امکان‌سنجی پیش‌بینی جمعیت استان فارس و شهرستان‌های آن با استفاده از مدل شبکه‌ی عصبی مصنوعی

امکان‌سنجی پیش‌بینی جمعیت استان فارس و شهرستان‌های آن با استفاده از مدل شبکه‌ی عصبی مصنوعی

مجری طرح

اکبر بیگلریان

همکاران طرح

محمد اصغری جعفرآبادی

احمدرضا باغستانی

قدرت‌اله روشنایی

مرتضی سدهی

جواد فردمال

فاطمه ربیعی

پژوهشکده‌ی آمار

گروه پژوهشی طرح‌های فنی و روش‌های آماری

خرداد ۱۳۸۹

به نام خداوند جان و خرد

پیش‌گفتار

در کشور ایران به جهت تغییرات عمده در ساختار اجتماعی، اقتصادی، سیاسی، فرهنگی و تحولات سریع جمعیتی پیش‌بینی جمعیت از پیچیدگی خاصی برخوردار است. از آنجا که با توجه به تحولات سیاسی و اجتماعی که در هر جامعه‌ای رخ می‌دهد، استفاده از روش‌های دقیق می‌تواند در پیش‌بینی دقیق جمعیت یک ناحیه، استان و یا کشور نقش بسزایی داشته باشد، لذا استفاده از روش‌ها و مدل‌های دقیق که بتواند ساختار و پیچیدگی متغیرهای تاثیرگذار بر پیش‌بینی جمعیت را شناسایی و درنظر گیرد، بسیار مناسب است.

با توجه به اینکه ساختار پیچیده‌ای برای پیش‌بینی جمعیت در سطوح پایین جغرافیایی وجود دارد که نیاز به معرفی مدل‌هایی است که بتواند ساختار واقعی‌تری از داده‌ها را در تحلیل مشارکت دهد، پژوهشکده‌ی آمار انجام یک طرح پژوهشی با هدف امکان‌سنجی پیش‌بینی جمعیت استان فارس و شهرستان‌های آن با استفاده از مدل شبکه-ی عصبی مصنوعی و مقایسه‌ی آن با روش‌های معمول مورد استفاده را در دستور کار قرار داد.

در گروه مطالعاتی طرح مذکور، آقای دکتر اکبر بیگلریان به عنوان مجری طرح و آقایان دکتر محمد اصغری جعفرآبادی، دکتر احمدرضا باغستانی، دکتر قدرت‌اله روشنایی، دکتر مرتضی سدهی، دکتر جواد فردمال و فاطمه ربیعی به عنوان همکار عضویت داشتند که بدین‌وسیله از زحمات یکایک این افراد تشکر و قدردانی می‌شود. داوری طرح نیز به عهده‌ی جناب آقای دکتر اشکان رحیمی کیان عضو هیات علمی دانشگاه تهران بوده است که بدین-وسیله از راهنمایی‌های ایشان تشکر و قدردانی می‌شود.

گروه پژوهشی طرح‌های فنی و روش‌های آماری

پیش‌گفتار مجری

پایه و اساس مطالعات جمعیت، آمارهای حیاتی است که مبتنی بر اعداد و ارقام به دست آمده از تحـولات تجدیـد پذیر و تجدیدناپذیر جمعیت است. چنان‌چه این اعداد و ارقام درست نبوده و یا نارسایی داشته باشد، محاسبات و به تبع آن شناخت تغییر و تحولات و مقایسه شاخص‌های جمعیتی که نقش کلیدی در توسعه دارند بی‌فایـده خواهـد بود.

اگرچه کارشناسان مسائل جمعیتی، متخصصان آمار، و سایر پژوهشگران علاقمند به مسایل جمعیتی هنوز قادر بـه پیش‌بینی دقیقی از جمعیت نیستند لیکن تلاش برای بهبود پیش‌بینی جمعیت کماکان ادامه دارد. پژوهش‌گـران و محققین، روش‌های مختلفی را برای برآورد پیش‌بینی جمعیت استفاده کرده‌اند و نقش متغیرهایی از جمله مهاجرت، ازدواج، مرگ و میر و ... را در پیش‌بینی جمعیت مورد بحث و بررسی قرار داده‌اند.

کشور ایران که جزو کشورهای در حال توسعه تلقی می‌شود، به دلیل تغییرات عمده در ساختار اجتماعی، اقتصادی، سیاسی و فرهنگی؛ تحولات سریع جمعیتی و به تبع آن تحولات در زمینـه‌هـای بـاروری، مـرگ و میـر و میـزان مهاجرت را تجربه می‌کند. علاوه بر این، روند جمعیت ایران و انجام پیش‌بینی آن به دلایل زیر نیز می‌تواند مهـم باشد:

- ✓ افزایش تعداد زنان در سن تجدید نسل (به رغم کاهش نرخ باروری)؛
- ✓ افزایش متوسط امید به زندگی زنان و مردان؛
- ✓ جوان بودن جمعیت ایران و بالا بودن جمعیت در سن ازدواج.

این موارد می‌تواند منجر به پیچیدگی ساختار جمعیتی شده و در عین حال امکان پیش‌بینی را با پیچیـدگی مواجـه سازد. در این پژوهش تلاش می‌شود تا به کمک مدل شبکه‌ی عصبی مصنوعی، امکان پیش‌بینی جمعیت اسـتان فارس (و نیز شهرستان‌های آن) مورد بررسی قرار گیرد.

این گزارش شامل بخش‌های زیر است:

فصل اول، که در آن مقدمات و مطالب کلی طرح بیان شده است. در فصل دوم، روش‌شناسی پـژوهش معرفـی و شرح داده شده است. فصل سوم، یافته‌های پژوهش ارائه شده و در فصل چهارم به نتیجه‌گیری پرداخته شده است.

فهرست مطالب

عنوان ..... صفحه

فصـل اول: مقدمـه

۱-۱- مقدمه ۲
۱-۲- تاریخچه و نظریه‌های مربوط به موضوع پژوهش ۳
۱-۳- نوشتگان موضوع پژوهش ۵
۱-۴- بیان مسأله و اهمیت پژوهش ۵
۱-۵- هدف‌های پژوهش ۷
۱-۶- پرسش‌های پژوهش ۸
۱-۷- تعریف مفاهیم ۸

فصل دوم: روش‌شناسی پژوهش

۲-۱- روش‌های برآورد جمعیت استان‌ها ۱۳
۲-۲- روش‌های پیش‌بینی جمعیت ۱۳
۲-۲-۱- میزان‌های رشد ۱۴
۲-۲-۲- پیش‌بینی‌های کوتاه‌مدت ۱۶
۲-۲-۲-۱- سری‌های زمانی ۱۶
۲-۲-۲-۲- تبدیلات در سری‌های زمانی ۱۷
۲-۲-۲-۳- مدل‌های ARIMA ۱۷
۲-۲-۲-۴- پیش‌بینی میزان‌های خام ۱۸
۲-۳- روش جدول عمر ۱۹
۲-۴- مبانی نظری شبکه‌های عصبی مصنوعی ۲۰
۲-۴-۱- یادگیری در سیستم‌های بیولوژیک ۲۲
۲-۴-۲- یادگیری در ماشین‌ها ۲۳
۲-۴-۳- مزایا و انتظارات ۲۴

۲-۴-۳-۱- قابلیت یادگیری ۲۴
۲-۴-۳-۲- پراکندگی اطلاعات ۲۵
۲-۴-۳-۳- قابلیت تعمیم ۲۵
۲-۴-۳-۴- پردازش موازی ۲۵
۲-۴-۳-۵- مقاوم بودن ۲۶
۲-۴-۳-۶- غیر خطی بودن ۲۶
۲-۴-۴- معایب ۲۶
۲-۴-۵- مدل ریاضی یک نورون ۲۶
۲-۴-۶- آموزش شبکه‌های عصبی ۲۸
۲-۴-۶-۱- اهداف کلی شبکه عصبی مصنوعی ۲۹
۲-۴-۷- پرسپترون چند لایه ۳۰
۲-۴-۸- الگوریتم پس‌انتشار خطا ۳۰
۲-۴-۹- شیوه‌ی تحلیل به کمک شبکه‌ی عصبی مصنوعی ۳۱
۲-۴-۱۰- مدل نورون چند ورودی با لایه‌ی میانی (پرسپترون چند لایه) ۳۲
۲-۴-۱۰-۱- فرایند یادگیری ۳۴
۲-۴-۱۰-۲- الگوریتم یادگیری پس‌انتشار خطا ۳۵
۲-۴-۱۰-۳- معیار توقف ۳۶
۲-۴-۱۰-۴- تنظیم پارامترهای لایه‌ی خروجی ۳۷
۲-۴-۱۰-۵- اندازه‌ی حرکت ۳۸
۲-۴-۱۰-۶- مبنای ریاضی الگوریتم پس‌انتشار ۳۸
۲-۴-۱۱- دیدگاه مدل خطی تعمیم یافته (شبکه‌ی عصبی، تعمیمی از رگرسیون خطی) ۴۱
۲-۴-۱۲- روش انجام پژوهش ۴۵
۲-۴-۱۳- روش تحلیل داده‌ها ۴۵

فصل سوم: یافتــه‌های پــژوهش

۳-۱- مقدمه ۴۷
۳-۲- تحلیل داده‌ها ۴۹

فصل چهارم: نتیـجه‌گیـری

۴-۱- مقدمه ۷۱
۴-۲- نتیجه‌گیری ۷۸

مرجع‌ها ۷۹

فهرست شکل‌ها و نمودارها

شکل ۲-۱. مشخصات اصلی یک نورون بیولوژیکی ۲۱

شکل ۲-۲. مشخصات کلی یک نورون مصنوعی ۲۷

شکل ۲-۳. پرسپترون یک لایه (MLP) با چندین نود ۳۲

شکل ۲-۴. شبکه پرسپترون چند لایه (MLP) با یک لایه مخفی ۳۳

شکل ۲-۵. نمایش شبکه به صورت مدل‌های شبکه عصبی مصنوعی (الف) و رگرسیونی خطی (ب) ۴۲

نمودار ۴-۱. مقایسه‌ی نرخ رشد جمعیت شهرستان استهبان به دو روش رشد و شبکه‌ی عصبی طی سال‌های ۱۳۸۶ تا ۱۴۰۰ ..... ۷۳

نمودار ۴-۲. مقایسه‌ی نرخ رشد جمعیت شهرستان اقلید به دو روش رشد و شبکه‌ی عصبی طی سال‌های ۱۳۸۶ تا ۱۴۰۰ ..... ۷۴

نمودار ۴-۳. مقایسه‌ی نرخ رشد جمعیت شهرستان جهرم به دو روش رشد و شبکه‌ی عصبی طی سال‌های ۱۳۸۶ تا ۱۴۰۰ ..... ۷۴

نمودار ۴-۴. مقایسه‌ی نرخ رشد جمعیت شهرستان سپیدان به دو روش رشد و شبکه‌ی عصبی طی سال‌های ۱۳۸۶ تا ۱۴۰۰ ..... ۷۵

نمودار ۴-۵. مقایسه‌ی نرخ رشد جمعیت شهرستان شیراز به دو روش رشد و شبکه‌ی عصبی طی سال‌های ۱۳۸۶ تا ۱۴۰۰ ..... ۷۵

نمودار ۴-۶. مقایسه‌ی نرخ رشد جمعیت شهرستان نی‌ریز به دو روش رشد و شبکه‌ی عصبی طی سال‌های ۱۳۸۶ تا ۱۴۰۰ ..... ۷۶

نمودار ۴-۷. نرخ رشد جمعیت استان فارس در سال‌های سرشماری ۱۳۶۵-۱۳۸۵ ..... ۷۷

فهرست جدول‌ها

جدول ۱-۱. فهرست متغیرها و شاخص‌های مورد نیاز برای پیش‌بینی جمعیت ۱۱

جدول ۳-۱. جمعیت و شاخص‌های مورد استفاده در پیش‌بینی جمعیت استان فارس ۴۸

جدول ۳-۲. جمعیت و شاخص‌های مورد استفاده در پیش‌بینی جمعیت شهرستان‌های منتخب استان فارس ۴۸

جدول ۳-۳. پیش‌بینی جمعیت شهرستان‌های استان فارس برای سال‌های ۱۳۸۶-۱۴۰۰ به روش میزان رشد ۴۹

جدول ۳-۴. پیش‌بینی جمعیت استان فارس برای سال‌های ۱۳۸۶-۱۴۰۰ به روش شبکه عصبی ۵۰

جدول ۳-۵. پیش‌بینی جمعیت شهرستان استهبان برای سال‌های ۱۳۸۶-۱۴۰۰ به روش شبکه عصبی ۵۲

جدول ۳-۶. مقایسه‌ی پیش‌بینی جمعیت شهرستان استهبان برای سال‌های ۱۳۸۶-۱۴۰۰ به روش‌های رشد و شبکه عصبی ۵۴

جدول ۳-۷. پیش‌بینی جمعیت شهرستان اقلید برای سال‌های ۱۳۸۶-۱۴۰۰ به روش شبکه عصبی ۵۵

جدول ۳-۸. مقایسه‌ی پیش‌بینی جمعیت شهرستان اقلید برای سال‌های ۱۳۸۶-۱۴۰۰ به روش‌های رشد و شبکه عصبی ۵۷

جدول ۳-۹. پیش‌بینی جمعیت شهرستان جهرم برای سال‌های ۱۳۸۶-۱۴۰۰ به روش شبکه عصبی ۵۸

جدول ۳-۱۰. مقایسه‌ی پیش‌بینی جمعیت شهرستان جهرم برای سال‌های ۱۳۸۶-۱۴۰۰ به روش‌های رشد و شبکه عصبی ۶۰

جدول ۳-۱۱. پیش‌بینی جمعیت شهرستان سپیدان برای سال‌های ۱۳۸۶-۱۴۰۰ به روش شبکه عصبی ۶۱

جدول ۳-۱۲. مقایسه‌ی پیش‌بینی جمعیت شهرستان سپیدان برای سال‌های ۱۳۸۶-۱۴۰۰ به روش‌های رشد و شبکه عصبی ۶۳

جدول ۳-۱۳. پیش‌بینی جمعیت شهرستان شیراز برای سال‌های ۱۳۸۶-۱۴۰۰ به روش شبکه عصبی ۶۴
جدول ۳-۱۴. مقایسه‌ی پیش‌بینی جمعیت شهرستان شیراز برای سال‌های ۱۳۸۶-۱۴۰۰ به روش‌های رشد و شبکه عصبی..... ۶۶
جدول ۳-۱۵. پیش‌بینی جمعیت شهرستان نی‌ریز برای سال‌های ۱۳۸۶-۱۴۰۰ به روش شبکه عصبی ۶۷
جدول ۳-۱۶. مقایسه‌ی پیش‌بینی جمعیت شهرستان نی‌ریز برای سال‌های ۱۳۸۶-۱۴۰۰ به روش‌های رشد و شبکه عصبی ۶۹
جدول ۴-۱. اختلاف جمعیت پیش‌بینی شده بر اساس مدل‌های ارائه شده توسط شبکه عصبی با جمعیت سرشماری ۷۲
جدول ۴-۲. رتبه‌بندی مدل‌های پیش‌بینی توسط شبکه عصبی بر اساس کمترین اختلاف جمعیت پیش‌بینی شده با جمعیت سرشماری شهرستان‌های منتخب ۷۳
جدول ۴-۳. نرخ رشد جمعیت استان فارس طی سه سال سرشماری ۱۳۶۵-۱۳۸۵ ..... ۷۶
جدول ۴-۴. برآورد جمعیت استان فارس برای سال‌های ۱۳۷۶-۱۳۷۹ به روش همبستگی نسبتی و میزان‌های حیاتی ۷۷
جدول ۴-۵. پیش‌بینی جمعیت استان فارس برای سال‌های ۱۳۷۶-۱۳۷۹ به روش میزان رشد ۷۷

فصل اول

مقدمه

۱-۱- مقدمه

مهم‌ترین منابع جمعیتی که از دیرباز در بررسی‌ها و مطالعات جمعیتی به کار می‌رفته است، عبارت‌اند از سرشماری و ثبت وقایع چهارگانه. در جمعیت شناسی، شناخت کم و کیف جمعیت، ابعاد جمعیت، افـزایش جمعیـت، حرکـات جمعیت، باروری، مرگ و میر، مهاجرت و... مورد توجه هستند. از این روی برخی سازمآن‌هـای رسـمی از قبیـل مراکز آمارهای ملی، مؤسسه‌های جمعیت شناسی، ادارات ثبت احوال و سازمآن‌هـای نمونـه‌گیـری آمـاری؛ بـرای گردآوری داده‌های مربوط به جمعیت و تبدیل آن‌ها به اطلاعات آماری در همه کشورهای جهان به وجود آمده‌اند.

بررسی در زمینه جمعیت و پرداختن به ابعاد مختلف جمعیتـی، گذشـته را مـنعکس و تحـول جمعیـت و چگـونگی حرکت‌های آینده جمعیت را ارزیابی می‌کند. به علاوه تحلیلگران جمعیت امـروزه بـا ارائـه شـاخص‌هـا و ارزیـابی داده‌های جمعیتی، سهم عمده‌ای در امر توسعه به عهده دارند.

اگرچه شناخت تعداد مطلق جمعیت، در جمعیت شناسی، دارای اهمیت است ولی در تجزیه و تحلیل اغلب از میزان و نسبت پدیده‌های جمعیتی چون میزان رشد سالانه جمعیت، میـزان مـرگ و میـر، میـزان موالیـد و... اسـتفاده می‌شود تا قابلیت مقایسه امکان‌پذیر باشد.

پیش‌بینی جمعیت یک ناحیه، استان یا کشور، با توجه به تحولات سیاسـی و اجتمـاعی کـه در هـر جامعـه‌ای رخ می‌دهد می‌تواند بسیار مهم باشد. لذا اگر روش یا روش‌هایی دقیق‌تر و یا معادل با روش‌های موجود وجـود داشـته باشد مورد استقبال محققین خواهد بود.

امروزه، به کارگیری مدل‌های آماری نظیر مدل‌های رگرسـیون خطـی و غیـر خطـی، تحلیـل خوشـه‌ای، تحلیـل تشخیصی، تحلیل سری‌های زمانی و ... در امور پژوهشی که توسط محققین رشته‌هایی آماری و سایر رشته‌های انجام می‌گیرد امری اجتناب ناپذیر است. این مدل‌ها با دو هدف تعیین روابط بین متغیرها و پیش‌بینی انجام می‌شوند. انتخاب روش مدل‌سازی و تحلیل آن به ماهیت متغیرها و شرایط حاکم بر مسئله بستگی دارد. بـرای مثـال زمانی که پیش‌بینی یک پاسخ مورد نظر است از روش‌های رگرسیونی می‌توان برای دستیابی بـه اهـداف اسـتفاده کرد.

به‌کارگیری اکثر مدل‌ها مستلزم در نظر گرفتن و برقراری پذیره‌هایی است که گاهی به آسانی برقرار نیستند. بدین معنی که ساختار و پیچیدگی متغیرها به‌گونه‌ای است که می‌تواند بر مناسبت مدل اثرگذار باشد. ایـن وضـعیت بـه دلایل زیر می‌تواند رخ دهد:

- تعداد بالای متغیرهای پیش‌بین؛

- وجود اثر متقابل بین متغیرهای پیش‌بین؛
- وجود همبستگی بین برخی از متغیرهای پیش‌بین مورد بررسی؛
- ناهمگنی و نرمال نبودن توزیع متغیر پاسخ؛
- غیرخطی بودن رابطه‌ی بین متغیر پاسخ و متغیرهای پیش‌بین.

یکی از راه‌های برطرف کردن چنین مشکلاتی، به‌کارگیری شبکه‌های عصبی مصنوعی[1] است که به عنوان یک روش معادل با مدل‌های آماری می‌توانند به کار گرفته شوند.

۱-۲- تاریخچه و نظریه‌های مربوط به موضوع پژوهش

بعضی از پیش زمینه‌های شبکه‌های عصبی را می‌توان به اواخر قرن نوزدهم و اوایل قرن بیستم برگرداند. در این دوره کارهای اساسی در فیزیک، روان شناسی و نروفیزیولوژی توسط دانشمندانی چون هرمان فاندر هلمهلتز، ارنست ماخ و ایوان پاولف صورت گرفت. این کارهای اولیه عموماً بر تئوری‌های کلی یادگیری، بینایی و شرطی تأکید داشته‌اند و اصلاً به مدل های مشخص ریاضی عملکرد نورون‌ها اشاره‌ای نداشته‌اند.

دیدگاه جدید شبکه‌های عصبی در دهه ۴۰ قرن بیستم شروع شد، زمانی که وارن مک کولاچ و والتر پیتز نشان دادند که شبکه‌های عصبی در اصل می‌توانند هر تابع حسابی و منطقی را محاسبه نمایند. کار این افراد را می‌توان نقطه شروع حوزه علمی شبکه‌های عصبی مصنوعی نامید و این موضوع با دونالد هب ادامه یافت. کسی که عمل شرط گذاری مکانیسمی را جهت یادگیری نورون‌ها ارائه داد.

نخستین کاربرد عملی شبکه‌های عصبی در اواخر دهه ۵۰ قرن بیستم مطرح شد. در آن زمان فرانک روزنبلات در سال ۱۹۵۸ شبکه پرسپترون را معرفی نمود. وی و همکارانش شبکه‌ای ساختند که قادر بود الگوها را از هم شناسایی کند. در همین زمان بود که برنارد ویدرو در سال ۱۹۶۰ شبکه عصبی تطبیقی خطی آدالاین را با قانون یادگیری جدید مطرح نمود که از لحاظ ساختار شبیه شبکه پرسپترون می‌باشد.

هر دوی این شبکه‌های پرسپترون و آدالاین دارای این محدودیت ذاتی بودند که تنها توانایی طبقه‌بندی الگوهایی را داشتند که به طور خطی از هم متمایز باشند. روزنبلات و ویدرو، هر دو از این امر آگاه بودند، چون آن‌ها قانون یادگیری را برای شبکه‌های عصبی تک لایه مطرح نموده بودند که توانایی محدودی جهت تخمین توابع داشتند.

---

[1] Artificial Neural Network

هر چند آن‌ها توانستند شبکه‌های عصبی چند لایه را مطرح کنند ولی نتوانستند الگوریتم‌های یادگیری شبکه‌های تک‌لایه خود را بهبود بخشند.

پیشرفت شبکه‌های عصبی تا دهه ۷۰ قرن بیستم ادامه یافت. در سال ۱۹۷۲ تئو کوهونن و جیمز اندرسون به طور مستقل و بدون اطلاع از هم شبکه‌های عصبی جدید را معرفی نمودند که قادر بودند به عنوان «عناصر ذخیره ساز» عمل نمایند. اشتفان گروسبرگ در این دهه روی شبکه‌های «خود سازمان‌ده» فعالیت می‌کرد. علاقه به فعالیت در زمینه شبکه‌های عصبی در دهه ۸۰ در مقایسه با دهه‌های پیشین به علت عدم بروز ایده‌های جدید و نبود رایانه‌های سریع جهت پیاده‌سازی، کمرنگ شد. با این وجود، در خلال دهه ۸۰ رشد تکنولوژی میکرو پروسسورها روند صعودی طی می‌کرد و تحقیقات روی شبکه‌های عصبی فزونی یافت و ایده‌های بسیار جدیدی مطرح شدند.

ایده‌های نو و تکنولوژی بالا برای رنسانس در شبکه‌های عصبی کافی بود. در این زایش مجدد شبکه‌های عصبی دو نگرش جدید قابل تأمل می‌باشد. ایده اول، استفاده از مکانیسم تصادفی جهت توضیح عملکرد یک طبقه وسیع از «شبکه‌های بازخوردی[۱]» که می‌توان از آن ها برای ذخیره سازی اطلاعات استفاده نمود. این ایده توسط جان هاپفیلد فیزیک دان آمریکایی در سال ۱۹۸۲ مطرح شد. دومین ایده مهم که کلید توسعه شبکه‌های عصبی در دهه ۸۰ شد، «الگوریتم پس انتشار خطا[۲]» می‌باشد که توسط دیوید راملهارت و جیمز مک‌کلند در سال ۱۹۸۶ مطرح گردید.

با بروز این دو ایده شبکه‌های عصبی متحول شدند. در ده سال اخیر هزاران مقاله درباره شبکه‌های عصبی نوشته شده است و کاربردهای زیادی در رشته‌های مختلف علوم پیدا کرده‌اند. شبکه‌های عصبی در هر دو قسمت تئوری و عملی پیشرفت خوبی داشته‌اند. آن چه که باید به آن توجه داشت این است که روند رشد شبکه‌های عصبی آهسته و مطمئن نبوده است، دوره‌هایی بوده که خیلی رشد نموده و دوره‌هایی هم رشد خوبی نداشته و دچار رکود شده است. بیشتر پیشرفت ها در شبکه‌های عصبی به ساختار های نوین و روش های یاد گیری جدید مربوط می‌شود.

ممکن است این سؤال به ذهن برسد که در آینده چه اتفاقی رخ خواهد داد؟ برای مثال، آیا در ده سال آینده شبکه‌های عصبی جایگاه خاصی به عنوان یک ابزار علمی– مهندسی خواهند یافت یا بر عکس کمرنگ شده و دیگر چندان مطرح نیست؟ آن‌چه که در حال حاضر می‌توان گفت این است که شبکه‌های عصبی جایگاه مهمی

---

[۱] Feed forward Networks
[۲] Back Propagation Algorithm

خواهند داشت، نه به عنوان یک جواب و راه حل برای هر نوع مسأله، بلکه به عنوان یک ابزار علمی که بتواند برای راه حل‌های خاص و مناسب مورد استفاده قرار گیرد. باید به یاد داشت که هم اکنون اطلاعات موجود درباره عملکرد مغز بسیار محدود است، در نتیجه مهم‌ترین پیشرفت‌ها در شبکه‌های عصبی در آینده مطرح خواهند شد چراکه انتظار می‌رود اطلاعات بیشتری از چگونگی عملکرد مغز و نورون‌های بیولوژیکی به دست آید.

۱-۳- نوشتگان موضوع پژوهش

محققین مختلف روش‌های مختلفی را برای برآورد و پیش‌بینی جمعیت استفاده کرده‌اند و نقش متغیرهایی از جمله مهاجرت، ازدواج، مرگ و میر و ... را در پیش‌بینی جمعیت مورد بحث و بررسی قرار داده‌اند.

راجل(۱۹۹۲) نقش مهاجرت، ازدواج و مرگ و میر را به عنوان منابع تصحیح تورش تبیین کرد. کیلمن (۲۰۰۲)، توضیح داد که چگونه ساختارهای همبسته در داده‌ها در تغییر جمعیت از کشوری به کشور دیگر اثر دارد. یوآن (۲۰۰۶) شبکه‌ی عصبی را در تحلیل اثر جمعیت بر ساختار صنعتی به کار گرفت. سوراجیت (۲۰۰۶) از یک مدل پرسپترون چند لایه با قانون یادگیری دلتای تعمیم یافته برای پیش‌بینی جمعیت هند استفاده کرد. آن‌ها برای پیش‌بینی از جمعیت زن و مرد، زنان متأهل، متوسط امید به زندگی در زنان و مردان، نرخ باروری کل استفاده کردند. در عین حال از زیرگروه‌های سنی با فواصل متقاطع نیز استفاده کردند. آن‌ها با به‌کارگیری یک شبکه‌ی عصبی مصنوعی با الگوریتم پس‌انتشار خطا و مبتنی بر داده‌های سال‌های ۱۹۹۵-۱۹۹۷ پیش‌بینی جمعیت برای سال ۲۰۰۰ را انجام دادند.

۱-۴- بیان مسأله و اهمیت پژوهش

پیش‌بینی روند جمعیت، مسأله‌ی پیچیده‌ای است و جمعیت هرکشوری به متغیرهای مختلفی وابسته است. الگو و ساختار جمعیت در هر کشوری تحت تأثیر عوامل اجتماعی، سیاسی و فرهنگی قرار دارد به طوری که هر یک از این عوامل می‌تواند موجب تغییراتی در الگوی پیش‌بینی کننده‌های جمعیت شود.

تا کنون مطالعات انجام شده در مورد پیش‌بینی جمعیت، مبتنی بر روش‌های آماری یا تکنیک‌های ریاضی بوده است. همان‌طور که می‌دانید، به‌کارگیری مدل‌های مختلف پیش‌بینی مستلزم برقرای برخی پذیره‌ها است و اگر برقرار نباشند استفاده از مدل‌های پیش‌بینی نادرست خواهد بود و در صورت استفاده، دارای خطای پیش‌بینی بالا خواهد بود. در مورد مدل‌های رگرسیونی که برای پیش‌بینی استفاده می‌شوند بیان این نکته لازم است که زمانی

نتایج مدل‌های رگرسیونی معتبر است که پیش‌فرض‌هایی نظیر نرمالیتی، هم‌واریانسی و استقلال مانده‌ها برقرار باشد و همچنین مدل باید خطی باشد و بین متغیرهای مستقل هم‌خطی وجود نداشته باشد. این پیش‌فرض‌ها در عمل بسیار محدود کننده هستند و ممکن است در وضعیت‌های عملی برقرار نباشد. تکنیک ANN نیاز به برقراری این پیش‌فرض‌های دست و پا گیر ندارد. در عین حال، مدل‌های پیشگو نیاز دارند که از توان پیش‌گویی بالایی برخوردار باشند که در مدل‌های رگرسیونی برای این منظور، لازم است متغیرهای کمکی با درجات بالاتر را در مدل وارد نمود و این کار را باید به صورت آزمون و خطا انجام داد تا مدل به پیچیدگی لازم رسیده و توان پیش‌بینی لازم را داشته باشد. ولی مدل شبکه‌ی عصبی مصنوعی به طور خودکار در قالب مدل‌های خطی و غیر خطی قابل فراگیری (آموزش) است تا به پیچیدگی لازم برای پیش‌بینی مطلوب دست یابد.

با توجه به این که فرآیند پیش‌بینی یک فرآیند دو مرحله‌ای (پیش‌بینی متغیرهای کمکی و سپس جمعیت) است؛ بنابراین نیاز به توان بالای پیش‌بینی مدل و خطای کوچک است تا خطای ایجاد شده در این فرآیند دو مرحله‌ای، به اندازه‌ی کافی کوچک و نتایج حاصل از آن دقت لازم را داشته باشد.

در عین حال، برخی از مدل‌های ریاضی و جمعیت‌شناسی که برای پیش‌بینی استفاده می‌شوند فاقد به‌کارگیری متغیرهای کمکی هستند. ولی مدل شبکه‌ی عصبی مصنوعی مانند مدل‌های رگرسیونی با لحاظ کردن متغیرهای کمکی، سعی در بهترین پیش‌بینی دارد با این تفوت که شبکه‌ی عصبی مصنوعی اثرات پیچیده و غیر خطی بین متغیرهای کمکی را نیز در نظر می‌گیرد.

در واقع، شبکه‌های عصبی مصنوعی جزء آن دسته از سیستم‌های دینامیکی هستند که با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل می‌کنند. به همین خاطر به این سیستم‌ها، هوشمند می‌گویند. چرا که بر اساس محاسبات روی داده‌های عددی یا مثال‌ها، قوانین کلی را فرا می‌گیرند.

مدل‌های شبکه‌ی عصبی مصنوعی با در نظر گرفتن کمترین مفروضات به بررسی متغیرها و ساختار داده‌ها می‌پردازد و لذا به عنوان روشی ناپارامتری برای مدل‌سازی داده‌ها در نظر گرفته می‌شوند و می‌توانند برای مدل‌های با ارتباطات پیچیده مورد استفاده قرار گیرند. از نقطه نظر آماری، شبکه‌های عصبی به این دلیل مورد توجه هستند که به عنوان یک سامانه‌ی داده‌پردازی به طور بالقوه‌ای توانایی پیش‌بینی و دسته‌بندی اطلاعات را دارا هستند. برخی از حوزه‌های کاربرد شبکه عصبی در موارد زیر است:

- وجود همبستگی ناشناخته بین متغیرهای موجود؛
- عدم وجود راه حل الگوریتمی برای مسائل مورد بررسی؛

- وجود داده‌های پیچیده در مسئله‌ی مورد بررسی؛
- پیش‌بینی سری‌های زمانی؛
- پیش‌بینی شاخص‌های اقتصادی؛
- طبقه‌بندی، شناسایی و تشخیص الگو؛
- پیش‌بینی آب و هوا و . . . .

در پیش‌بینی جمعیت یک ناحیه، استان یا کشور، اگر روش یا روش‌هایی دقیق‌تر و یا معادل با روش‌های موجود وجود داشته باشد باید بررسی شده و با روش‌های موجود مقایسه شود. چراکه پیش‌بینی‌های بهتر و دقیق‌تر مورد نیاز و استقبال دستگاه‌های استفاده کننده است. در هر حال، استفاده از شبکه‌ی عصبی مصنوعی در زمینه‌ی پیش‌بینی می‌تواند مفید و در مواردی که مشکل استفاده از مدل‌های رگرسیونی وجود دارد، راه‌گشا باشد.

در این پژوهش، سعی شده است که امکان به‌کارگیری از مدل شبکه‌ی عصبی مصنوعی در زمینه‌ی پیش‌بینی جمعیت (استان فارس) بررسی شود. با توجه به تعدد شهرستآن‌ها در استان فارس و همچنبن به دلیل اینکه مسایل جمعیتی در این استان در سطح کمی بالاتر از میانه جمعیت کشور است، از داده‌های این استان برای اعمال مدل پیشنهادی استفاده شده است. لذا هدف این پژوهش آن است که بر اساس متغیرهای پیش‌بین سال‌های قبل، امکان پیش‌بینی جمعیت آتی استان فارس (و همچنین شهرستان‌های این استان) را بر اساس مدل شبکه‌ی عصبی مصنوعی مورد بررسی قرار دهد.

۱-۵- هدف‌های پژوهش

الف) هدف‌های کلی

۱- امکان‌سنجی پیش‌بینی جمعیت استان فارس تا سال ۱۴۰۰ با استفاده از شبکه‌ی عصبی مصنوعی؛

۲- امکان‌سنجی پیش‌بینی جمعیت شهرستان‌های استان فارس تا سال ۱۴۰۰با استفاده از شبکه‌ی عصبی مصنوعی.

ب) هدف ویژه

۱-مقایسه‌ی پیش‌بینی انجام شده با شبکه‌ی عصبی مصنوعی و پیش‌بینی‌های انجام شده توسط مرکز آمار ایران.

۱-۶- پرسش‌های پژوهش

۱- آیا جمعیت استان فارس را می‌توان با استفاده از روش شبکه‌ی عصبی مصنوعی پیش‌بینی کرد؟

۲- آیا جمعیت شهرستان‌های استان فارس را می‌توان با استفاده از روش شبکه‌ی عصبی مصنوعی پیش‌بینی کرد؟

۱-۷- تعریف مفاهیم

داده: صفت یا اطلاعی، اغلب عددی، که از طریق مشاهده گردآوری می شود (واژه نامه OECD).

داده‌های ثبتی[۱]**:** داده‌هایی هستند که در حین انجام فعالیت‌های عینی، واقعی و جاری یک سازمان و با استفاده از فرم‌ها، اسناد، مدارک و فایل های کامپیوتری جاری در داخل سازمان ثبت پیش‌بینی.

اطلاعات[۲]**:** به اشکالی از داده‌های مرتب، خلاصه و به‌صورت گزارشات مفید درآمده‌ای که در نظام ذخیره شده است اطلاق می‌شود.

نماگر: متغیر کمی است که بر اساس روابط تابعی، داده‌های آماری اعم از داده‌های خرد وکلان را برای زمان مشخص، مکان و سایر ویژگی‌ها بیان می‌کند.

شاخص: نماگری است که تغییرات را به‌صورت نسبی (نسبت به یک مبداء یا زمان پایه یا در فضای اندازه‌ی آن) اندازه‌گیری و به‌صورت یک کمیت بدون واحد بیان می‌شود.

سرشماری[۳]**:** سرشماری عبارت است از بررسی خصوصیات و ویژگی‌های موردنیاز تمامی واحدهای جامعه آماری موردنظر. از جمله نمونه‌های آن می‌توان به «سرشماری عمومی نفوس و مسکن»، «سرشماری کشاورزی» و «سرشماری کارگاهی» اشاره کرد. همان‌گونه که در تعریف مشخص شد، در آمارگیری به روش سرشماری به تمامی اعضای جامعه مورد نظر مراجعه، مشخصات و صفات لازم از آن‌ها پرسش و بررسی می‌شود.

از جمله ویژگی‌های مهم این روش تهیه آمار می‌توان به موارد زیر اشاره کرد:

- دقت زیاد،

---

[۱] Register Data

[۲] Information

[۳] Census

– هزینه بالا،

– مقطعی بودن آمار حاصل،

– وجود خطاهای غیرنمونه‌ای،

– چارچوب سازی برای مطالعات نمونه‌ای،

– اجرا در فواصل زمانی ۵ یا۱۰ ساله.

نمونه‌گیری احتمالی[۱]: به روش علمی انتخاب برخی از واحدهای آماری، بررسی خصوصیات، صفات و ویژگی‌های موردنظر در آن‌ها و تعمیم نتایج به جامعه با پذیرش درصدی خطا، نمونه‌گیری احتمالی گویند.

از جمله مثال‌های طرح‌های نمونه‌گیری در کشور می‌توان به «طرح هزینه و درآمد خانوار»، «طرح نیروی کار» و «طرح آمارگیری صنعتی» اشاره کرد.

از جمله ویژگی‌های مهم این روش تهیه آمار عبارت‌اند از:

– افزایش دقت با افزایش تعداد نمونه،

– افزایش هزینه با افزایش تعداد نمونه،

– کمتر بودن زمان انجام پردازش و اعلام نتایج،

– کلیه نتایج حاصله جنبه براورد (علمی) دارد،

– وجود خطاهای نمونه‌ای، و

– وجود برخی خطاهای غیرنمونه‌ای (اما با اندازه‌ای به مراتب کمتر نسبت به سرشماری).

برآورد و پیش‌بینی:

برای دو واژه‌ی برآورد و پیش‌بینی باید تمایز قائل شد؛ برآورد برای سال‌های گذشته به کار می‌رود که اطلاعات هر چند محدودی از بعضی شاخص‌های جمعیتی یا متغیرهای کمکی مرتبط با جمعیت موجودند. این اطلاعات می‌توانند از آمارهای ثبتی یا طرح‌های نمونه‌گیری به دست آمده باشند. در مقابل، پیش‌بینی به آینده اشاره دارد و بنا بر ماهیت خود هیچ اطلاع مطمئنی از شاخص‌های جمعیتی برای آن موجود نیست و فقط براساس مشاهدات و اطلاعات گذشته می‌توان مقادیر محتملی را برای شاخص‌های مورد نظر فرض کرد و تحت این مفروضات

---

[۱] Sampling