نظر گرفتن سایر عوامل موثر بر مقدار مصرف در ماه های آتی است.

#### ۴- درک و شناخت و آماده سازی داده ها

بر اساس شناخت حاصل شده نسبت به فرایند کنترل موجودی قطعات یدکی، عوامل تاثیرگذار بر پیش بینی و طراحی اولیه مدل پیش بینی برای قطعات یدکی کم مصرف، پرمصرف و تعمیرپذیر، سه دسته اصلی از داده ها در سیستم اطلاعاتی فولاد مورد نیاز است:

۱. داده های مربوط به تاریخچه گردش کالا شامل حواله از انبار، بازگشت کالا به انبار، خریدهای اضطراری و مستقیم و اطلاعات مربوط به گردش کالاهای تغییر ماهیت یافته.

۲. داده های مربوط به نگهداری و تعمیرات شامل اطلاعات مربوط به درختواره تعمیرات، تاریخچه درخواست های تعمیراتی عادی، اضطراری و پیشگیرانه به همراه تعداد قطعات حواله شده از انبار به ازای درخواست تعمیراتی، اطلاعات BOM تجهیزات

۳. داده های مربوط به تاریخچه تعمیرات قطعات تعمیرپذیر در تعمیرگاه مرکزی به همراه اطلاعات نتایج تعمیر، مدت زمان، نحوه انجام و ... و اتصال درخواست تعمیرگاه مرکزی به درخواست تعمیراتی مرتبط

در حال حاضر اطلاعات مربوط به گردش کلیه قطعات یدکی، اطلاعات نگهداری و تعمیرات و تعمیرگاه مرکزی در سیستم اطلاعاتی موجود فولاد مبارکه تحت عنوان IS\_Suite ذخیره سازی می شود. اطلاعات مربوطه از زمان تاسیس شرکت فولاد مبارکه تا زمان استقرار سیستم IS\_Suite در سیستمی با عنوان MIS ذخیره سازی می شده است. در زمان جایگزینی سیستم اطلاعاتی مربوط به مدیریت کالا در سال ۱۳۹۴ و نگهداری و تعمیرات در سال ۱۳۹۶ تنها بخشی از اطلاعات مورد نیاز از سیستم قبل به سیستم جدید انتقال داده شد.

با توجه به این که انجام فرایند داده کاوی و اعتبار سنجی نتایج آن وابستگی زیادی به حجم و اعتبار داده های موجود دارد و تنها اطلاعات مربوط به سال ۱۳۸۸ به بعد موجود است، لذا در روند انجام پروژه موضوع کانورت داده های مربوط به قبل از سال ۱۳۸۸ که در Tape هایی از سیستم گزارش گیری ETL موجود بود مطرح

گردید که پس از برگزاری جلسات متعدد و انجام فعالیت زمان بر کانتور داده و تست و بررسی اعتبار داده ها نتایج زیر حاصل شد:

۱. اطلاعات گردش کالا در زمان MIS هر ۲ سال یک بار پاک شده و اطلاعات به صورت آماری نگه داشته می شده است. هم چنین به دلیل نبودن ETL از ابتدای MIS، داده های گردش کالای موجود در ETL مربوط به ۲ سال است و بدلیل کلی بودن گردش ها امکان اتصال داده ها به گردش های IS\_Suite و بالانس موجودی امکان پذیر نیست.

۲. طبق بررسی انجام شده موجودی فیزیکی با گردش به صورت تاریخیچه ای ارتباط نداشته و در لحظه به روز می شده است، بنابراین نقطه افتتاحیه موجودی در زمان پاک کردن داده های MIS و انتقال به Tape ها عملاً تعریف نمی شده است، لذا نمی توان به مجموعه گردش های آماری قدیم اعتماد و داده ها را به کاردکس جدید IS\_Suite اضافه و بالانس موجودی را انجام داد و برای هیچ یک از دسته قطعات اطلاعات کامل گردش از سال ۷۱ تا ۸۸ قابل دسترسی نیست.

۳. کانتور داده های نگهداری و تعمیرات که به صورت پایلوت از ناحیه ۰۷ که به نظر می رسید داده های کامل تری داشته باشد، توسط تیم نت شرکت ایریسا انجام شد و مشخص گردید که برای حدود ۶۰ درصد از داده های ناحیه فولاد سازی امکان کانتور و لینک به موقعیت های موجود وجود داشته است که برای برخی از اطلاعات کانتور شده قطعه یدکی معادل قابل استخراج نیست. هم چنین عدم تعریف تمامی قطعات در درختواره تجهیزات موجود و کامل نبودن لیست BOM تجهیزات از جمله ریسک های ممکن برای دستیابی به داده های دقیق و کامل مورد نیاز از سیستم نگهداری و تعمیرات است.

با توجه به موارد مطرح شده مقرر شد فرایند داده کاوی بر روی داده های موجود در سیستم IS\_Suite و با وزن دهی کمتر به داده های سال ۱۳۹۴ از نظر اعتبار داده ادامه یابد. با توجه به رویکردی که برای مدل سازی

مدنظر قرار گرفته آماده سازی هایی بر روی داده های ورودی انجام شد که در قسمت مدل سازی بسته به هر رویکرد به آن اشاره خواهد شد.

## ۵- طراحی و مدل سازی

همان گونه که بیان گردید در این پروژه به منظور دسته بندی و پیش بینی نرخ مصرف قطعات یدکی از ترکیب تکنیک های مختلف داده کاوی استفاده می شود. با بهره گیری از داده کاوی می توان به استخراج قواعد و الگوهای مناسب از درون داده های گذشته پرداخت که در این راستا می توان از روش ها و تکنیک های توصیف و پیش بینی استفاده نمود.

علم داده، یک زمینه میان رشته ای است که از روش ها، فرآیندها، الگوریتم ها و سیستم های علمی برای استخراج دانش و بینش از داده ها در اشکال گوناگون (ساختار یافته و ساختار نیافته) استفاده می کند.

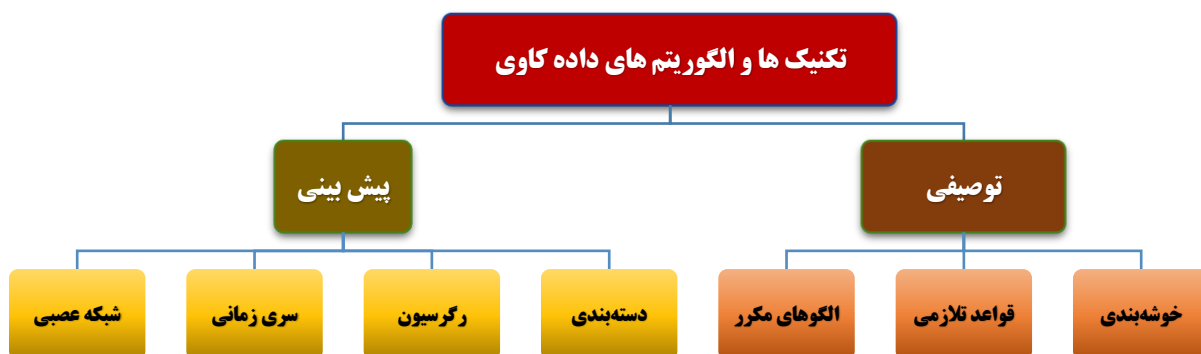


ابزارهای داده کاوی به دو دسته اصلی طبقه بندی می شود:

- ابزارهای توصیفی
- ابزارهای پیش بینانه

ابزارهای توصیفی خواص عمومی داده‌ها را مشخص می‌سازد. هدف از توصیف، یافتن الگوهایی در مورد داده‌هاست که برای انسان قابل تفسیر باشد.

ابزارهای پیش‌بینانه، به منظور پیش‌بینی رفتارهای آینده استفاده می‌شود. منظور از پیش‌بینی، به کارگیری چند متغیر در پایگاه داده برای پیش‌بینی مقادیر آینده برای سایر متغیرها می‌باشد. در شکل زیر طبقه‌بندی تکنیک‌ها و الگوریتم‌های داده‌کاوی نشان داده شده است.



در این پروژه از روش‌های مختلف داده‌کاوی برای دسته‌بندی و پیش‌بینی نرخ مصرف قطعات یدکی استفاده شده است که در ادامه به شرح این روش‌ها پرداخته شده است.

## ۵-۱ روش‌های داده‌کاوی مورد استفاده

### ➤ خوشه‌بندی

خوشه‌بندی<sup>۱</sup> فرآیندی است که به کمک آن می‌توان مجموعه‌ای از اشیاء را به گروه‌های مجزا افراز کرد. هر افراز یک خوشه نامیده می‌شود. اعضاء هر خوشه با توجه به ویژگی‌هایی که دارند به یکدیگر بسیار شبیه هستند و میزان شباهت بین خوشه‌ها کمترین مقدار است. در چنین حالتی هدف از خوشه‌بندی، نسبت دادن

<sup>۱</sup> Clustering

برچسب‌هایی به اشیاء است که نشان دهنده عضویت هر شیء به خوشه است. با توجه به روش‌های مختلف اندازه‌گیری شباهت یا الگوریتم‌های تشکیل خوشه، ممکن است نتایج خوشه‌بندی برای مجموعه داده ثابت متفاوت باشند. روش‌های خوشه‌بندی استفاده شده در این پروژه شامل روش‌های زیر است:

### • k\_Means

الگوریتم خوشه‌بندی K\_Means روشی کمی برای دسته‌بندی  $n$  مشاهده در  $k$  خوشه براساس میانگین آن‌ها است. در این روش دسته‌بندی مشاهدات براساس مشابهت میانگین آن‌ها با یکدیگر صورت می‌گیرد. بدست آوردن نقاطی به عنوان مراکز خوشه‌ها این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند. نسبت دادن هر نمونه داده به یک خوشه که آن داده کمترین فاصله تا مرکز آن خوشه را دارا باشد. در نوع ساده‌ای از این روش ابتدا به تعداد خوشه‌های مورد نیاز نقاطی به صورت تصادفی انتخاب می‌شود. سپس در داده‌ها با توجه با میزان نزدیکی (شباهت) به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها مراکز جدیدی برای آن‌ها محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود.

بهترین خوشه‌بندی آن است که مجموع تشابه بین مراکز خوشه و همه اعضای خوشه را حداکثر و مجموع تشابه بین مراکز خوشه‌ها را حداقل کند. برای انتخاب بهترین خوشه ابتدا براساس نظرات خبره و مطالعات قبلی یک محدوده پیشنهادی برای تعداد خوشه‌ها مشخص می‌شود. بنابراین تعداد خوشه به عنوان پارامتر ورودی این روش در نظر گرفته می‌شود. فرض کنید مشاهدات  $(x_1, x_2, \dots, x_n)$  که دارای  $d$  بعد هستند را باید به  $k$  بخش یا خوشه تقسیم کنیم. این بخش‌ها یا خوشه‌ها را با مجموعه‌ای به نام  $S = \{S_1, S_2, \dots, S_k\}$  می‌شناسیم. اعضای خوشه‌ها باید به شکلی از مشاهدات انتخاب شوند که تابع

مجموع مربعات درون خوشه‌ها<sup>۱</sup> «که در حالت یک بعدی شبیه واریانس است، کمینه شود. بنابراین، تابع هدف در این الگوریتم به صورت زیر نوشته می‌شود.

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_S \sum_{i=1}^k |S_i| \text{Var } S_i$$

در اینجا منظور از  $\mu_i$  میانگین خوشه  $S_i$  و  $|S_i|$  تعداد اعضای خوشه  $i$  ام است. البته می‌توان نشان داد که کمینه کردن این مقدار به معنی بیشینه‌سازی میانگین مربعات فاصله بین نقاط در خوشه‌های مختلف است زیرا طبق قانون واریانس کل، با کم شدن مقدار WCSS، مقدار BCSS افزایش می‌یابد، زیرا واریانس کل ثابت است.

## • DB\_SCAN

روش DB\_SCAN، معروف‌ترین روش خوشه‌بندی مبتنی بر چگالی<sup>۲</sup> است. در خوشه‌بندی مبتنی بر چگالی، خوشه‌ها به عنوان نواحی چگال از مجموعه داده، تعریف می‌شوند. اشیای موجود در نواحی کم‌تراکم، جداکننده‌ی خوشه‌ها از هم هستند.

این روش نقاطی که در محدوده‌ی معینی (در یک شعاع همسایگی) از هم قرار دارند را به هم وصل می‌کند. این الگوریتم، تنها نقاطی را متصل می‌کند که چگالی کمینه‌ای داشته باشند. که این امر به عنوان حداقل تعداد اشیای موجود<sup>۳</sup> در شعاع همسایگی<sup>۴</sup> تعریف شده است و بر خلاف بسیاری از روش‌های دیگر، می‌تواند خوشه‌های دارای اشکال دلخواه را شناسایی کند. این روش نقاطی که در محدوده‌ی معینی (در یک شعاع همسایگی) از هم قرار دارند را به هم وصل می‌کند. این الگوریتم، تنها نقاطی را متصل می‌کند

<sup>۱</sup> Within-Cluster Sum of Squares- WCSS

<sup>۲</sup> Density Based Clustering

<sup>۳</sup> MinPoints

<sup>۴</sup> Epsilon

که چگالی کمینه‌ای داشته باشند. که این امر به عنوان MinPoints در شعاع همسایگی Epsilon تعریف شده است و بر خلاف بسیاری از روش‌های دیگر، می‌تواند خوشه‌های دارای اشکال دلخواه را شناسایی کند. در این روش نیازی به مشخص بودن تعداد خوشه‌ها، بر خلاف روش‌های خوشه‌بندی مبتنی بر مرکز مانند K-Means نیست و می‌تواند خوشه‌های با اشکال مختلف را بیابد. همچنین می‌تواند خوشه‌ای را که کاملاً توسط خوشه‌ی دیگر احاطه شده است (ولی به آن متصل نیست)، شناسایی کند. مشکل تک‌یالی (خوشه‌های مختلف با مسیر نازکی از نقاط به هم متصل باشند) نیز با پارامتر MinPoints کاهش یافته است.

دو پارامتر Min\_Samples و Eps در الگوریتم DB\_SCAN وجود دارد. هر نقطه داده، از دیگر نقاط داده فاصله‌ای دارد. هر نقطه‌ای که فاصله‌اش با یک نقطه مفروض کمتر از Eps باشد، به عنوان همسایه آن نقطه در نظر گرفت می‌شود. هر نقطه داده مفروضی که min\_samples همسایه داشته باشد، یک نقطه مرکزی<sup>۱</sup> محسوب می‌شود. نمونه‌های مرکزی<sup>۲</sup> که نسبت به یکدیگر نزدیک‌تر از فاصله Eps هستند، در خوشه مشابهی قرار می‌گیرند.

<sup>۱</sup> Core

<sup>۲</sup> Core Samples



## • Expectation Maximization

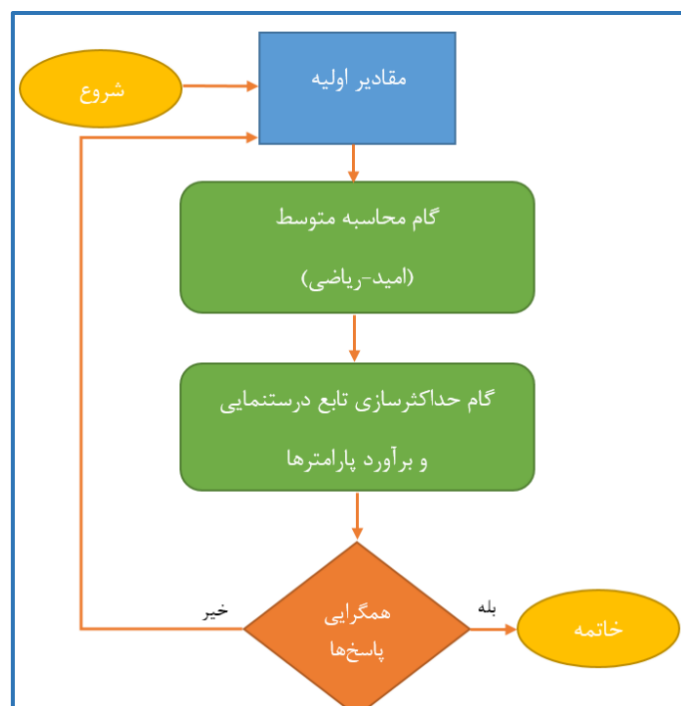
این مدل خوشه بندی که دقیقاً مربوط به آمار می باشد، بر اساس مدل های توزیع است. خوشه ها به راحتی می توانند به عنوان اشیایی تعریف شوند که به احتمال زیاد توزیع یکسانی دارند. یک ویژگی خوب این رویکرد این است که نمونه برداری از اشیاء تصادفی از یک توزیع، دقیقاً شبیه نحوه تولید مجموعه داده های مصنوعی است. الگوریتم خوشه حداکثر انتظار یا EM دارای دو بخش است. بخش اول Expectation یا همان انتظار است، که در این قسمت، الگوریتم بررسی می کند که هر کدام از نمونه ها (نقاط) به کدام توزیع گوسی بیشتر نزدیک هستند و در واقع احتمال عضویت یک نمونه (نقطه) را به تابع گوسی پیدا می کند. بحث EM در الگوریتم K\_Means نیز وجود دارد. در هر دور Iteration از این الگوریتم، هر کدام از نمونه ها (نقاط) به نزدیک ترین مرکز خوشه تعلق پیدا می کردند. این یعنی الگوریتم انتظار دارد که این نمونه (نقطه) به یک خوشه ی خاص تعلق داشته باشد. حال در GMM<sup>۱</sup> در هر بار تلاش، الگوریتم انتظار دارد تا احتمال عضویت یک نمونه (نقطه) را به هر کدام از توزیع های گوسی مورد نظر نسبت دهد. بخش دوم روش EM در واقع Maximization یا بیشینه سازی است. در الگوریتم K\_Means در هر دور نیاز بود که مرکز خوشه را تغییر داده شود تا شباهت مرکز خوشه با تمامی نمونه ها (نقاط) داخل آن خوشه بیشینه شود. در الگوریتم GMM نیز پارامترها که شامل وزن، میانگین و کواریانس است در هر دور آپدیت می شود تا در نهایت توزیع گوسی برای هر کدام از خوشه ها تشکیل شود و در واقع شباهت توزیع داده ها به صورت گوسی بیشینه شود.

در این الگوریتم پارامترها را به دو گروه تفکیک می کنیم. در مرحله اول برای پارامترهای گروه اول مقادیر اولیه<sup>۲</sup> در نظر گرفته و براساس این پارامترها، مقدار پارامترهای گروه دوم را برآورد می کنیم. حال براساس پارامترهای گروه دوم و داده های مشاهده شده پارامترهای گروه اول را برآورد

<sup>۱</sup> Gaussian Mixture Model

<sup>۲</sup> Initial Values

کرده و آن‌ها را بهبود می‌دهیم. این کار تا زمانی که الگوریتم به همگرایی برسد تکرار می‌شود. نتیجه نهایی از برآورد پارامترهای مدل در الگوریتم EM، تابع درستنمایی را به حداکثر مقدار خود می‌رساند.



## ➤ طبقه بندی

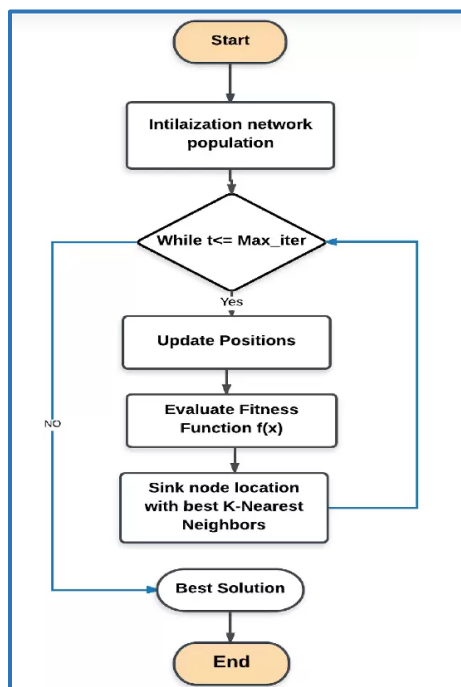
طبقه بندی<sup>۱</sup> یکی از زیرشاخه های اصلی داده کاوی و یادگیری ماشین است که می توان به صورت هوشمند اشیا مختلف را در یک تصویر شناسایی کرد. در واقع طبقه بندی، ارزشیابی ویژگی های مجموعه ای از داده ها و سپس اختصاص دادن آن ها به مجموعه ای از گروه های از پیش تعریف شده است. مسائل طبقه بندی به شناسایی خصوصیات منجر می شوند که مشخص می نمایند هر مورد به کدام گروه تعلق دارد. این الگو هم می تواند برای فهم داده موجود و هم برای پیش بینی این که هر نمونه جدید چگونه کار می کند استفاده شود. روش های طبقه بندی استفاده شده در این پروژه شامل روش های زیر است:

### • KNN

الگوریتم K-نزدیک ترین همسایگی<sup>۲</sup> از پرکاربردترین روش های طبقه بندی در پروژه های گوناگون یادگیری ماشین و داده کاوی است. در این الگوریتم، هر کدام از نمونه های جدید با تمامی نمونه های قبلی مقایسه می شوند و به هر کدام از نمونه های قبلی که نزدیک تر باشند، به آن دسته از نمونه ها تعلق می گیرند. در واقع این الگوریتم از تشابه ویژگی برای پیش بینی مقادیر نقاط داده جدید استفاده می کند به این معنی که به نقطه داده جدید بر اساس میزان مطابقت آن با نقاط مجموعه آموزشی، یک مقدار تخصیص می دهد. تغییر پارامتر K می تواند نتایج را به کلی تغییر دهد و خروجی الگوریتم وابستگی کامل به مقدار K دارد. تغییر K باعث تغییر مرز های دسته بندی می شود و با افزایش K مرز بین دسته بندی ها، روان تر می شود. اگر K برابر با ۱ باشد نام الگوریتم به نزدیک ترین همسایه تغییر می یابد که به این معنی است که تنها به سراغ نزدیک ترین همسایه می رود. در شکل پایین نحوه عملکرد این الگوریتم به شکل فلوچارت آمده است که به درک بهتر موضوع به شما کمک خواهد کرد.

<sup>۱</sup> Classification

<sup>۲</sup> K-Nearest Neighbors



## • Decision Tree

یکی از پرکاربردترین الگوریتم‌های داده‌کاوی، الگوریتم درخت تصمیم است. در داده‌کاوی، درخت تصمیم یک مدل پیش‌بینی کننده است به طوری که می‌تواند برای هر دو مدل رگرسیون و طبقه بندی مورد استفاده قرار گیرد. زمانی که درخت برای کارهای طبقه‌بندی استفاده می‌شود، به عنوان درخت طبقه‌بندی<sup>۱</sup> و هنگامی که برای فعالیت‌های رگرسیونی به کار می‌رود، درخت رگرسیون<sup>۲</sup> نامیده می‌شود. الگوریتم درخت تصمیم یکی از الگوریتم‌های پرکاربرد برای زمانی که حجم داده‌ها بسیار بالا است، می‌باشد. در ساختار درخت تصمیم، پیش‌بینی به دست آمده از درخت در قالب یک سری قواعد توضیح داده می‌شود. هر مسیر از ریشه تا یک برگ درخت تصمیم، یک قانون را بیان می‌کند و در نهایت برگ با کلاسی که بیشترین مقدار رکورد در آن تعلق گرفته برچسب می‌خورد.

<sup>۱</sup> Classification Tree

<sup>۲</sup> Regression Decision Tree

اجزای اصلی درخت تصمیم شامل برگ<sup>۱</sup>، ریشه<sup>۲</sup> و شاخه<sup>۳</sup> است. برگ گره‌هایی که تقسیم‌های متوالی در آن‌جا پایان می‌یابد. برگ‌ها با یک کلاس مشخص می‌شوند. منظور از ریشه، گره آغازین درخت است و در هر گره داخلی به تعداد جواب‌های ممکن شاخه ایجاد می‌شود. معمولاً پیچیدگی درخت توسط یکی از معیارهای کل گره‌ها، کل برگ‌ها، عمق درخت و تعداد مشخصه‌های به کار رفته تعیین می‌شود.

معیارهای انتخاب مشخصه برای انشعاب درخت تصمیم، به طوری که بهترین تفکیک داده‌های آموزش را به کلاس‌های برچسب‌دار داشته باشد اهمیت بسیار دارد. سه معیار انتخاب مشخصه شناخته شده عبارتند از:

❖ Information gain

❖ Gain ratio

❖ Gini index

## • Deep learning

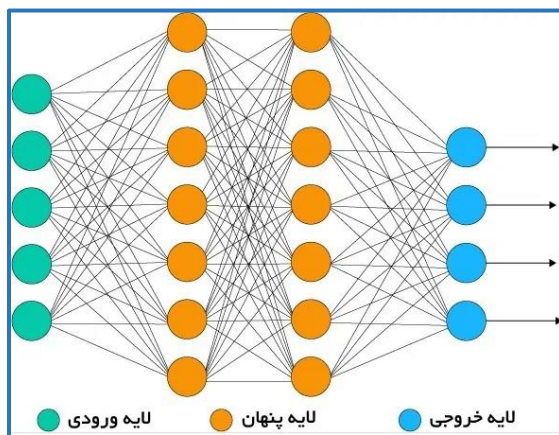
یادگیری عمیق یک روش یادگیری ماشینی الگوریتمی چندلایه است که از شبکه‌ی عصبی مغز انسان الهام گرفته شده است. مانند شبکه‌های عصبی عمیق موجود در مغز، معماری یادگیری عمیق در تکمیل اطلاعات پردازش شده در زمینه‌ی هوش مصنوعی نقش مهمی دارد. کلمه‌ی "عمیق" به تعداد لایه‌هایی اطلاق می‌شود که از طریق آن‌ها داده‌ها در طی فرایندهایی تبدیل می‌شوند. با استفاده از یادگیری عمیق، می‌توان کارهایی که معمولاً انسان‌ها انجام می‌دهند را به ماشین‌ها یاد داد. در واقع یادگیری عمیق به شکلی نظارت شده (مثل طبقه بندی) و یا بدون نظارت (مثل تحلیل الگو) یادگیری می‌کنند. یادگیری عمیق دربرگیرنده شبکه‌های عصبی مصنوعی است که روی شبکه‌هایی مشابه با آنچه در مغز انسان وجود

<sup>۱</sup> Leaf Nodes

<sup>۲</sup> Root Node

<sup>۳</sup> Branches

دارد مدل شده‌اند. با جابه‌جایی داده در این میس مصنوعی<sup>۱</sup>، هر لایه یک جنبه از داده‌ها را پردازش، دورافتادگی‌ها را فیلتر، موجودیت‌های مشابه را علامت‌گذاری و خروجی نهایی را تولید می‌کند.



### • لایه ورودی

لایه ورودی<sup>۲</sup> شامل نورون‌هایی است که کاری به جز دریافت ورودی‌ها و پاس دادن آن‌ها به دیگر لایه‌ها انجام نمی‌دهند. تعداد لایه‌ها در لایه ورودی باید مساوی خصیصه‌ها<sup>۳</sup> یا ویژگی‌های موجود در مجموعه داده باشد.

### • لایه خروجی

لایه خروجی<sup>۴</sup> ویژگی پیش‌بینی شده است، این لایه اساساً به نوع مدلی که ساخته می‌شود بستگی دارد.

### • لایه پنهان

<sup>۱</sup> Artificial Mesh

<sup>۲</sup> Input Layer

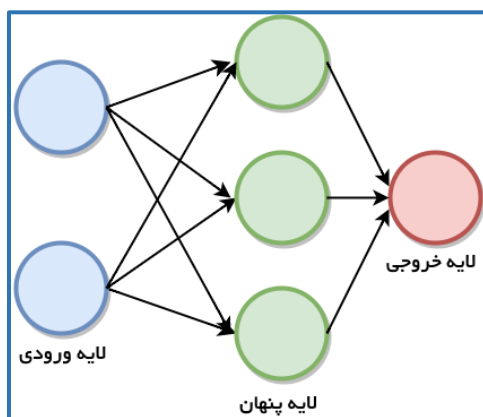
<sup>۳</sup> Attributes

<sup>۴</sup> Output Layer

در میان لایه ورودی و خروجی لایه‌های پنهان<sup>۱</sup> بسته به نوع مدل قرار دارند. لایه‌های پنهان شامل گستره وسیعی از نورون‌ها می‌شوند. نورون‌ها در لایه پنهان، تبدیل‌ها<sup>۲</sup> را پیش از پاس دادن ورودی‌ها روی آن‌ها اعمال می‌کنند. با آموزش دیدن شبکه، وزن‌ها به روز رسانی می‌شوند تا پیش‌بینی‌ها بهتر باشند.

## • وزن نورون‌ها

شبکه‌های عصبی نظارت شده پیش‌خور<sup>۳</sup> از جمله اولین و موفق‌ترین الگوریتم‌های یادگیری به شمار می‌آید. به این الگوریتم شبکه‌های عمیق<sup>۴</sup> پرسپترون چند لایه<sup>۵</sup> یا به صورت ساده‌تر شبکه‌های عصبی نیز گفته می‌شود و معماری متداول آن با یک لایه پنهان نمایش داده می‌شود. هر نورون با وزنی به نورون دیگر مرتبط می‌شود. شبکه، به منظور تولید یک مقدار خروجی، ورودی را با فعال‌سازی نورون‌ها پردازش می‌کند. به این کار مسیر رو به جلو<sup>۶</sup> در شبکه گفته می‌شود.



<sup>۱</sup> Hidden Layers

<sup>۲</sup> Transformations

<sup>۳</sup> Feedforward Supervised Neural Networks

<sup>۴</sup> Deep Networks

<sup>۵</sup> Multi-Layer Perceptron | MLP

<sup>۶</sup> Forward Pass

## ➤ سری های زمانی<sup>۱</sup>

یک سری زمانی مجموعه ای از مشاهدات درباره ی یک متغیر است که در نقاط گسسته ای از زمان که معمولاً فاصله های مساوی دارند، اندازه گیری شده و بر حسب زمان مرتب شده اند. سری زمانی مجموعه مشاهداتی است که به ترتیب زمان انجام شده است. تحلیل سری زمانی در مورد یک پدیده، ایجاد یک مدل آماری برای داده های وابسته به زمان براساس اطلاعات گذشته ی آن پدیده است. با این کار امکان پیش بینی در مورد آینده پدیده مورد بحث میسر می شود. به بیان دیگر تحلیل سری زمانی، ایجاد مدلی گذشته نگر است تا امکان تصمیمات آینده نگر را فراهم سازد.

## • ARIMA<sup>۲</sup>

روش میانگین متحرک خودهمبسته یکپارچه ARIMA یکی از مدل های آماری به منظور تحلیل و پیش بینی داده های سری زمانی است. مخفف Autoregressive Integral Moving Average است. این یک تعریف از میانگین متحرک Autoregressive است به طوری که مفهوم یکپارچه سازی را نیز در خود جای داده است. این اختصار توصیفی است و جنبه های کلیدی ذیل را شامل می شود:

AR، خودگرایی: روشی که از رابطه وابسته بین یک متغیر وابسته و تعدادی مشاهد بهره می برد.

I، یکپارچه: استفاده از تفکیک مشاهدات خام (به عنوان مثال کم کردن متغیر از مشاهده در مرحله زمان قبلی) به منظور ثابت ماندن سری زمانی.

MA، میانگین متحرک: روشی که از وابستگی بین یک مشاهده و یک خطای باقیمانده از مدل میانگین متحرک بهره می برد.

<sup>۱</sup> Time series

<sup>۲</sup> Autoregressive Integrated Moving Average



هر یک از این مولفه ها به طور مشخص در مدل به عنوان یک پارامتر در نظر گرفته شده اند. همچنین این مدل برای هریک از پارامترها از علامت استاندارد از  $ARIMA(p,d,q)$ ، استفاده می کند که در آن پارامترها با مقادیر صحیح جایگزین می شوند تا مدل  $ARIMA$  اجرا شود. پارامترهای مدل  $ARIMA$  به شرح زیر تعریف می شوند:

$p$ : تعداد مشاهدات موجود در مدل که ترتیب تأخیر نیز نامیده می شود.

$d$ : تعداد دفعاتی که مشاهدات خام از یکدیگر متمایز می شوند، درجه تمیز نیز نامیده می شود.

$q$ : اندازه پنجره میانگین متحرک است.

یک مدل  $ARIMA(p,d,q)$  فاکتورهای این چندجمله ای را به صورت زیر بیان می کنند:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

و بنابراین می تواند به صورت یک نمونه خاص  $ARMA("p+d", "q")$ ، که دارای بخش خودگردانی با ریشه های واحد است دیده شود. به همین دلیل هر  $ARIMA$  با  $d > 0$  به طور کلی ثابت<sup>۱</sup> نیست.

مدل های  $ARIMA$  برای داده های با فرایندهای غیر ثابت<sup>۲</sup> که روندهایی کاملاً قابل تشخیص دارند به کار می روند.

- یک روند ثابت (با میانگین صفر) مدل شده به صورت  $d=0$
- یک روند خطی (برای مثال رفتار رشد خطی) مدل شده به صورت  $d=1$
- یک روند مربعی (برای مثال رفتار رشد مرتبه دوم) مدل شده به صورت  $d=2$

<sup>۱</sup> Stationary

<sup>۲</sup> Non-Stationary

در این موارد مدل ARIMA را می توان به صورت ترکیبی از دو مدل دید. اولی غیر ثابت است:

$$Y_t = (1 - L)^d X_t$$

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) Y_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t.$$

در این حالت تکنیک های استاندارد پیش بینی می تواند برای فرموله کردن فرایند  $Y_t$  به کار رود و سپس با داشتن تعداد کافی مشاهدات اولیه به پیش بینی  $X_t$  پرداخت.

## ۲-۵ رویکرد مدل سازی قطعات یدکی

نحوه دسته بندی کالاهای موجود بر اساس مدل مدیریتی معایبی به شرح زیر به همراه دارد:

- دسته بندی کلان قطعات بدون در نظرگیری مشابهت و الگوی مصرف تنها بر اساس شاخص مصرف در نظر نگرفتن فاکتورهای مانند نوع کالا، قیمت، تاریخچه تعمیرات و ...
- در نظرگیری مصرف ۳ سال اخیر به صورت کلی و صرف نظر نمودن از سابقه اطلاعات مصرف به صورت ماهیانه و فصلی
- دینامیک نبودن تغییر مدل مدیریتی قطعات تعمیر پذیر بر اساس میزان مصرف

پس از انجام تحلیل های اولیه بر روی داده های در دست از گردش کالا، داده های تعمیرات و تعمیرگاه مرکزی و بررسی مشابهت رفتاری و وجود روند در بین کالاهای مختلف الگوی زیر برای دسته بندی اولیه کالاها بر اساس داده های موجود تعیین گردید. لذا تمامی قطعات یدکی فعال، غیرراکد و با قابلیت خرید و دارای سابقه مصرف در فولاد مبارکه در یک یا بیش از یک دسته به صورت زیر قرار می گیرند:

دسته کلی	دسته بندی بر اساس رویکرد مدل سازی	تعداد قطعات
پرمصرف	قطعات پرمصرف با مصرف ۱۲ سال اخیر غیرصفر	۱,۷۱۲
	قطعات پرمصرف با مصرف ۵ سال اخیر غیرصفر	۴,۵۹۶
	قطعات پرمصرف با مصرف ۳ سال اخیر غیرصفر	۷,۱۱۶
کم مصرف	قطعات کم مصرف (در ۳ سال اخیر مصرف صفر داشته اند)	۵۱,۵۷۷
تعمیر پذیر	قطعات تعمیرپذیر دارای سابقه تعمیرات	۱۱,۰۱۳

به منظور پیش بینی قطعات یدکی، دسته بندی اولیه روی این قطعات انجام شد و فرآیند پیش بینی به ترتیب روی قطعات یدکی پر مصرف، کم مصرف و قطعات تعمیرپذیر انجام شد. در ادامه به شرح نحوه مدل سازی در هر یک از دسته های تعیین شده پرداخته می شود.

### ۵-۳ مدل سازی دسته پرمصرف

مطابق با دسته بندی ارائه شده در جدول بالا قطعات یدکی پرمصرف در ابتدا به قطعاتی اطلاق شد که در ۳، ۵ یا ۱۲ سال اخیر مصرف خالص غیرصفر داشته باشند. به منظور مشخص نمودن دسته مناسب تر، مدل سازی قطعات برای هر سه دسته ۳ و ۵ و ۱۲ سال مصرف خالص غیر صفر انجام و بر اساس نتایج ارزیابی دسته مصرف ۳ سال به عنوان دسته بندی مناسب تر انتخاب گردید. لذا در این پروژه قطعات یدکی پرمصرف به قطعاتی اطلاق می شود که مصرف خالص ۳ سال اخیر غیرصفر باشد.

### ➤ آماده سازی داده:

محاسبه مقدار مصرف خالص با در نظرگیری مقادیر زیر:

- مقدار حواله از انبار
- میزان بازگشتی به انبار
- میزان خریدهای مستقیم و اضطراری

- تغییر ماهیت و انتقال موجودی

### ➤ پاک سازی داده:

- حذف داده های پرت و نامعتبر
- دسته بندی اقلام براساس مجموع مصرف خالص کالا در کلیه دوره ها
- متعادل سازی داده های با مصرف خالص منفی

### ➤ خوشه بندی:

- تعیین بهترین روش خوشه بندی
- بهینه سازی پارامترهای روش منتخب
- تعیین تعداد خوشه برای هر دسته کالا
- محاسبه نرخ ضریب هر کالا در خوشه

### ➤ پیش بینی:

- اجرای الگوی سری زمانی در هر دسته
- بهینه سازی پارامترهای اتورگرسیون
- پیش بینی دوره های آتی هر دسته
- پیش بینی دوره آتی کالاهای موجود در هر دسته

## ۴-۵ مدل سازی دسته کم مصرف

براساس آنچه در قسمت قبل برای قطعات یدکی پرمصرف تعریف شد، قطعات یدکی کم مصرف عبارتند از مجموعه قطعات یدکی که در حداقل یکی از ۳ سال اخیر مصرف خالص صفر دارند. علاوه بر میزان خالص مصرف، پارامترهایی مانند مقدار میانگین مصرف، مجموع مصرف و واریانس مصرف، مقدار فاصله زمانی

مصارف، میانگین فاصله زمانی، مجموع فاصله زمانی و واریانس فاصله زمانی برای هر قطعه یدکی در نظر گرفته شده است.

با توجه به اینکه ماهیت این نوع قطعات، کم مصرف است. برای این دسته از قطعات دو پارامتر متوسط مقدار مصرف آتی<sup>۱</sup> (MQTY) و متوسط فاصله زمانی تا مصرف آتی<sup>۲</sup> (MTBD) پیش بینی خواهد شد. به این معنی که مقدار مصرف و فاصله زمانی مصرف پیش بینی خواهد شد.

### ➤ آماده سازی داده:

محاسبه مقدار مصرف خالص با در نظرگیری مقادیر زیر:

- مقدار حواله از انبار
  - میزان بازگشتی به انبار
  - میزان خریدهای مستقیم و اضطراری
  - تغییر ماهیت و انتقال موجودی
- محاسبه پارامترهای دیگر شامل:
- میانگین، مجموع و واریانس مصرف
  - میانگین، مجموع و واریانس فاصله زمانی مصرف

### ➤ خوشه بندی:

- نرمالایز کردن برخی از فیلدها
- خوشه بندی اقلام بر اساس روش K\_Means

<sup>۱</sup> Mean Quantity

<sup>۲</sup> Mean Time Between Demand

### ➤ پیش بینی مقدار مصرف با برچسب MQTY:

- در نظر گرفتن فیلد متوسط مقدار مصرف به عنوان Label
- بهینه سازی و پیش بینی مقدار مصرف قطعات یدکی با روش های Decision Tree, KNN, Deep Vote learning

### ➤ پیش بینی زمان مصرف با برچسب MTBD:

- در نظر گرفتن فیلد متوسط فاصله زمانی مصرف به عنوان Label
- بهینه سازی و پیش بینی مقدار مصرف قطعات یدکی با روش های Decision Tree, KNN, Deep Vote Learning

### ۵-۵ مدل سازی دسته تعمیرپذیر

قطعات یدکی تعمیرپذیر آن دسته از قطعاتی هستند که دارای سابقه تعمیرات در سیستم تعمیرگاه مرکزی باشند. با توجه به اینکه قطعات یدکی تعمیر پذیر در طی فرآیند تعمیرات ممکن است منجر به اسقاط شوند، لذا نتیجه تعمیر بعدی (اسقاط یا تعمیر) و همچنین فاصله زمانی تا اسقاط بعدی با استفاده از فرآیند داده کاوی برای این قطعات پیش بینی خواهد شد.

در ابتدای فرآیند داده کاوی، فیلدهای دیگری از جمله تعداد کل درخواست های هر قطعه، تعداد اسقاط هر قطعه، تعداد تعمیرات هر قطعه، بیسیک کد کالا، متوسط زمان تعمیر هر قطعه، تعداد کل تعمیرات هر قطعه، نتیجه تعمیر بعدی قطعه (اسقاط یا تعمیر)، تعداد تعمیر بین اسقاط هر قطعه، فاصله زمانی بین اسقاط، فاصله زمانی بین تعمیرات، نرخ اسقاط و نسبت اسقاط به تعمیرات برای هر قطعه تعریف شده است.

برای این دسته از قطعات دو پارامتر نتیجه تعمیر بعدی (NextResult) و متوسط فاصله زمانی تا اسقاط آتی<sup>۱</sup> (MTBS) پیش بینی خواهد شد.

### ➤ آماده سازی داده:

- در نظرگیری زمان خالص تعمیر از زمان بعد از AC
  - محاسبه مدت زمان تعمیر قطعه با در نظر گرفتن زمان تعلیق به علت انتظار قطعه
  - تخصیص فلگ از نظر ویژگی برون سپاری بودن یا نبودن
- محاسبه پارامترهای دیگر شامل:
- تعداد کل درخواست، تعداد اسقاط، تعداد تعمیرات هر قطعه
  - بیسیک کد کالا، متوسط زمان تعمیر هر قطعه، تعداد کل تعمیرات هر قطعه، نتیجه تعمیر بعدی قطعه (اسقاط یا تعمیر)
  - تعداد تعمیر بین اسقاط هر قطعه، فاصله زمانی بین اسقاط، فاصله زمانی بین تعمیرات، نرخ اسقاط و نسبت اسقاط به تعمیرات

### ➤ پاک سازی داده:

- اصلاحات داده های ثبت شده در سیستم CEW به علت حجم بالای داده های نامعتبر
- تک رکورد کردن کلیه درخواست های مربوط به یک ایتm
- خوشه بندی

<sup>۱</sup> Mean Time Between Scrap

➤ دسته بندی و پیش بینی با برچسب **NextResult**:

- در نظر گرفتن فیلد نتیجه تعمیر آتی (اصلاح/اسقاط) به عنوان Label
- کلاسه بندی قطعات یدکی با روش های مختلف مانند Decision Tree, KNN, Deep Learning

➤ دسته بندی و پیش بینی با **MTBS**:

- در نظر گرفتن فیلد متوسط زمان بین دو اسقاط به عنوان Label
- کلاسه بندی قطعات یدکی با روش های مختلف مانند Decision Tree, KNN, Deep Learning



## ۶- ارزیابی و استقرار

به منظور مقایسه نتایج مقادیر پیش بینی با مقادیر واقعی، شاخص های ارزیابی زیر تعریف شده است. دسته ای از این معیارها شامل معیارهای شناخته شده برای ارزیابی نتایج مدل پیش بینی استفاده شده است و تعدادی از شاخص های مربوط به Availability در جریان پروژه و توسط افراد خبره مطرح گردیده است. شاخص های خبرگی میزان کارآمدی مدل از نظر در دسترس بودن قطعات یدکی و جلوگیری از مازاد و کمبود قطعات یدکی را اندازه گیری می نماید. شاخص های جدول زیر برای پارامترهای قابل پیش بینی از نوع عدد صحیح استفاده می شود. در فرمول های جدول زیر  $P_{test}(t)$  مقدار پیش بینی شده و  $A_{test}(t)$  مقدار واقعی است.  $N$  نیز تعداد ارقام در هر دسته جهت محاسبه شاخص برای آن دسته می باشد.

معیار	عنوان	شرح	فرمول شاخص
RMSE	Root mean square error	مجذور میانگین مربعات خطا	$RMSE = \sqrt{\frac{\sum_{t=1}^N (P_{test}(t) - A_{test}(t))^2}{N}}$
MAPE	Mean absolute percentage error	میانگین درصد خطای مطلق	$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{ P_{test}(t) - A_{test}(t) }{ A_{test}(t) }$
SMAPE	Symmetric mean absolute percentage error	میانگین متقارن درصد خطای مطلق	$SMAPE = \frac{1}{N} \sum_{t=1}^N \frac{ P_{test}(t) - A_{test}(t) }{( A_{test}(t)  +  P_{test}(t) )/2}$
MAD	Median Absolute Difference	میانگین قدر مطلق خطا	$MAD = \frac{1}{N} \sum_{t=1}^N  P_{test}(t) - A_{test}(t) $
Availability(۱۰۰)	-	درصد مواقعی که میزان پیش بینی شده از واقعی بیشتر است	$\frac{1}{N} \sum_{t=1}^N (P_{test}(t) - A_{test}(t) \geq 0, 1, 0)$

معیار	عنوان	شرح	فرمول شاخص
Availability(L <sup>۱۰۰</sup> )	-	درصد مواقعی که میزان پیش بینی شده از واقعی کمتر است	$\frac{1}{N} \sum_{t=1}^N (P_{test}(t) - A_{test}(t) \leq 0, 1, 0)$
Availability(L <sup>۵۰</sup> )	-	درصد مواقعی که میزان پیش بینی شده از واقعی کمتر است و از ۵۰ درصد مقدار واقعی بیشتر است.	$\frac{1}{N} \sum_{t=1}^N (P_{test}(t) - A_{test}(t) \leq 0 \ \&\& \ P_{test}(t) \geq 0.5 * A_{test}(t), 1, 0)$

در صورتی که پارامتر قابل پیش بینی باینری باشد، شاخص های ارزیابی زیر استفاده می شود.

Predict			Actual
False	True		
False Negative (تعداد مواردی که مقدار true دارند و به نادرستی تشخیص داده شده اند)	True positive (تعداد مواردی که مقدار true دارند و به درستی تشخیص داده شده اند)	True	
True Negative (تعداد مواردی که مقدار false دارند و به درستی تشخیص داده شده اند)	False positive (تعداد مواردی که مقدار false دارند و به نادرستی تشخیص داده شده اند)	False	

معیار	شرح	فرمول
Accuracy	معیار صحت که برابر است با تعداد مواردی که درست پیش بینی کردیم تقسیم بر تعداد کل پیش بینی هایی که انجام شده است.	$Accuracy = \frac{sum(TP) + sum(TN)}{sum(TP) + sum(FP) + sum(FN) + sum(TN)}$
Precision	مواردی که برنامه به غلط پیش بینی کرده است که به آن False Positive می گوئیم	$precision = \frac{sum(TP)}{sum(TP) + sum(FP)}$

معیار	شرح	فرمول
	نسبت به پیش بینی‌های درست یا True Positive بیشتر باشد مقدار Precision کمتر خواهد شد.	
<b>Recall</b>	مواردی که ما انتظار داشتیم پیش بینی شوند ولی برنامه پیش بینی نکرده است که به False Negative آن می‌گوییم نسبت به پیش بینی‌های درست یا True Positive بیشتر باشد مقدار Recall کمتر خواهد شد.	$\text{recall} = \frac{\text{sum}(TP)}{\text{sum}(TP) + \text{sum}(FN)}$
<b>F-Measure</b>	زمانی که می‌خواهید معیار ارزیابی شما میانگینی از دو مورد قبلی یعنی همان Precision یا Recall باشد می‌توانید از میانگین هارمونیک این دو معیار استفاده کنید که به آن F-measure می‌گویند.	$\text{F-measure} = \frac{2 * (\text{precision} * \text{recall})}{\text{precision} + \text{recall}}$

## ۶-۱ ارزیابی نتایج دسته پرمصرف

همان گونه که در بخش مدل سازی شرح داده شده ، فرایند مدل سازی قطعات پرمصرف برای سه دسته از داده ها بر اساس میزان مصرف خالص آن ها انجام و نتایج مورد ارزیابی قرار گرفت. در این پروژه به منظور پیاده سازی مدل و ارزیابی نتایج از نرم افزار Rapidminer استفاده شده است. نمونه ای از دیتاست ورودی به نرم افزار، فرآیند داده کاوی و همچنین نتایج مربوط به ۳، ۵ و ۱۲ سال اخیر در ادامه آورده شده است. لازم به ذکر است که شرح فرآیند داده کاوی انجام شده برای ۳ سال اخیر آورده شده است.

➤ داده ورودی به نرم افزار Rapidminer:

در شکل زیر میزان مصرف قطعات یدکی که در سه سال اخیر مصرف غیر صفر دارند ماه به ماه به صورت مثال در ۳۶ دوره آورده شده است.

ITEM_ITEM_ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
630992.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	0.0	0.0	5.0	0.0
631000.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0
631022.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	3.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0
631057.0	0.0	12.0	0.0	0.0	10.0	0.0	0.0	4.0	0.0	8.0	5.0	0.0	0.0	0.0	0.0	4.0	0.0	4.0	4.0	0.0	4.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	2.0	0.0	0.0	6.0	0.0
631235.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	
631236.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	
631237.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	
631238.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	
631461.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	3.0	0.0
631480.0	0.0	2.0	0.0	1.0	0.0	4.0	12.0	6.0	8.0	4.0	5.0	4.0	6.0	8.0	4.0	8.0	0.0	14.0	2.0	18.0	4.0	12.0	4.0	0.0	0.0	16.0	0.0	12.0	0.0	0.0	0.0	10.0	0.0	12.0	0.0	0.0
631500.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0
631604.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	2.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
631725.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	8.0	0.0	0.0	1.0	0.0	2.0	0.0	0.0	0.0	0.0	
631739.0	0.0	9.0	0.0	0.0	3.0	2.0	0.0	0.0	1.0	1.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0
631797.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	0.0	0.0	1.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
631824.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
632046.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
632143.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
632149.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
632172.0	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.0	0.0	0.0	0.0	30.0	0.0	50.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.0	50.0	6.0	0.0	30.0	0.0	0.0
632173.0	70.0	0.0	0.0	0.0	2.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	0.0	0.0	0.0
632187.0	0.0	2.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0

➤ آماده سازی قطعات یدکی پرمصرف جهت پیش بینی:

جهت پیش بینی قطعات یدکی، ابتدا کلیه قطعات یدکی براساس مجموع مصرف در طی سه سال اخیر در ۱۱ دسته تقسیم بندی می شوند. حد پایین و حدبالای دسته ها به شرح زیر است.

دسته	حد پایین	حد بالا
اول	-infinity	۳
دوم	۳	۵.۵
سوم	۵.۵	۷.۵
چهارم	۷.۵	۱۱.۵
پنجم	۱۱.۵	۱۵.۵
ششم	۱۵.۵	۲۲.۵
هفتم	۲۲.۵	۳۲.۵
هشتم	۳۲.۵	۵۲.۵
نهم	۵۲.۵	۹۴.۵
دهم	۹۴.۵	۲۲۵.۵
یازدهم	۲۲۵.۵	infinity

سپس در هر یک از دسته ها، جهت پیش بینی بهتر برای قطعات یدکی مربوط به یک مجموعه، خوشه بندی با استفاده از روش های DBSCAN، EM و K-Means بر روی هر دسته انجام شد. لازم به ذکر است که پارامترهای هریک از روش ها بهینه سازی شده است. از میان روش های پیاده سازی شده، K-Means به عنوان بهترین روش انتخاب شد. نتایج خوشه بندی و تعداد خوشه ها در هر دسته در جدول زیر آمده است.

دسته	تعداد در دسته	تعداد خوشه در دسته	تعداد در خوشه
(-infinity, ۳)	۱۵۰	۲	۱۳۰
			۲۰
(۳, ۵, ۵)	۵۷۱	۲	۲۵۳
			۳۱۸
(۵, ۵, ۷, ۵)	۵۷۷	۲	۴۸۲

دسته	تعداد در دسته	تعداد خوشه در دسته	تعداد در خوشه
			۹۵
(۷,۵,۱۱,۵)	۸۸۴	۲	۴۸۸
			۳۹۶
(۱۱,۵,۱۵,۵)	۶۲۷	۲	۳۵۳
			۲۷۴
(۱۵,۵,۲۲,۵)	۷۳۲	۲	۳۷۵
			۳۵۷
(۲۲,۵,۳۲,۵)	۶۹۰	۲	۴۱۲
			۲۷۸
(۲۲,۵,۵۲,۵)	۷۴۰	۲	۴۱۲
			۳۲۸
(۵۲,۵,۹۴,۵)	۷۲۰	۲	۴۱۵
			۳۰۵
(۹۴,۵,۲۲۵,۵)	۷۰۸	۲	۴۱۸
			۲۹۰
(۲۲۵,۵,infinity)	۷۱۲	۲	۷۰۸
			۴

پس از خوشه بندی داده ها، پیش بینی برای هر خوشه طی ۱۲ ماه آتی به صورت ماه به ماه با استفاده از روش سری های زمانی Arima انجام شد. پارامترهای روش Arima برای کلیه خوشه ها بهینه سازی شده است و میزان خطای مربوط به پیش بینی با حذف مقدار واقعی ۱۲ ماه گذشته و مقدار پیش بینی شده برای این دوره ها محاسبه شده است.

دسته	نام خوشه	p	d	q	MAD	MAPE	MSE	ARIMA (RMSE)
اول	-infinityCluster_۰	۱	۰	۰	۰,۰۴۹	Infinity	۰,۰۰	۰,۰۲ +/- ۰,۰۳۷
	infinityCluster_۱	۳	۰	۲	۰,۱۱۳	Infinity	۰,۰۲	۰,۱۸۴ +/- ۰,۳۲۳
دوم	۲.۵Cluster_۰	۳	۰	۰	۰,۰۵۲	Infinity	۰,۰۰۴	۰,۰۲۰ +/- ۰,۰۳۷
	۲.۵Cluster_۱	۲	۰	۰	۰,۰۵	Infinity	۰,۰۰۴	۰,۰۴۵ +/- ۰,۰۴۰
سوم	۵.۵Cluster_۰	۳	۰	۰	۰,۰۶	Infinity	۰,۰۰۶	۰,۰۳۴ +/- ۰,۰۲۲
	۵.۵Cluster_۱	۲	۰	۲	۰,۱۶	Infinity	۰,۱۳	۰,۰۵۰ +/- ۰,۰۵۶
چهارم	۷.۵Cluster_۰	۱	۱	۰	۰,۰۶	۰,۲۳۴	۰,۱۱	۰,۰۶۰ +/- ۰,۰۴۵
	۷.۵Cluster_۱	۱	۰	۱	۰,۰۶۷	۰,۳۴۸	۰,۰۰۶	۰,۰۷۰ +/- ۰,۰۹۳
پنجم	۱۱.۵Cluster_۰	۳	۰	۱	۰,۰۸	۰,۴۰۵	۰,۰۱	۰,۰۸۰ +/- ۰,۰۵۹
	۱۱.۵Cluster_۱	۲	۰	۱	۰,۰۷	۰,۲۳۹	۰,۰۰۸	۰,۰۹۳ +/- ۰,۰۴۵
ششم	۲۲.۵Cluster_۰	۲	۰	۰	۰,۱۱	۰,۲۵۷	۰,۰۱	۰,۰۸۱ +/- ۰,۱۰۰
	۲۲.۵Cluster_۱	۱	۰	۰	۰,۰۷	۰,۱۹۵	۰,۰۰۹	۰,۰۹۳ +/- ۰,۱۳۹
هفتم	۳۲.۵Cluster_۰	۱	۰	۰	۰,۱۰۲	۰,۱۹۲	۰,۱۲۶	۰,۲۱۰ +/- ۰,۱۶۹
	۳۲.۵Cluster_۱	۲	۰	۰	۰,۱۸۳	۰,۲۹۵	۰,۰۴۵	۰,۲۳۲ +/- ۰,۱۷۵
هشتم	۵۲.۵Cluster_۰	۳	۰	۲	۰,۱۲۲	۰,۱۵۱	۰,۰۲۴	۰,۲۰۹ +/- ۰,۱۶۴
	۵۲.۵Cluster_۱	۲	۰	۰	۰,۱۴۱	۰,۱۲۷	۰,۰۲۹	۰,۳۱۵ +/- ۰,۳۴۴
نهم	۹۴.۵Cluster_۰	۲	۰	۲	۰,۲۸۳	۰,۱۹۷	۰,۰۹۷	۰,۳۷۶ +/- ۰,۳۲۱
	۹۴.۵Cluster_۱	۱	۰	۰	۰,۴۶۳	۰,۲۴۹	۰,۲۸۸	۰,۵۳۵ +/- ۰,۵۸۹
دهم	۲۲۵.۵Cluster_۰	۱	۰	۰	۰,۴۰۶	۰,۱۵۹	۰,۲۴۸	۰,۷۱۴ +/- ۰,۸۷۳
	۲۲۵.۵Cluster_۱	۳	۰	۰	۰,۸۴۸	۰,۲۰۶	۱,۱۲۲	۱,۰۸۹ +/- ۱,۲۷۲
یازدهم	۲۲.۵Cluster_۰	۱	۰	۰	۵,۱۷	۰,۲۲	۵۰,۹۹	۸,۴۱۸ +/- ۱۰,۵۴۰
	۲۲.۵Cluster_۱	۱	۱	۱	۱۴۰,۲۰۸	۰,۹۵	۲۳۷۴۹۶۸	۳۵۶,۵۸۸ +/- ۳۲۱,۷۰۲

مقدار پیش بینی انجام شده به صورت ماهیانه برای هر خوشه در شکل زیر ارائه شده است.

id	att_37	att_38	att_39	att_40	att_41	att_42	att_43	att_44	att_45	att_46	att_47	att_48
-Infinitycluster_0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
-Infinitycluster_1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3cluster_0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.1	0.1
3cluster_1	0.1	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1
5.5cluster_1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
5.5cluster_0	0.0	0.0	0.2	0.0	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2
7.5cluster_1	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
7.5cluster_0	0.2	0.1	0.2	0.2	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.2
11.5cluster_1	0.2	0.1	0.4	0.3	0.2	0.1	0.2	0.3	0.3	0.3	0.3	0.3
11.5cluster_0	0.3	0.3	0.2	0.4	0.2	0.4	0.3	0.3	0.3	0.3	0.3	0.3
15.5cluster_0	0.4	0.3	0.4	0.5	0.3	0.4	0.3	0.4	0.4	0.4	0.4	0.4
15.5cluster_1	0.3	0.2	0.5	0.5	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4
22.5cluster_0	0.4	0.4	0.7	0.6	0.4	0.7	0.5	0.6	0.6	0.6	0.6	0.6
22.5cluster_1	0.5	0.4	0.7	0.7	0.5	0.8	0.5	0.7	0.7	0.7	0.7	0.7
32.5cluster_1	0.8	0.6	0.9	0.8	0.7	0.9	0.7	0.8	0.8	0.8	0.8	0.8
32.5cluster_0	0.7	1.0	1.1	1.3	0.9	1.0	1.2	1.1	1.1	1.1	1.1	1.1
52.5cluster_0	1.2	1.3	1.7	2.0	1.6	2.3	1.4	1.8	1.8	1.8	1.8	1.8
52.5cluster_1	1.3	0.8	1.7	1.7	1.2	1.7	1.2	1.4	1.4	1.4	1.4	1.4
94.5cluster_1	3.7	4.0	5.9	4.6	3.6	4.4	4.3	4.3	4.3	4.3	4.3	4.3
94.5cluster_0	1.9	2.1	3.2	2.7	3.5	2.8	3.3	2.8	2.7	2.7	2.7	2.7
225.5cluster_0	14.7	28.5	30.3	32.4	17.3	31.3	21.6	25.7	26.4	26.6	26.6	26.6
225.5cluster_1	775.0	2125.0	1662.5	2062.5	1250.0	2225.0	1500.0	1169.2	1202.1	1198.9	1199.2	1199.1

نتایج ارائه شده در جدول بالا مربوط به هر خوشه است. به منظور محاسبه مقدار پیش بینی شده برای هر قطعه یدکی، ابتدا سهم مصرف هر قطعه یکی از مجموع مصارف هر خوشه محاسبه می شود و در نهایت حاصل ضرب مصرف هر خوشه در سهم مصرف هر قطعه از مصارف کل خوشه میزان مقدار پیش بینی شده برای ماه های آتی محاسبه خواهد شد. در شکل زیر به عنوان نمونه سهم هر قطعه یدکی از مجموع مصارف خوشه مربوطه اش آورده شده است.

ITEM_ITEM_ID	id	RatioTotal
630992.0	22.5cluster_0	1.1
631000.0	3cluster_0	1.0
631022.0	15.5cluster_0	1.0
631057.0	52.5cluster_0	1.0
631235.0	5.5cluster_1	1.1
631236.0	5.5cluster_1	1.1
631237.0	5.5cluster_1	1.1
631238.0	5.5cluster_1	1.1
631461.0	7.5cluster_1	1.0
631480.0	94.5cluster_1	1.2
631500.0	5.5cluster_0	1.1
631604.0	7.5cluster_1	1.0
631725.0	15.5cluster_0	1.1
631739.0	15.5cluster_0	1.0
631797.0	15.5cluster_1	0.9
631824.0	5.5cluster_1	1.1



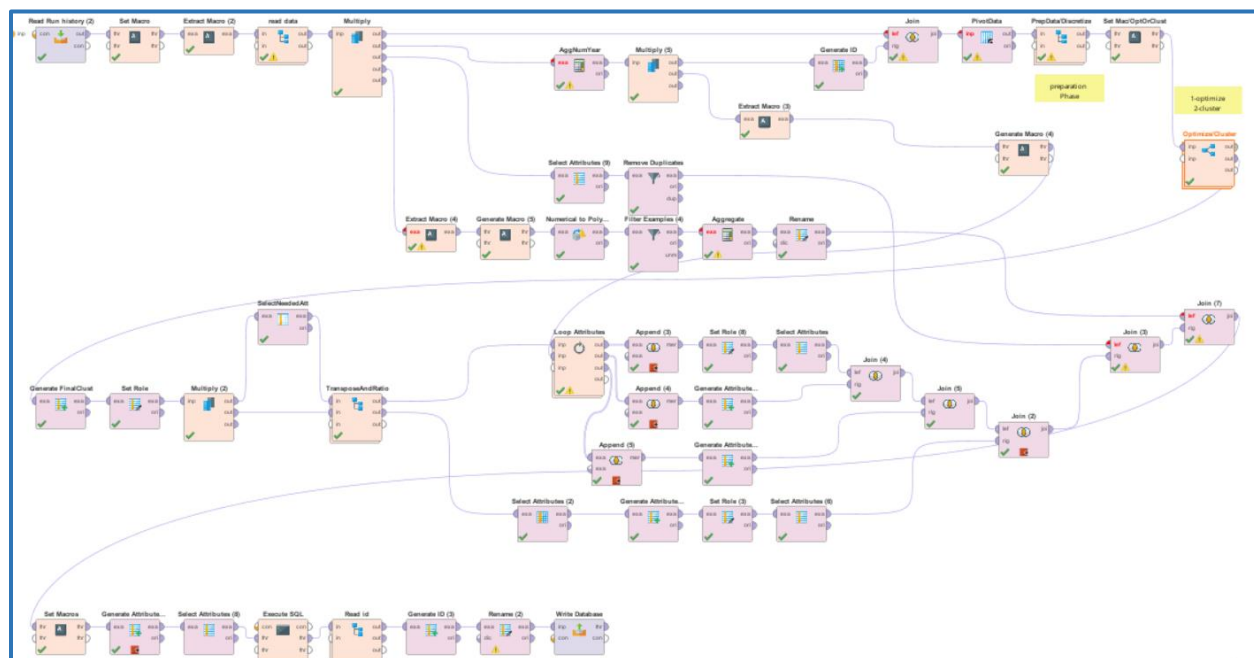
نتایج حاصل از سهم هر خوشه و میزان پیش بینی شده برای هر خوشه در ۱۲ ماه آتی به عنوان نمونه برای تعدادی از قطعات یدکی به شرح زیر است.

ITEM_ITEM_ID	forecast
630992.0	8.0
631000.0	1.0
631022.0	6.0
631057.0	23.0
631235.0	2.0
631236.0	2.0
631237.0	2.0
631238.0	2.0
631461.0	3.0
631480.0	62.0
631500.0	3.0
631604.0	3.0
631725.0	6.0
631739.0	6.0
631797.0	5.0
631824.0	2.0
632046.0	1.0

کلیه فرآیند انجام شده برای تعاریف دیگر قطعات یدکی پر مصرف در طی ۵ و ۱۲ سال مصرف غیرصفر دارند پیاده سازی شده است که نتایج مربوطه در جدول زیر آمده است. نتایج نشان می دهد که تعریف قطعات یدکی با مصرف ۳ سال اخیر غیر صفر نتایج بهتری داشته است و به عنوان تعریف نهایی قطعات پرمصرف انتخاب شده است.

تعاریف مختلف قطعات پرمصرف	MAD	MAPE	RMSE
مصرف ۳ سال اخیر غیر صفر	۰.۳۵	۰.۲۹	۴.۴۷
مصرف ۵ سال اخیر غیر صفر	۰.۵۴	۰.۸۹	۱۷.۵۶
مصرف ۱۲ سال اخیر غیر صفر	۰.۹۶	۰.۷۸	۴۳.۴۷

تصویری از مدل گرافیکی پیش بینی قطعات یدکی پرمصرف در نرم افزار Rapidminer در شکل زیر ارائه شده است.



## ۶-۲ ارزیابی نتایج دسته کم مصرف

شرح فرایند داده کاوی انجام شده و نتایج برای قطعات یدکی کم مصرف به شرح زیر است:

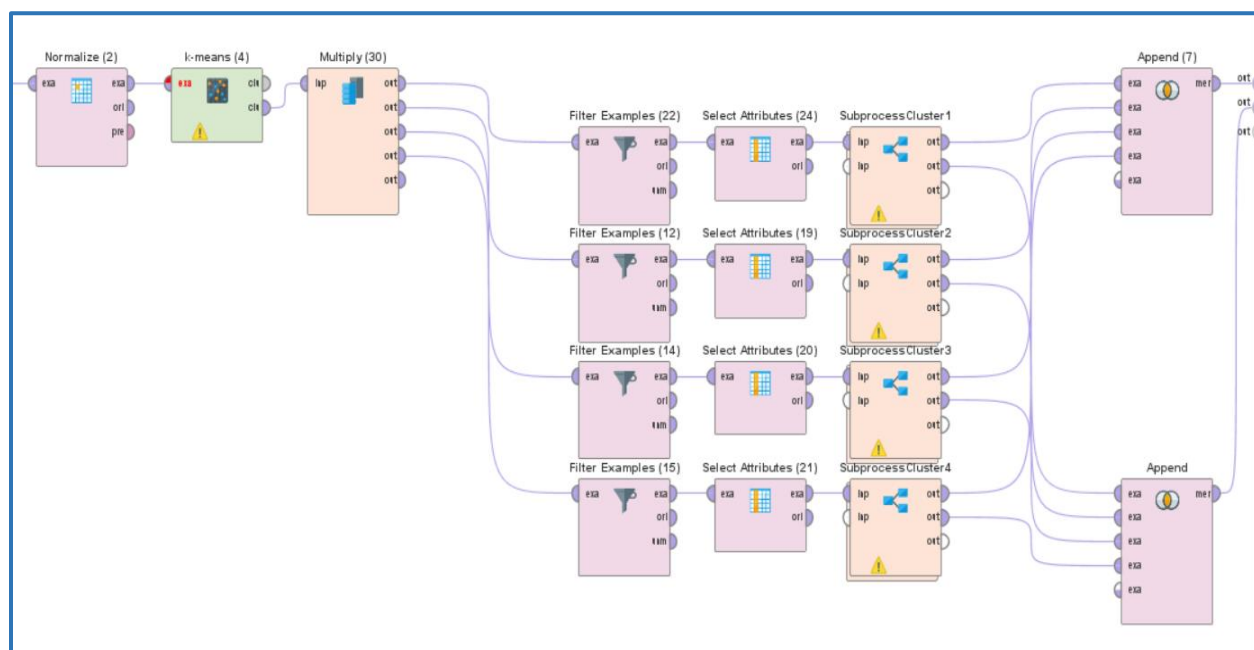
➤ داده ورودی به نرم افزار Rapidminer:

اطلاعات قطعات یدکی کم مصرف شامل بیسیک کد، تعداد حواله انبار مربوط به هر قطعه، مجموع، متوسط و واریانس مصرف، متوسط، مجموع و واریانس فاصله زمانی بین حواله ها به منظور پیش بینی پارامترها در دیتاست به شکل زیر تهیه شده است.

ITEM_ITEM_ID	پیسیک کد	تعداد حواله انبار	مجموع مصرف	متوسط میزان مصرف	واریانس مصرف	متوسط فاصله زمانی بین حواله ها	مجموع فاصله زمانی بین حواله ها	واریانس فاصله زمانی بین حواله ها
630997.0	10.0	1.0	1.0	1.0	0.0	72.0	72.0	0.0
631003.0	10.0	8.0	9.0	1.1	0.1	122.7	122.7	57.9
631004.0	10.0	4.0	8.0	2.0	0.0	84.2	84.2	105.8
631005.0	10.0	1.0	4.0	4.0	0.0	72.0	72.0	0.0
631006.0	10.0	8.0	3642.0	455.3	45410.5	101.4	101.4	78.4
631008.0	10.0	3.0	26.0	8.7	41.3	59.9	29.9	120.4
631012.0	10.0	1.0	2.0	2.0	0.0	72.0	72.0	0.0
631016.0	10.0	2.0	2.0	1.0	0.0	53.7	53.7	721.8
631017.0	10.0	2.0	2.0	1.0	0.0	53.7	53.7	721.8
631018.0	10.0	3.0	3.0	1.0	0.0	131.9	66.0	1672.4
631025.0	10.0	1.0	2.0	2.0	0.0	72.0	72.0	0.0
631027.0	10.0	1.0	1.0	1.0	0.0	72.0	72.0	0.0
631031.0	10.0	1.0	1.0	1.0	0.0	72.0	72.0	0.0
631032.0	10.0	1.0	2.0	2.0	0.0	72.0	72.0	0.0
631036.0	10.0	2.0	2.0	1.0	0.0	2.1	2.1	1.1
631039.0	10.0	2.0	5.0	2.5	0.5	28.4	28.4	202.1
631045.0	10.0	1.0	1.0	1.0	0.0	72.0	72.0	0.0
631056.0	10.0	1.0	5.0	5.0	0.0	72.0	72.0	0.0
631059.0	10.0	2.0	2.0	1.0	0.0	68.0	68.0	1154.9
631060.0	10.0	5.0	10.0	2.0	1.5	32.5	8.1	61.1

### ➤ خوشه بندی قطعات یدکی کم مصرف

پس از آماده سازی داده ها، کلیه قطعات با استفاده از روش های خوشه بندی K\_Means، DBSCAN، EM و بهینه سازی پارامترهای مربوطه در چهار خوشه تقسیم بندی شده است. از میان روش های خوشه بندی، روش K\_Means به عنوان بهترین روش انتخاب شده است و خوشه بندی نهایی براساس این روش انجام شد.



### ➤ پیش بینی متوسط مقدار مصرف و متوسط فاصله زمانی بین مصارف

پس از خوشه بندی داده ها، پارامترهای متوسط مقدار مصرف و متوسط فاصله زمانی بین مصارف برچسب شده و با استفاده از روش های کلاسه بندی از جمله KNN، Decision Tree، Deep learning و روش Vote که ترکیبی از روش های کلاسه بندی است، پیش بینی شده است. لازم به ذکر است که کلیه نتایج به دست آمده پس از بهینه سازی پارامترهای روش های کلاسه بندی به دست آمده است. ابتدا مقادیر شاخص های ارزیابی با استفاده از کلیه روش های کلاسه بندی برای لیبل MQTY و سپس MTBD آورده شده است.

روش KNN با لیبل MQTY							
تعداد اقلام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)	Availability(L۱۰۰)	Availability(L۵۰)
۱۶,۰۰۹	۳,۲۷۷	۰,۶۴۷	۲۸۲۴,۳۶۷	۵۳,۱۴۵	۰,۷۷۸	۰,۲۲۲	۰,۱۳۸
۶,۱۹۵	۰,۷۷۶	۰,۲۶۷	۲۳,۷۷۰	۴,۸۷۵	۰,۷۹۹	۰,۲۰۱	۰,۱۶۰
۱۷,۶۹۴	۲,۲۳۷	۰,۵۱۱	۴۳۳,۲۵۸	۲۰,۸۱۵	۰,۷۶۰	۰,۲۴۰	۰,۱۶۱
۱۱,۶۳۰	۲,۹۳۰	۰,۵۵۶	۶۷۶,۷۵۱	۲۶,۰۱۴	۰,۷۶۴	۰,۲۳۶	۰,۱۵۵

روش Deep Learning با لیبل MQTY							
تعداد اقلام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)	Availability(L۱۰۰)	Availability(L۵۰)
۱۶,۰۰۹	۳,۵۱۶	۱,۹۴۲	۵۴۶,۸۹۳	۲۳,۳۸۶	۰,۵۹۱	۰,۴۰۹	۰,۲۴۵
۶,۱۹۵	۰,۳۴۹	۰,۱۸۱	۴,۱۸۱	۲,۰۴۵	۰,۵۲۵	۰,۴۷۵	۰,۴۲۷
۱۷,۶۹۴	۱,۲۲۵	۰,۴۳۷	۲۷۹,۶۰۹	۱۶,۷۲۲	۰,۵۶۷	۰,۴۳۳	۰,۳۰۳
۱۱,۶۳۰	۱,۸۵۲	۰,۶۸۴	۲۴۸,۸۲۵	۱۵,۷۷۴	۰,۵۶۵	۰,۴۳۵	۰,۲۵۴

روش Decision Tree با لیبل MQTY							
تعداد ارقام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)	Availability(L۱۰۰)	Availability(L۵۰)
۱۶,۰۰۹	۱.۱۶۷	۰.۰۹۷	۲۲۵۴.۲۲۷	۴۷.۴۷۹	۰.۷۵۱	۰.۲۴۹	۰.۲۴۸
۶,۱۹۵	۰.۲۰۳	۰.۰۳۳	۴.۹۱۰	۲.۲۱۶	۰.۸۷۳	۰.۱۲۷	۰.۱۲۷
۱۷,۶۹۴	۰.۸۶۴	۰.۰۸۶	۳۴۲.۵۳۹	۱۸.۵۰۸	۰.۷۸۶	۰.۲۱۴	۰.۲۱۴
۱۱,۶۳۰	۱.۲۹۵	۰.۰۹۶	۳۶۹.۴۸۵	۱۹.۲۲۲	۰.۷۷۵	۰.۲۲۵	۰.۲۲۴

روش Vote با لیبل MQTY							
تعداد ارقام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)	Availability(L۱۰۰)	Availability(L۵۰)
۱۶,۰۰۹	۰.۶۲۲	۰.۱۸۹	۵۰۲.۳۸۷	۲۲.۴۱۴	۰.۸۶۱	۰.۱۳۹	۰.۱۳۹
۶,۱۹۵	۰.۲۱۶	۰.۰۸۳	۲.۵۱۹	۱.۵۸۷	۰.۶۳۸	۰.۳۶۲	۰.۳۶۲
۱۷,۶۹۴	۰.۳۶۱	۰.۱۱۶	۴۰.۳۵۹	۶.۳۵۳	۰.۵۶۷	۰.۴۳۳	۰.۴۲۶
۱۱,۶۳۰	۰.۴۸۱	۰.۱۰۴	۳۲.۵۸۴	۵.۷۰۸	۰.۵۶۳	۰.۴۳۷	۰.۴۲۰

براساس نتایج به دست آمده روش Decision Tree به عنوان بهترین روش با کمترین خطا برای پیش بینی مقدار MQTY انتخاب شده است. در ادامه جداول مربوط به نتایج MTBD آورده شده است.

روش KNN با لیبل MTBD							
تعداد اقدام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)	Availability(L۱۰۰)	Availability(L۵۰)
۱۷,۶۹۴	۲.۰۴۱	۰.۰۷۸۶۸۱	۳۱.۰۹۷	۵.۵۷۶	۱۷,۶۹۴	۰.۲۵۹	۰.۲۵۸
۱۶,۰۰۹	۲.۱۶۳	۰.۰۸۸۲۴۲	۳۰.۴۳۶	۵.۵۱۷	۱۶,۰۰۹	۰.۲۷۷	۰.۲۷۷
۶,۱۹۵	۲.۸۱۳	۰.۱۲۲۲۰۵	۴۴.۹۹۵	۶.۷۰۸	۶,۱۹۵	۰.۲۶۷	۰.۲۶۵
۱۱,۶۳۰	۲.۲۸۷	۰.۰۸۹۵۸۸	۳۴.۳۵۸	۵.۸۶۲	۱۱,۶۳۰	۰.۲۷۸	۰.۲۷۸

روش Deep Learning با لیبل MTBD							
تعداد اقدام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)	Availability(L۱۰۰)	Availability(L۵۰)
۱۷,۶۹۴	۰.۶۳۲	۰.۰۳۶۲۷۷	۰.۹۹۳	۰.۹۹۶	۰.۴۸۰	۰.۵۲۰	۰.۵۱۸
۱۶,۰۰۹	۰.۶۹۳	۰.۰۴۳۱۸۵	۱.۲۰۹	۱.۱۰۰	۰.۵۴۴	۰.۴۵۶	۰.۴۵۴
۶,۱۹۵	۰.۶۵۷	۰.۰۴۸۰۵۸	۱.۳۶۱	۱.۱۶۷	۰.۶۶۲	۰.۳۳۸	۰.۳۳۶
۱۱,۶۳۰	۰.۸۸۴	۰.۰۵۷۴۹۱	۱.۹۲۸	۱.۳۸۸	۰.۴۱۶	۰.۵۸۴	۰.۵۸۳

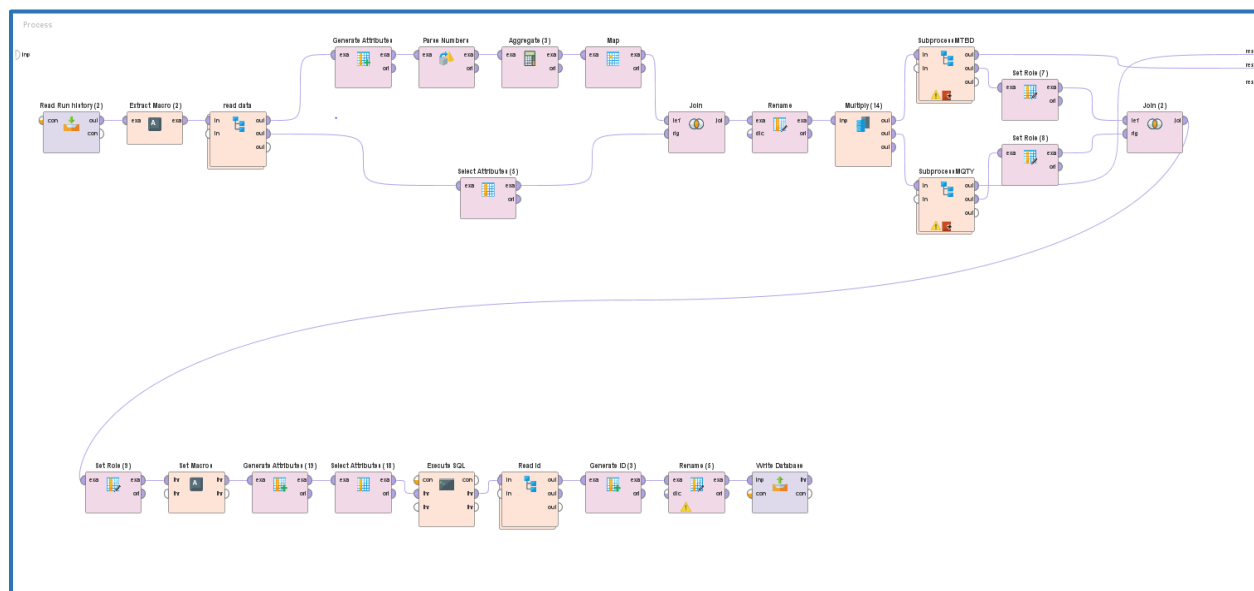
روش DecisionTree با لیبل MTBD

تعداد اقدام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)	Availability(L۱۰۰)	Availability(L۵۰)
۱۷,۶۹۴	۰.۷۷۱	۰.۰۵۵۹۱۲	۲.۰۷۶	۱.۴۴۱	۰.۷۰۲	۰.۲۹۸	۰.۲۹۷
۱۶,۰۰۹	۰.۹۳۰	۰.۰۶۸۶۱۵	۲.۷۶۱	۱.۶۶۲	۰.۶۷۲	۰.۳۲۸	۰.۳۲۷
۶,۱۹۵	۰.۹۲۰	۰.۰۶۷۱۹۶	۲.۶۶۳	۱.۶۳۲	۰.۶۶۸	۰.۳۳۲	۰.۳۳۱
۱۱,۶۳۰	۰.۸۴۵	۰.۰۵۷۸۳	۲.۶۰۲	۱.۶۱۳	۰.۷۰۸	۰.۲۹۲	۰.۲۹۲

روش Vote با لیبل MTBD

تعداد اقدام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)	Availability(L۱۰۰)	Availability(L۵۰)
۱۷,۶۹۴	۰.۴۲۳	۰.۰۲۶۰۸۷	۰.۴۴۳	۰.۶۶۶	۰.۵۳۷	۰.۴۶۳	۰.۴۶۳
۱۶,۰۰۹	۰.۶۴۵	۰.۰۲۲۳۱۲	۰.۵۵۶	۰.۷۴۶	۰.۵۰۱	۰.۴۹۹	۰.۴۹۹
۶,۱۹۵	۰.۴۱۷	۰.۰۲۴۶	۰.۴۱۱	۰.۶۴۱	۰.۵۹۰	۰.۴۱۰	۰.۴۱۰
۱۱,۶۳۰	۰.۳۷۹	۰.۰۳۱۱۶	۰.۳۴۳	۰.۵۸۵	۰.۶۴۹	۰.۳۵۱	۰.۳۵۰

براساس نتایج به دست آمده روش Vote به عنوان بهترین روش با کمترین خطا برای پیش بینی مقدار MTBD انتخاب شده است. در ادامه تصویری از فرآیند پیش بینی قطعات یدکی کم مصرف در نرم افزار Rapidminer آورده شده است.



### ۳-۶ ارزیابی نتایج دسته تعمیر پذیر

قطعات یدکی تعمیر پذیر، در مجموع قطعات یدکی کم مصرف و پرمصرف هستند که مقدار پیش بینی آنها در یکی از روش های بالا محاسبه شده است. برای این دسته از قطعات دو پارامتر نتیجه تعمیر بعدی (NextResult) و متوسط فاصله زمانی تا اسقاط آتی (MTBS) پیش بینی خواهد شد. فرآیند داده کاوی این قطعات با رویکرد خوشه بندی به چند دسته تقسیم شد و سپس با استفاده از روش های کلاسه بندی مانند KNN، Deep Learning، Decision Tree، پیش بینی پارامترهای MTBS، NextResult انجام شد. شرح فرآیند داده کاوی در ادامه آورده شده است.

➤ داده ورودی به نرم افزار Rapidminer:

اطلاعات قطعات یدکی تعمیر پذیر شامل کد کالا، تعداد کل درخواست های هر قطعه، تعداد اسقاط هر قطعه، تعداد تعمیرات هر قطعه، بیسیک کد کالا، متوسط زمان تعمیر هر قطعه، تعداد کل تعمیرات هر قطعه، نتیجه

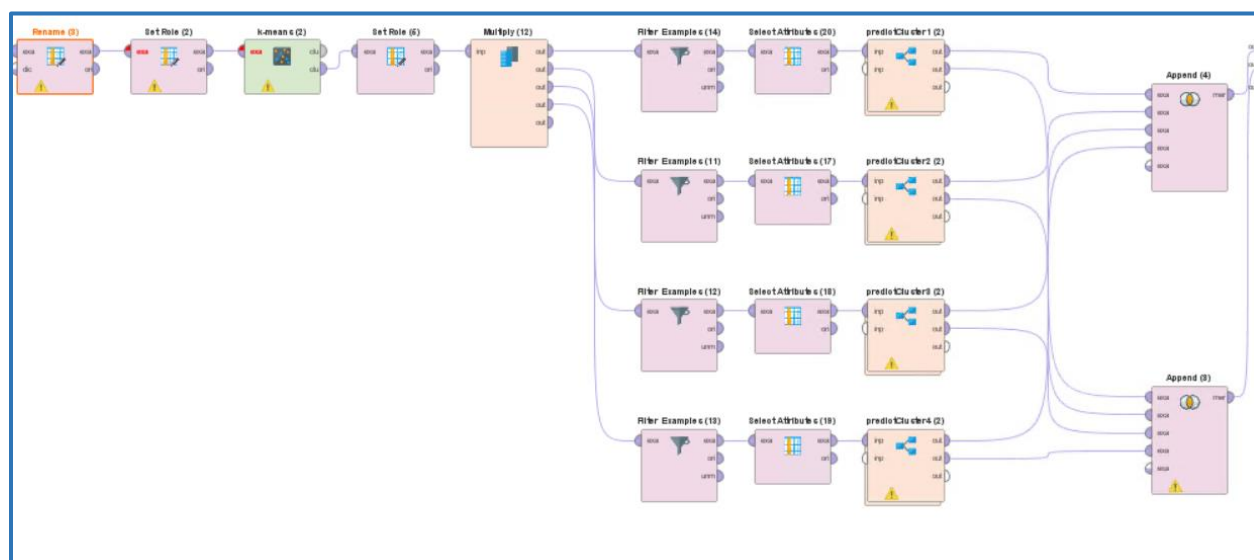


تعمیر بعدی قطعه (اسقاط یا تعمیر)، تعداد تعمیر بین اسقاط هر قطعه، فاصله زمانی بین اسقاط، فاصله زمانی بین تعمیرات، نرخ اسقاط و نسبت اسقاط به تعمیرات برای هر قطعه است.

[illegible]

➤ خوشه بندی قطعات یدکی

همان طور که در بخش قبلی توضیح داده شد، برای قطعات یدکی تعمیر پذیر مقادیر MTBS و NextResult پیش بینی خواهند شد. به منظور پیش بینی مقدار NextResult کلیه قطعات با هم در یک دسته قرار دارند و پیش بینی می شوند اما در پیش بینی مقدار MTBS پس از آماده سازی داده ها، کلیه قطعات با استفاده از روش های خوشه بندی K\_Means ، DBSCAN ، EM و بهینه سازی پارامترهای مربوطه در چهارخوشه تقسیم بندی شده است. از میان روش های خوشه بندی، روش K\_Means به عنوان بهترین روش انتخاب شده است و خوشه بندی نهایی براساس این روش انجام شد.



➤ پیش بینی نتیجه تعمیر بعدی:

پس از خوشه بندی داده ها، پارامتر نتیجه تعمیر بعدی برچسب شده و با استفاده از روش های کلاسه بندی از جمله KNN، Decision Tree، Deep Learning و روش Vote که ترکیبی از روش های کلاسه بندی است، پیش بینی شده است. لازم به ذکر است که کلیه نتایج به دست آمده پس از بهینه سازی پارامترهای روش های کلاسه بندی به دست آمده است. در جدول زیر مقادیر شاخص های ارزیابی با استفاده از کلیه روش های کلاسه بندی برای لیبل NextResult آورده شده است.

پیش بینی نتیجه تعمیر بعدی					
روش	قطعه در هر تعداد دسته	Accuracy	Precision	Recall	F-Measure
KNN	۵۱۷۵۵	۰٫۹۳	۰٫۶۱	۰٫۱۳	۰٫۲۱
Decision Tree	۵۱۷۵۵	۰٫۹۳	۰٫۷۱	۰٫۲۳	۰٫۳۵
Deep Learning	۵۱۷۵۵	۰٫۹۱	۰٫۴۳	۰٫۴۹	۰٫۴۶
Vote	۵۱۷۵۵	۰٫۹۳	۰٫۷۵	۰٫۲۶	۰٫۳۹

براساس نتایج به دست آمده، روش Vote به عنوان بهترین روش برای پیش بینی مقدار نتیجه تعمیر بعدی هر قطعه انتخاب شده است.

### ➤ پیش بینی متوسط فاصله زمانی بین اسقاط:

پس از خوشه بندی داده ها، پارامترهای متوسط فاصله زمانی بین اسقاط لیبل شده و با استفاده از روش های کلاسه بندی از جمله KNN، Decision Tree، Deep learning و روش Vote که ترکیبی از روش های کلاسه بندی است، پیش بینی شده است. لازم به ذکر است که کلیه نتایج به دست آمده پس از بهینه سازی پارامترهای روش های کلاسه بندی به دست آمده است. مقادیر شاخص های ارزیابی با استفاده از کلیه روش های کلاسه بندی برای لیبل MTBS آورده شده است.

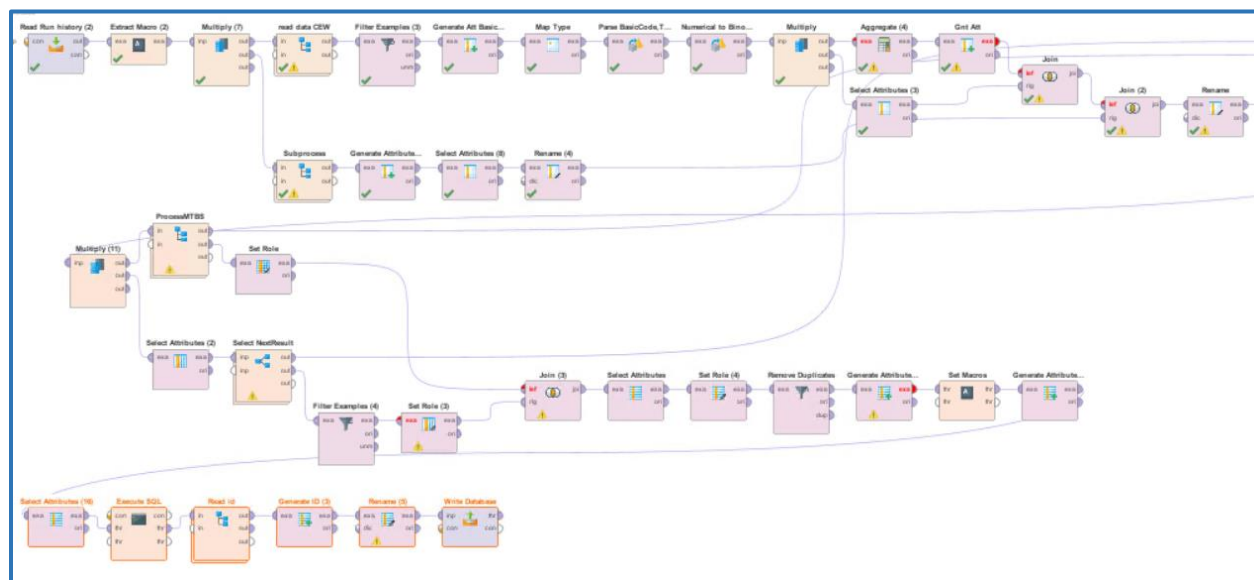
روش Deep Learning با لیبل MTBS					
تعداد اقسام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)
۱۴۵۴	۲۷۶,۸۶	۰.۰۸	۰,۰۸	۳,۷۰	۰,۴۴
۲۲۹۹	۱۹۲,۰۰	۰.۱۰	۰,۱۰	۲۴,۹۸	۰,۴۲
۲۸۲۰	۴۶,۸۴	۰.۱۳	۰,۰۹	۷,۹۳	۰,۵۶
۲۱۹	۱۲,۷۴	۰.۰۰	۰,۰۰	۱۶,۱۸	۰.۴۹

روش KNN با لیبل MTBS					
تعداد اقسام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)
۱۴۵۴	۳۶۷,۹۹	۰,۱۱	۰,۱۰	۴۳,۴۳	۰,۴۹
۲۲۹۹	۱۸۲,۱۹	۰,۱۰	۰,۱۰	۲۴,۲۹	۰,۴۸
۲۸۲۰	۴۶,۳۵	۰,۱۴	۰,۱۵	۷,۱۲	۰,۵۷
۲۱۹	۳۳۹,۱۸	۰,۱۲	۰,۱۱	۵۳,۲۹	۰,۴۵

روش Decision Tree با لیبل MTBS					
تعداد اقلام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)
۱۴۵۴	۱۴۲,۹۳	۰,۰۴	۰,۰۴	۲۲,۹۰	۰,۵۲
۲۲۹۹	۱۵۸,۷۲	۰,۰۹	۰,۰۸	۲۱,۹۳	۰,۴۹
۲۸۲۰	۱۴,۱۳	۰,۰۳	۰,۰۳	۲۳,۷۸	۰,۵۰
۲۱۹	۲۴۲,۸۵	۰,۰۹	۰,۰۸	۲۱,۲۲	۰,۵۰

روش Vote با لیبل MTBS					
تعداد اقلام در هر دسته	MAD	MAPE	MSE	RMSE	Availability(۱۰۰)
۱۴۵۴	۱۹۷,۶	۰,۱	۰,۱	۲۴,۱	۰,۵
۲۲۹۹	۱۰۴,۵	۰,۱	۰,۱	۱۳,۲	۰,۶
۲۸۲۰	۴۲,۴	۰,۱	۰,۱	۳۸,۳	۰,۶
۲۱۹	۵۵,۴	۰,۰۲	۰,۰	۳۶,۷	۰,۴

براساس نتایج به دست آمده روش vote به عنوان بهترین روش با کمترین خطا برای پیش بینی مقدار MTBS انتخاب شده است. در ادامه تصویری از فرآیند پیش بینی قطعات یدکی تعمیر پذیر در نرم افزار Rapidminer آورده شده است.



#### ۴-۶ استقرار در سیستم مدیریت کالا

با توجه به پیش بینی های انجام شده در قطعات یدکی کم مصرف، پر مصرف و تعمیر پذیر، در سیستم مدیریت کالا IS-suite تب مقادیر پیش بینی به فرم پرکاربرد "کنترل دوره ای سطح ذخیره اقلام" اضافه شده و تصویری از این فرم در ادامه آورده شده است. هر زمان که سطح ذخیره موجودی پایین تر از سطح ذخیره ایمن قرار گیرد اطلاعات کالا در این فرم نمایش داده می شود.

[illegible]

## ۷- جمع بندی و پیشنهادات آتی

در نظام های کنترل موجودی قطعات و کالاها، نرخ مصرف قطعات به عنوان یک پارامتر بسیار مهم در محاسبات مقدار اقتصادی سفارش و نقطه سفارش دهی شناخته می شود. لذا وجود داده کافی، تعیین الگوی مناسب مصرف و محاسبات دقیق نرخ مصرف حائز اهمیت است. در این پروژه سعی بر این بوده است که با شناخت دقیق داده و با به کارگیری نظر افراد خبره و استفاده از تکنیک های داده کاوی، مدل های مناسب و معتبر برای پیش بینی نرخ مصرف قطعات یدکی طراحی و پیاده سازی گردد، اما همان گونه که بیان گردید وجود داده صحیح و معتبر و کافی از الزامات اصلی پروژه های داده کاوی است.

در جریان شناخت و تحلیل داده های موجود در سیستم های اطلاعاتی فولاد مبارکه، اشکالات داده ای و فرایندی که منجر به ثبت نادرست یا ناقص داده ها می شود، مشاهده گردید. ارتباط تنگاتنگ بین سه واحد کنترل موجودی قطعات یدکی، نگهداری و تعمیرات و تعمیرگاه مرکزی الزام وجود یک فرایند یکپارچه و منظم بین این سه واحد به منظور ثبت کافی و دقیق داده را ایجاد می نماید. در جریان انجام پروژه به صورت عمده اشکالات داده ای و فرایندی زیر شناسایی گردید:

- نبود ارتباط فرایندی بین واحدهای انبار، تعمیرگاه مرکزی و تعمیرات و در نتیجه اشکال در ردیابی قطعات
- وجود انبارهای پای کار و برگشت حجم بالای قطعات به انبار مرکزی بدون اتصال به حواله انبار
- نبود سیستم کدینگ یکپارچه و منسجم و وجود قطعات تکراری و مشابه
- مشکلات در فرایند اسقاط و به روزرسانی موجودی انبار در ارتباط انبار و تعمیرگاه مرکزی
- عدم تعریف کل قطعات تعمیرپذیر در درختواره تجهیزات

- دستی بودن و کمبود کنترل های لازم برای ورود دقیق و صحیح داده و عدم دسترسی به تاریخچه داده کافی

در راستای حل این مشکلات و ایجاد بسترهای ثبت داده ها در آینده پیشنهادات زیر ارائه گردید:

- ایجاد مکانیزم حذف انبارهای پای کار و کاهش حجم قطعات بازگشتی
- ایجاد مکانیزم کدینگ یکپارچه قطعات و قابلیت دسته بندی قطعات براساس ماهیت
- ایجاد مکانیزم ثبت حواله برای خریدهای مستقیم و اضطراری
- بازبینی فرایند موجود کنترل موجودی قطعات
- طراحی و ساخت بستر داده ای به منظور ایجاد قابلیت ردیابی قطعات بین فرایندها
- ایجاد مکانیزم یکپارچگی محل نصب و داده های قطعات یدکی با درختواره تجهیزات نت