دوماهنامه علمی-تخصصی مطالعات کاربردی در علوم مدیریت و توسعه
سال پنجم، شماره ۴ (پیاپی: ۲۴)، مرداد ۱۳۹۹

ارائه مدل جهت خوشه‌بندی کشورهای درگیر ویروس کووید-۱۹ با به‌کارگیری روش‌های خود سازمان‌ده

زهره مؤمنی [1*]

تاریخ دریافت: ۹۹/۰۴/۲۷
تاریخ پذیرش: ۹۹/۰۵/۲۷
کد مقاله: ۸۴۵۳۰

چکـیده

در اواخر دسامبر سال ۲۰۱۹، ویروس جدیدی، با نام کروناویروس جدید-۲۰۱۹، باعث آغاز شیوع پنومونی از ووهان، به سراسر کشور چین شد که در حال حاضر تهدیدات بهداشتی بزرگی را برای سلامتی عمومی جهان ایجاد کرده است. بیماری همه گیر کووید- ۱۹ ناشی از کرونا ویروس جدید- ۲۰۱۹ در سراسر جهان، در حال گسترش است و تا اول مارس ۲۰۲۰ تعداد ۶۷ کشور، از جمله ایران را مبتلا و درگیر کرده است. در این پژوهش، داده‌های مرتبط با کرونا ویروس جدید-۲۰۱۹ که از سایت بهداشت جهانی اخذ شده، به منظور خوشه بندی، از روش های خود سازمان‌ده از جمله الگوریتم‌های کوهونن و کا از میانگین و خوشه بندی دومرحله ای بهره گرفته شده است که هدف تمامی روش ها این است که کدام رکورد در کدام خوشه جای می گیرد، سپس با انجام آزمون، کارایی این سه تکنیک تعیین شده است. نتایج حاکی از آن است که روش خوشه بندی دومرحله ای کارایی بالاتری نسبت به دو روش دیگر دارد و در خوشه بندی شاهد نتایج مطلوب تری با بهره گیری از این روش هستیم.

واژگـان کلـیدی: کووید-۱۹، خوشه بندی، روش‌های خود سازمان‌ده

۱- دانشجوی دکتری تخصصی، مدیریت صنعتی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران؛
zohrehmomeni21@gmail.com

۱- مقدمه

خبر ابتلای چندین نفر به یک ذات الریه غیرمعمول در ابتدای سال نو میلادی ۲۰۲۰ به سازمان بهداشت جهانی از سوی چین باعث معرفی نوع جدیدی از کروناویروس به عنوان عامل ایجاد یک بیماری تنفسی جدید گردید. با گسترش بسیار سریع این بیماری در چین و پس از آن به سایر نقاط دنیا، کروناویروس جدید با نام علمی *SARS-CoV-2* و بیماری حاصل از آن به نام کووید-۱۹[1] نگرانی و وحشت زیادی را در بین مردم جهان به وجود آورد و سازمان بهداشت جهانی نیز طی اطلاعیه ای، شیوع این ویروس را عامل وضعیت اضطراری بهداشت عمومی در سرتاسر جهان اعلام نمود. بیماری کروناویروس[2] ۲۰۱۹ یا کووید-۱۹ که به آن بیماری تنفسی حاد ان‌کاو-۲۰۱۹[3] نیز گفته می‌شود، بیماری‌ای عفونی است که بر اثر کروناویروس سندرم حاد تنفسی ۲ ایجاد می‌شود. این بیماری دلیل دنیاگیری ۲۰-۲۰۱۹ کروناویروس است. علایم معمول آن تب، سرفه، تنگی نفس و به تازگی نابویایی هستند. درد عضلانی، گلودرد، ناچشایی و سرخی چشم از جمله نشانه‌های کمتر معمول آن هستند. با این که اکثریت موارد این بیماری باعث علایم خفیف می‌شود، بعضی از موارد به سینه‌پهلو و نارسایی چند اندامی پیشرفت می‌کند. نرخ مرگ و میر بین ۱٪ و ۵٪ تخمین زده می‌شود ولی بر حسب سن و دیگر شرایط سلامتی تغییر می‌کند. این بیماری اساساً از طریق قطرات ریز تنفسی افراد مبتلا، وقتی سرفه یا عطسه می‌کنند، به سایر افراد سرایت می‌کند. زمان مابین در معرض بیماری قرار گرفتن و بروز نشانه‌ها، بین ۲ و ۱۴ روز است. از طریق شستن دست‌ها و دیگر تدابیر بهداشتی، می‌توان از پخش آن جلوگیری کرد (توکلی و همکاران، ۱۳۹۸ و Thompsom, 2020 & Lai et al., 2020). محمدی و روحانی (۱۳۹۶) در مقاله ای تحت عنوان "به کارگیری روش های خوشه بندی کا از میانگین، میانگین فازی و گوستافسون کسل در تلفیق نتایج وارون سازی داده های توموگرافی لرزهای انکساری و مقاومت ویژه الکتریکی برای ارزیابی آبرفت و سنگ بستر" به مطالعه پرداخته اند. در این مطالعه با استفاده از روش شاخص دان برای بهینه سازی تعداد خوشه ها عدد ۱۲ به دست آمد که با توجه به نقشه های به دست آمده برای مقاومت ویژه الکتریکی و توموگرافی لرزهای انکساری، تعداد خوشه مناسبی است. با توجه به بررسی های انجام شده در این مطالعه در بستر سد، محدوده آبرفت و سنگ بستر و همچنین لایه بندی، روش خوشه بندی گوستافسون کسل نتیجه بهتری را نشان داده است. با به کار بردن نتایج میانگین فازی در آغاز الگوریتم گوستافسون-کسل، گام های تکرار کاهش و سرعت همگرایی افزایش می یابد (محمدی و روحانی، ۱۳۹۶). توکلی و همکاران (۱۳۹۸) در مقاله ای تحت عنوان "کروناویروس جدید ۲۰۱۹: بیماری عفونی نوظهور در قرن ۲۱" به بررسی پرداخته اند. در این مطالعه، در جستجوی اولیه، تعداد ۱۴۱۶ مقاله استخراج شد که پس از حذف موارد تکراری و ارزیابی عنوان و چکیده، ۵۳ مقاله برگزیده شد. پس از بررسی متن کامل مقالات در نهایت تعداد ۳۷ مقاله شرایط لازم برای شرکت در این مطالعه را دارا بودند. بر اساس گزارش سازمان بهداشت جهانی، دوره نهفتگی کرونا ویروس جدید بین ۲-۱۰روز می باشد. به طور کلی نرخ کشندگی این ویروس ۴,۳ درصد بوده و نتایج این مطالعه نشان می دهند که میزان مرگ و میر این ویروس در سالمندان و افراد مبتلا به بیماری های زمینه ای در مقایسه با افراد سالم به میزان قابل ملاحظه ای بالاتر می باشد. گروه های پرخطر برای این بیماری به ترتیب شامل بیماران قلبی- عروقی، دیابتی، مبتلایان به بیماری های تنفسی مزمن و فشار خون بالا می باشند. نرخ مرگ و میر در افراد سالم کمتر از یک درصد برآورد شده است (توکلی و همکاران، ۱۳۹۸). جعفری و همکاران (۱۳۹۹) در مقاله ای با عنوان "تحلیل موضوعی مطالعات کووید-۱۹ در پنج قاره بزرگ" به مطالعه پرداختند. هدف این مطالعه آشکارسازی موضوعات پژوهشی کووید-۱۹ در پنج قاره بر اساس آثار نمایه شده در وب آو ساینس است. نتایج این مطالعه نشان داد که پژوهشگران آسیایی برمباحث اپیدمیولوژیک، پژوهشگران اروپایی بر مباحث بیولوژیکی و پژوهشگران آمریکایی بر مباحث اپیدمیولوژیک و ژنتیک متمرکز هستند (جعفری و همکاران، ۱۳۹۹). علی پور و همکاران (۱۳۹۸) در مقاله ای تحت عنوان "مروری بر اثرات عصبی و شناختی کووید-۱۹" به بررسی پرداختند. در این بررسی گزارش ها نشان از حملات ویروسی به دستگاه اعصاب مرکزی و ایجاد التهاب ویروسی مغز داشته همچنین بر نقش واسطه ای سیستم ایمنی بدن در مقابله با عفونت تاکید گردیده است. فقدان حس بویایی و چشایی در افراد مبتلا و ارتباط آن با سیستم عصبی از دیگر نشانه های مهم این ویروس بوده و از اولین علایم حمله به دستگاه عصبی محسوب می گردد. در این بررسی بیماری حاد مغزی – عروقی و روند شکل گیری آن در اثر هیپوکسی و دیگر عوارض ناشی از آلودگی نیز تشریح گردیده است (علی پور و همکاران، ۱۳۹۸). فرنوش و همکاران (۱۳۹۹) در مقاله ای با عنوان "شناخت کروناویروس نوین-۲۰۱۹ و کووید-۱۹ بر اساس شواهد موجود- مطالعه مروری" به مطالعه پرداختند. در این مطالعه مروری، بر اساس شواهد منتشر شده تا اول مارس ۲۰۲۰، ویژگی های اپیدمی و اتیولوژیک کرونا ویروس نوین-۲۰۱۹، ویژگی های اساسی بیولوژیکی آن، از جمله گیرنده ها و مسیر انتقال آن، تشریح رویکردهای پیشگیری از بیماری و درمان کووید- ۲۳ ارائه شده است (فرنوش و همکاران، ۱۳۹۹). جوادی و همکاران (۱۳۹۷) در مقاله ای تحت عنوان "تحلیل چند پارامتره

1 COVID-19

2 Coronavirus disease 2019

3 nCoV acute respiratory disease-2019

آلودگی آبخوان قزوین بر مبنای نقشه کاربری اراضی و با استفاده از تکنیک خوشه بندی کا از میانگین" به بررسی پرداختند. در این تحقیق با به کارگیری تکنیک خوشه بندی و با توجه به کاربری اراضی وضع موجود، نقشه پهنه بندی کیفی بر اساس ترکیب چند پارامتر استخراج شده است. نتایج نشان می دهد با توجه به استاندارد جهانی آب شرب و در نظر گرفتن سه پارامتر منتخب، مناسب ترین خوشه (C1) با مساحت ۲۲ درصد در نواحی شمالی آبخوان بوده و به سمت نواحی مرکزی، که تمرکز فعالیت های کشاورزی و صنایع است، آلودگی در آبخوان (خوشه های C4 و C5) با مساحت ۳۵ درصد به عنوان نامناسب ترین منطقه شناخته شده اند (جوادی و همکاران، ۱۳۹۷). کریمی علویجه و همکاران (۱۳۹۵) در مقاله ای تحت عنوان "روش فرابتکاری در یکپارچه سازی مدل بخش بندی بازار مشتریان تلفن همراه تهران با استفاده از شبکه های خودسازمانده و روش میانگین کا" به مطالعه پرداختند. در این مطالعه، با هدف دستیابی به درکی صحیح از بازار هدف، رفتار مصرف و سبک زندگی خرده بازارهای تحقیق، پس از نقد روش های سنتی و معرفی تکنیک های بخش بندی مبتنی بر شبکه های عصبی، با بهره مندی از روش دلفی فازی، متغیرهای بخش بندی بازار هدف انتخاب شده است و از طریق شبکه های خودسازمانده کوهنن، تعداد خوشه های مناسبی از جامعهٔ آماری به دست آمد و در نهایت با استفاده از تکنیک میانگین کا و تکنیک تجمعی، به تدقیق خوشه بندی ها و بخش بندی بازار پرداخته شده است. پس از جمع آوری داده ها از طریق پرسشنامه و تجزیه وتحلیل آن ها، یافته ها نشان داد بازار تلفن همراه تهران در پنج خوشه دسته بندی می شود که هر یک می تواند قابلیت پیاده سازی استراتژی های بازاریابی مجزا، با در نظر گرفتن مزیت های رقابتی شرکت های حوزهٔ ICT، برای حداکثرسازی تقاضا و حاشیهٔ سود، داشته باشد (کریمی علویجه و همکاران، ۱۳۹۵). روش های خوشه بندی در برخی از تحقیقات توسعه یافته است یا با روش های دیگر جهت بهبود نتایج ادغام شده است. در این تحقیق روش خوشه بندی کا از میانگین توسه یافته و شاخص ارزیابی دان بهبود بخشیده شده و شاخص اعتبار جدیدی مبتنی بر ضریب نیمرخ برای روش خوشه بندی کا از میانگین معرفی شده است (Manochandar et al., 2020). در این مطالعه یک روش طبقه بندی درخت تصمیم چند متغیره خطی دودویی مبتنی بر الگوریتم کا از میانگین معرفی شده است. در این روش جدید، از حذف موارد اقلیت در موارد نامتعادل خودداری به عمل آمده است؛ علاوه بر این مدل پیشنهادی جدید مبتنی بر هایپرپلن[1] بوده و به روند طبقه بندی سرعت می بخشد (Wang et al., 2020). در تحقیقی دیگر با استفاده از روش خود سازمان ده کوهونن جهت بهبود شبکه های عصبی چندلایه بهره گرفته شده است. در این تحقیق علی رغم کاهش ابعاد مسئله به افزایش آن پرداخته شده است تا شبکه های عصبی پیکربندی شبکه های جدید را تشکیل دهند تا اطلاعات بیشتری داشته باشند. برای تقویت اطلاعات، از روش خود سازمان ده کوهونن استفاده شده است، که می تواند ورودی های بیشتر را ارائه داده و به همان اندازه وزن مشابه تولید کند (Kamimura, 2019). پژوهش حاضر بر آن است که داده های مرتبط با کرونا ویروس جدید- ۲۰۱۹ را که از سایت بهداشت جهانی به آدرس – covid19.who.int– در تاریخ ۷ ژوئن اخذ نموده و به منظور خوشه بندی، از روش های خود سازمان‌ده از جمله الگوریتم های کوهونن و کا از میانگین و خوشه بندی دومرحله ای بهره گیرد و در نهایت با انجام آزمون، کارایی این سه تکنیک تعیین شده است.

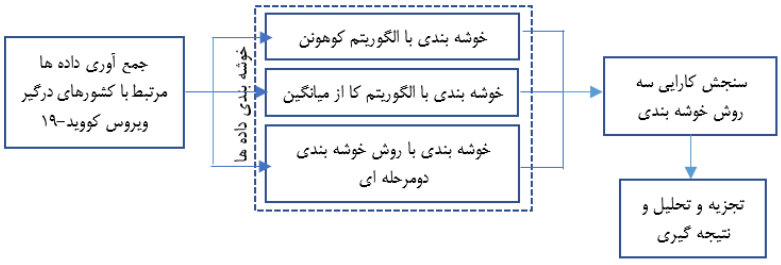
۲-اهداف پژوهش

هدف اصلی: خوشه بندی داده های کرونا ویروس جدید- ۲۰۱۹ می باشد که اهداف فرعی به شرح ذیل می باشند:

۱) خوشه بندی با الگوریتم کوهونن[2]

۲) خوشه بندی با الگوریتم کا از میانگین[3]

۳) خوشه بندی با روش خوشه بندی دومرحله ای[4]

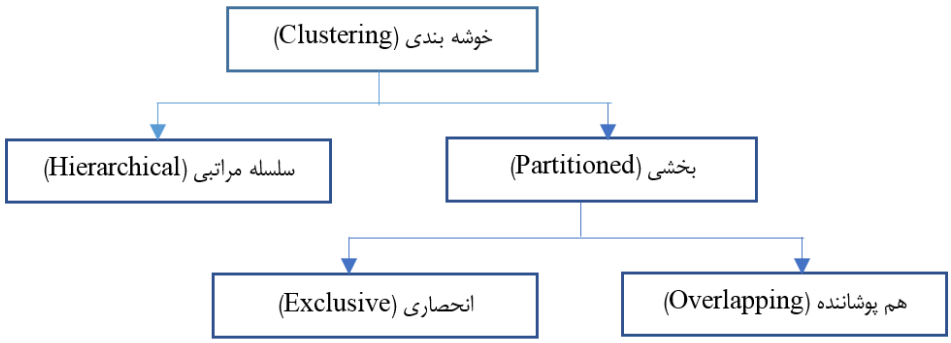


شکل ۱- مدل مفهومی پژوهش

1 hyperplane
2 Kohonen
3 K-means
4 Two-step

۳- روش تحقیق

امروزه، میزان داده های تولید شده و ذخیره شده به دلیل پیشرفت شگفت انگیز ابزارهای تولید، اندازه گیری و جمع آوری اطلاعات به طور چشمگیری افزایش یافته است. به علت وجود ظرفیت شناسایی دانش از داده های اندازه گیری شده، رشته جداگانه علمی ایجاد شد که به علم «کشف دانش از بانک داده ها۱» شناخته شده است. داده کاوی در واقع کاربرد یک الگوریتم خاص برای استخراج الگوها از داده هاست که امکان استخراج دانش جدید را در یک فرآیند علمی یا یک سیستم فراهم می سازد. انواع مختلفی از روش های داده کاوی ارائه و توسعه شده اند که پرکاربردترین آن ها خوشه بندی می باشد. در تحقیقات مهندسی و مدیریت، تحلیل خوشه ها ابزار بسیار کارآمدی به شمار می رود. «تحلیل خوشه‌بندی۲» بطور خلاصه خوشه‌بندی، 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 در عوض میزان شباهت بین خوشه‌ها کمترین مقدار است. در چنین حالتی هدف از خوشه‌بندی، نسبت دادن برچسب‌هایی۳ به اشیاء است که نشان دهنده عضویت هر شیء به خوشه است. به این ترتیب تفاوت اصلی که بین تحلیل خوشه‌بندی و طبقه‌بندی وجود دارد، نداشتن برچسب‌های اولیه برای مشاهدات است. در نتیجه براساس ویژگی‌های مشترک و روش‌های اندازه‌گیری فاصله یا شباهت بین اشیاء، باید برچسب‌هایی بطور خودکار نسبت داده شوند. در حالی که در طبقه‌بندی برچسب‌های اولیه موجود است و باید با استفاده از الگوی‌های پیش‌بینی قادر به برچسب گذاری برای مشاهدات جدید باشیم. با توجه به روش‌های مختلف اندازه‌گیری شباهت یا الگوریتم‌های تشکیل خوشه، ممکن است نتایج خوشه‌بندی برای مجموعه داده ثابت متفاوت باشند. شایان ذکر است که تکنیک‌های خوشه‌بندی در علوم مختلف مانند، گیاه‌شناسی، هوش مصنوعی، تشخیص الگوهای مالی و ... کاربرد دارند. اگر چه بیشتر الگوریتم‌ها یا روش‌های خوشه‌بندی مبنای یکسانی دارند ولی تفاوت‌هایی در شیوه اندازه‌گیری شباهت یا فاصله و همچنین انتخاب برچسب برای اشیاء هر خوشه در این روش‌ها وجود دارد. به طور کلی روش های خوشه بندی به دو دسته روش های خوشه بندی سلسله مراتبی و خوشه بندی بخشی تقسیم می گردد. خوشه بندی بخشی نیز شامل دو گروه روش های انحصاری و هم پوشاننده است (شکل ۲). در خوشه بندی انحصاری، یک جزء منحصراً عضو یک مجموعه است. تکنیک هـای کا میـانگین و شبکهٔ عصبی خودسازمان (SOM۴) نمونه ای از ایـن نـوع روش هـا هسـتند؛ امـا روش هـای هم پوشاننده، بر اساس تئوری مجموعه های فازی بنا شده اند که در آن ها هر جزء مـی توانـد عضـو یک یا چند مجموعه باشد (جوادی و همکاران، ۱۳۹۷ و کریمی علویجه و همکاران، ۱۳۹۵).



شکل ۲- طبقه بندی از انواع خوشه بندی (کریمی علویجه و همکاران، ۱۳۹۵)

۳-۱- الگوریتم کوهونن

الگوریتم کوهونن توسط تیوو کوهونن۵، استاد دانشگاه هلسینکی فنلاند و نویسنده کتاب Self-organizing maps برای اولین بار ارائه شد. فرض این الگوریتم، یک ساختار توپولوژیکی در بین واحدهای خوشه مشابه آنچه در مغز انسان می باشد که در سایر روش ها دیده نمی شود. شبکه خود سازمانده کوهونن، یک شبکه عصبی بدون سرپرست می باشد که هدفش کاهش بعد و خوشه بندی است.کوهونن به علت استفاده از تابع همسایگی به منظور حفظ ویژگی های مکانی فضای ورودی با سایر شبکه های

1 (KDD) Knowledge Discovery in Database
2 Cluster Analysis
3 Subject
4 Self-Organization Map (SOM)
5 *Teuvo Kohonen*

عصبی متفاوت است. هر شبکه کوهونن از تعدادی نود (گره) تشکیل شده است. هر نود دارای برداری از وزن ها می باشد. ابعاد این بردار با ابعاد فضای ورودی برابر است. پس از آموزش شبکه هر ناحیه از نودهای شبکه به الگوهای خاصی از داده های ورودی واکنش نشان می دهند. روش آموزش شبکه یادگیری رقابتی است. وقتی نمونه آموزشی جدید به شبکه اعمال می شود فاصله اقلیدسی آن از بردار وزن تمام نودهای شبکه حساب می شود. در خوشه بندی کوهونن، گام هایی به قرار زیر طی می شود :

مرحله صفر : مقداردهی اولیه به وزن ها، پارامترهای مربوط به ساختار همسایگی را تعیین نمایید، پارامترهای نرخ یادگیری را تعیین نمایید.

مرحله ۱ : تا زمانی که شرایط توقف برقرار نیست، مراحل ۲ تا ۸ را انجام دهید.

مرحله ۲ : برای هر بردار ورودی X، مراحل ۳ تا ۵ را انجام دهید.

مرحله ۳ : برایی هر j، مقدار زیر را محاسبه نمایید :

$$D(j) = \sum_i (w_{ij} - x_i)^2 \qquad (1)$$

مرحله ۴ : اندیس j را طوری بدست آورید که D(j) کمینه گردد.

مرحله ۵ : برای تمام واحدهای j که در همسایگی مشخصی از j قرار دارند و برای تمام iها قرار داده شود :

$$w_{ij}(new) = w_{ij}(old) + \alpha\left[x_i - w_{ij}(old)\right] \qquad (2)$$

مرحله ۶ : نرخ یادگیری را به روز نمایید.

نرخ یادگیری : تابعی از زمان (یا دورهای آموزش) است که با گذشت زمان به آرامی کاهش می یابد و کاهش این پارامتر به صورت خطی، برای کاربردهای ملی مناسب تر است.

مرحله ۷ : شعاع همسایگی ساختار توپولوژیک را در زمان های مشخص کاهش دهید.

شعاع همسایگی ساختار توپولوژیک : شعاع ناحیه اطراف یک خوشه نیز با پیشروی فرآیند خوشه بندی کاهش می یابد.

مرحله ۸ : شرایط توقف را بررسی نمایید (سلطانی و همکاران، ۱۳۸۹ و Ebrahim ghadiri & Mazlumi, 2019).

۳-۲- الگوریتم کا از میانگین

روش کا میـانگین، کاربردی ترین روش خوشه بندی داده هاست. این روش اولین بار توسط (مک کویین، ۱۹۶۷) ارائه شد. تعداد خوشه ها در این روش ثابت و از پیش تعیین شده است. این روش برای خوشه بندی داده هایی طراحی شد که به صورت عددی

(کمی) باشند و خوشه دارای مرکزی به نام «میانگین» باشد. در این روش ابتدا اشیاء به صورت تصادفی به K خوشه تقسیم می شوند. در گام بعد، فاصله هر یک از اشیاء از مرکز خوشه خود محاسبه می شود. در صورتی که فاصله شیء مورد نظر از میانگین خوشه خود زیاد و به خوشه دیگری نزدیک تر باشد، این شی به خوشه ای که نزدیک تر است اختصاص می یابد. این کار آنقدر تکرار می شود تا تابع خطا حداقل شود و یا اعضای خوشه ها تغییر نیابند. تابع خطا (EF) مجموع فواصل هر شیء از مرکز خوشه خودش تعریف می شود که طبق فرمول ذیل محاسبه می گردد :

$$EF = \sum_{i=1}^{k} \sum_{X \in C_i} d(X, \mu(C_i)) \qquad (3)$$

که در آن، μ نشاندهنده مرکز (میانگین) خوشه، و $d(X, \mu(C_i))$ فاصله هر شیء از مرکز خود است. به دلیل این که در این نوع خوشه بندی تابع خطایی داریم که باید آن را حداقل نماییم، می توان به این مسائل به دید مسائل بهینه سازی نگریست. در این نوع خوشه بندی تابع هدفی وجود دارد که تابع خطاست و به دنبال حداقل سازی آن هستیم و محدودیت هایی چون : الف) تعداد خوشه ها از پیش تعیین شده استو نمی توان آن ها را کم یا زیاد کرد؛ و ب) تعداد اعضای هر یک از خوشه ها نمی تواند صفر باشد، وجود دارد.

در خوشه بندی کا میـانگین، گام هایی به قرار زیر طی می شود :

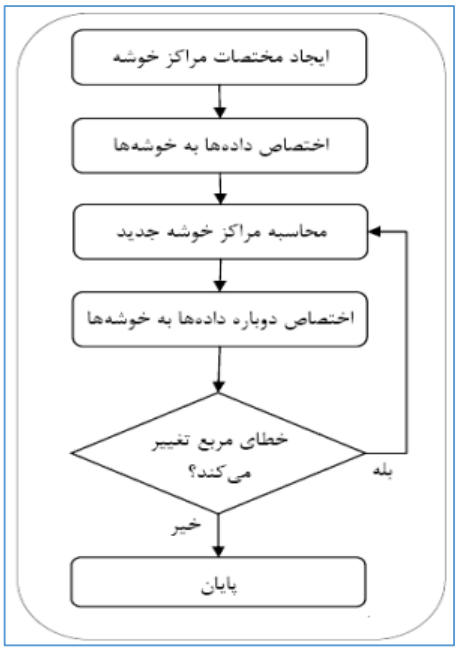
گام آغازین : تفکیک داده های اولیه به K خوشه به صورت دلخواه،

گام تکراری : الف) محاسبه فاصله هر شئ از مرکز خود، ب) محاسبه تابع خطا،

گام بهبود : جابجایی عضوی که بیشترین فاصله را با مرکز خوشه خودش دارد، به خوشه ای که کمترین فاصله را با آن دارد.

دستور توقف : تغییر نیافتن اعضای خوشه ها یا کاهش نیافتن مقدار تابع خطا یا عدم تغییر در خوشه ها (مؤمنی، ۱۳۹۸ و Kantardzic, 2011).

در نهایت، متدولوژی روش خوشه بندی مطابق الگوریتم ذیل می باشد (شکل ۳) :



شکل ۳- مراحل انجام الگوریتم خوشه بندی کا از میانگین (جوادی و همکاران، ۱۳۹۷)

۳-۳- خوشه‌بندی دومرحله‌ای

جهت تحلیل و گروه بندی مجموعه داده های بزرگ از الگوریتم های خوشه بندی استفاده می شود. این نوع الگوریتم ها با استفاده از روش خوشه بندی سلسله مراتبی، مشاهدات را در خوشه ها گروه بندی می کنند. در مقایسه با دیگر روش های خوشه بندی، الگوریتم خوشه بندی دومرحله ای دو مزیت اصلی دارد : اول اینکه امکان خوشه بندی هم معیارهای پیوسته و هم معیارهای گسسته را فراهم می کند. همچنین تعداد خوشه های بهینه توسط خود الگوریتم تعیین می شود، در نتیجه احتمال خطا تا حد امکان به صفر می رسد. یکی از الگوریتم های خوشه بندی، الگوریتم خوشه بندی دومرحله ای است. این روش، ابزاری اکتشافی است که برای آشکار نمودن خوشه های ذاتی و طبیعی موجود در مجموعه داده که به طور معمول دیده نمی شوند، طراحی شده است. وجه تمایز الگوریتم موجود در این روش با فنون سنتی خوشه بندی عبارت است از :

قابلیت خوشه بندی بر اساس متغیرهای گسسته (رسته ای) و پیوسته

انتخاب خودکار تعداد خوشه ها

قابلیت تحلیل کارآمد فایل داده های بسیار بزرگ

الگوریتم تحلیل خوشه‌ای دومرحله‌ای بدین صورت است که با فرض مستقل بودن متغیرها در مدل خوشه بندی، اندازه درستنمایی[1] را به عنوان فاصله موجودیت ها درنظر می گیرد. هم چنین، فرض بر این است که هر متغیر پیوسته دارای توزیع نرمال و هر متغیر گسسته دارای توزیع نرمال چندمتغیری است. البته آزمون های تجربی نشان می دهد که الگوریتم دومرحله ای نسبت به تغییر فرض های استقلال و نرمال بودن، به اندازه کافی نیرومند[2] است. الگوریتم خوشه بندی دومرحله ای به صورت زیر خلاصه می شود:

مرحله اول : در این مرحله که مرحله پیش- خوشه[3] می باشد، بر اساس رویکرد خوشه بندی سلسله مراتبی تجمعی[4]، هر رکورد به عنوان یک خوشه درنظر گرفته می شود. تمامی رکوردها یک به یک بررسی شده و بر اساس معیار درنظرگرفته شده برای فاصله موجودیت ها، در مورد ادغام آن با خوشه های قبل یا آغاز یک خوشه جدید تصمیم گیری می شود. بنابراین در این مرحله به زیرخوشه هایی دست می یابیم.

مرحله دوم : در این مرحله زیرخوشه های مرحله اول به عنوان ورودی ها درنظر گرفته شده و بر اساس خوشه بندی سلسله مراتبی تجمعی به تعداد خوشه های موردنظر تبدیل می شوند. در صورت عدم تعیین تعداد خوشه ها، این الگوریتم به صورت

1 Likelihood Measure
2 Robust
3 Pre-cluster
4 Agglomerative hierarchical clustering method

خودکار و با معیارهایی هم چون BIC[1] و AIC[2] زیرخوشه ها را به تعداد بهینه‌ای از خوشه ها تبدیل می‌کند (زعفریان و همکاران، ۱۳۹۷ و Hassannezhad & Clarkson, 2017).

نتایج و بحث

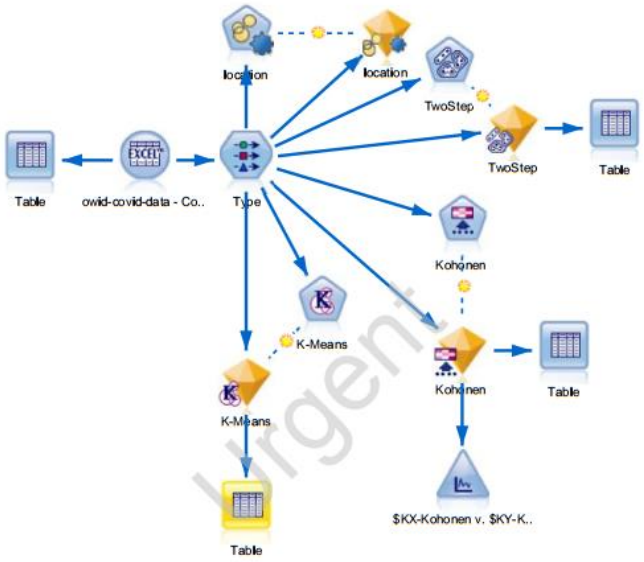
پژوهش حاضر داده های مرتبط با کرونا ویروس جدید- ۲۰۱۹ را که از سایت بهداشت جهانی در تاریخ ۷ ژوئن اخذ نموده و به منظور خوشه بندی، از روش های خود سازمان‌ده از جمله الگوریتم های کوهونن و کا از میانگین و خوشه بندی دومرحله ای بهره گرفته است که هدف تمامی تکنیک ها این است که کدام رکورد در کدام خوشه جای می گیرد و در نهایت با انجام آزمون، کارایی این سه تکنیک تعیین شده است. داده ها در قالب فایل اِکسل از سایت بهداشت جهانی در بازه ماه اول (January) سال ۲۰۲۰ الی ماه ششم (June) سال ۲۰۲۰ در تاریخ های متفاوت در طول این بازه برای ۸۰ کشور (۷۹ کشور و رکورد آخر مربوط به آمار کل کشورهای جهان است) جمع آوری شده است که شامل ۲۴ ستون (متغیر) به شرح جدول ۱ می باشد:

جدول ۱- نام ستون های داده های اخذ شده و توضیحات مربوط به آن ها

ردیف	نام ستون	توضیحات
۱	نام قاره	نام ۵ قاره از جمله آسیا، اروپا، آفریقا، استرالیا، آمریکا
۲	نام کشور	نام ۸۰ کشور از ۵ قاره در تاریخ های متفاوت جهت جمع آوری آمار
۳	تاریخ	تاریخ های متفاوت جهت جمع آوری آمار بیماری کووید-۱۹ به تفکیک هر کشور
۴	نرخ کل مبتلایان	نرخ کل افراد مبتلا به بیماری کووید-۱۹
۵	نرخ مبتلایان جدید	نرخ افراد جدید مبتلا شده به بیماری کووید-۱۹
۶	نرخ کل مرگ و میر	نرخ کل مرگ و میر بر اثر بیماری کووید-۱۹
۷	نرخ مرگ و میر جدید	نرخ مرگ و میر جدید بر اثر بیماری کووید-۱۹
۸	نرخ کل مبتلایان در هر میلیون	نرخ کل افراد مبتلا به بیماری کووید-۱۹ در هر میلیون به تفکیک هر کشور
۹	نرخ مبتلایان جدید در هر میلیون	نرخ افراد جدید مبتلا شده به بیماری کووید-۱۹ در هر میلیون به تفکیک هر کشور
۱۰	نرخ کل مرگ و میر در هر میلیون	نرخ کل مرگ و میر بر اثر بیماری کووید-۱۹ در هر میلیون به تفکیک هر کشور
۱۱	نرخ مرگ و میر جدید در هر میلیون	نرخ مرگ و میر جدید بر اثر بیماری کووید-۱۹ در هر میلیون به تفکیک هر کشور
۱۲	شاخص شدت	عددی از ۰ تا ۱۰۰ است که این شاخص ها را منعکس می کند. نمره شاخص بالاتر نشان دهنده سطح شدت بالاتر است
۱۳	جمعیت	میزان جمعیت به تفکیک هر کشور
۱۴	تراکم جمعیت	میزان تراکم جمعیت به تفکیک هر کشور با واحد نفر بر کیلومتر مربع
۱۵	متوسط سن افراد مبتلا	متوسط سن افراد مبتلا به بیماری کووید-۱۹
۱۶	افراد مبتلا بالای ۶۵ سال	نرخ افراد مبتلا به بیماری کووید-۱۹ که بالای ۶۵ سال سن دارند
۱۷	افراد مبتلا بالای ۷۰ سال	نرخ افراد مبتلا به بیماری کووید-۱۹ که بالای ۷۰ سال سن دارند
۱۸	سرانه تولید ناخالص داخلی	کل ارزش پولی محصولات نهایی تولیدشده توسط واحدهای اقتصادی مقیم کشور در دوره زمانی معین
۱۹	شدت فقر	این شاخص شدت فقر را با در نظر گرفتن تعداد فقرا، عمق فقر و نابرابری بین فقرا اندازه گیری می کند
۲۰	نرخ مرگ و میر بیماران قلبی و عروقی مبتلا	نرخ مرگ و میر بر اثر بیماری کووید-۱۹ که دارای بیماری زمینه ای قلبی و عروقی می باشند
۲۱	نرخ مرگ و میر بیماران دیابتی مبتلا	نرخ مرگ و میر بر اثر بیماری کووید-۱۹ که دارای بیماری زمینه ای دیابت می باشند
۲۲	نرخ زن های سیگاری	نرخ زن های مبتلا به بیماری کووید-۱۹ که سیگار می کشند
۲۳	نرخ مردهای سیگاری	نرخ مردهای مبتلا به بیماری کووید-۱۹ که سیگار می کشند
۲۴	نرخ تخت بیمارستان در هر هزار	نرخ تخت بیمارستان در هر هزار به تفکیک هر کشور

خوشه بندی، با روش های خود سازمان‌ده از جمله الگوریتم های کوهونن و کا از میانگین و خوشه بندی دومرحله ای صورت گرفته است که مدل کلی آن مطابق شکل ۴ است.

1 Schwarz's Bayesian Criterion
2 Akaike Information Criterion



شکل ٤- شمای کلی از مدل با نرم افزار IBM Modeler 18.0

در ابتدا بر روی داده های اخذ شده، الگوریتم کوهونن صورت گرفته است، بر اساس خروجی نرم افزار IBM Modeler 18.0 فضای ورودی به ۱۲ خوشه تقسیم بندی شده است که با توجه به مختصات هر رکورد می توانیم تصمیم بگیریم کدام رکورد در کدام خوشه قرار دارد که اطلاعات آن به شرح جدول ۲ می باشد. در مرحله بعد با استفاده از الگوریتم کا از میانگین، کلاستربندی روی داده ها صورت گرفته است که فضای ورودی مطابق انتخاب نویسنده به ۵ خوشه تقسیم بندی شده است که اطلاعات آن به شرح جدول ۳ می باشد. در نهایت، با استفاده از روش خوشه‌بندی دومرحله‌ای، کلاستربندی روی داده ها صورت گرفته است که اطلاعات آن به شرح جدول ۴ می باشد.

با ملاحظه جدول ۲، در الگوریتم کوهونن، با در نظر گرفتن معادلات شماره ۱ و ۲، رکوردها به ۱۲ خوشه تقسیم بندی شده اند که با توجه به ماهیت داده ها برخی از رکوردها در خوشه های دیگر تکرار شده اند. تعداد رکوردها در دو خوشه ی ۱ و ۳ با یکدیگر مشابه بوده و برابر با ۱۹ است، تعداد رکوردها در دو خوشه ی ۵ و ۶ با یکدیگر مشابه بوده و برابر با ۲ است و تعداد رکوردها در دو خوشه ی ۴ و ۸ نیز با یکدیگر مشابه بوده و برابر با ۱ است. در خوشه شماره ۸، رکورد مربوط به کل کشورهای جهان به تنهایی قرار گرفته است. بیشترین تعداد رکورد در خوشه آخر یعنی ۱۲ قرار دارد. با ملاحظه جدول ۳، در الگوریتم کا از میانگین، با در نظر گرفتن معادله شماره ۳، رکوردها به ۵ خوشه تقسیم بندی شده اند. در خوشه شماره ۲، رکورد مربوط به کل کشورهای جهان به تنهایی قرار گرفته است. بیشترین تعداد رکورد در خوشه ۳ قرار دارد. دو الگوریتم کوهونن و کا از میانگین، حداکثر تعداد رکورد در خوشه یکسان دارند که برابرست با ۲۳ رکورد. با ملاحظه جدول ۴، در روش خوشه‌بندی دومرحله‌ای، رکوردها به ۳ خوشه تقسیم بندی شده اند. بیشترین تعداد رکورد در خوشه آخر یعنی ۳ قرار دارد.

نمودار دایره ای به ترتیب مرتبط با الگوریتم کوهونن، الگوریتم کا از میانگین و روش خوشه‌بندی دومرحله‌ای در شکل های ۵، ۶ و ۷ نشان داده شده اند که درصد هر خوشه را به تفکیک هر روش در نمودار دایره ای نمایش می دهد و در جدول پایین هر نمودار اندازه کوچکترین خوشه، بزرگترین خوشه و نسبت بزرگترین خوشه به کوچکترین خوشه را به تفکیک هر روش نمایش می دهد.

جدول ۲- مشخصات خوشه های به دست آمده از روش کوهونن

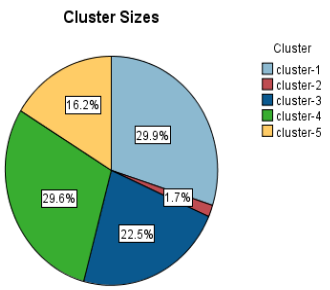
نام خوشه	C**1	C**2	C**3	C**4	C**5	C**6	C**7	C**8	C**9	C**10	C**11	C**12
نرخ کل مبتلایان	۱۹ رکورد از هر کدام از ستون ها از کشورهای اتریش، بوسنی و هرزگوین، یونان، ایرلند، ایتالیا، لتونی، لیتوانی، لوکزامبورگ، مالت، مونته نگرو، نروژ، پرتقال، رومانی، روسیه، اسلواکی، اسپانیا، سوئد، اوکراین، انگلیس	۱۷ رکورد از هر کدام از ستون ها از کشورهای آلبانی، اتریش، بوسنی و هرزگوین، یونان، ایرلند، ایتالیا، لیتوانی، لوکزامبورگ، نروژ، پرتقال، رومانی، روسیه، اسلواکی، اسپانیا، سوئد، اوکراین، انگلیس	۱۹ رکورد از هر کدام از ستون ها از کشورهای آلبانی، الجزایر، اتریش، بوسنی و هرزگوین، یونان، ایرلند، ایتالیا، لیتوانی، لوکزامبورگ، مولداوی، نروژ، پرتقال، رومانی، روسیه، اسلواکی، اسپانیا، سوئد، اوکراین، انگلیس	۱ رکورد از هر کدام از ستون ها از کشور مولداوی	۲ رکورد از هر کدام از ستون ها از کشورهای استرالیا، مولداوی	۲ رکورد از هر کدام از ستون ها از کشورهای استرالیا، آمریکا	۱۲ رکورد از هر کدام از ستون ها از کشورهای ارمنی، هند، اندونزی، ایران، اسرائیل، مالزی، نپال، کره جنوبی، سری لانکا، تایلند، ترکیه، ویتنام	۱ رکورد از هر کدام از ستون ها از رکورد مربوط به کل کشورهای جهان	۱۳ رکورد از هر کدام از ستون ها از کشورهای آرژانتین، کلمبیا، کاستاریکا، دومینیکا، اکوادور، السالوادور، هاییتی، مکزیک، پاناما، پاراگوئه، آمریکا، اروگوئه، رکورد مربوط به کل کشورهای جهان	۱۶ رکورد از هر کدام از ستون ها از کشورهای آذربایجان، بنگلادش، هند، اندونزی، ایران، اسرائیل، قزاقستان، مالزی، میانمار، نپال، پاکستان، کره جنوبی، سری لانکا، تایلند، ترکیه، ویتنام	۲۰ رکورد از هر کدام از ستون ها از کشورهای آذربایجان، بنگلادش، هند، اندونزی، ایران، اسرائیل، قزاقستان، قرقیزستان، لائوس، مالزی، مغولستان، میانمار، نپال، پاکستان، سری لانکا، تایلند، تیمور، ترکیه، ویتنام، یمن	۲۳ رکورد از هر کدام از ستون ها از کشورهای الجزایر، بنین، بورکینافاسو، الجزایر قمر، مصر، اتیوپی، گامبیا، غنا، کنیا، لیبریا، مالاوی، موریس، مراکش، موزامبیک، نیجر، سیشل، آفریقای جنوبی، تانزانیا، توگو، تونس، اوگاندا، زامبیا، زیمباوه
نرخ مبتلایان جدید												
نرخ کل مرگ و میر												
نرخ مرگ و میر جدید												
نرخ کل مبتلایان در هر میلیون												
نرخ مبتلایان جدید در هر میلیون												
نرخ کل مرگ و میر در هر میلیون												
نرخ مرگ و میر جدید در هر میلیون												
شاخص شدت												
جمعیت												
تراکم جمعیت												
متوسط سن افراد مبتلا												
افراد مبتلا بالای ۶۵ سال												
افراد مبتلا بالای ۷۰ سال												
سرانه تولید ناخالص داخلی												
شدت فقر												
نرخ مرگ و میر بیماران قلبی و عروقی مبتلا												
نرخ مرگ و میر بیماران دیابتی مبتلا												
نرخ زن های سیگاری												
نرخ مردهای سیگاری												
نرخ تخت بیمارستان در هر هزار												

جدول ۳- مشخصات خوشه های به دست آمده از روش کا از میانگین

نام خوشه	C*1	C*2	C*3	C*4	C*5
نرخ کل مبتلایان	۳۲ رکورد از هر کدام از ستون ها از کشورهای ارمنی، آذربایجان، بنگلادش، هند، اندونزی، ایران، اسرائیل، قزاقستان، قرقیزستان، لائوس، مالزی، منغولستان، میانمار، نپال، پاکستان، کره جنوبی، سری لانکا، تایلند، تیمور، ترکیه، ویتنام، یمن	۱ رکورد از هر کدام از ستون ها از رکورد مربوط به کل کشورهای جهان	۳۳ رکورد از هر کدام از ستون ها از کشورهای الجزایر، بنین، بورکینافاسو، الجزایر قمر، مصر، اتیوپی، گامبیا، غنا، کنیا، لیبریا، مالاوی، موریس، مراکش، موزامبیک، نیجر، سیشل، آفریقای جنوبی، تانزانیا، توگو، تونس، اوگاندا، زامبیا، زیمباوه	۲۱ رکورد از هر کدام از ستون ها از کشورهای آلبانی، اتریش، بوسنی و هرزگوین، یونان، ایرلند، ایتالیا، لتونی، لیتوانی، لوکزامبورگ، مالت، مولداوی، مونته نگرو، نروژ، پرتقال، رومانی، روسیه، اسلواکی، اسپانیا، سوئد، اوکراین، انگلیس	۱۳ رکورد از هر کدام از ستون ها از کشورهای آرژانتین، استرالیا، کلمبیا، کاستاریکا، دومینیکا، اکوادور، السالوادور، هاییتی، مکزیک، پانامای پاراگوئه، آمریکا، اروگوئه
نرخ مبتلایان جدید					
نرخ کل مرگ و میر					
نرخ مرگ و میر جدید					
نرخ کل مبتلایان در هر میلیون					
نرخ مبتلایان جدید در هر میلیون					
نرخ کل مرگ و میر در هر میلیون					
نرخ مرگ و میر جدید در هر میلیون					
شاخص شدت					
جمعیت					
تراکم جمعیت					
متوسط سن افراد مبتلا					
افراد مبتلا بالای ۶۵ سال					
افراد مبتلا بالای ۷۰ سال					
سرانه تولید ناخالص داخلی					
شدت فقر					
نرخ مرگ و میر بیماران قلبی و عروقی مبتلا					
نرخ مرگ و میر بیماران دیابتی مبتلا					
نرخ زن های سیگاری					
نرخ مردهای سیگاری					
نرخ تخت بیمارستان در هر هزار					

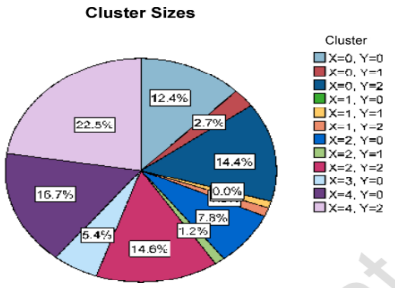
جدول ٤- مشخصات خوشه های به دست آمده از روش خوشه‌بندی دومرحله‌ای

نام خوشه	C1	C2	C3
نرخ کل مبتلایان	۲۴ رکورد از هر کدام از ستون ها از کشورهای آلبانی، استرالیا، اتریش، بوسنی و هرزگوین، یونان، ایرلند، ایتالیا، لتونی، لیتوانی، لوکزامبورگ، مالت، مولداوی، مونته نگرو، نروژ، پرتقال، رومانی، روسیه، اسلواکی، کره جنوبی، اسپانیا، سوئد، اوکراین، انگلیس، آمریکا	۱۴ رکورد از هر کدام از ستون ها از کشورهای ارمنی، اکوادور، ایرلند، اسرائیل، ایتالیا، لوکزامبورگ، پانامای پاراگوئه، پرتقال، اسپانیا، سوئد، انگلیس، آمریکا، رکورد مربوط به کل کشورهای جهان	۵۵ رکورد از هر کدام از ستون ها از کشورهای الجزایر، آرژانتین، ارمنی، آذربایجان، بنگلادش، بنین، بورکینافاسو، کلمبیا، الجزایر قمر، کاستاریکا، دومینیکا، اکوادور، مصر، السالوادور، اتیوپی، گامبیا، غنا، هاییتی، هند، اندونزی، ایران، اسرائیل، قزاقستان، کنیا، قرقیزستان، لائوس، لیبریا، مالاوی، مالزی، موریس، مکزیک، منغولستان، مراکش، موزامبیک، میانمار، نپال، نیجر، پاکستان، پانامای پاراگوئه، سیشل، آفریقای جنوبی، سری لانکا، تانزانیا، تایلند، تیمور، تیگو، تونس، ترکیه، اوگاندا، اروگوئه، ویتنام، یمن، زامبیا، زیمباوه
نرخ مبتلایان جدید			
نرخ کل مرگ و میر			
نرخ مرگ و میر جدید			
نرخ کل مبتلایان در هر میلیون			
نرخ مبتلایان جدید در هر میلیون			
نرخ کل مرگ و میر در هر میلیون			
نرخ مرگ و میر جدید در هر میلیون			
شاخص شدت			
جمعیت			
تراکم جمعیت			
متوسط سن افراد مبتلا			
افراد مبتلا بالای ۶۵ سال			
افراد مبتلا بالای ۷۰ سال			
سرانه تولید ناخالص داخلی			
شدت فقر			
نرخ مرگ و میر بیماران قلبی و عروقی مبتلا			
نرخ مرگ و میر بیماران دیابتی مبتلا			
نرخ زن های سیگاری			
نرخ مردهای سیگاری			
نرخ تخت بیمارستان در هر هزار			



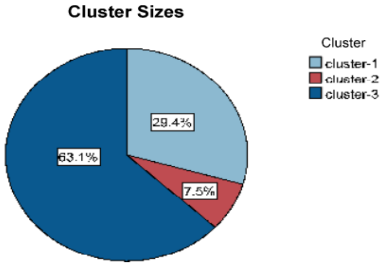
Size of Smallest Cluster	160 (1.7%)
Size of Largest Cluster	2758 (29.9%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	17.24

شکل ٦- نمودار دایره‌ای خوشه بندی با الگوریتم کا از میانگین



Size of Smallest Cluster	1 (0%)
Size of Largest Cluster	2078 (22.5%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	2,078.00

شکل ٥- نمودار دایره ای خوشه بندی با الگوریتم کوهونن



Size of Smallest Cluster	690 (7.5%)
Size of Largest Cluster	5813 (63.1%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	8.42

شکل ۷- نمودار دایره ای خوشه بندی با روش خوشه‌بندی دومرحله‌ای

به منظور سنجش کارایی روش های خوشه بندی مورد استفاده در این پژوهش، با استفاده از دستور Auto cluster در نرم افزار IBM Modeler 18.0 و اجرای آن روی داده ها معین می کنیم که با توجه به درجه اهمیت، از کدام الگوریتم استفاده نماییم. با در نظر گرفتن ضریب الگوریتم ها در سه حالت Build time، Silhouette و Importance، روش خوشه‌بندی دومرحله‌ای با توجه به این که وزن بالاتری برابر با ۰,۳۶۷ دارد، مطلوب تر از سایر الگوریتم ها جهت خوشه بندی می باشد.

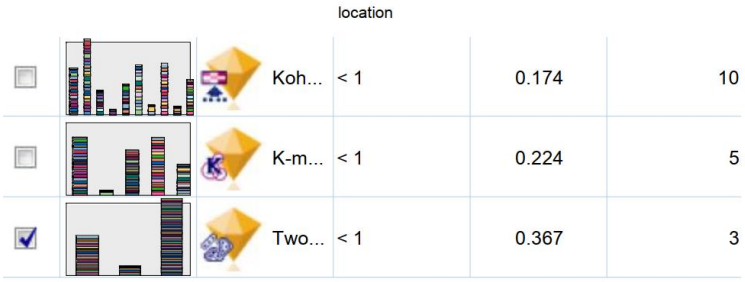
۱-برای حالت Build time (شکل ۸):

location

☑			Two...	< 1	0.367	3
☐			K-m...	< 1	0.224	5
☐			Koh...	< 1	0.174	10

شکل ۸- کارایی روش های خوشه بندی برای حالت Build time

۲-برای حالت Silhouette (شکل ۹):



شکل ۹- کارایی روش های خوشه بندی برای حالت Silhouette

۳-برای حالت Importance (شکل ۱۰):

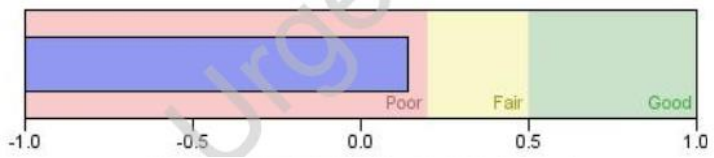
شکل ۱۰- کارایی روش های خوشه بندی برای حالت Importance

همچنین با توجه به مقدار ضریب نیمرخ[1]، برای سه روش خوشه بندی مورد استفاده در این پژوهش نیز به همین نتیجه می رسیم که روش خوشه‌بندی دومرحله‌ای با ضریب نیمرخ نزدیک به ۰،۵ کاراتر از سایر روش ها می باشد. ضریب نیمرخ برای الگوریتم کوهونن به شرح شکل ۱۱، برای الگوریتم کا از میانگین به شرح شکل ۱۲ و برای روش خوشه‌بندی دومرحله‌ای به شرح شکل ۱۳ می باشد. شاخص نیمرخ بر اساس ماتریس عدم تشابه بنا شده است. مقدار ضریب نیمرخ برای هر نقطه بر اساس میزان شباهت آن به خوشه آن نسبت به میزان شباهت به خوشه های دیگر محاسبه می شود :

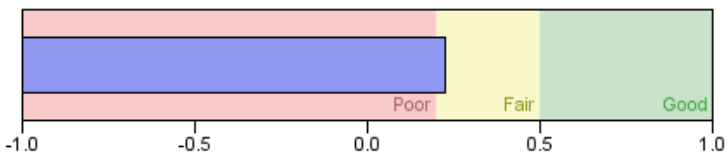
$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \qquad (4)$$

در این رابطه، a_i نشاندهنده میانگین فاصله از نقطه i ام تا سایر نقاط هم خوشه ای آن است و b_i نمادکمترین مقدار متوسط میان نقطه i ام تا سایر خوشه هاست. ضریب نیمرخ در بازه [۱ و ۱-] مقدار می گیرد که هرچه مقدار آن به یک نزدیک تر باشد، مطلوب تر است. تفسیر مقدار ضریب نیمرخ با توجه به معادله شماره ۴ به شرح ذیل می باشد :

ضریب نیمرخ :
- ۱- شیء اشتباه به خوشه ای تخصیص داده شده است.
- ۰ شیء به روشنی به خوشه خود یا همسایه تعلق ندارد.
- ۱ شیء از خوشه های همسایه بسیار دور است که بهترین وضعیت را دارد (مؤمنی، ۱۳۹۸ و فائضی راد و پویا، ۱۳۹۵).

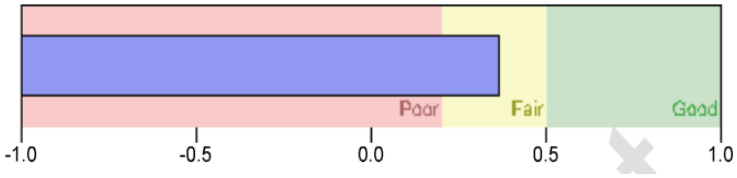


شکل ۱۱- ضریب نیمرخ برای الگوریتم کوهونن



شکل ۱۲- ضریب نیمرخ برای الگوریتم کا از میانگین

1 Silhouette



شکل ۱۳- ضریب نیمرخ برای روش خوشه‌بندی دومرحله‌ای

نتیجه‌گیری

شیوع کووید- ۱۹ به یک تهدید بالینی برای جمعیت عمومی و پرسنل مراقبت های بهداشتی در سراسر جهان تبدیل شده است. پژوهش حاضر داده های مرتبط با کرونا ویروس جدید- ۲۰۱۹ را از سایت بهداشت جهانی در تاریخ ۷ ژوئن اخذ نموده است که داده ها در قالب فایل اِکسل از سایت بهداشت جهانی در بازه ماه اول (January) سال ۲۰۲۰ الی ماه ششم (June) سال ۲۰۲۰ در تاریخ های متفاوت در طول این بازه برای ۸۰ کشور (۷۹ کشور و رکورد آخر مربوط به آمار کل کشورهای جهان است) جمع آوری شده است که شامل ۲۴ ستون (متغیر) می باشد. به منظور خوشه بندی، از روش های خود سازمان‌ده از جمله الگوریتم های کوهونن و کا از میانگین و خوشه بندی دومرحله ای بهره گرفته و در نهایت با انجام آزمون، کارایی این سه تکنیک تعیین شده است. با توجه به نتایج ملاحظه می گردد که روش خوشه بندی دومرحله ای با توجه به مقدار ضریب نیمرخ تقریباً نزدیک به ۰,۵ و انجام دستور Auto Cluster با وزن ۰,۳۶۷ (با توجه به شکل های ۸-۱۳) از سایر روش ها جهت خوشه بندی کاراتر می باشد. این پژوهش با در نظر گرفتن روش های متفاوت در خوشه بندی داده ها نسبت به تحقیقات مشابه کاربردی تر می باشد و نتایج دقیق تری ارائه داده است. برای تحقیقات آتی پیشنهاد می گردد از روش های خوشه بندی فازی از تلفیق رویکرد فازی در بحث خوشه بندی و یا تلفیق شبکه عصبی مصنوعی و روش های خود سازمان‌ده بهره گرفته شود.

منابع

۱. زعفریان، طاهره، اندبیلی، محمد، مومنی، حسین، نجفی، سید اسماعیل (۱۳۹۷)، «بخش بندی قیمتی بازار خودروی سواری ایران و رتبه بندی خودروها در بخش های قیمتی با استفاده از روش ترکیبی دیمتل-خوشه بندی دومرحله ای-تاپسیس و وزندهی دو مرحله ای آنتروپی شانون»، فصلنامه علمی پژوهشی مطالعات مدیریت صنعتی، شماره ۵۰، صص ۱۵۹-۱۹۲

۲. محمدی، بهمن و کامکار روحانی، ابوالقاسم (۱۳۹۶)، «به کارگیری روش های خوشه بندی کا از میانگین، میانگین فازی و گوستافسون کسل در تلفیق نتایج وارون سازی داده های توموگرافی لرزهای انکساری و مقاومت ویژه الکتریکی برای ارزیابی آبرفت و سنگ بستر»، علوم زمین خوارزمی، شماره ۲، صص ۱۸۳-۱۹۸

۳. توکلی، احمد، وحدت، کتایون و کشاورز، محسن (۱۳۹۸)، «کروناویروس جدید ۲۰۱۹ (COVID-19): بیماری عفونی نوظهور در قرن ۲۱»، پژوهشکده زیست-پزشکی خلیج فارس، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی بوشهر، شماره ۶، صص ۴۳۲-۴۵۰

۴. جعفری، سمیه، فرشید، راضیه و جباری، لیلا (۱۳۹۹)، «تحلیل موضوعی مطالعات کووید-۱۹ در پنج قاره بزرگ»، دوفصلنامه علمی ـ پژوهشی دانشگاه شاهد، صص ۱-۲۰

۵. علی پور، احمد، اورکی، محمد و خرامان، آزیتا (۱۳۹۸)، «مروری بر اثرات عصبی و شناختی کووید-۱۹»، فصلنامه علمی ـ پژوهشی عصب روانشناسی، شماره ۴، صص ۱۳۵-۱۴۶

۶. فرنوش، غلامرضا، علیشیری، غلامحسین، حسینی ذیجود، سید رضا، درستکار، روح الله و جلالی فراهانی، علیرضا (۱۳۹۹)، «شناخت کروناویروس نوین-۲۰۱۹ و کووید-۱۹ بر اساس شواهد موجود- مطالعه مروری»، مجله طب نظامی، دوره ۲۲، شماره ۱، صص ۱-۱۱

۷. جوادی،سامان، هاشمی، مهدی و سوخته زاری، مهدی (۱۳۹۷)، «تحلیل چند پارامتره آلودگی آبخوان قزوین بر مبنای نقشه کاربری اراضی و با استفاده از تکنیک خوشه بندی کا از میانگین»، اکوهیدرولوژی، دوره ۵، شماره ۱، صص ۲۹۳-۳۰۵

۸. کریمی علویجه، محمدرضا، خدنگی، سعید و صالح ترکستانی، محمد (۱۳۹۵)، «روش فرا ابتکاری در یکپارچه سازی مدل بخش بندی بازار مشتریان تلفن همراه تهران با استفاده از شبکه های خودسازمانده و روش میانگین کا»، مدیریت فناوری اطلاعات، دانشکده مدیریت دانشگاه تهران، دوره ۸، شماره ۲، صص ۳۵۱-۳۷۲

۹. مؤمنی، منصور، (۱۳۹۸)، «خوشه بندی (تحلیل خوشه ای)»، چاپ پنجم

۱۰. سلطانی، آزاده، صدوقی یزدی، هادی، اشک زری طوسی، سهیلا و روحانی، مجتبی (۱۳۸۹)، «بهبود شبکه خود سازمانده کوهونن با هدف خوشه بندی داده های فازی»، دهمین کنفرانس سیستم های فازی ایران، دانشگاه شهید بهشتی، شماره ۲۲-۲۴، صص ۱۱۷-۱۲۳

۱۱. فائضی راد، محمدعلی و پویا، علیرضا (۱۳۹۵)، «خوشه بندی فروشگاه های آنلاین از نگاه تأمین کننده با کمک بهینه یابی تعداد خوشه ها در الگوریتم دومرحله ای SOM»، فصلنامه علمی پژوهشی مطالعات مدیریت صنعتی، شماره ۴۳، صص ۱۳۴-۱۰۹

12. Thompson, R. (2020). Pandemic potential of 2019-nCoV. Lancet Infect Dis .20 (3), 280.
13. Lai, C.C., Shih, T.P., KO, W.C., Tang, H.J. and Hsueh, P.R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents. Mar; 55(3),105924.
14. Kantardzic M. (2011). DATA MINING Concepts, Models, Methods, and Algorithms, Second Edition.
15. Manochandar, S., Punniyamoorthy, M. and Jeyachitra, R.K. (2020). Development of new seed with modified validity measures for k- means clustering, Computers & Industrial Engineering, 1-40, https://doi.org/10.1016/j.cie.2020.106290.
16. Wang, F., Wang, Q., Nie, F., Li, Zh., Yu, W. and Ren, F. (2020). A linear multivariate binary decision tree classifier based on K-means splitting, Pattern Recognition, 107, pp. 1-13.
17. Ebrahim Ghadiri, S.M. and Mazlumi, K. (2019). Adaptive protection scheme for microgrids based on SOM clustering technique, Applied Soft Computing Journal, 1-46, https://doi.org/10.1016/j.asoc.2020.106062.
18. Kamimura, R. (2019). SOM-based information maximization to improve and interpret multi-layered neural networks: From information reduction to information augmentation approach to create new information, Expert Systems with Applications, 125, pp. 397-411.
19. Hassannezhad, M. and Clarkson, P.J. (2017). Internal and External Involvements in Integrated Product Development: A Two-Step Clustering Approach, Procedia CIRP 60, pp. 253 – 260.