فصلنامه علمی برنامه‌ریزی منطقه‌ای
سال ۱۰، شمارهٔ پیاپی۳۷، بهار ۱۳۹۹
شاپای چاپی: ۶۷۳۵- ۲۲۵۱ – شاپای الکترونیکی: ۷۰۵۱- ۲۴۲۳
http://jzpm.miau.ac.ir

مقاله پژوهشی

ارزیابی و تحلیل نرخ بیکاری در سطح استان‌های کشور با استفاده از خوشه‌بندی مبتنی بر چگالی پیش‌بینی

رامین خوچیانی[1]: استادیار گروه اقتصاد، دانشگاه آیت ا... العظمی بروجردی، بروجرد، ایران
سید محمد حسینی: استادیار گروه ریاضی، دانشگاه آیت ا... العظمی بروجردی، بروجرد، ایران

دریافت:۱۳۹۷/۸/۲۰ صص ۱۶-۱ پذیرش: ۱۳۹۸/۲/۶

چکیده

آگاهی از وضعیت نرخ بیکاری استانهای کشور در افق‌های زمانی معین، برای برنامه‌ریزان منطقه‌ای و سیاستگزاران اقتصادی کشور بسیار مهم می‌باشد. در این مقاله خوشه‌بندی سری‌های زمانی بر اساس چگالی پیش‌بینی آنها تا افق مشخص بررسی شده است. در این روش به منظور تقریب توزیع پیش‌بینی‌ها از فرآیند بوت‌استرپ غربالی استفاده می‌شود. تفاوت‌های بین هر زوج از چگالی‌های بوت‌استرپ یک ماتریس عدم تشابه تولید می‌کند که برای خوشه‌بندی استفاده می‌شود. به همین منظور از داده‌های فصلی نرخ بیکاری استانی در بازه زمانی بهار ۱۳۸۴ تا پاییز ۱۳۹۶ استفاده گردید و با توجه به الگوریتم چگالی پیش‌بینی، نرخ بیکاری استان‌های کشور را برای دو افق زمانی ۴ فصل (یک سال) و ۱۰ فصل (دو سال و نیم) خوشه‌بندی کردیم. بهترین وضعیت را چه در ۴ گام و چه در ۱۰ گام (دو سال و نیم)، استان‌های سمنان و زنجان خواهند داشت و بدترین وضعیت استان‌های لرستان و کرمانشاه دارا می‌باشند. همچنین در دو افق زمانی مورد مطالعه، به جز چند استان، بقیه در خوشه‌های اصلی خود ثابت بودند. همچنین نمودار پراکنش فضایی بیکاری در ایران بر اساس خوشه‌بندی مبتنی بر چگالی پیش‌بینی نشان می‌دهد که همچنان استان‌های غرب و جنوب غربی بالاترین نرخ بیکاری را خواهند داشت. بنابراین لزوم برنامه‌ریزی منطقه‌ای و توجه جدی به اشتغال استان‌های فوق‌الذکر توصیه می‌شود. در ضمن استان‌هایی که در وضعیت نامطلوبی قرار دارند، دارای همسایگان با نرخ بیکاری بالا هستند و استان‌های با نرخ بیکاری پایین عمدتاً دارای استانهای همجوار با نرخ بیکاری پایین هستند. به عبارت دیگر یک همبستگی مکانی مثبت بین استانهای همجوار و نرخ بیکاری وجود دارد.

واژه‌های کلیدی: خوشه‌بندی مبتنی بر چگالی پیش‌بینی، سری‌های زمانی، بوت‌استرپ غربالی، نرخ بیکاری.

[1] نویسنده مسئول: *khochiany@abru.ac.ir*. ۰۹۱۳۳۷۲۸۲۳۱

مقدّمه:

با پیشرفت علم و تکنولوژی، تولید داده‌ها در عصر حاضر به مراتب بیشتر از گذشته است. مدیریت داده‌های خام و استخراج اطلاعات و دانش مفید از آنها، نقش مهمی در تصمیم‌گیری‌ها دارد. تحلیل داده‌ها در اکثر حوزه‌های پژوهش از جمله مدیریت، اقتصاد، مهندسی و سایر رشته‌ها استفاده شده است *(Radmehr and Alamolhodai, 2014)*. خوشه‌بندی، روشی بسیار قدرتمند برای کشف طبقات و وابستگی‌های طبیعی در یک مجموعه داده و همچنین شناخت الگوهای ساختاری موجود در آن، بدون داشتن هرگونه پیش‌زمینه در مورد ویژگی‌های داده است. در مورد داده‌های سری زمانی، روش‌های خوشه‌بندی را می‌توان بطور کلی به خوشه‌بندی‌های مبتنی بر داده‌های خام[1]، مبتنی بر ویژگی‌های مستخرج از داده‌ها[2] و مبتنی بر مدل[3] تقسیم‌بندی کرد *(Liao, 2005: 1858)*. در روش‌های مبتنی بر داده‌های خام، خوشه‌بندی بر اساس اطلاعات مقطعی انجام می‌شود که چندان مطلوب نیست زیرا اطلاعات مقطعی به طور ذاتی ایستا هستند و تحول سری‌های زمانی را لحاظ نمی‌کند. در روش‌های مبتنی بر ویژگی‌ها ابتدا ویژگی‌های هر یک از سری‌های زمانی استخراج می‌شود و سپس بر اساس ویژگی‌ها خوشه‌بندی انجام می‌گیرد. لذا نویز زیاد داده‌های خام در خوشه‌بندی تاثیرگذار نیست. در روش‌های مبتنی بر مدل، ابتدا برای هر سری زمانی، مدل یا ترکیبی از توزیع‌های احتمال که آنرا تولید می‌کند شناسایی می‌شود و سپس بر اساس مدل‌های بدست آمده، خوشه‌بندی انجام می‌گیرد. در این روش‌ها می‌توان با وابستگی بین داده‌های زمانی مقابله کرد.

در این مقاله از هر دو دیدگاه مبتنی بر ویژگی و مبتنی بر مدل برای خوشه‌بندی داده‌های سری زمانی استفاده شده است. هدف اصلی از انجام این پژوهش، خوشه‌بندی داده‌های سری زمانی بر اساس چگالی‌های پیش‌بینی کامل برای هر یک از سری‌های موجود در مجموعه، بجای تمرکز بر پیش‌بینی‌های نقطه‌ای است. در این جا با الگوریتم جدید مبتنی بر چگالی پیش‌بینی، سری‌های زمانی در یک توزیع چگالی پیش‌بینی در افق زمانی آینده مشخص با دیگر سری‌های زمانی مشابه در همان افق زمانی در یک خوشه قرار می‌گیرند. در روش ارایه شده، نه تنها مدل‌ها در خوشه‌بندی تاثیر می‌گذارند بلکه پیش‌بینی‌ها در بازه‌ی زمانی مشخص، نیز لحاظ خواهد شد. بنابراین نویز زیاد داده‌های خام و همچنین وابستگی بین داده‌های زمانی بر این روش خوشه‌بندی تاثیرگذار نخواهند بود. مهم ترین منبع اقتصادی هر کشور، نیروی انسانی آن است. پدیده‌ی بیکاری، معیاری مناسب برای سنجش وضعیت توازن بین ارکان اساسی اقتصاد کشور است. لذا پرداختن به مساله بیکاری و عوامل کاهش یا افزایش آن برای مسولین و برنامه‌ریزان کشور به منظور هدف‌گزاری و اتخاذ سیاست‌های مناسب برای مقابله با افزایش آن، اهمیت ویژه‌ای دارد. ویژگی‌های مختلفی از پدیده‌ی بیکاری و بخصوص نرخ بیکاری را می‌توان به عنوان سری زمانی در مقاطع سالیانه، فصلی، ماهیانه و غیره ذخیره کرد. استفاده از خوشه‌بندی‌های متعارف که یا مبتنی بر ویژگی و یا مبتنی بر ساختار داده‌هاست نمی‌تواند چندان مفید واقع شوند. در این پژوهش به منظور نمایش کارآیی روش خوشه‌بندی سری‌های زمانی ارایه شده، مساله‌ی خوشه‌بندی نرخ بیکاری در سطح استان‌های کشور مورد توجه قرار گرفت و استان‌های کشور از لحاظ نرخ بیکاری در افق چهار سال و ده سال آینده خوشه‌بندی شدند. در تحقیق حاضر، یک مفهوم جدید عدم تشابه،[4] که با پیش‌بینی‌های آینده سر و کار دارد، معرفی خواهد شد. بدیهی است، یک خوشه‌بندی مبتنی بر این مفهوم ممکن است با مواردی که از روش‌های مبتنی بر مدل[5] یا مبتنی بر ویژگی[6] بدست می‌آیند، نتایج کاملا متفاوتی داشته باشد. خوشه‌بندی و عدم تشابه مبتنی بر پیش‌بینی برای اهداف قابل تفسیر به آسانی قابل استفاده است. هدف واقعی این نیست که مجموعه‌ای از سری‌های زمانی که شکل یا ساختار مشابه دارند در یک خوشه قرار بگیرند؛ بلکه هدف گروه‌بندی مجموعه‌ای از سری‌هایی که پیش‌بینی‌های مشابهی در یک زمان مشخص از آینده دارند، است. بدین منظور سوالات تحقیق زیر مطرح می شود:

۱)آیا بین بیکاری و موقعیت استانها، همبستگی معناداری وجود دارد؟

۲)وضعیت توزیع بیکاری در بین استانهای کشور به چه صورت است؟

۱) بین بیکاری استانها و موقعیت مکانی آنها همبستگی معنادار وجود دارد.

۲) توزیع بیکاری در استان‌های کشور به صورت ناهمگن توزیع شده است به طوری که در استان‌های غرب کشور شدت بیکاری بالاتر از شرق کشور است.

از نتایج خوشه‌بندی حاصل می‌توان به موارد زیر اشاره کرد. شناسایی استان‌هایی که پیش‌بینی نرخ بیکاری آنها در خوشه‌های با نرخ بیکاری بالا یا پایین قرار می‌گیرند. به این ترتیب می‌توان استان‌ها را بر اساس این معیار اولویت‌بندی کرد و به استان‌هایی که در افق مورد نظر، در خوشه‌ی افزایش نرخ بیکاری قرار می‌گیرند در سیاست‌گزاری‌های بیشتر توجه نمود. همچنین در مورد استان‌هایی که در یک خوشه قرار می‌گیرند می‌توان وجوه اشتراکی که منجر به این خوشه‌بندی شده است را شناسایی و با اعمال سیاست‌های کلی‌تر روند نرخ بیکاری را بهتر کنترل کرد. شایان ذکر است تحقیقاتی که در زمینه‌ی بازار کار در کشور انجام گرفته است به دلیل محدودیت آمار، تنها به توصیف یا تجزیه و تحلیل آمار بیکاری بر اساس اطلاعات سرشماری‌های عمومی نفوس و مسکن و بخصوص

---

1 Raw data based approach
2 Feature based approach
3 Model based approach
4 Dissimilarity.
5 . Model-Based Clustering.
6 . Feature-Based Clustering.

بر پایه‌ی تعریف مرکز آمار از بیکاری محدود می‌باشد و این تحقیق نیز ناگزیر از این محدودیت می‌باشد. چنانچه آمار و داده‌های بیشتر از منابع مختلف و بخصوص با تعاریف متفاوت از بیکاری در اختیار قرار گیرد به طور قطع نتایج بدست آمده نیز به واقعیت موجود در جامعه نزدیک‌تر خواهد بود.

پیشینه و مبانی نظری تحقیق:

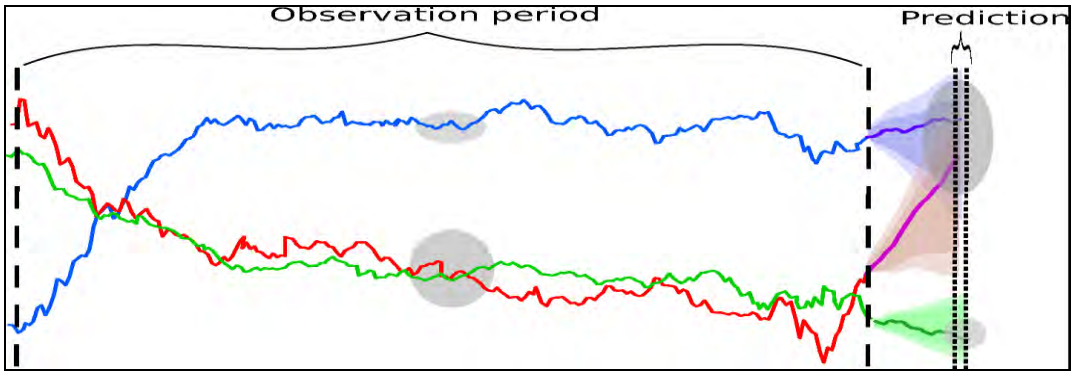
خوشه‌بندی به‌عنوان یکی از روش‌های داده‌کاوی توصیفی، به‌دنبال سازماندهی مجموعه‌ای از داده‌ها با استفاده از اندازه‌گیری فاصله یا انحراف، در تعدادی خوشه است به‌طوری‌که داده‌های قرار گرفته در هر خوشه، بالاترین درجه شباهت را دارند و داده‌های متعلق به خوشه‌های متفاوت دارای حداکثر میزان تفاوت هستند *(Kaufman and Rousseeuw, 2005: 33)*. مطالعات گسترده‌ای بر روی مسایل خوشه‌بندی انجام گرفته است اما از روش‌های چند متغیره در مورد سری‌های زمانی نمی‌توان بهره برد. علاوه بر این، روش‌های خوشه‌بندی عام، همبستگی ساختار سری‌های زمانی را در نظر نمی‌گیرد. این نکته، انگیزه‌ای برای توسعه‌ی روش‌های خوشه‌بندی سری‌های زمانی است. مقاله فروهویرس و کافمن (۲۰۰۴) *(Fruhwirth-Schnatter and Kaufmann, 2004)* یک مدل بیزین مرکب همراه با یک الگوریتم *MCMC* برای تخمین پارامتر و دسته‌بندی داده‌ها پیشنهاد می‌کند. پاتارین و همکاران *(Pattarin et al, 2004)* در سال ۲۰۰۴ از روش‌های چندمتغیره کلاسیک بعد از الگوریتم ژنتیک، استفاده کرده است. مطالعه‌ای جامع در مورد سایر تکنیک‌های چندمتغیره در مقاله گالیانو و پنا (۲۰۰۰) *(Galeano and Peña, 2000)* و مروری کامل بر روش‌های خوشه‌بندی سری‌های زمانی در مقاله‌ی آقابزرگی و همکاران (۲۰۱۵) *(Aghabozorgi et al, 2015)* ارایه شده است.

در خوشه‌بندی سری‌های زمانی که در این مقاله مورد نظر است نه تنها مدل‌هایی که مشاهدات را تولید می‌کنند در خوشه‌بندی تاثیر می‌گذارند بلکه پیش‌بینی‌ها در بازه‌ی زمانی مشخص، نیز لحاظ خواهد شد. چنانچه سری‌های زمانی را بر اساس مدل‌ها یا بر اساس آخرین مقادیر مشاهدات، خوشه‌بندی کنیم بوضوح نتایج عمیقاً متفاوتی با حالتی که خوشه‌بندی بر اساس پیش‌بینی‌های تا افق مشخص انجام می‌گیرد نتیجه می‌شود. تکنیک استفاده شده در این مقاله بخصوص زمانی که پدیده‌ی مورد بررسی بر سری‌های طولانی مدت همگرا یا واگرا قرار می‌گیرد مناسب می‌باشد. در بررسی یک پدیده، ممکن است هدف این باشد که تعیین کنیم آن پدیده یا سیستم می‌تواند در افقی مشخص به یک سطح معلومی برسد یا خیر. در اینصورت این روش کارآمد و موثر واقع می‌شود. چنین موضوع‌هایی اغلب در بسیاری از وضعیت‌های واقعی رخ می‌دهد. برای نمونه، در هر مساله‌ی توسعه‌ی پایدار، پیش‌بینی رسیدن به یک سطح مورد انتظار در آینده ظاهر می‌شود. در چنین وضعیت‌هایی مناسب است که یک روش خوشه‌بندی بکار برده شود که بطور مستقیم بر ویژگی‌های مقادیر پیش‌بینی تکیه دارد. به عنوان نمونه آلونسو و همکاران *(Alonso et al. 2006)* در سال ۲۰۰۶ مساله و موضوع کاهش انتشار گاز کربن دی‌اکسید کشورها به منظور رسیدن به مفاد پروتکل کیوتو تا سال ۲۰۱۲ را بررسی کردند. در چنین مساله‌ای بوضوح بکار بستن خوشه‌بندی بر اساس مقادیر پیش‌بینی تا افق ۲۰۱۲ کشورها، مناسب و معقول بنظر می‌رسد.

در تحقیق حاضر، یک مفهوم جدید عدم تشابه، که با پیش‌بینی‌های آینده سر و کار دارد، معرفی خواهد شد. بدیهی است، یک خوشه‌بندی مبتنی بر این مفهوم ممکن است با مواردی که از روش‌های مبتنی بر مدل یا مبتنی بر ویژگی بدست می‌آیند، نتایج کاملا متفاوتی داشته باشد. خوشه‌بندی و عدم تشابه مبتنی بر پیش‌بینی برای اهداف قابل تفسیر به آسانی قابل استفاده است. هدف واقعی این نیست که مجموعه‌ای از سری‌های زمانی که شکل یا ساختار مشابه دارند در یک خوشه قرار بگیرند؛ بلکه هدف گروه‌بندی مجموعه‌ای از سری‌هایی که پیش‌بینی‌های مشابهی در یک زمان مشخص از آینده دارند، است. شکل شماره ۱ این مسأله را بیشتر روشن می‌کند. همانطور که در شکل مشاهده می‌شود، خوشه‌بندی‌های مبتنی بر شکل[1] مبتنی بر ساختار داده‌ها[2] و مبتنی بر چگالی پیش‌بینی[3] برای ۳ سری زمانی انجام شده است. جالب این است که هر کدام از خوشه‌بندی‌ها، نتایج کاملاً متفاوتی داشته است.

---

1 *Shape Based Clustering.*
2 *Structure Based Clustering.*
3 *Clustering Based On Forecast Densities.*



نمودار ۱- سه خوشه‌بندی متفاوت مبتنی بر شکل، مبتنی بر ساختار داده و مبتنی بر چگالی پیش‌بینی. - منبع: *(Montero and Vilar, 2006)*

دو سری زمانی که توسط یک توزیع مشابه تولید شده‌اند، می‌توانند در یک افق از قبل مشخص‌شده[1] پیش‌بینی‌های متفاوتی داشته باشند و از این رو ممکن است این سری‌ها با استفاده از این معیار متمایز جدید با یکدیگر همخوانی نداشته باشند. خوشه‌بندی به طور مستقیم بر نوع و ویژگیهای پیش‌بینی‌ها مبتنی است؛ برای نمونه می‌توان گفت که هرگونه شرایطی که در مورد رسیدن به مقادیر هدف مشخص در دوره زمانی آتی از پیش‌تعیین‌شده وجود دارد، از این دسته خوشه‌بندی‌هاست(*Montero and Vilar, 2014: 12*) آلونسو و همکاران در سال ۲۰۰۶ ، یک مفهوم اندازه عدم مشابهت را بر پایه مقایسه چگالی پیش‌بینی برای هر سری در یک افق آینده مشخص پیشنهاد دادند. آنها استدلال کردند که با استفاده از تمام چگالی‌های پیش‌بینی، تغییرپذیری پیش‌بینی‌ها مدنظر قرار می‌گیرد. چیزی که در پیش‌بینی‌های نقطه‌ای (یعنی پیش‌بینی نقاط داده، و نه پیش‌بینی چگالی احتمال وقوع داده) کاملا نادیده گرفته می‌شود. در واقعیت، چگالی پیش‌بینی نامعلوم است و می‌بایست به کمک داده‌ها تقریب زده شود. آلونسو و همکاران (۲۰۰۶) این تقریب را با استفاده از ترکیب روش بوت استرپ غربالی هموار شده و تکنیک‌های تخمین چگالی هسته (کرنل) به دست آورده‌اند.

البته این روش مستلزم پذیرفتن این فرض است که سری زمانی، یک رفتار خودتوضیح مرتبه اول را از خود نشان داده است، چرا که بوت استرپ غربالی بر اساس بازنمونه‌گیری پسماندهایی است که از تقریب‌های خودتوضیح[2] به دست می‌آیند. ویلار و همکاران (۲۰۱۰) *(Vilar et al, 2010)* این روش را توسعه داده و مدلهای خود توضیح غیرمقید ناپارامتریک را نیز پوشش دادند. در این روش، چگالی پیش‌بینی‌ها با روش بوت استرپی که فرآیندهای تولید داده را بدون هرگونه فرض مدل پارامتریک (برای ساختار اتورگرسیو حقیقی سری‌ها) ایجاد می‌کند؛ تخمین زده می‌شود.[3]فرآیند خوشه‌بندی بکار رفته در این مقاله، بر اساس چگالی‌های پیش‌بینی کامل برای هر یک از سری‌های موجود در مجموعه، بجای تمرکز بر پیش‌بینی‌های نقطه‌ای است. یک فرآیند بوت‌استرپ غربالی بکار گرفته شده است که با ایده‌های تخمین چگالی هسته ناپارامتریک ترکیب شده است تا توزیع پیش‌بینی‌ها را تقریب بزند. تفاوت‌های بین هر زوج از چگالی‌های بوت‌استرپ، یک ماتریس عدم تشابه تولید می‌کند که در تشخیص ساختارهای خوشه‌بندی، استفاده می‌شود.

یکی از مزیت‌های فوق‌العاده‌ی فرآیند جدید ارایه شده در این مقاله، کاهش ابعاد مساله است. زیرا مساله‌ی مورد بررسی یک فضای سه بعدی متشکل از تعدادی سری زمانی متفاوت در n ویژگی در T گام زمانی دارد که دیدگاه مذکور ساختار آنها را به یک ساختار دو بعدی تبدیل می‌کند. در این ساختار p پیش‌بینی حاصل از n ویژگی در افق مشخص $T+h$ انجام می‌پذیرد. بخاطر اینکه این پیش‌بینی‌ها، از اطلاعات حال و مقادیر گذشته‌ی این سری‌ها نتیجه می‌شود لذا هیچگونه اطلاعات ارزشمندی حذف نمی‌گردد. در نهایت با در نظر گرفتن چگالی‌های پیش‌بینی کامل بجای اینکه فقط پیش‌بینی‌های نقطه‌ای لحاظ شود، اجازه می‌دهد سری‌های زمانی مربوط به مدل‌های مشابه،

---

1 Pre-Specified Horizon.
2 Autoregressive.

[3] . روش عمومی که توسط ویلار و همکاران در سال ۲۰۱۰ پیشنهاد شد در پکیج *Tsclust* نرم افزار R، اجرا شده است. و به همین دلیل امکان طبقه‌بندی مدلهای خود توضیح عمومی، از جمله گستره وسیع مدلهای پارامتریک مورد مطالعه همانند مدل‌های خودتوضیح آستانه‌ای (*TAR*)، خودتوضیح نمایی (*EXPAR*)، مدل خودتوضیح انتقال هموار (*STAR*) و خطی دو گانه وجود خواهد داشت. نگاه کنید به (*Tong , Yeung 2000* و منابع در آن)

به خوشه‌های مختلف تفکیک شوند. به عنوان مثال، مدل‌هایی که تنها در تغییرپذیری مشاهدات یا در توزیع متفاوتند اما چگالی‌های پیش‌بینی مختلفی تولید می‌کنند. با توجه به اینکه در این مقاله به موضوع خوشه‌بندی بیکاری نیز پرداخته شده است در ادامه در مورد تحقیقات انجام گرفته بطور مختصر اشاره خواهد شد. نگاه به پدیده‌ی بیکاری، ملاک مناسبی برای سنجش وضعیت ساز و کار میان ارکان اساسی اقتصاد یک کشور است *(Gharavi Nakhjavani, 2002: 179)*. علاوه بر نرخ بالای بیکاری در سطح ملی، پراکندگی فضایی آن نیز مورد توجه قرار گرفته است. تحلیل تفاوت‌های منطقه‌ای بیکاری توجه فزاینده‌ای را در ادبیات جغرافیای اقتصادی جلب کرده است *(Rezvani et al., 2013: 41)*.

در مورد ساختار فضایی اختلافات بیکاری منطقه‌ای در سطح شهرستان‌ها، سعادت و همکاران (۱۳۸۷) با استفاده از مدل‌های اقتصادسنجی فضایی تحقیقی را انجام دادند. حق‌شناس *(Haghshenas, 2002)* به بررسی روند اشتغال و بیکاری جمعیت فعال استان‌های کشور پرداخت. شهیادی (۱۳۸۳) تأثیر مجاورت استان‌ها بر عرضه و تقاضای نیروی کار استان همجوار و میزان وابستگی فضایی بازار کار هر استان از همسایگان خود را با استفاده از روش و مدل‌های اقتصادسنجی فضایی، به صورت مقطعی، در سال ۱۳۷۵ بررسی می‌کند. صیدایی و همکاران (۱۳۹۰) به شناخت و ریشه‌یابی وضع بیکاری و اشتغال کشور از سال ۱۳۳۵ تا ۱۳۸۹ پرداخته‌اند. *(Abbasi Nejad et al, ۲۰۱۲)* در قالب یک الگوی اقتصادسنجی با استفاده از داده‌های تلفیقی، تأثیرگذاری مستقیم اما با وقفه های یک یا دو دوره‌ای بیکاری و سایر متغیرهای اجتماعی بر روی جرایم قتل عمد، سرقت و خودکشی را نشان دادند. رضوانی و همکاران (۱۳۹۲) با تأکید بر رویکرد تحلیل اکتشافی داده‌های مکانی، توزیع فضایی نرخ بیکاری را در نواحی شهری و روستایی ایران در سال ۱۳۹۰ مطالعه کرده‌اند. نتایج تحلیل اکتشافی داده‌های مکانی بیانگر وجود خوشه‌بندی (خود همبستگی مکانی مثبت) و ناهمگنی فضایی در توزیع شاخص‌های نرخ بیکاری شهری و روستایی در سطح شهرستان‌های ایران است. اخباری و طایی (۱۳۹۶) با استفاده از نسل‌های اول و دوم آزمون ریشه واحد پانل با تأکید بر نسل دوم، وجود اثر تأخیری در نرخ بیکاری را مورد بررسی قرار دادند. در مقاله حسینی و همکاران (۱۳۹۷) به واکاوی الگوهای مهاجرت داخلی در ایران در سه دهه‌ی اخیر پرداخته شده است.

نرخ بیکاری منطقه‌ای در کشور ترکیه، در سال‌های ۱۹۸۰ و ۲۰۰۰ توسط فیلیزتکین (۲۰۰۸) با استفاده از تکنیک‌های فضایی و ناپارامتریک، تحلیل شد. مطالعات بسیاری نیز در این حوزه قابل ذکر می‌باشند که به عنوان نمونه مقالات ذیل را اشاره می‌کنیم. *(Elhorst,2000; Garcilazo,2006; Cracolici et al,2007; Lopez-Bazo,2011)* در این پژوهش به منظور نمایش کارآیی روش خوشه‌بندی سری‌های زمانی ارایه شده، مساله‌ی خوشه‌بندی نرخ بیکاری در سطح استان‌های کشور مورد توجه قرار گرفت. به کمک این روش، استان‌های کشور از لحاظ نرخ بیکاری در افق چهار سال و ده سال آینده خوشه‌بندی شدند. به همین منظور از داده‌های فصلی نرخ بیکاری استانی در بازه زمانی بهار ۱۳۸۴ تا پاییز ۱۳۹۶ استفاده گردید و با توجه به الگوریتم چگالی پیش‌بینی، نرخ بیکاری استان‌های کشور را برای دو افق زمانی ۴ فصل (یک سال) و ۱۰ فصل (دو سال و نیم) خوشه‌بندی شدند.

مواد و روش تحقیق:

تحقیق حاضر به روش تحلیل کمی انجام شده است. البته به منظور تکمیل تحلیل، همراه ساختن تحلیل کیفی در یافته های پژوهش نیز نیاز است. ابتدا در بخش توصیفی تحلیل، درصد بیکاری استان‌های کشور به همراه آمارهای توصیفی، همچون میانگین، میانه، بیشینه، کمینه و چولگی آورده شده است. ابزار تحلیل و سنجش، خوشه‌بندی مبتنی بر چگالی پیش‌بینی است که در ادامه به تفصیل در مورد آن بحث خواهد شد. این روش تحلیل داده از طریق نرم افزار R و پکیج *TSclust* انجام و سپس خروجی‌های نرم افزار در بخش یافته‌های پژوهش تحلیل می‌شود. در این بخش، الگوریتم جدید خوشه‌بندی مبتنی بر پیش‌بینی که از طریق تخمین چگالی پیش‌بینی انجام می‌شود؛ معرفی خواهد شد و در پایان یک مطالعه موردی از خوشه‌بندی نرخ بیکاری در سطح استانهای کشور بر اساس این الگوریتم انجام می‌شود.

الگوریتم فوق بدین صورت است که فرض کنید X_T و Y_T فرآیند های مانایی باشند که به صورت فرم خودتوضیح زیر نمایش داده شوند.

$S_t = \varphi(S_{t-1}) + \varepsilon_t$	(۱)

اگرچه رشد اقتصادی در هر کشوری از اهمیت فوق‌العاده برخوردار است، اما از بسیاری جهات، ایجاد اشتغال و کاهش بیکاری از آن نیز مهم‌تر است. در اقتصادهای متعارف دنیا، مسئله رشد اقتصادی و ایجاد اشتغال دو مسئله مجزا تلقی نشده و حل یکی به معنای حل دیگری هم هست. با این حال، تجربه اقتصاد ایران به خصوص در سال‌های اخیر نشان می‌دهد که هر نوع رشد اقتصادی به رفع معضل بیکاری در اقتصاد ایران منجر نمی‌شود. بنابراین آگاهی از وضعیت نرخ بیکاری استان‌های کشور در افق‌های زمانی معین، برای برنامه‌ریزان و سیاست‌گزاران

اقتصادی کشور بسیار مهم می‌باشد. خوشه‌بندی‌های متعارف که یا مبتنی بر شکل سری زمانی و یا مبتنی بر ساختار داده‌هاست نمی‌تواند در این امر کمک کند؛ چرا که یا با توجه به اطلاعات گذشته (مبتنی بر مدل) و یا اطلاعات زمان حال (مبتنی بر شکل) گروه‌بندی استان‌ها انجام می‌شود. بنابراین محدوده مطالعه کلیه استان‌های کشور در بازه زمانی بهار ۱۳۸۴ تا پاییز ۱۳۹۶ می‌باشد. در این بخش با توجه به الگوریتم چگالی پیش‌بینی، نرخ بیکاری استانهای کشور را برای دو افق زمانی ۴ فصل (یک سال) و ۱۰ فصل (دو سال و نیم) خوشه‌بندی خواهیم کرد. به همین منظور از داده‌های فصلی نرخ بیکاری استانی استخراج شده از مرکز آمار استفاده گردید. خلاصه آمار استانی بیکاری در جدول شماره ۱ آمده است. ابتدا با توجه به آمار جدول ۱ وضعیت موجود در حوزه بیکاری استانی کشور شناسایی می‌شود.

جدول ۱- خلاصه آماری نرخ بیکاری استانهای کشور

نام استان	میانگین نرخ بیکاری	کمترین نرخ بیکاری	بیشترین نرخ بیکاری	چولگی
آذربایجان شرقی	۸/۶	۳/۹	۱۵/۷	-۰/۲۳۲
آذربایجان غربی	۱۰/۷۸	۶	۲۱/۵	۰/۸۵۶
اردبیل	۱۲/۳	۸/۵	۱۸/۳	۰/۰۶۲
اصفهان	۱۲/۴۲	۸	۱۹/۴	۰/۵۸۰
بوشهر	۱۰/۸	۶/۹	۱۵/۳	-۰/۵۲۹
تهران	۱۱/۳۱	۵/۱	۱۴/۹	-۰/۰۰۷
چهارمحال و بختیاری	۱۴/۲۹	۷	۲۲/۶	-۰/۶۴۰
ایلام	۱۳/۶۶	۸/۵	۲۱/۶	۰/۶۳۱
خراسان جنوبی	۹/۲۶	۵/۴	۱۴/۷	-۰/۷۳۴
خراسان رضوی	۱۰/۰۲	۶/۶	۱۵/۲	-۰/۷۸۵
خراسان شمالی	۹/۶۳	۴/۲	۱۷/۹	-۰/۲۷۴
خوزستان	۱۲/۴۲	۶/۹	۲۰/۸	۱/۴۷۳
سیستان و بلوچستان	۱۱/۳۷	۶/۱	۱۴/۲	۰/۳۳۷
کردستان	۱۲/۴۵	۵/۶	۲۳/۷	۰/۴۲۶
کرمان	۱۰/۴۱	۵/۲	۱۷/۲	-۰/۸۲۲
کرمانشاه	۱۶/۰۶	۱۰/۲	۲۵/۶	-۰/۵۰۹
سمنان	۹/۴۹	۵/۹	۱۷/۴	۲/۴۳
زنجان	۹/۷۵	۶/۱	۱۷/۹	۲/۷۲
مازندران	۹/۹۸	۵/۸	۱۵/۳	-۰/۵۵۰
مرکزی	۹/۹	۵	۱۸/۴	۱/۰۱۳
همدان	۱۱/۶۱	۵/۸	۲۳	۰/۸۹۸
هرمزگان	۹/۶۸	۳/۶	۱۵/۶	۰/۵۸۳
یزد	۸/۸۷	۴/۳	۱۳/۹	-۰/۷۰۵
لرستان	۱۶/۴	۱۰/۷	۲۵/۵	-۰/۴۴۲
فارس	۱۳/۹۶	۷/۱	۲۳/۹	۰/۳۲۷
گیلان	۱۳/۵۴	۸/۴	۲۱	-۰/۲۲۰
گلستان	۹/۴	۴	۱۵	-۰/۴۰۱
قزوین	۱۱/۰۹	۷/۴	۱۶/۸	۱/۵۴۹
قم	۱۰/۳۲	۷/۳	۱۴/۳	۰/۰۴۹
کهگیلویه و بویراحمد	۱۴/۹۸	۹/۸	۲۳	۰/۷۹۱

منبع: محاسبات پژوهش، ۱۳۹۷.

کرمانشاه با ۲۵/۶درصد و لرستان با ۲۵/۵ درصد بالاترین نرخ بیکاری در بین استانهای کشور در سالهای اخیر را دارند. از سوی دیگر هرمزگان با ۳/۶ درصد و آذربایجان شرقی با ۳/۹ درصد کمترین نرخ بیکاری در کشور در همین سالها را داشته‌اند. از نظر میانگین نیز استانهای لرستان با ۱۶/۴ و کرمانشاه با۱۶/۰۶درصد بالاترین و استانهای آذربایجان شرقی با ۸/۶ و یزد با ۸/۸۷ کمترین میانگین نرخ بیکاری را در بین استانهای کشور دارا هستند. سوالی که در اینجا مطرح می‌شود این است که چگونه می‌توان وضعیت بیکاری در افق‌های آتی مشخص را شناسایی و تحلیل کرد. یکی از راهکارها، استفاده از مدلهای پیش بینی روند است. اما چند نکته قابل توجه است؛ اول اینکه یا مانند مدل‌های پارامتریک می‌بایست فرض خاصی در مورد توزیع داده‌ها داشت و یا در غیر این صورت نیز پیش‌بینی‌های نقطه‌ای، همراه با خطای پیش‌بینی خواهد بود چرا که نتایج پیش‌بینی‌ها با تغییر پارامترها تغییر پذیرند. راه حل این مساله استفاده از مدل‌های چگالی پیش‌بینی با استفاده از مدل-

های ناپارامتریک است چرا که با استفاده از تمام چگالی‌های پیش‌بینی، تغییرپذیری پیش‌بینی‌ها مدنظر قرار می‌گیرد. چیزی که در پیش‌بینی‌های نقطه‌ای کاملا نادیده گرفته می‌شود. در واقعیت، چگالی پیش‌بینی نامعلوم است و می‌بایست به کمک داده‌ها تقریب زده شود. خوشه‌بندی مبتنی بر چگالی پیش‌بینی سری‌های زمانی نرخ بیکاری استانی را اگرچه از یک فرایند تولید مشابه به وجود آمده‌اند، با استفاده از تخمین چگالی در افق زمانی معین(که ممکن است با این معیار همخوانی نداشته باشند)در خوشه‌های مختلف قرار می‌دهد. پس از آنکه سیاستگزار دانست که در افق زمانی مشخصی وضعیت بیکاری استانهای خاصی مانند هم بوده، اقدام به سیاستگزاری می‌کند. در ادامه، تحلیل داده‌ها از طریق نمودارهای مربوط به چگالی پیش‌بینی ارائه شده است.

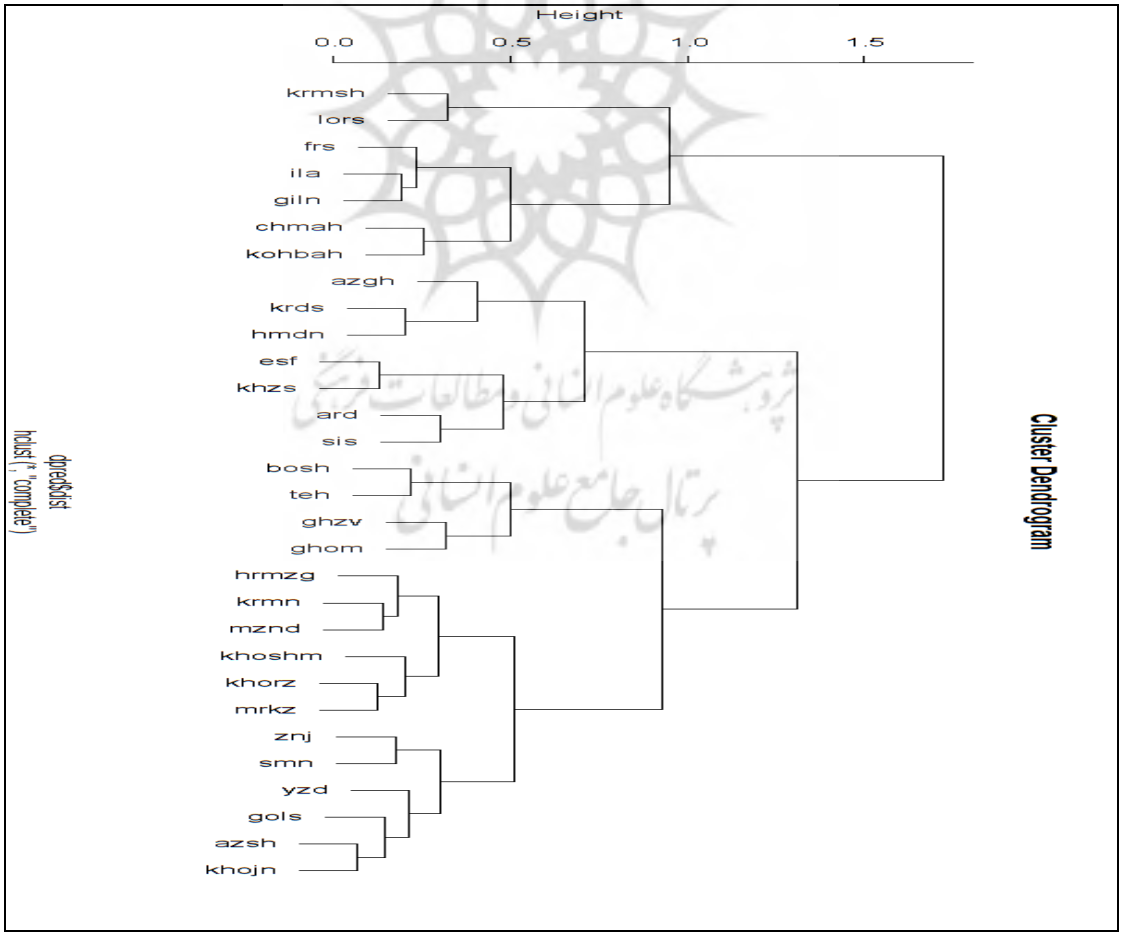
بحث و ارائه یافته‌ها:

نمودار درخت سلسله مراتبی که دندروگرام نیز گفته می‌شود، در شکل ۲-الف مشاهده می‌شود. این نمودار در جایی استفاده می شود که تعداد خوشه ها قبل از گروه بندی مشخص نباشد. فاصله‌ای که در این روش استفاده می شود عبارت است از فاصله بین دو سری زمانی از لحاظ عدم تشابه بین مقادیرشان در افق $T+h$. و یا به عبارت دیگر همان رابطه ۲ که در بالا به آن اشاره شد.

نمودار ۲-ب و ۳-ب نیز چگالی‌های پیش‌بینی نرخ بیکاری استانها در دو افق زمانی ۱۰ و ۴ گام به جلو است. با توجه به نمودارهای خوشه‌بندی نرخ بیکاری، برخی از مهمترین نتایجی که می‌توان به آن اشاره کرد به صورت ذیل است.

الف) در خوشه‌بندی چگالی پیش‌بینی ۴ گام به جلو (یکسال آینده) استانهای کشور بدین صورت خوشه‌بندی می‌شوند.

خوشه اصلی ۱) ، خوشه فرعی ۱-۱)زنجان، سمنان، کرمان، مازندران، خراسان رضوی، خراسان شمالی، خراسان جنوبی، یزد، آذربایجان شرقی، گلستان، مرکزی، هرمزگان. خوشه فرعی ۲-۱) قزوین، قم، سیستان و بلوچستان، بوشهر و تهران. خوشه اصلی ۲) ، خوشه فرعی ۱-۲) چهارمحال و بختیاری، کهکیلویه و بویراحمد، کرمانشاه و لرستان،. خوشه فرعی ۲-۲) کردستان، آذربایجان غربی، همدان، ایلام، گیلان، اردبیل، فارس، اصفهان، خوزستان.

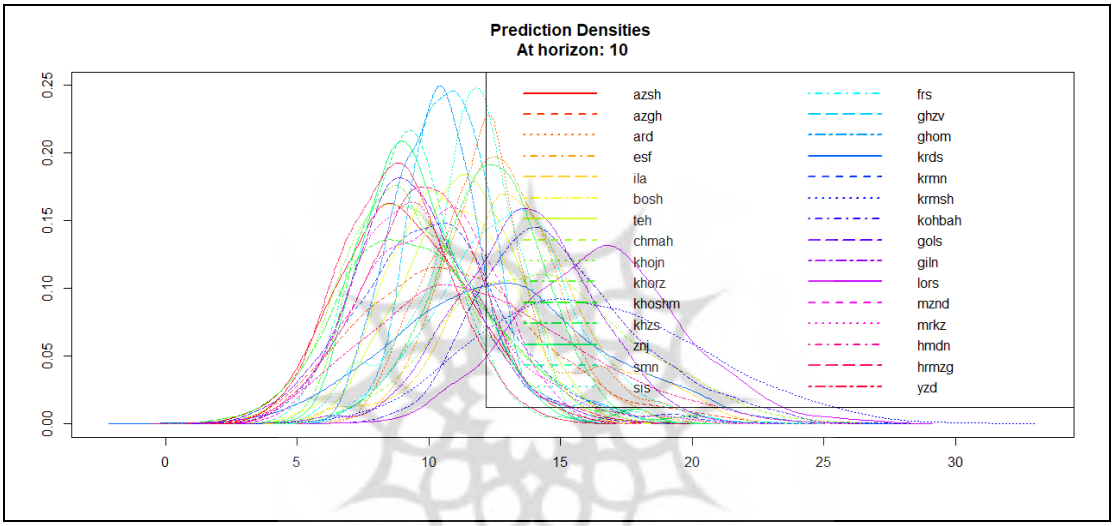


نمودار۲-الف) نمودار دندروگرام(درخت سلسله مراتبی) خوشه‌بندی نرخ بیکاری استانها در ۱۰ گام به جلو.- منبع: یافته‌های پژوهشگر، ۱۳۹۷.

در نمودار فوق محور عمودی بر فاصله اتصال یا ترکیب دلالت می‌کند. در هر برآمدگی نمودار، یعنی جایی که خوشه جدیدی تشکیل می‌شود، می‌توانیم فاصله ملاکی را ببینیم که در آن اجزای نماینده به هم متصل شده‌اند و یک خوشه منفرد را به وجود آورده‌اند.به طور کلی فاصله بین همه خوشه‌های آغازین محاسبه شده سپس دو تا از مشابه‌ترین خوشه‌ها ترکیب می‌شوند و فواصل دوباره محاسبه می‌شوند. این کار همچنان ادامه خواهد داشت تا زمانی که در نهایت همه در یک خوشه کلی قرار بگیرند.

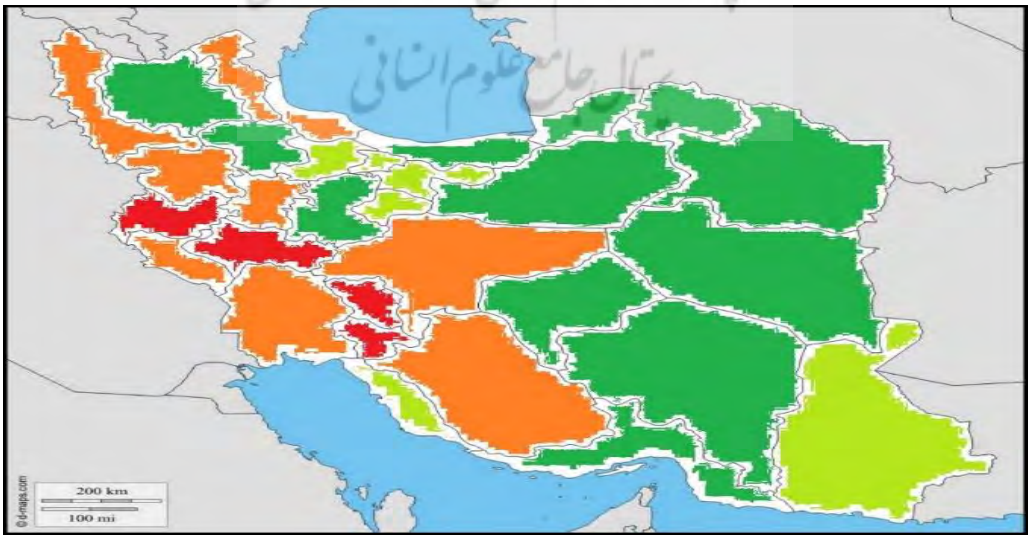
ب)از پیش بینی ۴ گام تا ۱۰ گام، تغییر عمده ای که اتفاق افتاده است؛ این است که استانهای آذربایجان غربی، همدان، کردستان، اردبیل، خوزستان و اصفهان از خوشه اصلی ۲ به خوشه اصلی ۱ منتقل شده‌اند. از آنجایی که وضعیت استانهای خوشه ۱ نسبت به خوشه ۲ بهتر است؛ بنابراین می‌توان گفت که این استانها در دوسال آینده وضعیت بهتری نسبت به یکسال آینده خواهند داشت.

ج) کرمانشاه و لرستان در هر دو افق زمانی ۴ و ۱۰ گام، بدترین وضعیت را دارا هستند. با توجه به اینکه، این دو استان در میانگین وضع موجود نیز همین وضعیت را داشته‌اند؛ لزوم توجه جدی به وضع اشتغال در این دو استان بیش از پیش دارای اهمیت است. لازم به ذکر است که چگالی پیش‌بینی تغییرپذیری پیش‌بینی‌ها را در خود لحاظ می‌کند به عبارت دیگر نتایج این خوشه‌بندی مطمئن‌تر از نتایج خوشه‌بندی مبتنی بر پیش‌بینی نقطه‌ای است.



نمودار۲-ب) نمودار چگالی‌های پیش‌بینی خوشه‌بندی نرخ بیکاری استانها در ۱۰ گام به جلو.- منبع: یافته‌های پژوهشگر، ۱۳۹۷.

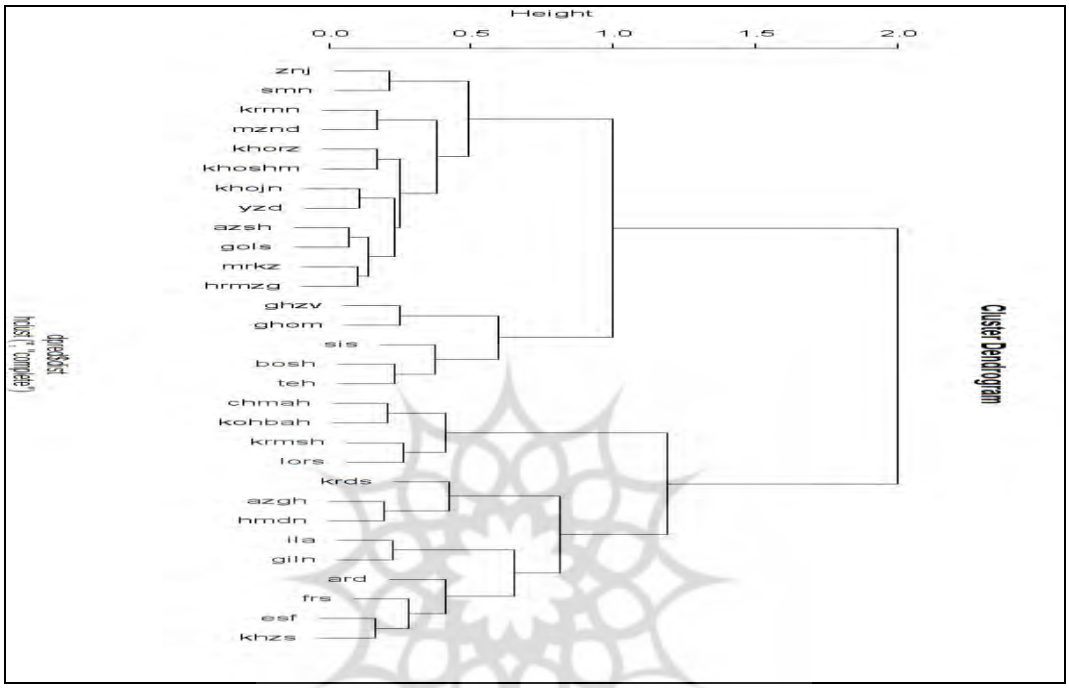
د) در وضعیت موجود (جدول ۱) استانهای آذربایجان شرقی و یزد کم‌ترین میانگین نرخ بیکاری در بین استانهای کشور را داشتند؛ اما در افق‌های زمانی مذکور اگر چه در خوشه ۱(استانهای با وضعیت مناسب) قرار گرفته اند، اما چگالی پیش‌بینی دو استان زنجان و سمنان وضعیت مناسب‌تری را نشان می‌دهد. به نظر سیاستهای اشتغال‌زایی در سالهای اخیر، نویدبخش وضعیت مطلوب برای این استانها است.



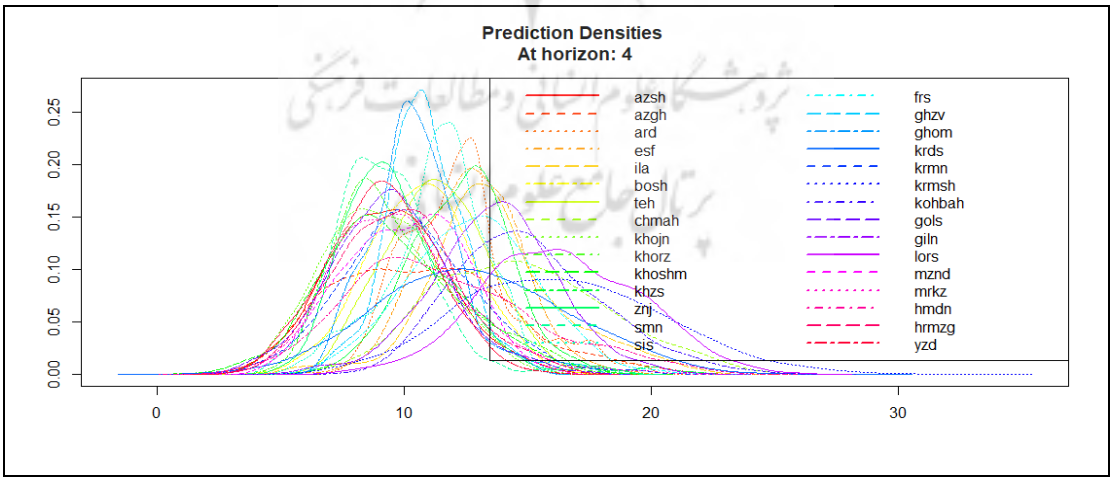
نمودار۲-ج) نمودار پراکنش فضایی بیکاری در استانهای کشور- منبع: یافته‌های پژوهشگر، ۱۳۹۷.

پراکنش فضایی بیکاری در ایران بر اساس خوشه‌بندی مبتنی بر چگالی پیش‌بینی در نمودار۲-ج ارائه شده است. رنگ قرمز و نارنجی استانهای با نرخ بیکاری بالا (تفاوت رنگها به علت زیرخوشه‌های متفاوت است) و رنگ سبز پررنگ و کم رنگ بیانگر نرخ بیکاری پایین در افق ۱۰ گام می‌باشد.

ه) طبق نتایج خوشه‌بندی همچنان استان‌های غرب و جنوب غربی بالاترین نرخ بیکاری را خواهند داشت. بنابراین لزوم برنامه‌ریزی منطقه‌ای و توجه جدی به اشتغال استان‌های فوق‌الذکر توصیه می‌شود. نتیجه جالب توجه در نمودار پراکنش فضایی، این است که استانهایی که در وضعیت نامطلوبی قرار دارند، دارای همسایگان با نرخ بیکاری بالا هستند و استانهای با نرخ بیکاری پایین عمدتا دارای استانهای همجوار با نرخ بیکاری پایین هستند. به عبارت دیگر یک همبستگی مکانی مثبت بین استانهای همجوار و نرخ بیکاری وجود دارد.



نمودار۳-الف) نمودار دندروگرام(درخت سلسله مراتبی) خوشه‌بندی نرخ بیکاری استانها در ٤ گام به جلو.- منبع: یافته‌های پژوهشگر، ۱۳۹۷.



نمودار۳-ب) نمودار چگالی‌های پیش‌بینی خوشه‌بندی نرخ بیکاری استانها در ٤ گام به جلو.- منبع: یافته‌های پژوهشگر، ۱۳۹۷.

نتیجه‌گیری و ارائه پیشنهادها:

در این پژوهش، الگوریتم جدید خوشه‌بندی مبتنی بر چگالی پیش‌بینی که نه جزء خوشه‌بندی مبتنی بر شکل و نه مبتنی بر ساختار سری‌های زمانی است، معرفی گردید. هدف، ارائه روش جدید خوشه‌بندی مبتنی بر چگالی پیش‌بینی و توجه برنامه‌ریزان منطقه‌ای به این روش حتی در سایر حوزه‌هایی همچون مسکن، جرم و جنایت و ... می‌باشد. این روش در سیاستگزاری عمومی و آینده‌پژوهی کاربرد خواهد

داشت. به عنوان یک نمونه از کاربردهای فراوان آن، سری زمانی نرخ بیکاری استان‌های کشور در دو افق زمانی ۴ فصل و ۱۰ فصل خوشه‌بندی گردید. برای انجام خوشه‌بندی سری زمانی مذکور از نرم‌افزار برنامه‌نویسی R و بسته نرم‌افزاری $TSclust$ که بر روی R قابل نصب خواهد بود، استفاده شد. نتایج نشان می‌دهد که بهترین وضعیت را چه در ۴ گام و چه در ۱۰ گام (دو سال و نیم)، استانهای سمنان و زنجان خواهند داشت و بدترین وضعیت استانهای لرستان و کرمانشاه دارا می‌باشند. همچنین در دو افق زمانی مورد مطالعه، به جز چند استان، بقیه در خوشه های اصلی خود ثابت بودند.

اگر چه وجود زیرساخت‌های مختلف در مناطق مختلف و تفاوت کارآمدی در حوزه‌های جغرافیایی متفاوت، باعث شده تا نیروی کار در کشور به صورت ناهمگن پخش شده باشد؛ اما این وضعیت در حدود سه سال آینده نیز اصلاح نخواهد شد. البته برای به دست آوردن نتایج مبسوط‌تر، نیازمند اتخاذ و به کارگیری انواع تکنیک‌های داده‌کاوی و مدل‌های طبقه‌بندی می‌باشد. همچنین نمودار پراکنش فضایی بیکاری در ایران بر اساس خوشه‌بندی مبتنی بر چگالی پیش‌بینی نشان می دهد که همچنان استان‌های غرب و جنوب غربی بالاترین نرخ بیکاری را خواهند داشت. بنابراین لزوم برنامه‌ریزی منطقه‌ای و توجه جدی به اشتغال استان‌های فوق‌الذکر توصیه می‌شود. در ضمن استانهایی که در وضعیت نامطلوبی قرار دارند، دارای همسایگان با نرخ بیکاری بالا هستند و استانهای با نرخ بیکاری پایین عمدتا دارای استانهای همجوار با نرخ بیکاری پایین هستند. به عبارت دیگر یک همبستگی مکانی مثبت بین استانهای همجوار و نرخ بیکاری وجود دارد. نتایج تحقیق رضوانی و همکاران (۱۳۹۲) که به تحلیل مکانی بیکاری در نواحی شهری و روستایی ایران در سال ۱۳۹۰ می‌پردازد از نرخ بالای بیکاری در نواحی جنوبی و غربی کشور حکایت می‌کند. همانطور که از نتایج تحقیق حاضر حاصل می‌گردد می‌توان انتظار داشت که روند نرخ بالای بیکاری در همان نواحی به استثنای استان‌های جنوب شرقی مانند سیستان و بلوچستان، کرمان و هرمزگان و همچنین بوشهر پابرجا بماند.

نتایج حاصل از این مطالعه، توجه به سیاست‌های اقتصادی استان محور را برای کاهش نابرابری‌های اشتغال توصیه می‌کند. به سیاستگزاران اقتصادی پیشنهاد می‌شود، در تدوین برنامه‌های اقتصادی میان‌مدت و بلندمدت به ویژه برنامه‌هایی در جهت رفع معضل بیکاری به نتایج این نوع از خوشه بندی توجه نمایند. همچنین پیشنهاد می‌شود سایر متغیرهای کلان اقتصادی و سری‌های زمانی اقتصادی نیز با این روش جدید خوشه‌بندی شوند تا افق روشن‌تری در سیاستگزاری اقتصادی ایجاد شده باشد. بعلاوه، همانگونه که نتایج تحقیقات نوغانی‌دخت‌بهمنی و میرمحمدتبار (۱۳۹۴) *(Noghani Dokht Bahmani and Mir Mohamad Tabar, 2015)* و فیض‌پور و لطفی (۱۳۹۴) *(Feizpour and Lotfi, 2015)* و حسینی و همکاران (۱۳۹۷) نشان می‌دهند بیکاری تاثیر بسیار زیادی مهاجرت، جرم و مساله خودکشی می‌گذارد. لذا پرداختن به حل مشکل نرخ بالای بیکاری در افق ۴ یا ۱۰ گام آینده در استان‌هایی که با این مشکل درگیر خواهند شد می‌تواند تاثیر این چنین پدیده‌هایی را خنثی یا به حداقل برساند.

References:

1. *Abbasi Nejad H, Ramezani H, Sadeghi M.(2013). A Study of the Relation between Unemployment and Crime in Iran Provincial Consolidated Data Approach. 3. 20 (64) :65-86 (In Persian).*
2. *Aghabozorgi, S. Shirkhorshidi, A.S. and Wah, T.Y. (2015). Time-series clustering – A decade review. Information Systems, Vol. 53, pp. 16-38.*
3. *Akhbari, R., Taee, H. (2017). Identifying hysteresis effect in unemployment rate with emphasis on second generation panel unit root and PANIC method. journal of applied economics studies in iran, 6(22), 1-31. doi: 10.22084/aes.2017.11136.2209 (In Persian).*
4. *Alonso, A.M. Berrendero J.R. Hernandez A. and Justel A. (2006): Time Series Clustering Based on Forecast Densities. Computational Statistics and Data Analysis, Vol. 51, pp. 762-776.*
5. *Cowpertwait, P.S.P. Cox, T.F. (1992): Clustering population means under heterogeneity of variance with an application to a rainfall time series problem. The Statistician, Vol. 41, pp. 113–121.*
6. *Cracolici, M.F. Cuffaro, M. and Nijkamp, P. (2007): A spatial analysis on Italian unemployment defferences. Statistical Methods and Applications, Vol. 18, Issue 2, pp. 275-291.*
7. *Feizpour, M., Lotfi, E. (2015). Economical Distinctions and Social Problems of Iran: Rates of Unemployment and Suicide. strategic rssearch on social problems in iran university of isfahan, 4(1), 153-166 (In Persian).*

8. *Filiztekin, A. (2009): Regional unemployment in Turkey, Regional Science, Vol. 88, Issue 4, pp. 863-878.*
9. *Fruhwirth-Schnatter, S., Kaufmann, S. (2004): Model-based clustering of multiple time series. CEPR Discussion Paper No. 4650.*
10. *Galeano, P. and Peña, D. (2000): Multivariate analysis in vector time series, Resenhas, Vol. 4, pp. 383–403.*
11. *Garcilazo, J. E. (2006): Regional unemployment clusters: neighborhood and state effects in Europe and North America, The Review of Regional Studies, Vol. 37, Issue 3, pp. 282-302.*
12. *Gharavi Nakhjavani, A. (2002): "The Unemployment Crisis in the Iranian Economy", Economic Quarterly, No. 6, 171-184 (In Persian).*
13. *Hosseini, G., Sadeqi, R., Ghasemi, A., Rostamalizadeh, V. (2018). Trends and Patterns of Internal Migration in Iran. Journal of Regional Planning , 8(31), 1-18 (In Persian).*
14. *Imani, M. (2012): Clustering of Persian texts, Master's thesis, Faculty of Computer Engineering, Sharif University of Technology. (In Persian).*
15. *Iran Statistical Center, "Employment and Unemployment Indicators of the country in 1376-1391", Tehran (*In Persian*).*
16. *Izadparast, M. (2011): Classification of Insurance Customers Using Data Mining, Payame Noor University of Tehran, New World Insurance Monthly, No. 161.(In Persian).*
17. *Kakizawa, Y. Shumway, R.H. and Taniguchi, M. (1998): Discrimination and clustering for multivariate time series, Journal of the American Statistical Association, Vol. 93, pp. 328–340.*
18. *Kaufman, L. and Rousseeuw, P.J. (2005): Finding groups in data: an imtroduction to cluster analysis, Hoboken, NJ: Wiley.*
19. *Liao, W. T. (2005): Clustering of time series data-a survey. Pattern recognition, Vol. 38, Issue 11, pp. 1857-1874.*
20. *Lopez-Bazo, E. and Motellon, E. (2011): The regional distributions of unemployment. What do micro-data tell us?, Research Institute of Applied Economics.*
21. *Macchiato, M.F. La Rotonda, L. Lapenna,V. and Ragosta, M. (1995): Time modelling and spatial clustering of daily ambient temperature: an application in Southern Italy, Environmetrics, Vol. 6, pp. 31–53.*
22. *Maharaj, E.A. (1996): A significance test for classifying ARMA models, Journal of Statistical Computation and Simulation, Vol. 54, pp. 305–331.*
23. *Montero P. and Vilar, J.A. (2015): TSclust: An R Package for Time Series Clustering, Journal of Statistical Software, Vol. 62, Issue 1, pp. 1-43.*
24. *Motii-Haghshenas, N. (2002): "A Comparative Study of the Employment and Unemployment Rate of Population in Different Provinces of Iran", First Conference of Iranian Demographic Association, Tehran, Iranian Demographic Association (In Persian).*
25. *Noghani Dokht Bahmani, M., Mir Mohamad Tabar, S. (2016). Study of Economic Factors Affecting the Crime (Meta-analysis of Research Conducted in Iran). strategic rssearch on social problems in iran university of isfahan, 4(3), 85-102 (In Persian).*
26. *Pattarin, F. Paterlini, S. and Minerva, T. (2004): Clustering financial time series: an application to mutual funds style analysis, Computational Statistics and Data Analysis, Vol. 47, pp. 353–372.*
27. *Piccolo, D., (1990). A distance measure for classifying ARIMA models. J. Time Ser. Anal. 11, 153–164.*
28. *Radmehr, F. and Alamolhoda, S. H. (2014). Cluster analysis: a tool for analyzing data in quantitative and mixed method studies. psychological methods and models, fourth year, No. 15, pp. 36-13. (In Persian).*
29. *Rezvani, M., Mansourian, H., Mahmoudian zamaneh, M., Heydarian Mohammadabadi, R. (2013). Spatial analysis of unemployment in Urban and Rural Areas in Iran With*

exploratory Spatial Data Analysis Approach. physical sacial planning, 1(3), 37-48 (In Persian).

30. *Saadat, M., Mila Elmi, Z. and Akbari, N. (2009). Analysis of Spatial Unemployment in Iran, Nameye Mofid,19(69), 151-170 (In Persian).*
31. *Seidaii, S., Bahari, E. and Zarei, A.(2011): "Investigation of the Employment and Unemployment Status in Iran during the Years 1335-1389". Jasmine Strategy, No. 25, 241-216 (In Persian).*
32. *Shahyadi, S (2004): "Spatial Analysis and Labor Demand in Iran", MSc Thesis, Faculty of Administrative Sciences and Economics, University of Isfahan (In Persian).*
33. *Tong, H. Yeung, I. (2000): On Tests for Self-Exciting Threshold Autoregressive-Type Non- Linearity in Partially Observed Time Series, Journal of the Royal Statistical Society C, Vol. 40, Issue 1, pp. 43-62.*
34. *Vilar, J.A. Alonso, A.M. and Vilar, J.M. (2010): Non-Linear Time Series Clustering Based on Non-Parametric Forecast Densities, Computational Statistics and Data Analysis, Vol. 54, Issue 11, pp. 2850-2865.*

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

Journal of Regional Planning

ISSN (Print): 2251-6735 Vol. 10/ No. 37/ Spring 2020
ISSN (Online): 2423-7051

Research Paper

Clustering Based on Forecasting Density: Case Study of Unemployment Rate in Iran's Provinces

Ramin Khochiani*[1]*: *Assistant Professor of Economics, Ayatollah Borujerdi University, Borujerd, Iran*
Seyed Mohammad Hosseini: *Assistant Professor of Mathematics, Ayatollah Borujerdi University, Borujerd, Iran*

Received: 2018/11/11 *PP: 13- 16* *Accepted:2019/4/26*

Abstract

It is important for regional planners and policymakers to be aware of the unemployment rate of the provinces in specified time horizons. In this paper, clustering of time series based on their forecasting density to a specified horizon is investigated. In this algorithm, we use the bootstrap process to approximate the distribution of predictions. The differences between each pair of bootstrap densities generate a dissimilarity matrix that is used for clustering. For this purpose, seasonal unemployment data was used in the spring of 2005 to fall of 2017, and according to the forecasting density algorithm, we will cluster the unemployment rate of Iran's provinces for two horizons of 4 steps (one year) and 10 steps (two and a half years). The best situation will be in the 4 steps or 10 steps (two and a half years), the provinces of Semnan and Zanjan, and the worst situation in the provinces of Lorestan and Kermanshah. Also, in the two horizons studied, except for some provinces, the rest were fixed in their main clusters. The spatial distribution of unemployment in Iran, based on forecasting density clustering, shows that western and southwestern provinces will have the highest unemployment rates. Therefore, the need for regional planning and serious attention to the employment of these provinces is recommended. At the same time, the provinces that are in an unfavorable situation have high unemployment neighbors, and the provinces with low unemployment rate have predominantly neighborhoods with a low unemployment rate. In other words, there is a positive spatial correlation between the neighboring provinces and the unemployment rate.

Key words: *Forecast Density Clustering, Time Series, Sieve Bootstrap, Unemployment Rate.*

Extended Abstract

Introduction:

Managing raw data and extracting useful information plays an important role in decision making. Clustering as one of the descriptive data mining methods is followed by organizing the data into a number of clusters in such a way that objects in the same cluster are more similar to each other than to those in other clusters. Alonso et al. (2006) proposed a concept of dissimilarity measure based on the forecast densities, for each one of the observed series in the sample for a given future horizon. They combined a smoothed sieve bootstrap procedure with nonparametric kernel density estimation ideas to approximate the distribution of the predictions. Villar et al. (2010) developed this method and also covered nonparametric nonlinear autoregressive models.

[1]- *Corresponding Author's Email: khochiany@abru.ac.ir, Tel: +989133728231*

In this paper, both feature-based and model-based approaches are used to cluster time series data. The main purpose of this study is to cluster time series data based on complete prediction densities for each series in the set, rather than focusing on point predictions. Here, time series clustering is performed based on full forecast densities. Time series fall into a cluster in a forecast density distribution in the future time horizon with other similar time series in the same future time horizon. Clustering and dissimilarity based on the forecast densities can be easily interpreted. Unemployment is a good measure for the state of the balance between the key pillars of the country's economy. Therefore, it is important for the authorities to address it. In this study, in order to show the efficiency of the mentioned clustering method, the problem of unemployment rate clustering in the provinces of Iran was considered and the provinces were clustered in terms of unemployment rate in the next four and ten steps horizon.

Methodology:

In this paper, clustering is performed based on full forecast densities instead of focusing on point predictions. Suppose X_T and Y_T are two stationary processes with the following autoregressive representation:

$S_t = \varphi(S_{t-1}) + \varepsilon_t$ *Where ε_t is a white noise and $\varphi(.)$ is a smooth function which is not constrained by any predetermined parametric model. At a specified future time T+h, let*

$$d_{PRED,h}(X_T, Y_T) = \int \left|\hat{f}_{X_{T+h}}(u) - \hat{f}_{Y_{T+h}}(u)\right| du$$

where $\hat{f}_{X_{T+h}}$ and $\hat{f}_{Y_{T+h}}$ denote the density function of the forecasts X_{T+h} and Y_{T+h}, respectively. $d_{PRED,h}(X_T, Y_T)$ denotes the distance between X_T and Y_T at the specified future time T+h. The correct forecast densities are replaced by kernel (nonparametric) estimates based on bootstrap predictions.

Results and discussions:

According to the proposed algorithm, the unemployment rate of the Iran's provinces will be clustered for two time horizons of 4 seasons and 10 seasons. Seasonal data of provincial unemployment rate during spring 2005 to autumn 2017 were extracted from the statistics center. Some of the results are:

Clustering for 4 future seasons:

- *cluster 1) Zanjan, Semnan, Kerman, Mazandaran, Khorasan Razavi, North Khorasan, South Khorasan, Yazd, East Azarbaijan, Golestan, Markazi, Hormozgan, Qazvin, Qom, Sistan and Baluchestan, Bushehr and Tehran.*
- *cluster 2) Chaharmahal & Bakhtiari, Kohkiluyeh & Boyerahmad, Kermanshah, Lorestan, Kurdistan, West Azarbaijan, Hamedan, Ilam, Gilan, Ardabil, Fars, Isfahan, Khuzestan.*

Forecasting at 10 future seasons, the provinces of West Azerbaijan, Hamedan, Kurdistan, Ardebil, Khuzestan and Isfahan were moved from main cluster 2 to main cluster 1. Since the situation in the provinces of cluster 1 is better than cluster 2, it can be said that these provinces will have better situation in the next two years compared to the next one. Kermanshah and Lorestan have the worst situation in both 4 and 10 time horizons.

In the current situation, the provinces of East Azerbaijan and Yazd have the lowest average unemployment rate among the provinces of the country. However, in the above time horizons, although they are located in cluster 1 (provinces with good status), the prediction density of Zanjan and Semnan proves better. According to the results of the clustering, west and southwestern provinces will also have the highest unemployment rate.

Conclusion:

A new density-based clustering method is represented for forecasting. The time series of unemployment rates of the provinces were clustered in two time horizons of 4 seasons and 10 seasons. The results show that Semnan and Zanjan provinces will have the best situation in 4 or 10 steps and

the worst situation will be in Lorestan and Kermanshah provinces. Also in the two time horizons studied, with the exception of a few provinces, the rest were fixed in their original clusters.
The spatial distribution of unemployment in Iran based on the forecast density-based clustering shows that the western and southwestern provinces will still have the highest unemployment rate. There is also a positive spatial correlation between the neighboring provinces and the unemployment rate. The results of the research of Razvani et al (2013) indicate high unemployment rate in the south and west of the country. As can be seen from the results of the present study, one can expect the high unemployment rate to remain in the same areas with the exception of southeastern provinces such as Sistan and Baluchestan, Kerman and Hormozgan as well as Bushehr.

The results of this study recommend attention to provincial-based economic policies to reduce employment inequalities. It is also suggested that other macroeconomic variables be clustered with this new approach to provide a clearer horizon in economic policy making. In addition, the literature results show that the unemployment has a huge impact on immigration, crime and suicide. Therefore, addressing the problem of high unemployment rates in the next 4 or 10 time horizons in the provinces involved can minimize the impact of such phenomena.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی