فصلنامه نظریه‌های کاربردی اقتصاد/ سال نهم/ شماره ۴/ زمستان ۱۴۰۱/ صفحات ۳۰۶-۲۷۷

تحلیل رفتار اضافه‌کاری در بازار کار ایران با رویکرد درخت تصمیم[۱]

سمیه سادات نقوی

دانشجوی دکترای اقتصاد دانشگاه فردوسی مشهد، somayenaghavi@gmail.com

محمود هوشمند*

استاد اقتصاد دانشگاه فردوسی مشهد، m-hoshmand@um.ac.ir

سید سعید ملک الساداتی

استادیار اقتصاد دانشگاه فردوسی مشهد، msadati@um.ac.ir

تاریخ دریافت: ۱۴۰۱/۰۶/۲۹ تاریخ پذیرش: ۱۴۰۱/۱۲/۱۳

چکیده

درک رفتار نیروی‌کار در مواردی که مربوط به تصمیم افراد به انجام اضافه‌کاری است، در ارزیابی دقیق اثرات سیاست‌های اشتغال‌زایی، از اهمیت اساسی برخوردار می‌باشد؛ با توجه به اینکه موضوع اضافه‌کاری در تحلیل‌های مربوط به منابع انسانی و نظام بازارکار در اقتصاد کلان، تا حدودی نادیده گرفته شده است، این مقاله به بررسی این موضوع از طریق تحلیل رفتار نیروی‌کار ایران با استفاده از آمار هزینه و درآمد خانوارهای شهری برای دوره زمانی ۹۹-۱۳۸۴ پرداخته و عواملی که بر تصمیم‌گیری فرد در خصوص اضافه‌کاری تأثیرگذار هستند را مورد مطالعه و بررسی قرار می‌دهد. روش مورد استفاده در این مطالعه، الگوریتم داده‌کاوی درخت تصمیم می‌باشد که به ما این امکان را می‌دهد تا با پی‌بردن به توزیع زیربنایی داده‌های اخذ شده از مجموعه افراد مورد بررسی، بتوانیم رفتار آن‌ها را مطالعه کرده و در مورد آن استنتاج منطقی و مفید داشته باشیم. نتایج حاکی از این است که اضافه‌کاری به سختی با ویژگی‌های سطح فردی مانند سن، تحصیلات، جنسیت، نگرش‌های کاری و یا هر عامل غیرقابل مشاهده‌ای که این متغیرها نماینده آن هستند، تعریف می‌شود، اما می‌توان آن را به وضعیت شغلی و مهارتی که نیروی‌کار دارد، ویژگی‌های ساختاری بازار کار و همین‌طور دهک هزینه‌ای که خانوار در آن قرار دارد نسبت داد. این نتایج می‌تواند پیامدهای قابل توجهی هم برای افراد و هم برای سیاست‌گذاران، به جهت تخصیص بهینه نیروی‌کار داشته باشد.

واژه‌های کلیدی: اضافه‌کاری، بازار کار، رفتار نیروی‌کار، مدل یادگیری ماشین (ML)، درخت تصمیم

طبقه‌بندی JEL: M54، J82، J30، J22

---

[۱] این مقاله برگرفته از رساله دکترای نویسنده اول می‌باشد.

* نویسنده مسئول

۱-مقدمه

مطالعه مباحث مربوط به مدیریت بازار کار در ایران، بیانگر وجود شکاف تحقیقاتی در زمینه مدیریت اضافه‌کاری[1] و موضوعات مربوط به الگوی حاکم بر تصمیم‌گیری نیروی‌کار در خصوص اضافه‌کاری در اقتصاد ایران است. بر این اساس اضافه‌کاری در تحلیل‌های مربوط به منابع انسانی و نظام بازار کار در اقتصاد کلان، نه تنها به کار گرفته نشده، بلکه تا حدودی نیز نادیده گرفته شده است. با توجه به اینکه تغییر رویکرد به استفاده از اضافه‌کاری موجود، به جای بکارگیری نیروی‌کار جدید در شرایط خاص می‌تواند گزینه اجتناب‌ناپذیر بازار کار در ایران باشد، انجام این دسته مطالعات در بخش عرضه نیروی‌کار ضرورت دارد. از طرفی دیگر، دستیابی به بیکاری پایین توأم با رشد ضعیف تولید، یک چالش مهم سیاستی است، لذا در بخش عرضه، توزیع بیشتر و بهینه ساعات کار نیروی‌کار می‌تواند راه‌حلی مفید باشد؛ یکی از بهترین رویکردهای ممکن برای نیل به این امر، بررسی کامل و ساماندهی مجدد مدت زمان کار و بررسی کامل جوانب این امر است؛ بر این اساس اطلاع از اضافه‌کاری و شناخت کاربردهای آن در بازار کار، به بازبینی این موضوع می‌تواند کمک کند.

به طور معمول بخش عرضه نیروی‌کار به مراتب کمتر از طرف تقاضا مورد مطالعه قرار گرفته و به همین نسبت با محدودیت مبانی نظری مواجه است؛ با این وجود، بدیهی است که اشتغال و بیکاری هم از طریق عرضه و هم تقاضای کار تحت تاثیر قرار می‌گیرند. عمدتاً دولت و به تبع آن کارفرمایان دولتی، به‌ویژه در بخش خدمات، به دنبال ایجاد بسترهای اشتغال‌زایی به منظور بکارگیری نیروی‌کار بیکار هستند، در حالی‌که کارایی و بهره‌وری نیروی‌کار مغفول واقع می‌شود، به این شکل که هرچند ممکن است اشتغال بیشتر در بخش خدمات عارضه جدی ایجاد نکند، اما در سایر بخش‌ها به‌ویژه صنعت، بکارگیری فراتر از حد بهینه نیروی‌کار با هر انگیزه‌ای که صورت پذیرد، اعم از اشتغال‌زایی و ...، قطعاً بهره‌وری را تحت تاثیر قرار خواهد داد و ورود نیروی‌کار جدید بدون نگاه به بهره‌گیری از حداکثر توان نیروی‌کار موجود، باعث افزایش هزینه بنگاه و نیروی‌کار شامل آموزش، بیمه، مالیات و سایر هزینه‌ها خواهد شد. بنابراین، با هدف ارزیابی دقیق اثرات سیاست‌های

[1] Overtime Working

اشتغال‌زایی، درک رفتار نیروی‌کار در تصمیم‌گیری به انجام اضافه‌کاری، از اهمیت قابل توجهی برخوردار است.

هدف اصلی این مقاله، تحلیل رفتار نیروی‌کار در موضوع اضافه‌کاری به منظور دستیابی به نتایج نظری مربوط به سمت عرضه نیروی‌کار است؛ اما با توجه به اینکه هرگونه اعمال سیاست در زمینه تسهیم کار نیاز به شناخت الگوی حاکم بر تصمیم‌گیری فرد در خصوص انجام اضافه‌کاری و درک عوامل موثر بر آن در خانوارهای با سطوح متفاوت هزینه و درآمد دارد، نتایج این مطالعه می‌تواند به سیاست‌گذاران در اتخاذ سیاست‌های مناسب برای مدیریت اضافه‌کاری در بازار کار کمک نماید. این مقاله بر تحلیل رفتار نیروی‌کار تمرکز دارد و برای این امر، از یک الگوریتم یادگیری ماشین برای جمع‌آوری نتایج جدید از یک مجموعه داده در مقیاس بزرگ استفاده شده است.

در این زمینه لازم به توضیح است، با توجه به اینکه اشتغال امری پیچیده و چندوجهی است، از این رو سیاست‌های معطوف به اشتغال نیز از گستره وسیعی برخوردار هستند. برخی رویکردهای تجربه شده با محور قراردادن سیاست‌های رشد اقتصادی، موضوع اشتغال را بررسی نموده‌اند؛ اما تجارب کشورهایی که بحران‌های اقتصادی را سپری کرده‌اند، منجر به تغییر این نوع نگرش شده است. به گونه‌ای که در دهه‌های گذشته، گرایش چشم‌گیری به سیاست‌های ایجاد اشتغال از طریق توزیع فرصت‌های شغلی شکل گرفته که عموماً از این سیاست‌ها با عنوان سیاست‌های تسهیم کار[1] یا گسترش کار[2] یاد می‌شود. این سیاست‌ها، سیاست‌هایی هستند که به بازتوزیع فرصت‌های شغلی موجود از طریق محدودسازی ساعات کاری، به ویژه ساعات اضافه‌کاری[3]، به روش‌های گوناگون می‌پردازند که یک عامل پنهان در بازار کار می‌باشد. در نظریه‌های بازار کار استدلال می‌شود که این توزیع دوباره، با کاستن از میزان پرکاری[4] و اضافه‌کاری، سبب ایجاد فرصت‌های شغلی بیشتر برای بیکاران می‌شود.

با این‌حال، اضافه‌کاری هم از سمت تقاضا و هم از سمت عرضه نیروی‌کار دارای مطلوبیت است. از سمت عرضه حائز اهمیت است چون افراد معمولاً برای کسب درآمد بیشتر و

---

[1] Work Sharing
[2] Spreading the Work
[3] Overtime Hours
[4] Overemployment

جبران شکاف بین درآمدها و هزینه‌ها، اقدام به اضافه‌کاری می‌کنند و در وضعیت فعلی اقتصاد ایران به دلیل عدم همخوانی سطح درآمد افراد با تورم، انگیزه انجام کار اضافه و اضافه‌کاری وجود دارد و این تقاضا در بیشتر بخش‌ها دیده می‌شود.

از سمت تقاضای نیروی‌کار نیز، هم دولت و هم بخش خصوصی -با درصدی بالاتر از بخش دولتی- تمایل دارند تقاضای کار خود را با به‌کارگیری نیروی‌کار موجود، از طریق انجام اضافه‌کاری و پرداخت دستمزد آن تأمین کنند، زیرا هزینه‌های سربار کمتری دارد. حال تقاضاکننده کار باید تصمیم بگیرد که آیا برای تأمین نیروی‌کار خود، از استخدام نیروی‌کار جدید استفاده کند و یا از ظرفیت نیروی‌کار موجود استفاده کرده و ساعات بیشتری آن‌ها را به‌کار گیرد؛ اما، این حقیقت وجود دارد که در شوک‌های طرف تقاضا، نادیده گرفتن رفتار متقابل سمت عرضه کار می‌تواند به طور جدی گمراه‌کننده باشد. ساعات کارکرد نیروی‌کار در سمت عرضه، می‌تواند نسبتاً سریع و انعطاف‌پذیر تغییر کند و این به شدت به اشتغال، بهره‌وری نیروی‌کار و نتایج بیکاری مربوط می‌شود. از طرفی هزینه‌های کوتاه‌مدت برای تنظیم اشتغال نیز نسبتاً زیاد است، بنابراین تغییر در میزان ساعات کاری باعث می‌شود اقتصادهای ملی با سرعت بیشتری نسبت به شوک‌های تقاضا واکنش نشان دهند.

با توجه به اهمیت ساعات کاری در تصمیم‌گیری‌های سمت عرضه و تقاضای کار، در این مقاله درصدد هستیم به مطالعه رفتار نیروی‌کار ایران در سمت عرضه کار بپردازیم؛ تمرکز این پژوهش، بر روی ساعات کار نیروی‌کار شاغل و به طور خاص ساعات اضافه‌کاری در بازار کار ایران است. از جمله عوامل تأثیرگذار بر اضافه‌کاری، وضعیت معیشت و مخارج خانوار می‌باشد و این مقاله در نظر دارد با بررسی هزینه‌های خانوار، الگوی مناسبی را برای تحلیل و کنترل این رفتار ارائه دهد؛ از طرفی ترکیب تفصیلی مشخصات خانوار را بررسی کرده و عواملی مانند سن، جنس، تحصیلات و تأهل و ارتباط آن با میزان اضافه‌کاری افراد را تحلیل می‌کنیم. همچنین به بررسی وضعیت درآمدی خانوارها و افراد پرداخته و درآمدهای غیرشغلی فرد، میزان دستمزد ساعات کار عادی و دستمزد ساعات کار اضافه‌کاری و سایر عوامل که بر روی میزان ساعات اضافه‌کاری فرد، تأثیر می‌گذارند را مورد مطالعه قرار خواهیم داد. علاوه بر این‌ها، ازآن‌جا که انجام اضافه‌کاری در میان گروه‌های مختلف هزینه‌ای با انگیزه‌های متفاوت صورت می‌گیرد، به بررسی رفتار

اضافه‌کاری در میان دهک‌های مختلف هزینه‌ای نیز پرداخته خواهد شد. همچنین نیروی-کار در شغلی اضافه‌کاری انجام می‌دهد که در آن مهارت دارد، لذا از متغیرهای کد شغلی و کد فعالیت نیز به عنوان متغیر سمت عرضه کار در مدل استفاده شده است.

در اینجا، چند سؤال از اهمیت اساسی برخوردار است. اول این که، عوامل تعیین‌کننده پدیده اضافه‌کاری چیست؟ چه عواملی بر رفتار اضافه‌کاری نیروی‌کار تأثیرگذار است؟ آیا عوامل تاثیرگذار بر رفتار نیروی‌کار در گروه‌های مختلف کاری از یکدیگر متمایز است؟ رفتار اضافه‌کاری نیروی‌کار در دهک‌های مختلف هزینه‌ای به چه صورت می‌باشد؟

انتظار می‌رود این مقاله بتواند با پاسخگویی به این سوالات به شناخت بهتر رفتار نیروی‌کار ایران برسد و نتایج این مطالعه بتواند به سیاست‌گذاران در اتخاذ سیاست‌های مناسب برای مدیریت اضافه‌کاری در بازار کار کمک نماید.

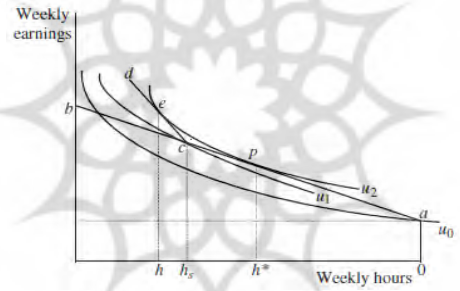
۲-ادبیات تحقیق

زمان کار از ابتدای پیدایش نظریه علمی اقتصاد موضوعی بسیار مورد بحث بوده است و امروزه همچنان موضوعی است که توجه دانشمندان علوم اجتماعی را به خود جلب می کند. از اواخر قرن نوزدهم تا آغاز قرن بیستم، موضوع اصلی نظریه‌های موجود در این بحث، این بود که آیا کاهش ساعت‌های طولانی کار تاکنون برای اقتصاد مفید بوده است یا خیر. اقتصاد سیاسی کلاسیک مبتنی بر نظریه ارزش کار است و توضیح می‌دهد که به چه طریقی همه واحدهای اقتصادی مستقل برای تولید و مبادله با یکدیگر متحد و هماهنگ می‌شوند؛ علاوه بر این، عقلانیت و لذت‌گرایی به عنوان ویژگی‌های انسان اقتصادی در نظر گرفته می‌شوند. عقلانیت ترجیحات کارگران را نیز توصیف می‌کند. اوقات فراغت یک کالای قابل مبادله برای کارگران است، درآمد نیز کالایی قابل مبادله است؛ ترجیحات کارگران بین اوقات فراغت و درآمد متعادل می‌شود که رابطه رقابتی با یکدیگر دارند. به همین دلیل، فراز و نشیب‌های زمان کار به درآمد یا اثرات جایگزینی برای کارگران نسبت داده می‌شود (پاپاگیناکی[1]، ۲۰۱۴).

تئوری استاندارد بازار کار در مورد اضافه‌کاری به پرلمن[2] (۱۹۶۶) برمی‌گردد (آنکسو و کارلسون[3]، ۲۰۱۹). در مدل نئوکلاسیک عرضه ساعات کاری، مطلوبیت فرد تحت قید

[1] Papagiannaki
[2] Perlman
[3] Anxo & Karlsson

محدودیت بودجه حداکثر می‌شود. تابع مطلوبیت نیز از تلفیق مصرف و فراغت، تحت محدودیت بودجه‌ای که شامل ترکیب دستمزد پولی و غیرپولی حقیقی است استخراج می‌شود. در ابتدا فرض می‌شود هیچ نوع مقررات اضافه‌کاری (شامل ساعات کاری استاندارد و به تبع آن ساعات اضافه‌کاری) وجود ندارد. با فرض وجود یک دستمزد ساعتی ثابت، محدودیت بودجه در شکل ۱ با خط $o-a-b$ نشان داده شده و منحنی بی‌تفاوتی u_o نیز نمایش‌دهنده تمامی ترکیبات مصرف- فراغتی است که مطلوبیت یکسانی را به فرد می‌بخشد. در این شکل، نقطه p میزان بهینه مصرف- فراغت را نشان می‌دهد که در آن میزان مطلوبیت ساعت کاری اختیاری با h^* و میزان مطلوبیت متناظر آن با u_2 نشان داده شده است. در این حالت با فرض کار کردن فرد به میزان h^*، به طور مستقیم میزان ساعات فراغت نیز معین می‌شود. در حقیقت، انتخاب p یک را حل میانی برای عرضه‌کننده نیروی‌کار است که در انتخاب ساعت کاری خود، با هیچ قیدی روبرو نیست.



شکل (۱): اضافه‌کاری و پرکاری

منبع: هارت[1]، ۲۰۰۴

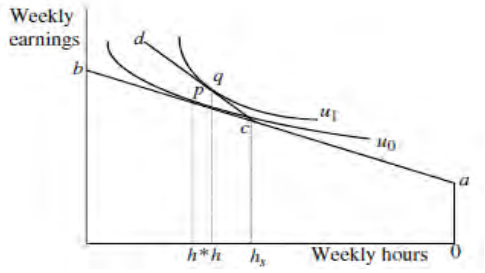
حال فرض می‌شود که اعمال قانون توسط دولت یا چانه‌زنی‌های اتحادیه‌های کارگری، منجر به اجرای مقررات مربوط به ساعت کاری استاندارد می‌شود که در نتیجه آن h_s، میزان ساعت کاری استاندارد (قانونی) تعریف شده است. در این حالت، فرد ملزم می‌شود از نقطه p به نقطه c حرکت کرده و به میزان $h_s > h^*$ کار کند. احتمالاً، فرد عرضه‌کننده کار مایل به قبول شرایط جدید و عرضه ساعت کار است زیرا میزان مطلوبیت u_1 متناظر با نقطه c، بیشتر از میزان مطلوبیت حالت جایگزین u_o است که در آن ساعت کاری فرد برابر با $h=o$ می‌باشد. در این شرایط، فرد به اصطلاح دچار "پرکاری[2]" شده است، بدین

[1] Hart
[2] Overemployment

معنا که ساعات کاری که موظف است کار کند از ساعات حداکثر کننده مطلوبیت وی بیشتر شده است. دو سناریو ممکن است اتفاق بیافتد؛ در وهله اول، کارگر ممکن است سعی کند تا حدودی موقعیت خود را با حرکت به سمت نقطه ترجیحی p از طریق غیبت[1] از محیط کار اصلاح کند؛ در حقیقت، هرچه شکاف $h_s - h^*$ بیشتر باشد، تمایل به غیبت، با توجه به کاهش مطلوبیت نهایی فراغت، بیشتر خواهد بود. این نشان می‌دهد ارتباط مثبتی بین تعداد ساعات استاندارد و غیبت وجود دارد. در وهله دوم، با فرض این‌که بنگاه می‌تواند نقش یک انحصارگر را ایفا کند، ممکن است درصدد جبران غیبت احتمالی و غیراخلاقی از طریق پیشنهاد دستمزد اضافه‌کاری برای ساعات کاری بیشتر از h_s باشد. در این حالت، دولت با اعمال مقررات اضافه‌کاری، بنگاه را ملزم به پرداخت دستمزد اضافه‌کاری برای ساعات کاری بیشتر از h_s می‌کند. در وضعیت جدید، افزایش شیب خط بودجه $c - d$، نشانگر این ضریب افزایش دستمزد اضافه‌کاری است و نقطه جدید e، رسیدن به تعادل قبلی را با ساعت کاری $h_s + (h - h_s)$ و مطلوبیت u_2 نشان می‌دهد. البته این مسئله می‌تواند منجر به نتایج متفاوتی شود و بسته به ترجیحات نیروی‌کار، ممکن است سبب افزایش یا کاهش ساعات اضافه‌کاری عرضه شده توسط آن‌ها شود.

یک سناریوی عرضه جایگزین در شکل ۲ نشان داده شده است. فرض کنید دولت حداقل دستمزد اضافه‌کاری را برای ساعت‌های بیشتر از h_s تعیین کند. نیروی‌کار در شکل ۲ از چنین قوانینی بهره‌مند خواهد شد، زیرا متوسط پرداخت ساعتی برای $h > h_s$ افزایش خواهد یافت. در سناریویی که به تصویر کشیده شده است، کارگر از نقطه p به q، با توجه به درآمد بیشتر نسبت به حالت جایگزین، جابه‌جا می‌شود و ساعت کاری h را ارائه خواهد داد که به میزان $h - h_s$ دستمزد اضافه‌کاری دریافت خواهد کرد؛ اگر بنگاه از پرداخت اضافه‌کاری تحت این شرایط خودداری کند، موظف است ساعات کاری h_s را در نقطه c ارائه دهد. این شرایط نیروی‌کار را "کم‌کار[2] "خواهد کرد که در آن ساعات کاری واقعی کمتر از ساعات مطلوب فرد می‌باشد. هزینه‌های با این نوع محدودیت‌ها ممکن است روحیه کم‌کاری همراه با جستجو برای شغل جدید را به همراه داشته باشد.

[1] Absenteeism
[2] Underemployed



شکل (۲): اضافه‌کاری و کم‌کاری

منبع: همان

حالت‌های مختلف گفته شده، عکس‌العمل‌های متفاوتی است که نیروی‌کار با توجه به ترجیحات خود در بازار ارائه می‌دهد و همین مسئله اهمیت انجام مطالعه تجربی را برای رسیدن به درک کامل‌تری از رفتار نیروی‌کار در بازار کار کشورمان، بیش از پیش نمایان می‌کند.

بخش چشم‌گیری از مطالعات تجربی، به تاثیر اعمال سیاست‌هایی بر ساعات اضافه‌کاری و همینطور دستمزد اضافه‌کاری و تاثیر آن بر اشتغال پرداخته و تحقیقات در مورد عوامل تمایل به اضافه‌کاری و تحلیل رفتار اضافه‌کاری نیروی‌کار به ندرت صورت گرفته است؛ با این حال با توجه به نوع روش پژوهش مورد استفاده در این مقاله، در این قسمت، ابتدا به بررسی مطالعات صورت گرفته در خصوص اعمال سیاست‌های بازار کار بر ساعات و دستمزد اضافه‌کاری پرداخته خواهد شد و در ادامه مطالعاتی که رفتار نیروی‌کار را با استفاده از مدل‌های یادگیری ماشین بررسی کرده‌اند ارائه خواهند گردید؛

بائر و زیمرمن[1] (۱۹۹۹) در مطالعه مربوط به کشور آلمان به بررسی عوامل موثر بر تمایل نیروی‌کار به انجام اضافه‌کاری پرداختند. طبق نتایج آنان سن، تاثیر مثبت بر اضافه‌کاری داشته، همچنین، افراد متاهل و کارکنان بنگاه‌های کوچک، درخواست بیشتری برای انجام اضافه‌کاری دارند. علاوه بر این، کارکنانی که با احتمال بیشتر در مواجهه با بیکاری هستند (عمدتاً افراد غیرماهر)، تمایل کمتر و در مقابل افراد ماهر، تقاضای بیشتری برای اضافه-کاری دارند.

[1] Bauer & Zimmermann

بل و هارت[1] (۲۰۰۳)، در بررسی روابط بین نرخ دستمزد ساعتی استاندارد، دستمزد اضافه‌کاری و مدت زمان اضافه‌کاری بریتانیا، دریافتند که با احتساب اضافه‌کاری، متوسط درآمد دستمزد ساعتی در میان بنگاه‌های یک صنعت معین، کاملاً یکنواخت است، زیرا بنگاه‌هایی که دستمزدهای خود را پایین‌تر از سطح دستمزدهای بازار پرداخت می‌کنند، تمایل دارند دستمزد اضافه‌کاری در سطح بالاتر از بازار را بپردازند و برعکس، بنگاه‌هایی که دستمزدهایی بالاتر از سطح بازار پرداخت می‌کنند، دستمزد اضافه‌کاری در سطح پایین‌تر از بازار را ارائه می‌دهند. رانجان[2] (۲۰۰۸)، در مقاله‌ای نقش داده‌کاوی[3] را در بهبود کیفیت فرایند تصمیم‌گیری در سیستم‌های مدیریت منابع انسانی نشان می‌دهد و با استفاده از مدل‌های طبقه‌بندی[4] در یادگیری ماشین نشان می‌دهد می‌توان پیش‌بینی کرد چه فعالیتی احتمالاً موفق خواهد بود و شرکت‌ها می‌توانند درباره عملکرد کارمندان (هم عملکرد گروهی و هم عملکرد فردی) تصمیم بگیرند و از این رو سیاست‌های خود را بهبود بخشند و استراتژی‌های جدیدی را برای برتری رقابتی تنظیم کنند. هارت و ما[5] (۲۰۱۰)، یک تئوری مبتنی بر قرارداد را برای توضیح تعیین ساعت‌های استاندارد، ساعات اضافه‌کاری و پرداخت دستمزد اضافه‌کاری ارائه می‌کنند. آن‌ها با تاکید بر عدم تقارن اطلاعات بین بنگاه و نیروی‌کار، ابتدا یک قرارداد ساعت- دستمزد ساده را بدون اضافه‌کاری بررسی کرده و نشان می‌دهند که درج ساعت‌ها در قرارداد می‌تواند خود باعث افزایش بهره‌وری شود، سپس نشان داده می‌شود که چگونه معرفی ساعت اضافه‌کاری، که با نرخ دستمزد پرداخت می‌شود، می‌تواند کارایی قرارداد را بیشتر بهبود بخشد. السلطانی[6] (۲۰۱۳)، برای پیش‌بینی نیازهای بازار کار، از رویکرد داده‌کاوی و پیاده‌سازی تکنیک‌های نایو بیزین[7] و درخت تصمیم[8] و قوانین تصمیم‌گیری[9] استفاده کرده است. نتایج نشان می‌دهد که سه روش داده‌کاوی یک روش عملی برای پیش‌بینی نیازهای بازار کار است و

[1] Bell & Hart
[2] Ranjan
[3] Data Mining
[4] The classification and prediction models
[5] Hart & Ma
[6] Alsultanny
[7] Naïve Bayes Classifiers
[8] Decision Trees
[9] Decision Rules

مقایسه بین سه روش نشان می‌دهد درخت تصمیم بالاترین دقت را دارد، همچنین نتیجه‌گیری می‌شود پیاده‌سازی روش‌های داده‌کاوی بر روی داده‌های منابع انسانی می‌تواند مقادیر زیادی داده را کنترل کند و آن را به اطلاعات مفید برای بهبود روند تصمیم‌گیری در سازمان‌ها تبدیل کند. گرونوف[1] (۲۰۱۶)، به مدل‌سازی اشتغال از طریق درختان طبقه‌بندی و رگرسیون پرداخته و نتیجه می‌گیرد اشتغال، از طریق متغیرهای سنتی جمعیت‌شناختی و نگرش کاری توضیح داده می‌شود که مهمترین آن‌ها سن، سال تولد و همچنین جنسیت، وضعیت تاهل و ارجحیت کارهای فکری در مقابل کارهای دستی است و این نتایج بسیار با ادبیات بازار کار و چرخه عمر اشتغال مطابقت دارد، اما بیکاری چنین نیست، به این معنی که بیکاری احتمالاً توسط عوامل دیگری از جمله ویژگی‌های ساختاری بازار کار و حتی تصادفی بودن هدایت می‌شود. ژانگ و همکاران[2] (۲۰۲۱)، با استفاده از یک الگوریتم خوشه‌بندی بر اساس الگوی تحقیق داده‌های بزرگ، به ارزیابی کمّی داده‌های بزرگ منابع انسانی پرداخته و بر اساس این بررسی سیستماتیک، یک چارچوب نظری کلی را ارائه می‌دهد که به عنوان الگوی پیش‌بینی (داده‌کاوی / تئوری‌سازی) در مقابل الگوی توضیحی (آزمایش نظریه) توصیف شده است.

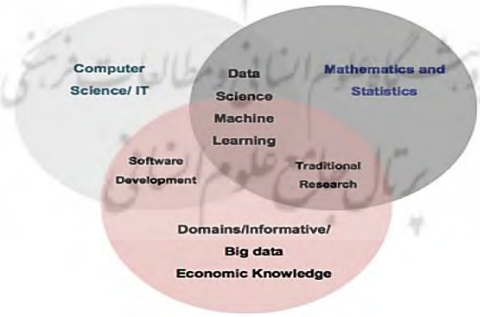
در ایران از جمله مطالعات سمت عرضه بازار کار می‌توان به مطالعه طائی[3] (۱۳۸۵) اشاره کرد که در آن به برآورد تابع عرضه نیروی کار مردان و زنان در مناطق شهری و روستایی برای دوره زمانی ۷۴-۱۳۷۱ با روش ترکیب داده‌های سری زمانی و مقطعی پرداخته است. نتایج این مطالعه نشان می‌دهد اثرگذارترین عامل بر روی تخصیص زمان بین ساعت کار بازاری و اوقات فراغت، متغیر درآمدهای کاری یا دستمزدها می باشد، همچنین یکی از مهمترین تفاوت‌های اساسی بین رفتار زنان و مردان در بازار کار مربوط به سطح تحصیلات است که باعث کاهش ساعات کاری مردان و افزایش ساعات کاری زنان در بازار کار می شود. کشاورز و خشابی[4] (۱۳۹۱) نیز، به مطالعه بررسی عوامل اصابت اضافه‌کاری و اعمال مقررات بر دستمزد اضافه‌کاری و تاثیر آن بر نرخ بیکاری پرداخته‌اند؛ نتیجه این مطالعه که برای داده‌های سال ۱۳۸۵ و با استفاده از روش هکمن دومرحله‌ای، توبیت و روش

[1] Gerunov
[2] Zhang et al.
[3] Taei (2006)
[4] Keshavarz & Khashabi (2012)

تصحیح خطای انتخاب مضاعف به‌کار برده شده، نشان می‌دهد اضافه‌کاری در ایران پدیده‌ای متمایل به سمت مردان بوده و با افزایش سن و تاهل میزان ساعت اضافه‌کاری افزایش می‌یابد؛ از طرفی، دستمزد اضافه‌کاری بر ساعات اضافه‌کاری انجام شده توسط نیروی‌کار اثر منفی داشته و این نکته، احتمال باز توزیع موفق فرصت‌های شغلی از طریق اعمال سیاست افزایش دستمزد اضافه‌کاری را تقویت می‌کند.

۳- روش‌شناسی تحقیق

بیش از صد سال است که اقتصاددانان تنها بر فرض تخمین سنتی ثابت فرض کردن متغیرهای غیرقابل مشاهده[1] متکی بوده‌اند؛ از طرفی، مشاهده و پیش‌بینی همزمان "داده‌های بزرگ" بسیار دشوار است و مطالعه بر روی محاسبات کلان داده اساساً به بیش از یک علم و یک نوآوری نیاز دارد. از نظر گرافیکی، این فرض تنها یکی از علوم مورد استفاده برای تجزیه و تحلیل داده‌ها است که در شکل ۵ نشان داده شده است. این امر کافی نیست و باعث می‌شود که بسیاری از محققان مجبور شوند برای تضمین این‌که مدل‌های تخمینی آن‌ها می‌توانند از نظر محاسباتی برای داده‌ها مناسب باشند، مفروضات بیشتری را قرار دهند. بر این اساس، برای رهایی از این مشکل، یادگیری ماشین (ML)[2] برای ترکیب کارآمد با تحلیل‌های کلان داده ابداع شده‌اند و این راه‌حلی برای شفاف‌سازی و پیش‌بینی روندها، نشانه‌های پنهان و بحران‌های اقتصاد در دوران مدرن تحقیقات دانشگاهی خواهد بود (چایبونسری و واناپان[3]، ۲۰۱۹).



شکل (۳): چارچوب مفهومی تحقیق

منبع: همان

[1] Non-Considerable Variables
[2] Machine Learning (ML)
[3] Chaiboonsri & Wannapan

الگوریتم‌های یادگیری ماشین می‌توانند بانظارت[۱] یا بدون نظارت[۲] باشند، اگرچه برخی از نویسندگان الگوریتم‌های دیگر را نیز به عنوان نیمه نظارتی[۳] و تقویتی[۴] نیز طبقه‌بندی می‌کنند، زیرا چنین تکنیک‌هایی داده‌ها را یاد می‌گیرند و الگو را با هدف واکنش به محیط، شناسایی می‌کنند. با این حال، اکثر مطالعات از الگوریتم‌های یادگیری ماشین بانظارت و بدون نظارت استفاده می‌کنند. تفاوت این دو طبقه اصلی، وجود برچسب‌ها در زیرمجموعه داده‌های آموزشی است (کوتسیانتیس[۵]، ۲۰۰۷).

مقالات بسیاری تکنیک‌های مختلف یادگیری ماشینی بانظارت و بدون نظارت را بدون این‌که لزوماً تمایزی قائل شوند، مورد بحث قرار داده‌اند. طبق گفته پراوینا[۶] (۲۰۱۷)، یادگیری بانظارت نیاز به راهنمایی دارد که از تجربه یا الگوهای اکتسابی در داده‌ها نشأت می‌گیرد و در بیشتر موارد، شامل یک متغیر خروجی تعریف شده است. مجموعه داده ورودی به زیرمجموعه‌های "آموزش" و "تست" تفکیک می‌شوند. همه الگوریتم‌هایی که از رویکرد یادگیری نظارت شده استفاده می‌کنند، الگوهایی را در مجموعه داده‌های آموزشی به دست می‌آورند و متعاقباً آن‌ها را با هدف پیش‌بینی یا طبقه‌بندی یک ویژگی، در زیرمجموعه تست اعمال می‌کنند. اکثر نویسندگان گردش کار یک یادگیری ماشین بانظارت را توصیف کردند و همان‌طور که از بررسی‌ها نیز مشخص شده، درخت تصمیم، نایوبیزین و ماشین‌های بردار پشتیبان متداول‌ترین الگوریتم‌های مورد استفاده هستند (علقانی[۷]، ۲۰۲۰). از بهترین الگوریتم‌های بررسی رفتار، الگوریتم درخت تصمیم می‌باشد که در این مطالعه از این الگوریتم استفاده شده و در ادامه به توضیح آن پرداخته خواهد شد.

۱-۳-درخت تصمیم

درخت تصمیم[۸] به دسته الگوریتم‌های یادگیری بانظارت تعلق دارد و یکی از محبوب‌ترین روش‌های طبقه‌بندی است. طبقه‌بندی در یادگیری ماشین دارای دو مرحله اصلی است،

---

[1] Supervised
[2] Unsupervised
[3] Semi- Supervised
[4] Reinforcement
[5] Kotsiantis
[6] Praveena
[7] Alloghani
[8] Decision Tree

مرحله یادگیری و مرحله پیش‌بینی؛ در مرحله‌ی اول مدل براساس داده‌های آموزشی گسترش می‌یابد و در مرحله پیش‌بینی مدل برای داده‌های تست مورد آزمایش قرار می‌گیرد. در مسائل مرتبط با درخت تصمیم، متغیرهای مستقل و وابسته می‌توانند از هر دو نوع پیوسته و یا گسسته باشند و این در حالی است که در مورد بسیاری دیگر از مدل‌های یادگیری ماشین این‌گونه نبوده و تمامی متغیرها باید از یک نوع خاص (ترجیحاً پیوسته) باشند، اما استثنائاً در مورد متغیر هدف، اگر گسسته باشد، درختی که در نهایت حاصل می‌شود همان درخت تصمیم، و اگر پیوسته باشد، درخت نهایی به اصطلاح درخت رگرسیون[1] نامیده خواهد شد. این درخت رگرسیون در واقع همان نسخه درختی الگوریتم کلاسیک رگرسیون خطی می‌باشد که در بسیاری از مسائل تخمین تابع و امثال آن همواره مورد استفاده بوده است. وظیفه یادگیری در درختان رگرسیون، شامل پیش‌بینی تقریبی اعداد حقیقی بجای مقادیر گسسته است و این عمل را با داشتن مقادیر حقیقی در گره‌های برگ خود نشان می‌دهند. این نوع از درختان، تفسیر آسان داشته و می‌توانند توابع ثابت تکه‌ای را تقریب بزنند. هدف درخت تصمیم ساخت مدل آموزشی به منظور پیش‌بینی کلاس و ارزش متغیر هدف با یادگیری قوانین تصمیم ساده از داده‌های آموزشی است و روش کار این الگوریتم با استفاده از روش «تقسیم و غلبه»[2] می‌باشد. الگوریتم، یک درخت را شبیه‌سازی می‌کند و ویژگی‌ها را از طریق گروه‌بندی بر اساس مقادیر داده مرتب می‌کند (لوید و همکاران[3]، ۲۰۱۶). در این درخت هر گره داخلی[4] (گره بدون برگ) نشان‌دهنده آزمایش روی یک ویژگی (متغیر) است، هر شاخه[5] نشان‌دهنده نتیجه آزمایش است و هر گره برگ[6] (یا گره پایانی) نشان‌دهنده برچسب کلاس است. بالاترین گره در درخت، گره ریشه[7] است. برخی از الگوریتم‌های درخت تصمیم فقط درخت‌های دودویی تولید می‌کنند (جایی که هر گره داخلی دقیقاً به دو گره دیگر منشعب می‌شود)، در حالی که برخی دیگر می‌توانند درختان غیردودویی تولید کنند. سوال این است که درخت تصمیم چگونه برای

---

[1] Regression Tree
[2] Divide-And-Conquer
[3] Lloyd et al.
[4] Internal Node
[5] Branch
[6] Leaf Node
[7] Root Node

طبقه‌بندی استفاده می‌شود؟ با توجه به یک عدد X، که برچسب کلاس مربوطه برای آن ناشناخته است، مقادیر ویژگی در برابر درخت تصمیم آزمایش می‌شود. یک مسیر از ریشه به یک گره برگ پیمایش می‌شود، که پیش‌بینی کلاس را برای آن رکورد داده منجر می‌شود. درختان تصمیم به راحتی می‌توانند به قوانین طبقه‌بندی تبدیل شوند.

در حین ساخت درخت، از معیارهای انتخاب ویژگی[1] برای انتخاب صفتی استفاده می‌شود که بهترین حالت را برای دسته‌بندی جداگانه دارد. هنگامی که درختان تصمیم ساخته می‌شوند، بسیاری از شاخه‌ها ممکن است نویز یا داده‌های پرت در داده‌های آموزشی باشند. هرس[2] درختان برای شناسایی و حذف چنین شاخه‌هایی، با هدف بهبود دقت طبقه‌بندی در داده‌های دیده نشده، تلاش می‌کند.

۲-۳-ارزش اطلاعاتی

ارزش اطلاعاتی[3] به عنوان *معیار انتخاب ویژگی*، بر اساس کار پیشگام کلود شانون[4] در نظریه اطلاعات است که ارزش یا «محتوای اطلاعات»[5] پیام‌ها را مورد مطالعه قرار می‌داد. این معیار به مفهومی به نام آنتروپی[6] اطلاعات وابسته است. آنتروپی معیاری است که سطح عدم قطعیت را در گروهی از نمونه‌ها نشان می‌دهد. ویژگی که دارای آنتروپی بالا است، در نتیجه اطلاعات کمتری دارد؛ شاخه‌ای در درخت با آنتروپی صفر، نمایان‌گر گره برگ است و آنتروپی بیشتر از صفر نیاز به تقسیم‌بندی بیشتر در درخت دارد. فرمول محاسبه آنتروپی به شرح زیر است:

$$E(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \qquad (۱)$$

که در آن pi احتمال غیر صفر است که یک رکورد دلخواه در مجموعه آموزشی D به کلاس Ci تعلق دارد و از یک تابع log بر پایه ۲ استفاده می‌شود.

انتخاب ویژگی صحیح در زمان ساخت درخت تصمیم بسیار در صحت نتیجه و کارایی درخت تصمیم تاثیرگذار است؛ لذا معیار ارزش اطلاعاتی به دنبال کاهش آنتروپی است. ارزش اطلاعاتی اختلاف آنتروپی قبل از تقسیم گره و آنتروپی بعد از تقسیم گره را بر

---

[1] Attribute Selection Method
[2] Pruning
[3] Information Gain
[4] Claude Shannon
[5] Information Content
[6] Entropy

اساس ویژگی خاص X محاسبه می‌کند. فرمول آن به صورت زیر است که در آن S معرف مجموعه داده‌هاست:

$$Information\ Gain\ (S,X) = Entropy(S) - Entropy(S,X) \quad (۳)$$

هر چه ارزش اطلاعاتی بیشتر باشد به این معنی است که بر اساس آن ویژگی انتخاب شده، داده‌ها بهتر تقسیم شده‌اند (هانو همکاران[1]، ۲۰۱۲).

۳-۳-جامعه آماری و پیش‌پردازش داده‌ها

آمار و داده‌های مورد استفاده در این مقاله، اطلاعات اعضای خانوارهای شهری حاضر در "نمونه‌گیری آماری از هزینه- درآمد خانوار" در سال‌های ۹۹-۱۳۸۴، منتشر شده توسط مرکز آمار ایران است. این مجموعه داده، اطلاعات کافی در مورد موضوعاتی مانند دستمزد، سطح تحصیلات، وضعیت شغلی، وضعیت افراد خانوار به لحاظ بعد خانوار، تعداد محصلین و تعداد افراد بادرآمد ارائه می‌کند، همچنین شامل اطلاعات دقیق در مورد زمان کار یک فرد، ساعات اضافه‌کاری و دستمزد مستمر و غیرمستمر حاصل از ساعات اضافه‌کاری است. با توجه به این‌که انجام اضافه‌کاری در میان گروه‌های مختلف هزینه‌ای با انگیزه‌های متفاوت صورت می‌گیرد، در تجزیه و تحلیل این مقاله، علاوه بر متغیرهای مدنظر در مطالعات تجربی، بررسی اثر دهک‌های هزینه‌ای بر میزان اضافه‌کاری افراد نیز مدنظر بوده است و از آن‌جا که این اطلاعات تنها برای سرپرست خانوار منتشر می‌گردد، لذا از جامعه اعضای خانوارهای شهری، اطلاعات مربوط به سرپرست خانوارها تفکیک و مورد استفاده قرار گرفته است. همچنین جهت پردازش داده‌ها، از آن جا که این پژوهش به بررسی عوامل تمایل به اضافه‌کاری می‌پردازد، برای تحلیل این موضوع، در قسمت درآمدی مزد و حقوق بگیران، تنها افرادی که شاغل هستند و در ماه گذشته مجموع درآمد آن‌ها بیشتر از صفر بوده جدا شده‌اند؛ همچنین با بررسی کد شغلی افرادی که دارای دو یا چند شغل حقوق بگیری بودند و یا در مشاغل آزاد و درآمدهای متفرقه خانوار نیز درآمد ماه قبل آن‌ها مثبت بوده، بنا به احتمال تک شغله نبودن، از مجموعه داده‌ها کنار گذاشته شده‌اند. در نهایت، پس از اعمال موارد ذکر شده و حذف مشاهدات با مقادیر بدون داده، تعداد کل نمونه (افراد سرپرست خانوار) در این مجموعه داده‌ها، ۷۶۰۰۵ نفر فرد شاغل تک شغله برای تحلیل رفتار اضافه‌کاری در نمونه وارد شده‌اند.

---

[1] Han et al.

پیش‌پردازش دیگری که در داده‌ها ایجاد شد، مربوط می‌شود به طبقه‌بندی شغلی و فعالیت افراد شاغل که به صورت کدهای ۴ و ۵ رقمی در سال‌های مختلف برای هر فرد استفاده شده است که در این پژوهش برای تحلیل بهتر کد شغلی، تقسیم‌بندی ۱۰ گروه اصلی ISCO[1] در داده‌ها مبنا قرار گرفته و کدهای شغلی ۴ رقمی برای هر نفر، به کدهای صفر تا ۹ تغییر کرده است؛

جدول (۱): طبقه‌بندی گروه‌های اصلی مشاغل

کد	عنوان گروه اصلی	کد	عنوان گروه اصلی
۱	قانونگذاران، مقامات عالی رتبه و مدیران	۶	کارکنان ماهر کشاورزی، جنگلداری و ماهیگیری
۲	متخصصان (علمی و فنی)	۷	صنعتگران و کارکنان مشاغل مربوط
۳	تکنسین ها و دستیاران	۸	متصدیان مونتاژ کاران ماشین آلات و دستگاه ها و رانندگان وسایل نقلیه
۴	کارمندان امور اداری و دفتری	۹	کارگران ساده
۵	کارکنان خدماتی و فروشندگان	۰	مشاغل نیروهای مسلح

منبع: طرح هزینه-درآمد خانوار

جهت رسته‌بندی فعالیت‌های اقتصادی برای حساب‌های ملی و هدف‌های تحلیل اقتصادی، از "طبقه‌بندی استاندارد بین المللی رشته فعالیت‌های اقتصادی" (ISIC)[2]، استفاده شده است؛ ساختار تفصیلی طبقه‌بندی فعالیت‌های اقتصادی ایران تا قبل از سال ۱۳۹۲ به صورت کدهای ۴ رقمی تعریف شده و از سال ۱۳۹۲ به بعد کدها به ۵ رقم تغییر کرده است؛ در این پژوهش پس از بررسی کد فعالیت‌های دو دوره مربوط به قبل و بعد از سال ۹۲، به همسان‌سازی و تطبیق کدها پرداخته شد و فعالیت‌های اقتصادی با بررسی نام تمام فعالیت‌ها، به کدهای مربوط انتساب داده شدند و در ۲۱ رده اصلی فعالیت‌های اقتصادی جای گرفتند؛

[1] International Standard Classification Occupations (ISCO)
[2] International Standard Industrial Classification (ISIC)

جدول (۲): طبقه‌بندی فعالیت‌های اقتصادی در ۲۱ رده اصلی

کد	عنوان فعالیت	کد	عنوان فعالیت
۱	کشاورزی، جنگلداری و ماهیگیری	۱۲	فعالیت‌های املاک و مستغلات
۲	استخراج معدن	۱۳	فعالیت‌های حرفه‌ای، علمی و فنی
۳	تولید صنعتی (ساخت)	۱۴	فعالیت‌های اداری و خدمات پشتیبانی
۴	تامین برق، گاز، بخار و تهویه هوا	۱۵	اداره امور عمومی و دفاع ؛ تامین اجتماعی اجباری
۵	آب رسانی؛ مدیریت پسماند، فاضلاب و فعالیت های تصفیه	۱۶	آموزش
۶	ساختمان	۱۷	فعالیت‌های مربوط به سلامت انسان و مددکاری اجتماعی
۷	عمده فروشی و خرده فروشی ؛ تعمیر وسایل نقلیه موتوری و موتور سیکلت	۱۸	هنر، سرگرمی و تفریح
۸	حمل و نقل و انبارداری	۱۹	سایر فعالیت‌های خدماتی
۹	فعالیت های خدماتی مربوط به تامین جا و غذا	۲۰	فعالیت‌های خانوارها به عنوان کارفرما، فعالیت‌های تفکیک‌ناپذیر تولید کالاها و خدمات توسط خانوارهای معمولی برای خود مصرفی
۱۰	اطلاعات و ارتباطات	۲۱	فعالیت‌های سازمان‌ها و هیئت‌های برون مرزی
۱۱	فعالیت‌های مالی و بیمه		

منبع: طرح هزینه-درآمد خانوار

جدول ۳، شامل خلاصه‌ای از آمارهای توضیحی برای حجم نمونه شاغلان دارای اضافه-کاری و نمادهای به‌کاررفته برای هر متغیر می‌باشد؛

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول (۳): معرفی و توضیحات مربوط به نحوه تولید متغیرها

متغیر	نماد	توضیحات
اضافه کاری	Overtime	متغیر هدف؛ تعداد ساعات اضافه کاری انجام شده توسط نیروی کار در طول یک هفته
سال	Year	دوره ۱۳۸۴ تا ۱۳۹۹
کد شغلی	Job_code	کد های ۴ رقمی فهرست موضوعی مشاغل به اعداد ۰ تا ۹ تغییر کرد
کد فعالیت	Act_code	کد های ۴ و ۵ رقمی طبقه بندی فعالیت های اقتصادی به کدهای ۱ تا ۲۱ تغییر کرد
وضع شغلی	Sector	۱ بخش عمومی ۲ بخش تعاونی ۳ بخش خصوصی
جنس	Gender	۱ مرد ۲ زن
سن	Age	۱ گروه سنی ۱۵ تا ۲۴ سال ۲ گروه سنی ۲۵ تا ۴۴ سال ۳ گروه سنی ۴۵ تا ۶۴ سال ۴ گروه سنی ۶۵ سال به بالا
آیا در حال حاضر تحصیل می کند؟	Edju_stat	۱ بله ۲ خیر
تحصیلات	Acad_degree	۱ سوادآموزی غیر رسمی و ابتدایی ۲ پایان دوره راهنمایی ۳ دیپلم، پایان متوسطه و پیش دانشگاهی ۴ مدرک کاردانی یا لیسانس ۵ مدرک کارشناسی ارشد و بالاتر ۶ سایر آموزش های غیر رسمی
وضع فعالیت	ActStatus	۱ شاغل ۲ بیکار جویای کار با درآمد ۳ بدون کار ۴ محصل ۵ خانه دار ۶ سایر
وضعیت تاهل	Maried	۱ دارای همسر ۲ بی همسر بر اثر فوت همسر ۳ بی همسر بر اثر طلاق ۴ هرگز ازدواج نکرده
نحوه تصرف محل سکونت	CaptHome	۱ ملکی عرصه و اعیان ۲ ملکی اعیان ۳ اجاری ۴ رهن ۵ در برابر خدمت ۶ رایگان
بعدخانوار	Family_size	تعداد افراد خانوار
تعدادمحصلین	Stud_no	تعداد افرادی که در خانوار مشغول به تحصیل هستند
تعدادشاغل	Empl_no	تعداد افرادی که در خانوار مشغول به کار هستند
تعداد با درآمد	Incom_no	تعداد افرادی که در خانوار درآمد کسب می کنند
دهک هزینه خالص	Decile	۱ دهک اول ۲ دهک دوم ۳ دهک سوم ۴ دهک چهارم ۵ دهک پنجم ۶ دهک ششم ۷ دهک هفتم ۸ دهک هشتم ۹ دهک نهم ۱۰ دهک دهم

منبع: یافته‌های تحقیق

۴-۳-برآورد الگوریتم درخت تصمیم

تمامی متغیرهای انتخاب شده در این مدل، طبق جدول ۳، ماهیت کیفی دارند؛ با این‌حال، تعدادی از داده‌ها ظاهر کمّی ولی خاصیت کیفی داشتند که برای این کار بایستی از طریق یکی از الگوریتم‌های هموارسازی داده‌ها به نام الگوریتم بسته‌بندی[1]

[1] Binning Algorithm

استفاده کرده و وضعیت آن‌ها از کمی به کیفی تغییر داد، مانند متغیر سال که به صورت عددی ذخیره می‌شود ولی در واقع مقدار آن کیفی می‌باشد. بدین ترتیب تمامی متغیرهایی با این ویژگی، از کمّی به کیفی تغییر کردند. با توجه به این‌که تمامی متغیرهای تاثیرگذار بر متغیر کلاس داده‌ها، که در اینجا هر گروه ساعت اضافه‌کاری می‌باشد، کیفی شدند، لذا خود متغیر کلاس نیز به صورت دسته‌بندی شده و کیفی تبدیل شد. در مدل مورد بررسی، تعداد بازه‌های تقسیم‌بندی داده‌های ساعات اضافه‌کاری به ۳ دسته کمتر از ۱۷ ساعت (کلاس a)، ۱۷ تا ۳۴ ساعت (کلاس b)و بیشتر از ۳۴ ساعت (کلاس c) در هفته تقسیم شده‌اند.

Classification of independent variable

Class Attribute (Overtime Working)		
Class a	Working Overtime less than 17 hours	a = (-inf-17]
Class b	Working Overtime between 17 to 34 hours	b = (17-34]
Class c	Working Overtime greater than 34 hours	c = (34-inf)

برای داده‌های مدل مورد بررسی با توجه به این‌که متغیر کلاس اضافه‌کاری به صورت کیفی درآمده، از الگوریتم‌های طبقه‌بندی درخت تصمیم می‌بایست استفاده گردد که در این مدل از الگوریتم REPTree[1] برای ساخت درخت تصمیم و پیش‌بینی مدل استفاده شده است که یک تکنیک یادگیری مبتنی بر الگوریتم $C4.5$ است و یک الگوریتم یادگیری درخت تصمیم سریع است و بر اساس معیار انتخاب ویژگی ارزش اطلاعاتی به دست آمده است؛ اطلاعات اضافی که مورد نیاز نیستند نیز با استفاده از یکی از تکنیک‌های هرس به نام هرس خطای کاهش یافته[2] حذف می‌شوند.

تغییر مهمی که در پارامترهای الگوریتم ایجاد شد، در پارامترهایی بود که تعداد حداقل داده‌ها در یک گره را جهت شکسته شدن، تعیین می‌کرد؛ در دو حالت یک گره می‌تواند تبدیل به گره برگ شده و دیگر شکسته نشود:

[1] Reduced Error Pruning Tree
[2] Reduced Error Pruning

- یا این که تمام داده‌هایی که در گره هستند همگی به یک کلاس تعلق داشته باشند که در این حالت عنوان می‌شود گره به خلوص[1] رسیده است و در این صورت دیگر نیازی به شکستن گره نیست و می‌توانیم تصمیم‌گیری کنیم.
- یا تعداد داده‌ها کمتر از یک حد نصاب باشند که در این‌صورت دیگر ارزش شکسته شدن ندارد؛ اگر حداقل تعداد نمونه در هر برگ تعیین نشده باشد، درخت تا زمانی که همه گره‌های برگ فقط به یک کلاس مرتبط شوند، شکافته می‌شود.

حد نصاب در پارامترهای الگوریتم REPTree، تعداد حداقل ۲۰ داده در یک گره، تعریف شد. طبق این محدودیت اعمال شده، اگر در یک گره ۲۰ داده یا کمتر باشد، آن گره به گره برگ تبدیل می‌شود؛ در این حالت برای تعیین کلاس داده‌ها، رای اکثریت[2] انجام می‌شود، بدین صورت که اگر بیشتر داده‌های گره، به عنوان مثال، متعلق به کلاس a باشند، نتیجه می‌گیریم که این گره مربوط به کلاس a است، چون بیشتر داده‌های آن گره هنگام آموزش به این کلاس تعلق داشته‌اند.

برای ارزیابی و تست گرفتن از الگوریتم مورد نظر، از میان حالت‌های تست[3] موجود، Cross-validation استفاده شده و تعداد fold نیز ۵ انتخاب شده است. بدین ترتیب Cross-validation مجموعه داده را به ۵ قسمت تقسیم می‌کند، ۴ قسمت اول برای آموزش الگوریتم و قسمت پنجم برای ارزیابی الگوریتم استفاده می‌گردد. این فرآیند تکرار می‌شود و به هر یک از ۵ قسمت مجموعه داده تقسیم شده فرصت می‌دهد تا مجموعه آزمایشی و تست باشند؛

۴-تحلیل‌های تجربی

پس از اعمال الگوریتم REPTree، خروجی مدل به صورت متنی ارائه می‌گردد که به صورت خلاصه، در شکل ۴ نتیجه الگوریتم به تصویر کشیده شده است.

[1] Purity
[2] Majority Voting
[3] Test Option

Overtime Working Hours	Decision
. Initial node	**create root node {act-code}**
. node {act-code}	
Branches: {act-code=1 to 5}	end with class {a=(-inf-17]}
. node {act-code}	
Branch: {act-code=6}	**create node {year}**
node {year} Branches: {year=84 to 89}	end with class {a=(-inf-17]}
node {year} Branch: {year=90}	**create node {decile}**
node {decile} Branches: {decile=1 to 5}	end with class {a=(-inf-17]}
node {decile} Branch: {decile=6}	**create node {job-code}**
node {job-code} Branches: {job-code=0 to 7}	end with class {a=(-inf-17]}
node {job-code} Branch: {job-code=8}	end with class {b=(17-34]}
node {job-code} Branch: {job-code=9}	end with class {a=(-inf-17]}
node {decile} Branches: {decile=7 to 10}	end with class {a=(-inf-17]}
node {year} Branches: {year=91 to 99}	end with class {a=(-inf-17]}
. node {act-code}	
Branch: {act-code=7}	end with class {a=(-inf-17]}
. node {act-code}	
Branch: {act-code=8}	**create node {year}**
node {year} Branches: {year=84 to 89}	end with class {a=(-inf-17]}
node {year} Branch: {year=90}	**create node {job-code}**
node {job-code} Branch: {job-code=0}	end with class {a=(-inf-17]}
node {job-code} Branch: {job-code=1}	end with class {b=(17-34]}
node {job-code} Branches: {job-code=2 to 7}	end with class {a=(-inf-17]}
node {job-code} Branch: {job-code=8}	**create node {decile}**
node {decile} Branches: {decile=1 to 5}	end with class {a=(-inf-17]}
node {decile} Branch: {decile=6}	end with class {b=(17-34]}
node {decile} Branches: {decile=7 to 8}	end with class {a=(-inf-17]}
node {decile} Branches: {decile=9}	end with class {b=(17-34]}
node {decile} Branches: {decile=10}	end with class {a=(-inf-17]}
node {job-code} Branch: {job-code=9}	end with class {a=(-inf-17]}
node {year} Branches: {year=91 to 99}	end with class {a=(-inf-17]}
. node {act-code}	
Branch: {act-code=9 to 21}	end with class {a=(-inf-17]}

شکل (۴): خروجی متنی درخت تصمیم با توجه به مهمترین متغیرهای طبقه‌بندی

منبع: یافته‌های تحقیق

طبق نتایج حاصل از درخت تصمیم، از میان تمام ویژگی‌های مستقل طبق جدول ۳، متغیرهای کد فعالیت، کد شغلی، سال و دهک هزینه‌ای بیشترین همبستگی را با ساعات

اضافه‌کاری افراد مورد بررسی داشته‌اند و از میان این متغیرهای تاثیرگذار، متغیر کد فعالیت (Act-code) به عنوان گره ریشه با توجه به آنتروپی پایین و ارزش اطلاعاتی بالایی که داشته، انتخاب شده است و این نشان می‌دهد مهمترین عاملی که بر میزان ساعات اضافه‌کاری نیروی‌کار تاثیر دارد، بستگی به نوع فعالیت شغلی افراد و مهارتی که در انجام کار کسب کرده، دارد. درخت به ازای تمام کد فعالیت‌ها به جز کد فعالیت ۶ و ۸ بیشتر گسترش پیدا نکرده و در این کد فعالیت‌ها تبدیل به گره برگ شده است و دیگر تقسیم-بندی پیدا نمی‌کند؛ همچنین همه گره‌های برگ برای این کد فعالیت‌ها، به کلاس a یعنی اضافه‌کاری کمتر از ۱۷ ساعت تعلق داشته‌اند. این می‌تواند به دو دلیل باشد، یا در خروجی داده آموزشی، همگی به کلاس a تعلق داشته‌اند و داده‌های این گره‌ها به خلوص رسیده‌اند (یعنی به عنوان مثال، از ۱۷۳۲ داده‌ای که در گره مربوط به Act-code=1 قرار گرفته‌اند، همگی به کلاس a یا کمتر از ۱۷ ساعت اضافه‌کاری در هفته ([-inf-17-]) تعلق داشته‌اند)، یا به دلیل این‌که طبق محدودیتی که برای حداقل داده‌های موجود در یک گره برای این‌که گسترش پیدا کند، ایجاد کردیم، تعداد داده‌های این گره‌ها کمتر از ۲۰ داده بوده‌اند که در این صورت، رای اکثریت انجام شده است، بدین معنا که چون بیشتر داده‌ها متعلق به کلاس a بوده‌اند، نتیجه می‌گیریم که این گره مربوط به کلاس a است.

از طرفی در کد فعالیت ۶ که مربوط به صنعت ساختمان می‌باشد و کد فعالیت ۸ که مربوط به حمل‌ونقل و انبارداری می‌باشد (جدول ۲)، داده‌هایی که به این گره‌ها رسیده‌اند، یا خلوص نداشتند و یا تعداد آن‌ها بیشتر از ۲۰ داده بوده است، لذا این گره‌ها شکسته شده و مدل مجبور شده از ابزار "انتخاب ویژگی" استفاده کرده تا ببیند کدام ویژگی از میان ویژگی‌های باقیمانده، به جز کد فعالیت، در اینجا میزان اطلاعات بیشتری در خود دارد.

از میان تمام ویژگی‌های باقیمانده برای تقسیم کد فعالیت ۶، ویژگی سال (Year) انتخاب شده است و این تقسیم‌بندی برای داده موجود در کد فعالیت ۶ در تمام سال‌های ۱۳۸۴ تا ۱۳۹۹ انجام شده است. همانطور که در شکل ۴ مشخص شده، در تمام سال‌ها به جز سال ۹۰، گره تبدیل به برگ شده و بیشتر گسترش پیدا نکرده است که باز هم به دلایل گفته شده مربوط به محدودیت تعداد داده کمتر از ۲۰ داده بوده و یا خلوص حاصل شده و همه داده‌های موجود در گره به یک کلاس تعلق داشته‌اند؛ اما در سال ۹۰ گره شکسته

شده و ویژگی که برای تقسیم‌بندی گره سال ۹۰ از طریق معیار ارزش اطلاعاتی انتخاب شده، دهک (Decile) است. در این قسمت نیز در دهک ۶ (Decile=6) دوباره گره شکسته شده و معیار تقسیم‌بندی نیز کد شغلی (Job-code) تعیین شده است.

کد فعالیت ۸ (Act-code=8) نیز مجدد با ویژگی سال (Year) شکسته شده است و تمام سال‌ها به جز سال ۹۰ به گره برگ تبدیل شده‌اند و محدودیت‌های مربوط به برگ شدن یک گره در مورد آن‌ها اتفاق افتاده است، ضمن این‌که همگی این سال‌ها به کلاس a تعلق داشته‌اند؛ اما متغیر سال ۹۰ با گره کد شغلی (Job-code) تقسیم شده و در این گره تبدیل به گره برگ شده است، به جز کد شغلی ۸ که با گره دهک (Decile) شکسته شده است.

همانطور که توضیح داده شد طبق نتایج حاصل از درخت تصمیم، مهمترین عاملی که بر میزان ساعات اضافه‌کاری نیروی‌کار تاثیر دارد، بستگی به نوع فعالیت شغلی افراد و مهارت آن‌ها دارد؛ همچنین با توجه به این که در چه سالی نیروی‌کار فعالیت داشته، بر میزان ساعات اضافه‌کاری تاثیر داشته است و در برخی سال‌ها نیز (به عنوان مثال سال ۹۰)، این که فرد در چه دهک هزینه‌ای قرار می‌گیرد، در نحوه عملکرد وی تاثیر داشته است.

پس از ارائه درخت به صورت متنی، اطلاعات آماری در مورد درخت ایجاد شده ارائه می‌گردد؛

جدول (۴): نتایج آماری درخت تصمیم

=== Stratified cross-validation ===		
Correctly Classified Instances	۶۵۹۷۵	۸۶.۸۰٪
Incorrectly Classified Instances	۱۰۰۳۰	۱۳.۲۰٪
Total Number of Instances	۷۶۰۰۵	
Size of the tree (node)	۹۴	

منبع: یافته‌های تحقیق

اندازه درخت ۹۴ گره به تعداد کل گره‌های درخت است که شامل گره‌های برگ و غیر برگ می‌شود. خروجی مدل نشان می‌دهد از ۷۶۰۰۵ داده‌ای که در مدل داشتیم، ۶۵۹۷۵ داده درست طبقه‌بندی شده‌اند، به عبارتی حدود ۸۷ درصد داده‌ها به درستی شناسایی شده‌اند. در واقع این عدد معیاری است که برای ارزیابی میزان صحت و دقت عملکرد مدل

به دست می‌آید و در اینجا به ما نشان می‌دهد که این مدل تا چه حد در تشخیص کلاس درست اضافه‌کاری برای داده‌ها موفق بوده است.

۱-۴- ماتریس درهم ریختگی[1]

در حوزه الگوریتم‌های هوش مصنوعی، به ماتریسی گفته می‌شود که نتیجه ارزیابی عملکرد الگوریتم‌ها در آن نمایش داده می‌شود. هر ستون از ماتریس، نمونه‌ای از مقدار پیش‌بینی-شده را نشان می‌دهد، در صورتی که هر سطر نمونه‌ای واقعی (درست) را در برمی‌گیرد. این ماتریس نشان‌دهنده این موضوع است که داده‌هایی که به عنوان مجموعه داده آزمایشی در نظر گرفته شده‌اند، تا چه اندازه توسط مدلی که روی داده آموزشی برازش داده شده است، به درستی دسته‌بندی شده‌اند. طبق جدول ۵، جمع تمام داده‌هایی که درست طبقه‌بندی شده‌اند روی قطر اصلی قرار دارند که برابر همان ۶۵۹۷۵ داده یا ۸۷ ٪ داده‌ای است که در قسمت قبل عنوان شد و بقیه داده‌ها در ماتریس، ۱۳ ٪ داده‌ای هستند که به درستی شناسایی نشده‌اند.

جدول (۵):نتایج ماتریس درهم‌ریختگی برای داده‌های آزمایشی

		مقدار پیش بینی شده		
	ساعت اضافه کاری در هفته	کمتر از ۱۷ ساعت (class a)	بین ۱۷ تا ۳۴ ساعت (class b)	بیشتر از ۳۴ ساعت (class c)
مقدار واقعی	کمتر از ۱۷ ساعت (class a)	۶۵۹۶۵	۲۳	۱
	بین ۱۷ تا ۳۴ ساعت (class b)	۷۳۵۲	۱۰	۰
	بیشتر از ۳۴ ساعت (class c)	۲۶۵۲	۲	۰

منبع: یافته‌های تحقیق

۵-نتیجه‌گیری

نتایج نشان می‌دهد تکنیک درخت تصمیم، رویکردی کاربردی برای پیش‌بینی نیازهای بازار کار ارائه می‌دهد و پیاده‌سازی روش‌های داده‌کاوی بر روی داده‌های نیروی‌کار می‌تواند حجم زیادی از داده‌ها در بخش عرضه نیروی‌کار را مدیریت کرده و آن‌ها را به اطلاعات

[1] Confusion Matrix

مفید برای بهبود فرآیند تصمیم‌گیری در بنگاه‌ها تبدیل کند. طبق مطالعات تجربی صورت گرفته، وضعیت اشتغال از طریق متغیرهای سنتی جمعیت‌شناختی و نگرش کاری توضیح داده می‌شود، اما اینکه چه عواملی فرد را ترغیب به اضافه‌کاری می‌کنند، جای ابهام دارد و به نظر می‌رسد که اشتغال و اضافه‌کاری شرایط متمایز داشته باشند و نیاز به مطالعه جداگانه دارند. طبق نتایج پژوهش حاضر که بر عرضه نیروی‌کار تاکید داشته است، وضعیت عرضه اضافه‌کاری به دلیل آن چیزی نیست که به طور متداول فرض می‌شود، یعنی سن، تحصیلات، جنسیت، نگرش‌های کاری یا هر چیز غیرقابل مشاهده‌ای که این متغیرها نماینده آن‌ها هستند، بلکه به دلایل متفاوتی است و از آنجایی که ورودی‌های مدل سنتی نمی‌توانند عرضه اضافه‌کاری را به شیوه‌ای رضایت‌بخش توضیح دهند، سایر عوامل توضیحی باید بیشتر مورد بررسی قرار گیرند. همانطور که عنوان شد، عرضه اضافه-کاری را به سختی می‌توان با ویژگی‌های سطح فردی توضیح داد، زیرا در ساخت درخت تصمیم که بر اساس میزان ارزش اطلاعاتی است که متغیرها در مورد متغیر هدف اضافه-کاری در اختیار قرار می‌دهند، این متغیرها دیده نمی‌شوند، اما می‌توان آن را به وضعیت شغلی که فرد در آن مشغول به فعالیت است و کسب دانش و تجربه کرده و مهارت انجام آن شغل را دارد و یا همین‌طور دهک هزینه‌ای که خانوار در آن قرار دارد، نسبت داد و این متغیرها بیشترین همبستگی را با متغیر هدف ساعت اضافه‌کاری افراد داشته‌اند. از طرفی، از سمت تقاضای کار نیز، این نتایج می‌تواند به سیاست‌گذاران در تصمیم‌گیری برای اعمال سیاست‌های اشتغال‌زایی کمک کند و برخی مشاغل مانند صنعت ساختمان و حمل‌ونقل، اعمال سیاست‌های اشتغال‌زایی کاربردی نخواهد بود، زیرا این مشاغل، به دلیل تمایل بیشتر سمت عرضه نیروی‌کار به انجام اضافه‌کاری، می‌توانند کمبود نیروی‌کار را با انجام اضافه‌کاری تامین کنند تا اینکه نیروی‌کار جدید استخدام کنند. لذا اگر این نتایج بیشتر مورد حمایت قرار گیرند، به جهت تخصیص بهینه نیروی‌کار، پیامدهای قابل توجهی هم برای افراد و هم برای سیاست‌گذاران دارند. چنین نتایجی می‌تواند مورد توجه

اولویت‌های پژوهشی در اقتصاد کار و هدایت آن به سمت درک بهتر عوامل تعیین‌کننده وضعیت شغلی افراد کمک کند.

تضاد منافع

نویسندگان نبود تضاد منافع را اعلام می‌دارند.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

فهرست منابع

۱. طائی، حسن (۱۳۸۵). تابع عرضه نیروی کار: تحلیلی بر پایه داده‌های خرد. *پژوهش-های اقتصادی ایران*، ۸(۲۹)، ۱۱۲-۹۳.

۲. کشاورز حداد، غلامرضا و خشابی، پویان (۱۳۹۱). تخمین عوامل تعیین اصابت و ساعات اضافه‌کاری در نیروی‌کار شهری ایران. *مجله تحقیقات اقتصادی*، ۹۸، ۲۰۳-۱۷۱.

1. Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and unsupervised learning for data science*, 3-21.

2. Alsultanny, Y. A. (2013). Labor market forecasting by using data mining. *Procedia Computer Science*, *18*, 1700-1709.

3. Anxo, D., & Karlsson, M. (2019). *Overtime work: A review of literature and initial empirical analysis*. International Labour Office, Inclusive Labour Markets, Labour Relations and Working Conditions Branch.

4. Bauer, T., & Zimmermann, K. F. (1999). Overtime work and overtime compensation in Germany. *Scottish Journal of Political Economy*, *46*(4), 419-436.

5. Bell, D. N., & Hart, R. A. (2003). Wages, hours, and overtime premia: evidence from the British labor market. *ILR Review*, *56*(3), 470-480.

6. Chaiboonsri, C., & Wannapan, S. (2019). Big data and machine learning for economic cycle prediction: application of Thailand's economy. In *Integrated Uncertainty in Knowledge Modelling and Decision Making: 7th International Symposium, IUKM 2019, Nara, Japan, March 27–29, 2019, Proceedings 7* (pp. 347-359). Springer International Publishing.

7. Ehrenberg, R. G., Smith, R. S., & Hallock, K. F. (2021). *Modern labor economics: Theory and public policy*. Routledge.

8. Gerunov, A. (2016). Employment modelling through classification and regression trees. *International Journal of Data Science*, *1*(4), 316-329.

9. Jiawei, H., Micheline, K., & Jian, P. (2012). *Data Mining: Concepts and Techniques.-3rd*. Morgan kaufmann.

10. Hart, R. A. (2004). *The economics of overtime working*. Cambridge University Press.

11. Hart, R. A., & Ma, Y. (2010). Wage–hours contracts, overtime working and premium pay. *Labour Economics*, *17*(1), 170-179.

12. Keshavarz Haddad, Gh. R., & Khashabi, P. (2012). Estimation of Factors Determining Impact and Overtime Hours in Iranian Urban Labor. *Journal of Economic Research*, *98*, 203-171 (in Persian).

13.Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3-24.
14.Lloyd, S., Garnerone, S., & Zanardi, P. (2016). Quantum algorithms for topological and geometric analysis of data. *Nature communications*, *7*(1), 10138.
15.Papagiannaki, E. (2014). Rising unpaid overtime: a critical approach to existing theories. *International Journal of Management Concepts and Philosophy*, *8*(1), 68-88.
16.Praveena, M., & Jaiganesh, V. (2017). A literature review on supervised machine learning algorithms and boosting process. *International Journal of Computer Applications*, *169*(8), 32-35.
17.Ranjan, J., Goyal, D. P., & Ahson, S. I. (2008). Data mining techniques for better decisions in human resource management systems. *International Journal of Business Information Systems*, *3*(5), 464-481.
18.Taei, H. (2006). Labor Supply Function: An Analysis Based on Microdata. *Iranian Journal of Economic Research*, *8*(29), 93-112 (in Persian).
19.Zhang, Y., Xu, S., Zhang, L., & Yang, M. (2021). Big data and human resource management research: An integrative review and new directions for future research. *Journal of Business Research*, *133*, 34-50.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی