فصلنامه پژوهشنامه بازرگانی، شماره ۱۰۵، زمستان ۱۴۰۱، ۲۲۰-۱۹۱

مقاله پژوهشی: ارائه یک سیستم نوین تخمین قیمت برای تعداد زیاد کالا در فروشگاه‌های آنلاین (مورد مطالعه: شرکت نگین پولاد نقش جهان)

امیرصادق میرگلوی بیات* علی شجاع**

بهلول ابراهیمی*** محمدحسین معماری**** کاظم نسب حاجی*****

دریافت: ۱۴۰۱/۰۶/۲۳ پذیرش: ۱۴۰۱/۰۷/۱۹

تخمین قیمت / رگرسیون ابتکاری / ماتریس همبستگی / فروشگاه‌های آنلاین / آهن آنلاین

چکیده

با توجه به تعداد زیاد کالاها در فروشگاه‌های آنلاین، تغییر روزانه قیمت کالاها و عدم امکان استعلام و ثبت این حجم از قیمت‌ها توسط اپراتورها، توسعه یک سیستم تخمین قیمت برای پیش‌بینی روزانه قیمت کالاها با دقت بالا بسیار ضروری می‌باشد. برای این منظور، در مقاله حاضر یک سیستم نوین تخمین قیمت با استفاده از ماتریس همبستگی و رگرسیون خطی، توسعه داده شده است. با توجه به خطای بالای رگرسیون خطی در پیش‌بینی قیمت‌ها، یک روش رگرسیون خطی ابتکاری جهت افزایش دقت پیش‌بینی و کاهش خطای میانگین بر مبنای تعیین اهمیت داده‌ها بر حسب زمان توسعه داده شده است. سیستم توسعه داده شده بر روی داده‌های واقعی شرکت نگین پولاد نقش جهان (آهن آنلاین) پیاده‌سازی شده است که نتایج حاکی از دقت

*. کارشناسی ارشد علوم کامپیوتر، دانشگاه امیرکبیر، تهران shoja.email@gmail.com.

**. دکترای مهندسی صنایع، شرکت نگین پولاد نقش جهان، o.rezvnaian@imps.ac.ir

***. دکترای مهندسی صنایع، شرکت نگین پولاد نقش جهان b.ebrahimi@aut.ac.ir

****. کارشناسی ارشد مدیریت بازرگانی، شرکت نگین پولاد نقش جهان hosein.memari@gmail.com

*****. کارشناسی ارشد مهندسی صنایع، دانشگاه صنعتی شریف، تهران ali.kazemnasabhaji@ie.sharif.edu

■ بهلول ابراهیمی، نویسنده مسئول.

بالای قیمت‌های تخمینی این مدل می‌باشد. همچنین نتایج مدل ارائه شده با در نظر گرفتن تغییرات نرخ دلار مورد بررسی و تحلیل قرار گرفته است. نتایج حاکی از آن است که در اکثر اوقات رگرسیون ابتکاری ارائه شده خطای کمتری نسبت به سایر روش‌های مورد بررسی دارد.

طبقه‌بندی C13, E64 :JEL

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مقدمه

با توجه به رشد چشمگیر خرید اینترنتی در سال‌های اخیر و اقبال مشتریان به استفاده از سایت و اپلیکیشن‌های خرید آنلاین و تنوع بسیار زیاد کالاها، لزوم تخمین قیمت کالاها برای فروشندگان بسیار حائز اهمیت است[1]. اهمیت این موضوع در زمینه‌ی قیمت‌گذاری آنلاین آهن آلات با توجه به تنوع کالاها و ارتباط قیمت این کالاها با یکدیگر و همچنین نوسان بالای قیمت آهن در بازار به خوبی حس می‌شود[2]. تنوع این کالاها شامل انواع میلگرد، ورق، لوله، پروفیل، نبشی، ناودانی، تیرآهن و استیل در سایزها، برندها و محل‌های بارگیری مختلف می‌باشد که طیف گسترده‌ای از کالاها را در بر می‌گیرد (بالغ بر ۵۰۰۰ کالا در بیش از ۸۰ دسته‌بندی کالای مختلف). تخمین قیمت تقریبی این کالاها می‌تواند به فروشگاه‌های آنلاین آهن که مرجع قیمت‌گذاری آهن می‌باشند کمک کند تا روزانه تخمین مناسبی از قیمت انواع کالاها را به مشتری ارائه دهند. با توجه به تغییرات لحظه‌ای و روزانه قیمت‌ها استعلام روزانه قیمت تمام این ۵۰۰۰ کالا از منابع و ثبت دستی آن‌ها برای اپراتورها کاری زمان بر و عملاً غیرممکن است؛ لذا تخمین قیمت می‌تواند در این زمینه کمک زیادی به اپراتورها بکند. در راستای مباحث مطرح شده پژوهش حاضر تلاش می‌کند تا به سوالات زیر پاسخ دهد:

- آیا نرخ دلار در تخمین قیمت کالاهای آهنی موثر است؟
- آیا ضریب اهمیت بازه زمانی داده‌ها تأثیری در دقت تخمین قیمت‌ها دارد؟
- از بین روش‌های رگرسیون، شبکه عصبی و درخت تصمیم کدام روش برای تخمین قیمت کالاهای شرکت نگین پولاد نقش جهان مناسب‌تر است؟

شایان ذکر است که پژوهش حاضر ضمن پاسخ‌گویی به سوالات فوق دو هدف مهم زیر را دنبال می‌کند:

- تعیین کالاهای پایه جهت استفاده برای تخمین قیمت سایر کالاها
- تخمین قیمت کالاهای مختلف (بالغ بر ۵۰۰۰ کالا) با استفاده از قیمت کالاهای پایه با دقت بالا

بنابراین در این مقاله یک مدل تخمین قیمت برای کالاها با تنوع بسیار بالا ارائه شده است.

1. Chen, W., et al., (2021).
2. Zhang, H. et al., (2021).

در هر یک از دسته‌بندی‌های کالاهای مربوط به آهن، تعدادی کالا هستند که روند قیمتی آن‌ها در سال‌های مختلف، همبستگی بالایی با قیمت سایر کالاهای آن دسته‌بندی داشته است. در این مقاله ابتدا برای هر دسته‌بندی، تعدادی کالا را که روند قیمتی آن‌ها همبستگی بالایی با سایر کالاهای آن دسته‌بندی داشته به‌عنوان کالاهای پایه انتخاب می‌کنیم؛ به‌طوریکه قیمت سهم قابل توجهی از کالاهای دیگر آن دسته‌بندی را بتوان با قیمت این کالاهای پایه تخمین زد. در ادامه ابتدا روش رگرسیون خطی برای تخمین قیمت پیشنهاد می‌شود. با توجه به تغییرات سریع قیمت‌ها، داده‌های قدیمی ارزش کمتری نسبت به داده‌های جدید دارند، لذا برای در نظر گرفتن اهمیت داده‌ها رگرسیون خطی را بهبود داده و یک رگرسیون خطی ابتکاری ارائه می‌دهیم که در اکثر موارد جواب بهتری نسبت به رگرسیون خطی ساده می‌دهد. مدل پیشنهادی، به منظور تخمین قیمت، بر روی داده‌های واقعی شرکت نگین پولاد نقش جهان (آهن آنلاین) اعمال شد. در ادامه نرخ دلار نیز به‌عنوان یک متغیر مستقل در محاسبات مدل پیشنهادی اضافه شد که نتایج حاکی از بهبود نتایج مدل می‌باشد. در نهایت می‌توان گفت که با در دست داشتن قیمت کالاهای پایه، می‌توان قیمت درصد قابل توجهی از کالاهای دیگر را با دقت مطلوب و به‌صورت روزانه تخمین زد. به‌طور خلاصه سهم علمی مقاله را می‌توان به‌صورت زیر بیان نمود.

۱. توسعه روش رگرسیون خطی ابتکاری برای کاهش خطای پیش‌بینی قیمت برحسب قیمت کالاهای پایه.
۲. تخمین قیمت تعداد بسیار زیاد کالا به‌صورت روزانه که می‌تواند در تمامی فروشگاه‌های آنلاین مشابه که دارای حجم زیادی از کالا و خدمات هستند مورد استفاده قرار گیرد.
۳. استفاده از ماتریس همبستگی جهت تعیین تعداد بهینه کالاهای پایه و حداقل همبستگی قابل قبول آن‌ها.
۴. در نظر گرفتن تأثیر تغییرات نرخ دلار در مدل پیشنهادی، جهت پیش‌بینی قیمت.

ادامه مقاله به‌صورت زیر سازماندهی شده است: در بخش دوم مبانی نظری تحقیق آورده شده است. در بخش سوم مرور ادبیات موضوع بیان شده است. در بخش چهارم ابتدا پیش‌پردازش داده‌ها انجام شده است و سپس انتخاب کالاهای پایه با در نظر گرفتن معیار همبستگی در دو

زیربخش «تعیین حداقل همبستگی قابل قبول» و «تعیین تعداد مناسب کالای پایه» آورده شده است. در بخش پنجم سیستم تخمین قیمت کالاها توضیح داده شده است. در این بخش ابتدا رگرسیون خطی ساده توضیح داده شده است، سپس یک روش ابتکاری برای کاهش خطای آن ارائه شده است. در بخش ششم مدل تخمین قیمت ارائه شده با داده‌های واقعی شرکت تشریح و نتایج آن با روش‌های دیگر مقایسه شده است. نتیجه‌گیری و مطالعات آتی نیز در بخش هفتم آورده شده است.

۱. مبانی نظری تحقیق

داده‌کاوی مجموعه‌ای از فنون است که به شخص امکان می‌دهد تا ورای داده‌پردازی معمولی حرکت کند و به استخراج اطلاعاتی که در انبوه داده‌ها مخفی یا پنهان است کمک می‌کند. یادگیری ماشین گاهی اوقات با داده‌کاوی ادغام می‌شود. تمرکز این زیرشاخه بر تحلیل اکتشافی داده‌ها است و با عنوان یادگیری بی‌نظارت شناخته می‌شود[۱].

توانایی دیجیتالی کردن جوامع که بتوان آن را در هر بخش از زندگی بشر به کار برد، یکی از عوامل تعیین‌کننده پیشرفت اجتماعی و اقتصادی در حال و آینده است[۲]. امروزه، معاملات آنلاین در فروشگاه‌ها، اقتصاد جهانی را هدایت می‌کند[۳]. بسیاری از شرکت‌های جدید با فناوری‌های دیجیتالی ظاهر می‌شوند. این شرکت‌ها خدماتی را با استفاده از اپلیکیشن‌های جدید برای خرید مشتریان ارائه می‌دهند. بازار تجارت الکترونیک یکی از ابزارهای خرید از طریق اینترنت است[۴]. دلایل اصلی افزایش تمایل به خرید آنلاین نفوذ اینترنت، افزایش تعداد خرده‌فروشان آنلاین، سهولت خرید، انعطاف‌پذیری در تحویل و افزایش قدرت خرید است[۵]. از آنجایی که خرید اینترنتی امکان خرید بدون مراجعه به فروشگاه را فراهم کرده است؛ امروزه اکثر مصرف‌کنندگان خرید اینترنتی را انتخاب می‌کنند. بنابراین هر شرکتی باید یک فروشگاه آنلاین به منظور افزایش خدمات مشتری داشته باشد[۶]. بنابراین صاحبان مشاغل می‌توانند از اینترنت برای فروش محصول

1. Witten, I. H., & Frank, E. (2002).
2. Ejdys, J., et al., (2019).
3. Hayton, J. W., & Blundell, M. (2021).
4. Budiantara, M., et al., (2019)
5. SEITZ, C., et al., (2017)
6. Chaffey, D. (2022).

خود از کسب و کارهای رسمی به یک مدل کسب و کار جدید، کسب و کارهای خرده‌فروشی الکترونیکی و جذب مشتریان بیشتر استفاده کنند[1].

با توجه به تغییرات سریع قیمت‌ها و تعداد زیاد کالاها در فروشگاه‌های آنلاین، نیاز است از روش‌های تخمین قیمت برای تخمین قیمت کالاها استفاده شود. از جمله این روش‌ها می‌توان به رگرسیون و شبکه عصبی اشاره نمود.

در تحلیل رگرسیونی رابطه‌ی متغیر یا متغیرهای مستقل با متغیر وابسته نشان داده می‌شود. ساده‌ترین مدل رگرسیون، مدل رگرسیون خطی است که در آن متغیر مستقل و وابسته دارای رابطه خطی می‌باشند. مدل رگرسیون را اولین بار ودربورن و فلدر در سال ۱۹۷۲ معرفی کردند.[2] با توسعه مدل رگرسیون خطی ساده، دیگر مدل‌های رگرسیون (شامل غیرخطی، بردار پشتیبان و ...) به‌دست می‌آید. همچنین با توجه به داده‌های مورد بررسی نیز می‌توان مدل‌های رگرسیون را دسته‌بندی کرد. مدل‌های این دسته‌بندی شامل رگرسیون سری زمانی، رگرسیون مقطعی، رگرسیون میانگین و رگرسیون تجمعی است.

در رگرسیون سری زمانی به بررسی جمعیت واحدی از داده‌ها در دوره‌های زمانی مختلف پرداخته می‌شود. در این رگرسیون، مقادیر داده‌های هر دوره متناسب با مقادیر داده‌های دوره‌های قبل است.[3] در روش رگرسیون مقطعی، معادله رگرسیون برآورد شده برای مقاطع زمانی مختلف متفاوت و مستقل از یکدیگر است. در روش‌های رگرسیون میانگین و مقطعی، مجموعه‌ای از داده‌های سری زمانی و مقطعی مورد مطالعه قرار می‌گیرد.[4] در بین این روش‌ها رگرسیون خطی به‌صورت گسترده مورد استفاده واقع شده است که در ادامه به برخی از آن‌ها اشاره می‌کنیم. در بخش بعدی منابع بیشتری مورد بررسی واقع می‌شود.

مدیرشانه‌چی و علی‌زاده (۱۳۸۵) در مطالعه خود با استفاده از شبکه عصبی و رگرسیون خطی، مدل هوشمندی را برای پیش‌بینی کوتاه مدت قیمت نفت خام ایران به‌صورت ماهیانه در طی دوره ۱۹۸۵-۲۰۰۳ ارائه کردند. نتایج این مدل شبیه‌سازی شده نشان داد که سیستم طراحی شده، پیش‌بینی بهتری در بازه‌های مختلف نسبت به مدل شبکه عصبی و رگرسیون خطی از

1. Riantini, R. E., et al., (2019).
2. Kedem B., & Fokianos K. (2004).
3. Lind D.A., et al., (2005).
4. Gujarati D.N. (1995)

خود نشان می‌دهد. خاشعی و همکارانش[1] به ارائه‌ی مدلی ترکیبی از دو روش پیش‌بینی یعنی مدل رگرسیون فازی و شبکه عصبی مصنوعی پرداخته و توانسته‌اند نقایص دو مدل را با هم ترکیب کنند و به نتایج بهتری دست یابند و مسایلی چون نیاز به میزان زیادی از داده‌ها در شبکه عصبی را با رگرسیون فازی که برای وضعیت داده‌های ناقص مناسب است حل کنند و از این روش در پیش‌بینی قیمت طلا و نرخ ارز استفاده برده‌اند. شایان ذکر است مدل ریاضی رگرسیون به‌طور مفصل در بخش ۵،۱ مورد بررسی واقع شده است.

۲. مروری بر پیشینه پژوهش

با توجه به اهمیت تخمین قیمت روش‌های مختلفی برای تخمین قیمت در ادبیات موضوع ارائه شده است که در این بخش مورد مطالعه و بررسی قرار می‌گیرند.

مسکوویچ و کیم[2] در مطالعه‌ای به بهبود قدرت کاربرد رگرسیون فازی پرداختند. در این مسیر آنان روابط میان H-Value، شکل تابع عضویت و گستره پارامترهای فازی را در مدل رگرسیون خطی فازی بررسی کردند. از کاربردهای مباحث داده کاوی و به‌طور خاص رگرسیون می‌توان به تحلیل رفتار مشتریان اشاره کرد. وو و تسنگ[3] روش رگرسیون فازی را در مدل‌سازی داده‌های اقتصادی و مالی استفاده کردند و نشان دادند که نتایج این مدل نسبت به رگرسیون کلاسیک بهتر است. تسنگ و لین[4] از رگرسیون لجستیک برای پیش‌بینی چرخه ی تجاری تولید و پیش‌بینی ورشکستگی شرکت‌ها در انگلستان استفاده کردند. پارک و چونگ[5] استفاده از داده‌های جریان‌های کلیک برای پیش‌بینی رفتارهای خرید کاربران را ارائه کردند. آن‌ها با استفاده از تحلیل رگرسیون سلسله‌مراتبی ثابت کردند که کاربران انتقال‌داده‌شده از پایگاه وب ارجاع‌دهنده، کمتر از کاربرانی که مستقیم وارد سایت می‌شوند، خرید می‌کنند و هرچه مدت زمان ماندن این کاربران در سایت بیشتر باشد و صفحاتِ مشاهده‌کرده کمتر باشد، احتمال خریدشان بیشتر است.

1. Khashei, M., et al., (2008).
2. Moskowitch, H, K. J, Kim(1993).
3. Wu, B., & Tseng, N. F. (2002).
4. Tseng, F. M., & Lin, L. (2005).
5. Park, J., & Chung, H. (2009).

آزاده و دیگران[1] به پیش‌بینی مصرف نفت برای چهار کشور ژاپن، آمریکا، استرالیا و کانادا پرداختند. بدین منظور از متغیرهای جمعیت، هزینه نفت خام وارداتی، تولید ناخالص داخلی (GDP) و میزان نفت تولیدی استفاده کردند. در این تحقیق از داده‌های سالانه برای دوره زمانی ۱۹۹۰-۲۰۰۵ استفاده شده است. سپس دقت دو مدل رگرسیون فازی و رگرسیون آماری را در پیش‌بینی مصرف نفت این کشورها با هم مقایسه کردند. نتایج به‌دست آمده حاکی از آن است که برای سه کشور کانادا، آمریکا و استرالیا رگرسیون فازی پیش‌بینی بهتری نسبت به رگرسیون آماری ارائه می‌دهد اما برای ژاپن نتایج رگرسیون آماری بهتر از رگرسیون فازی است. دشتی رحمت‌آبادی و همکاران (۱۳۹۰) تحقیقی با هدف معرفی الگوهای مطلوب پیش‌بینی برای قیمت نفت خام ایران انجام دادند. داده‌های مورد استفاده به‌صورت هفتگی و طی دوره بین سال‌های ۱۹۹۷ تا ۲۰۱۰ است و پیش‌بینی‌ها برای ۲۰، ۱۰ و ۳۰ درصد داده‌های یاد شده انجام شده است. آن‌ها در این مطالعه برای پیش‌بینی از چهار الگوی شبکه عصبی و الگوی خود رگرسیون میانگین متحرک استفاده کردند. نتایج این مطالعه نشان داد که برای پیش‌بینی ۱۰ درصد از داده‌های قیمت نفت خام، الگوهای شبکه رگرسیون تعمیم‌یافته و شبکه آبشاری پس انتشار با تابع آموزش شبه‌نیوتنی، به ترتیب با خطای کم‌تر از یک و کم‌تر از دو درصد دارای بهترین عملکرد هستند. همچنین به‌طور نسبی با افزایش درصد داده‌های مورد استفاده در پیش‌بینی، دقت پیش‌بینی‌ها به‌ویژه با افزایش از ۱۰ درصد به ۲۰ درصد رو به افول می‌رود. محققین در نهایت نشان دادند که دقت پیش‌بینی روش خود رگرسیون میانگین متحرک کم‌تر از الگوهای شبکه‌ی عصبی ارزیابی می‌شود. منصور زراءنژاد و همکاران (۱۳۹۱) در تحقیقی برای پیش‌بینی قیمت نفت خام اوپک از سه روش رگرسیون فازی، آریما و رگرسیون خودبازگشتی میانگین متحرک انباشته فازی (ترکیب دو روش مذکور) و از داده‌های روزانه قیمت نفت اوپک برای پیش‌بینی قیمت نفت خام اوپک استفاده کردند. نتایج حاکی از این است که مدل‌های رگرسیون فازی و رگرسیون خودبازگشتی میانگین متحرک انباشته فازی علاوه بر این که از نظر تمام معیارهای متداول خطای پیش‌بینی، عملکرد بهتری نسبت به مدل آریما دارند با فراهم کردن بهترین و بدترین حالت، تصمیم‌گیری را نسبت به مدل آریما تسهیل کرده است. همچنین مدل ترکیبی به مراتب پیش‌بینی بهتری نسبت به مدل رگرسیون فازی ارائه می‌دهد و فاصله بازه تصمیم‌گیری را کوتاه‌تر می‌کند.

1. Azadeh, A. , et al., (2009).

تسنگ و دیگران[1] عملکرد مدل رگرسیون خود بازگشتی انباشته فازی را با مدل‌های آریما، مدل سری زمانی چن و مدل سری زمانی واتاد به منظور پیش‌بینی نرخ ارز مقایسه کردند. آن‌ها بدین منظور از ۴۰ داده روزانه نرخ ارز (دلار تایوان به دلار آمریکا) طی دوره یک آگوست ۱۹۹۶ تا ۱۶ سپتامبر ۱۹۹۶ استفاده کردند. ۳۰ مشاهده اول برای مدل‌سازی و ۱۰ مشاهده آخر را برای ارزیابی عملکرد مدل به کار گرفتند. نتایج این مطالعه حاکی از این است که مدل رگرسیون خود بازگشتی انباشته فازی به مراتب پیش‌بینی دقیق‌تری نسبت به مدل‌های دیگر دارد.

در زمینه تخمین ارزش سهام در بازارهای مالی نیز تحقیقات گسترده‌ای انجام شده است. محمد حسن قلی‌زاده و همکاران (۱۳۹۰) در تحقیق پیش‌بینی قیمت سهام با روش رگرسیون فازی به روشی دست یافته‌اند که حاصل ادغام رگرسیون معمولی و رگرسیون فازی به همراه بهینه‌سازی و نافازی‌سازی پارامترها با الگوریتم ژنتیک می‌باشد. نتایج حاصل از به کارگیری این روش با روش رگرسیون معمولی نشان می‌دهد که خطا به‌طور قابل ملاحظه‌ای کاهش یافته است. همچنین این روش فوق العاده کم هزینه است و وقت کمتری نسبت به سایر روش‌ها می‌برد و برآورد نسبتاً بهتری انجام می‌شود. چن و همکاران[2] در تحقیقی تحت عنوان «آیا هموارسازی سود باعث تشدید یا کاهش ریسک سقوط قیمت سهام می‌شود» به بررسی ارتباط بین هموارسازی و ریسک سقوط قیمت سهام پرداختند. نتایج حاکی از وجود رابطه بین هموارسازی سود با درجه بالا و ریسک سقوط قیمت سهام در سطح بالا است. این ارتباط در شرکت‌هایی که کمتر مورد توجه تحلیلگران هستند و دارای درصد مالکیت نهادی کمتر و اقلام تعهدی انباشته مثبتی هستند، در تحلیل مقطعی تشدید می‌شود. صدرالسادات و همکاران (۱۳۹۶) در پژوهشی تحت عنوان «هموارسازی سود و خطر ریزش قیمت سهام در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران» به بررسی رابطه هموارسازی سود و خطر ریزش قیمت سهام پرداختند. نتایج نشان داد بین هموارسازی سود و خطر ریزش قیمت سهام در شرکت‌هایی با مالکیت شرکتی قوی ارتباط معناداری به‌صورت مثبت و در مورد مالکیت شرکتی ضعیف به‌صورت منفی بوده است. همچنین نتایج حاکی از نشان دادن ارتباط معناداری بین هموارسازی سود و خطر ریزش قیمت سهام در شرکت‌های با اقلام تعهدی منفی بوده است.

1. Tseng, F. M., et al., (2001).
2. Chen, C., et al., (2017).

همچنین در زمینه تخمین قیمت مسکن نیز تحقیقاتی صورت گرفته است. چاو و همکاران[1] در سال ۲۰۰۱ در هنگ کنگ و ون و همکاران[2] در چین مطالعاتی را درباره تأثیر عوامل مختلف بر قیمت مسکن انجام داده‌اند. نتایج این مطالعات حاکی از آن است که ۱۴ ویژگی از ۱۸ ویژگی مورد بررسی، اثر معناداری بر قیمت مسکن در شهر مورد مطالعه داشته است. روستایی و همکاران (۱۳۹۷) در پژوهش « تحلیل فضایی عوامل موثر بر افزایش قیمت مسکن در کلان شهر تبریز با استفاده از ضریب همبستگی و مدل برازش رگرسیونی» به بررسی متغیرهای تأثیر گذار بر قیمت مسکن در دو دسته واحدهای مسکونی ویلایی و آپارتمانی در کلان شهر تبریز با استفاده از روش همبستگی پیرسون و مدل برازش رگرسیونی پرداختند که نتایج حاصل نشان داد که متغیرهای چهارگانه، تأثیر مستقیم و معناداری بر افزایش قیمت مسکن در واحدهای مسکونی ویلایی و آپارتمانی داشته‌اند. کوه پیما و ارگانی (۱۳۹۹) در پژوهشی تحت عنوان «تخمین قیمت آپارتمان با استفاده از رگرسیون خطی و وزن دار جغرافیایی» به بررسی رابطه بین پارامترهای مختلف واحدهای مسکونی چون سن بنا، زیربنا، آسانسور، طبقه، پارکینگ و موقعیت جغرافیایی با قیمت تخمینی واحد پرداختند. نتایج نشان داد که با استفاده از مدل رگرسیون جغرافیایی می‌توان مدلی با دقتی بیش از ۹۵ درصد جهت پیش‌بینی قیمت ارائه داد.

همانگونه که از مطالعات پیشین مشخص است، در زمینه تخمین قیمت کالاها در فروشگاه‌های آنلاین مطالعات زیادی صورت نگرفته است و بیشتر مقالات در زمینه پیش‌بینی بازارهای مالی و مسکن است. از این رو در مقاله حاضر از کارهای انجام شده برای پیش‌بینی قیمت در سایر حوزه‌ها برای تخمین قیمت در فروشگاه‌های آنلاین ایده گرفته شده است.

۳. پیش پردازش داده‌ها و انتخاب کالاهای پایه

در این قسمت ابتدا پیش پردازش بر روی داده‌های خام انجام شده و داده‌های معتبر به‌عنوان ورودی به سیستم تخمین قیمت داده می‌شوند. همچنین مدل پیش‌بینی قیمت بر اساس قیمت تعدادی از کالاها، قیمت سایر کالاها را تخمین می‌زند. در این قسمت روش انتخاب این کالاها که به آن‌ها کالاهای پایه می‌گوییم تبیین می‌شود.

1. Chau, K., et al., (2001).
2. Wen, H. Z., et al. (2005).

۳-۱. پیش پردازش داده‌ها

قیمت ورودی کالاها ممکن است شامل مقادیری نامعتبر باشد که باید پیش از اعمال روش‌های تخمین قیمت، این مقادیر از داده‌ها حذف شده یا اصلاح شوند. این مقادیر نامعتبر شامل مقادیر منفی، داده‌های پرت یا مقادیر خالی می‌باشد. همچنین ممکن است برای یک کالا در یک روز چند قیمت ثبت شده باشد که در این صورت اگر تمام این قیمت‌ها معتبر باشد، میانگین آن‌ها به‌عنوان قیمت کالا در آن روز در نظر گرفته شده و اگر تعدادی از این مقادیر نامعتبر بود، اصلاح یا حذف آن‌ها صورت گرفته و سپس میانگین این مقادیر به‌عنوان قیمت روز کالا وارد سیستم تخمین قیمت می‌شود.

آنگونه که در بخش ۱،۲،۴ توضیح داده خواهد شد، برای اینکه سیستم تخمین قیمت بتواند قیمت کالا را تخمین بزند، باید حداقل در شش روز برای کالای مورد نظر قیمت ثبت شده باشد. به‌همین دلیل در این مرحله کالاهایی که کمتر از شش روز برایشان قیمت ثبت شده است از مدل حذف شده و به‌عنوان کالاهایی که سیستم تخمین قیمت نمی‌تواند قیمت آن‌ها را پیش‌بینی کند شناخته می‌شوند.

۳-۲. انتخاب کالاهای پایه

کالای پایه در هر دسته‌بندی، کالایی است که روند رشد قیمتی آن، رابطه مستقیمی با روند رشد قیمتی سایر کالاهای آن دسته‌بندی داشته باشد تا بتوان با استفاده از قیمت این کالا، قیمت سایر کالاهای آن دسته‌بندی را تخمین زد.

برای پیدا کردن کالاهای پایه از معیار همبستگی استفاده می‌کنیم؛ به این ترتیب که در هر دسته‌بندی، ماتریس همبستگی قیمت کالاها را تشکیل می‌دهیم[1]. طبق رابطه (۱) مقدار همبستگی را برای هر زوج کالا در هر دسته‌بندی حساب کرده و ماتریس همبستگی را تشکیل می‌دهیم.

$$Corr(x,y) = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (۱)$$

در این رابطه، x و y بردار داده‌های تاریخی قیمت برای زوج کالای دلخواه می‌باشند و n اندازه این بردارها است که معادل تعداد روزهایی است که برای این دو کالا قیمت ثبت شده است.

1. Whitehill, S. (2009).

دو پارامتر مهم برای تعیین کالاهای پایه وجود دارد که عبارتند از: تعداد کالاهای پایه و حداقل همبستگی قابل قبول بین کالای پایه و کالای دلخواه. در ادامه مقدار مطلوب این دو پارامتر تعیین می‌شود.

۳-۲-۱.تعیین حداقل همبستگی قابل قبول

هر چه مقدار حداقل همبستگی قابل قبول برای انتخاب کالاهای پایه بیشتر باشد، دقت تخمین بیشتر بوده و هر چه این مقدار کمتر باشد، قیمت کالاهای بیشتری قابل تخمین است. حداقل همبستگی قابل قبول به تعداد نمونه‌های ماتریس همبستگی وابسته است. برای حداقل تعداد شش نمونه، حداقل همبستگی ۰٫۹ قابل قبول است[1]. با فرض اینکه برای تمام کالاها حداقل شش روز قیمت ثبت شده و همچنین کالاهای پایه باید با کالاهای وابسته حداقل در شش روز مشترک قیمت داشته باشند، در این مقاله حداقل همبستگی قابل قبول، ۰٫۹ در نظر گرفته شده است.

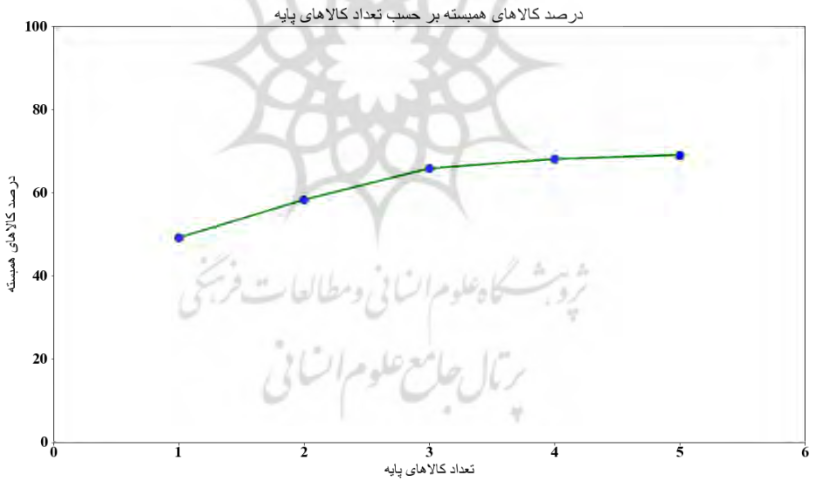
۳-۲-۲. تعیین تعداد مناسب کالای پایه

برای محاسبه تعداد کالاهای پایه مطلوب، حداقل میزان همبستگی قابل قبول بین رشد قیمت کالاهای پایه و سایر کالاها ۰٫۹ در نظر گرفته شده است و در ماتریس همبستگی اگر قدر مطلق همبستگی کمتر از این مقدار باشد صفر و اگر بزرگتر یا مساوی ۰٫۹ باشد یک در نظر گرفته می‌شود. هر چه تعداد کالاهای پایه کمتر باشد، با داشتن قیمت تعداد کالاهای کمتری می‌توان قیمت سایر کالاها را تخمین زد. هر چه قدر این تعداد بیشتر باشد، قیمت کالاهای بیشتری را می‌توان تخمین زد. همچنین سرعت اجرای الگوریتم نیز با کاهش تعداد کالاهای پایه به شکل چشم‌گیری کاهش می‌یابد. برای این منظور، در مدل پیشنهادی برای تعیین تعداد مناسب کالاهای پایه، نمودار درصد کالاهای قابل تخمین (یعنی کالاهایی که همبستگی آن‌ها با حداقل یکی از کالاهای پایه، یک در نظر گرفته شده است) بر حسب تعداد کالاهای پایه رسم می‌شود و قسمتی از نمودار که تغییرات درصد کالاهای قابل تخمین در آن کمتر از یک مقدار مشخص باشد؛ به‌عنوان تعداد کالای پایه مطلوب در نظر گرفته می‌شود. این مقدار در این مقاله پنج درصد در نظر گرفته شده است.

1. Bujang, M., & Baharum, N. (2016).

لازم به ذکر است که در تحلیل همبستگی، باید همبستگی پارامترهای مستقل با یکدیگر پایین باشد زیرا در صورت همبستگی بالای پارامترهای مستقل با یکدیگر می‌توان تنها یکی از آن‌ها را به‌عنوان برآیند تمام پارامترهای مستقل در نظر گرفت. در عمل با استفاده از روش پیش گرفته شده در انتخاب تعداد کالای پایه، این مورد که کالاهای پایه باید همبستگی پایینی داشته باشند در نظر گرفته شده است زیرا با افزایش تعداد کالای پایه نسبت به تعداد مطلوب، تغییر چندانی در تعداد کالاهای قابل تخمین دیده نمی‌شود (کمتر از ۵ درصد) و این به آن معناست که کالاهای پایه اضافی با کالاهای پایه مطلوب همبستگی بالایی داشته‌اند که از آن‌ها صرف نظر شده است.

برای نمونه در شکل یک از یک کالا به دو کالای پایه تغییرات درصد کالاهای قابل تخمین حدود نه درصد و از دو کالا به سه کالای پایه حدود هفت درصد و از سه کالا به چهار کالای پایه حدود دو درصد است. پس تعداد کالای پایه مناسب را سه در نظر می‌گیریم.



شکل ۱. درصد تعداد کالاهای همبسته برحسب تعداد کالاهای پایه

بعد از مشخص شدن تعداد کالاهای پایه، کالاهای پایه به گونه‌ای انتخاب می‌شوند که در مجموع تعداد بیشتری از کالاهای دسته‌بندی مورد نظر با حداقل یکی از این کالاهای پایه، همبستگی بیشتر یا مساوی ۰٫۹ داشته باشد.

۴. سیستم تخمین قیمت پیشنهادی

پس از مشخص شدن کالاهای پایه در هر دسته‌بندی، داده‌ها را به این صورت مدل می‌کنیم؛ ماتریس ورودی که ستون‌هایی شامل پارامترهای مستقل (قیمت کالاهای پایه، ضریب ثابت و قیمت دلار) دارد و هر ردیف ماتریس نیز معادل یک روز می‌باشد. همچنین بردار خروجی یا پارامتر وابسته که شامل قیمت کالایی است که می‌خواهیم قیمت آن را تخمین بزنیم.

در ادامه ابتدا با استفاده از روش رگرسیون خطی به برررسی و حل مسئله می‌پردازیم. در ادامه برای در نظر گرفتن اهمیت داده‌های جدید و کاهش خطای تخمین، یک روش رگرسیون خطی ابتکاری ارائه می‌شود. همچنین با توجه به تغییرات نرخ دلار و تأثیر آن بر روی قیمت محصولات مختلف آهن، نرخ دلار نیز به‌عنوان یک متغیر مستقل در محاسبات مدل‌های پیشنهادی در نظر گرفته شده است که نتایج حاکی از بهبود نتایج در تعدادی از موارد می‌باشد. برای بررسی دقت و روایی مدل باید تعدادی از داده‌ها را به‌عنوان داده تست در نظر گرفت تا پس از آموزش مدل، میزان دقت مدل ارائه شده در تخمین قیمت کالاها مورد بررسی قرار گیرد. به این منظور در تمام این روش‌ها، ۲۰ درصد از داده‌های جدید از نظر زمانی را به‌عنوان داده‌های تست در نظر گرفته‌ایم و بقیه داده‌ها برای آموزش مدل استفاده شده‌اند.

۴-۱. روش رگرسیون خطی

در روش رگرسیون خطی، هدف به دست آوردن تخمینی از قیمت کالا با کمینه کردن میانگین خطا می‌باشد[1]. به این منظور طبق رابطه (۲) رابطه‌ای بین قیمت کالای مورد بررسی و قیمت کالاهای پایه در روزهای مختلف نوشته می‌شود. در رابطه (۲)، w بردار ضرایب، y بردار قیمت کالایی که می‌خواهیم تخمین بزنیم و ماتریس قیمت کالاهای پایه است. هر ردیف از ماتریس و بردار y معادل قیمت کالاهای پایه و کالای مورد بررسی در یک تاریخ می‌باشد. با فرض اینکه تعداد k کالای پایه وجود دارد؛ برای k ستون اول از ماتریس ستون i شامل قیمت کالای پایه i در روزهای مختلف است. همچنین ستون آخر از ماتریس شامل درایه‌های ۱ است که معادل مقدار ثابت در روش رگرسیون خطی است. همچنین بردار ضرایب w نیز شامل ضرایب w_1 تا w_2 است که معادل ضرایب کالاهای پایه بوده و ضریب w_{const} نیز معادل ضریب مقدار ثابت می‌باشد.

1. Emioma, C. C., & Edeki, S. O. (2021).

$$y = A \times w\ ; \quad A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,k} & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,k} & 1 \end{bmatrix}, w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \\ w_{const} \end{bmatrix} \quad (۲)$$

در رابطه (۳)، هر ردیف از رابطه (۲) به‌صورت غیر ماتریسی آورده شده است.

$$y_i = \sum_{j=1}^{k} a_{i,j} w_j + w_{const}\ ; 1 \le i \le n \quad (۳)$$

برای به‌دست آوردن ضرایب مجهول در رابطه (۳)، مجموع مربعات خطای حاصل از اختلاف مقدار پیش‌بینی شده و مقدار واقعی را برای داده‌های آموزش، کمینه می‌کنیم. طبق رابطه (۴) با حل این معادله، ضرایب مجهول محاسبه می‌شود. برای حل معادله (۴) و محاسبه ضرایب مجهول، از مدل linearregression از کتابخانه سایکیت لرن در نرم افزار پایتون استفاده می‌کنیم.

$$\min_{w_j, w_{const}} \sum_{i=1}^{n} \sum_{j=1}^{k} \left[\left(a_{i,j}\, w_j + w_{const} \right) - y_i \right]^2 \quad (۴)$$

بعد از محاسبه ضرایب مجهول و برای سنجش میزان دقت مدل پیش‌بینی، میانگین درصد خطای مطلق را بر روی m داده تست محاسبه می‌کنیم. طبق رابطه (۵)، میانگین درصد خطای مطلق برابر اختلاف مقدار تخمینی و مقدار واقعی نسبت به مقدار واقعی می‌باشد.

$$MAPE\ Error = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{y_i - A'_i \times w}{y_i} \right| \quad (۵)$$

با به دست آمدن بردار ضرایب و با داشتن قیمت کالاهای پایه در هر روز، می‌توان با استفاده از رابطه (۶) قیمت کالا را در این روز محاسبه کرد.

$$y_{pred} = \sum_{i=1}^{k} a_i w_i + w_{const} \quad (۶)$$

در رابطه (۶)، y_{pred} معادل قیمت تخمین زده شده، معادل قیمت روز کالای پایه i ام و w_i و w_{const} معادل ضرایب محاسبه شده طبق روش رگرسیون خطی می‌باشند.

شایان ذکر است روش رگرسیون در تحقیقات مختلفی مورد استفاده واقع شده است که اکثر آنها در بخش (۳) توضیح داده شده است. در این قسمت به برخی از مهمترین آن‌ها اشاره می‌کنیم. فرج زاده و دارند (۱۳۸۸) از روش رگرسیون خطی برای تخمین میزان مرگ و میر ناشی از

آالودگی هوا در شهر تهران استفاده نمودند. قلی زاده (۱۳۸۹) از روش رگرسیون خطی ساده و فازی به منظور پیش‌بینی قیمت سهام استفاده کرده است. بهرامی و قربانی (۱۳۹۳) از روش رگرسیون برای پیش‌بینی تنوع گونه‌ای با در نظر گرفتن عوامل خاکی و فیزیوگرافی پرداختند. کوه‌پیما و همکاران (۱۳۹۹) از روش رگرسیون خطی برای تخمین قیمت آپارتمان در شهر تهران استفاده کرده‌اند.

۴-۲. روش رگرسیون خطی ابتکاری

با توجه به اینکه داده‌های جدیدتر باید اهمیت بیشتری در تخمین قیمت روز داشته باشند، در این روش به داده‌های جدیدتر اهمیت بیشتری داده می‌شود. برای این منظور داده‌های آموزش را بر اساس زمان به سه دسته تقسیم می‌کنیم: داده‌های جدید، داده‌های میان مدت و داده‌های قدیمی. برای هر یک از این سه دسته داده، مدل رگرسیون خطی را طبق روابط (۷)، (۸) و (۹) جداگانه اعمال می‌کنیم. با حل مسئله کمینه‌سازی طبق رابطه (۴) برای هر یک از این سه دسته، ضرایب رگرسیون را محاسبه می‌کنیم. از آنجایی که مدل، خطی می‌باشد، ضرایب مدل نهایی برابر با میانگین وزن دار ضرایب این سه مدل است.

$$y_i^{new} = \sum_{j=1}^{k} a_{i,j} w_j^{new} + w_{const}^{new} \; ; 1 \le i \le n_1 \quad (۷)$$

$$y_i^{mid} = \sum_{j=1}^{k} a_{i,j} w_j^{mid} + w_{const}^{mid} \; ; 1 \le i \le n_2 \quad (۸)$$

$$y_i^{old} = \sum_{j=1}^{k} a_{i,j} w_j^{old} + w_{const}^{old} \; ; 1 \le i \le n_3 \quad (۹)$$

برای به دست آوردن ضریب اهمیت هر مدل، برای هر کالا به‌صورت جداگانه ضرایب اهمیت مختلف را تست کرده و مدلی را که کمترین میانگین درصد خطای مطلق بر روی داده‌های تست به‌دست آمده از رابطه (۵) داشته باشد، به‌عنوان مدل برتر انتخاب می‌کنیم. همچنین پارامتر دیگری که باید تعیین شود، این است که چند درصد از داده‌ها را به‌عنوان داده جدید و چند درصد را به‌عنوان داده میان مدت و داده قدیمی در نظر بگیریم. این پارامترها را نیز با استفاده از کمینه کردن میانگین درصد خطای مطلق از رابطه (۵) برای مقادیر مختلف این پارامتر حساب می‌کنیم. برای اینکه تأثیر پارامتر ضریب اهمیت و حجم هر دسته را به‌طور هم‌زمان در نظر بگیریم، میانگین درصد خطای مطلق را برای حالت‌های مختلف این دو پارامتر محاسبه کرده و کمترین مقدار به‌دست آمده را به‌عنوان ضرایب مطلوب انتخاب می‌کنیم.

از آنجایی‌که این ضرایب اهمیت شامل تمامی مقادیر پیوسته در بازه صفر تا یک بوده و حجم

هر دسته نیز بر حسب درصد در بازه صفر تا صد قرار دارد؛ محاسبه این ضرایب برای تمام کالاها بسیار زمان‌بر خواهد بود. به این خاطر مقادیر ضرایب اهمیت را به‌صورت گسسته و با فاصله ۰٫۰۵ در نظر گرفته و مقادیر حجم هر دسته را نیز به‌صورت گسسته و با فاصله پنج درصد در نظر می‌گیریم.

اگر ضریب اهمیت داده‌های جدید، میان مدت و قدیمی را به ترتیب α، β و γ در نظر بگیریم؛ از آنجاییکه روابط خطی می‌باشند، ضرایب رگرسیون در این روش طبق رابطه (۱۰) محاسبه می‌شوند. همچنین طبق رابطه (۱۱) مجموع این ضرایب اهمیت باید برابر ۱ باشد.

$$w_i = \alpha w_i^{new} + \beta w_i^{mid} + \gamma w_i^{old} \quad ; 1 \le i \le k,$$
$$w_{const} = \alpha w_{const}^{new} + \beta w_{const}^{mid} + \gamma w_{const}^{old} \qquad (۱۰)$$
$$\alpha + \beta + \gamma = 1 \; ; \; \alpha, \beta, \gamma > 0 \qquad (۱۱)$$

با مشخص شدن بردار ضرایب w، مشابه روش رگرسیون خطی، قیمت تخمینی کالا در هر روز از رابطه (۶) محاسبه می‌شود.

طبق شکل (۲)، مراحل روش رگرسیون خطی ابتکاری به‌صورت زیر می‌باشد:

۱. تعیین ضرایب اهمیت و حجم هر دسته بر اساس کمترین میانگین درصد خطای مطلق برای مقادیر مختلف این دو پارامتر
۲. دسته‌بندی داده‌ها بر حسب زمان و بر اساس پارامتر حجم هر دسته به سه دسته
۳. محاسبه ضرایب رگرسیون در هر دسته طبق روابط (۷)، (۸) و (۹)
۴. محاسبه ضرایب رگرسیون نهایی که ترکیب خطی وزن دار ضرایب رگرسیون در هر دسته هستند طبق رابطه (۱۰)

شکل ۲- مراحل روش رگرسیون خطی ابتکاری

۴-۳. روش رگرسیون خطی با تأثیر نرخ دلار

این روش مشابه روش رگرسیون خطی است با این تفاوت که در ماتریس ورودی، پارامتر نرخ دلار نیز اعمال می‌شود. ماتریس ورودی و بردار ضرایب از رابطه (۱۲) محاسبه می‌شود که ستون آخر در ماتریس معادل قیمت دلار است.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,k} & 1 & d_1 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,k} & 1 & d_n \end{bmatrix}, w = \begin{bmatrix} w_2 \\ \vdots \\ w_k \\ w_{const} \\ w_d \end{bmatrix} \quad (۱۲)$$

محاسبه ضرایب، مشابه روش رگرسیون خطی بوده و بعد از محاسبه ضرایب، با استفاده از رابطه (۱۳) می‌توان قیمت کالای مورد نظر را تخمین زد. در رابطه (۱۳)، d معادل قیمت دلار، y_{predt} معادل قیمت تخمین زده شده، معادل قیمت روز کالای پایه i ام و w_{const} ،w_i و w_d معادل ضرایب محاسبه شده طبق روش رگرسیون خطی با تأثیر نرخ دلار می‌باشند.

$$y_{pred} = \sum_{i=1}^{k} a_i w_i + w_{const} + w_d d \quad (۱۳)$$

۴-۴. روش رگرسیون خطی ابتکاری با تأثیر نرخ دلار

این روش نیز مشابه روش رگرسیون خطی ابتکاری است با این تفاوت که در ماتریس ورودی، پارامتر نرخ دلار نیز اعمال می‌شود. به این ترتیب که ماتریس ورودی و بردار ضرایب از رابطه (۱۲) محاسبه شده و طبق روابط (۷)، (۸) و (۹) ضرایب هر دسته از داده‌ها جداگانه حساب شده و در نهایت ضرایب مدل از رابطه (۱۴) حساب می‌شوند.

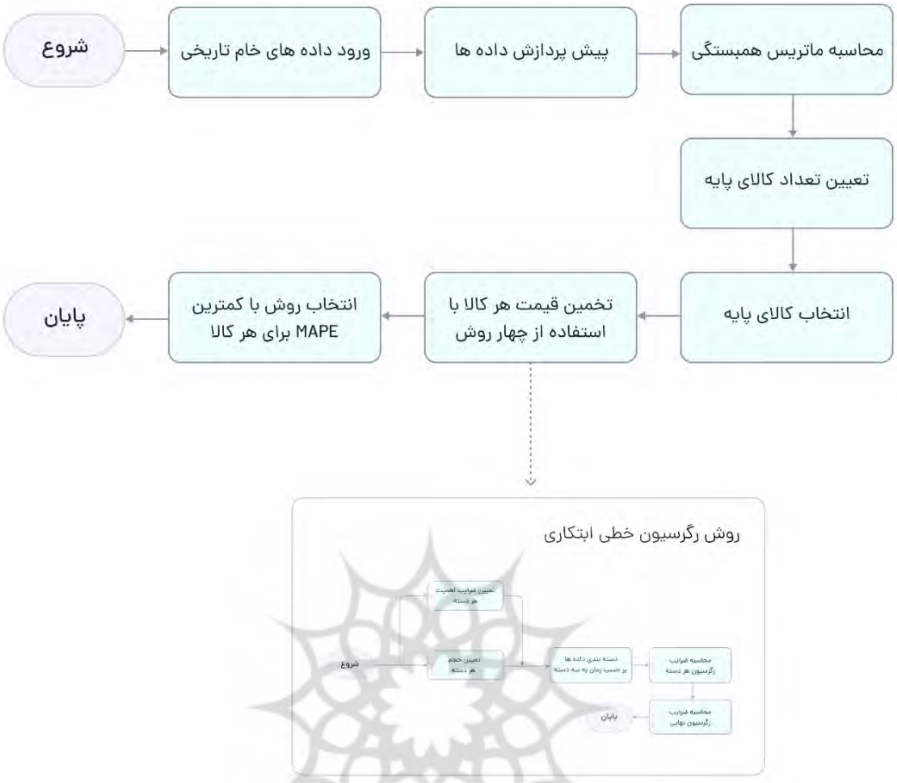
$$\begin{aligned} w_i &= \alpha w_i^{new} + \beta w_i^{mid} + \gamma w_i^{old} \quad ; 1 \le i \le k \\ w_{const} &= \alpha w_{const}^{new} + \beta w_{const}^{mid} + \gamma w_{const}^{old} \\ w_d &= \alpha w_d^{new} + \beta w_d^{mid} + \gamma w_d^{old} \end{aligned} \quad (۱۴)$$

در نهایت با استفاده از رابطه (۱۳) و جایگذاری ضرایب به‌دست آمده از رابطه (۱۴) می‌توان قیمت کالای مورد نظر را تخمین زد.

۴-۵. سیستم تخمین قیمت

مراحل اجرای سیستم تخمین قیمت در شکل (۳) آورده شده است. این مراحل به شرح زیر می‌باشد:

- **گام ۱) ورود داده‌های خام تاریخی:** تاریخچه قیمتی کالاهایی که می‌خواهیم تخمین بزنیم به‌عنوان ورودی داده می‌شود.
- **گام ۲) پیش پردازش داده‌ها:** پیش پردازش روی داده‌های خام انجام شده و داده‌های نامعتبر حذف یا اصلاح می‌شوند. همچنین کالاهایی که کمتر از شش روز قیمت دارند به‌عنوان کالاهایی که سیستم تخمین قیمت نمی‌تواند قیمت آن‌ها را تخمین بزند از مدل خارج می‌شوند.
- **گام ۳) محاسبه ماتریس همبستگی:** ماتریس همبستگی برای رشد قیمت کالاها تشکیل شده و کالاهایی که بیش از ۰٫۹ همبستگی دارند به‌عنوان زوج کالای همبسته شناخته می‌شوند. زوج کالای همبسته، کالاهایی هستند که با داشتن قیمت یکی از آن‌ها می‌توان قیمت دیگری را تخمین زد.
- **گام ۴) تعیین تعداد کالای پایه:** درصد تعداد کالاهای قابل تخمین را برای تعداد کالاهای پایه مختلف (شروع از یک کالای پایه) را حساب می‌کنیم. اگر با افزایش تعداد کالاهای پایه، درصد تعداد کالاهای قابل تخمین کمتر از پنج درصد افزایش یافت، تعداد کالای پایه مطلوب را به‌دست آورده‌ایم.
- **گام ۵) انتخاب کالای پایه:** با مشخص شدن تعداد کالای پایه، کالاهای پایه به گونه‌ای انتخاب می‌شوند که تعداد حداکثر کالا قابل تخمین باشند.
- **گام ۶) تخمین قیمت هر کالا با استفاده از چهار روش:** در این مرحله با استفاده از روش‌های «رگرسیون خطی»، «رگرسیون خطی ابتکاری»، «رگرسیون خطی با تأثیر نرخ دلار» و «رگرسیون خطی ابتکاری با تأثیر نرخ دلار» قیمت تمام کالاهای ورودی را تخمین می‌زنیم.
- **گام ۷) انتخاب روش با کمترین MAPE برای هر کالا:** برای هر کالا، روشی را که کمترین میانگین درصد خطای مطلق روی داده‌های تست داشته باشد به‌عنوان روش برگزیده انتخاب می‌کنیم. ضرایب رگرسیون این روش را برای کالای مورد نظر ذخیره کرده و قیمت هر کالا در هر روز را با استفاده از ضرایب رگرسیون و قیمت کالاهای پایه و قیمت دلار در آن روز تخمین می‌زنیم.



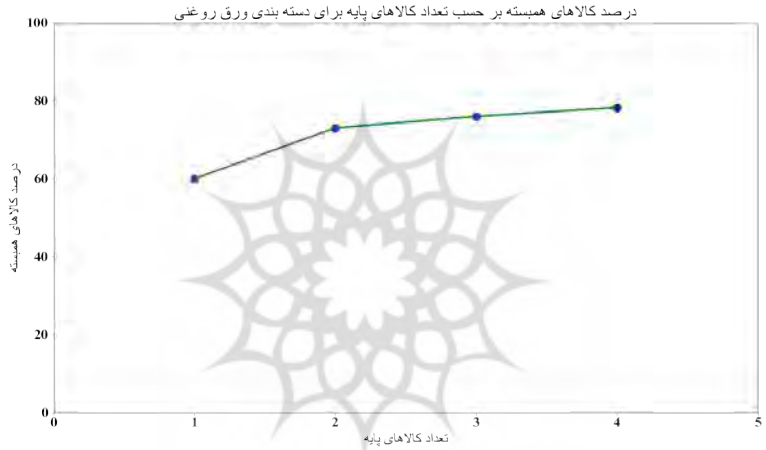
شکل ۳- سیستم تخمین قیمت

۵. مطالعه موردی و تحلیل نتایج (شرکت نگین پولاد نقش جهان)

داده‌های استفاده شده در این مطالعه، داده‌های شرکت نگین پولاد نقش جهان (آهن آنلاین) است. رسالت شرکت نگین پولاد نقش جهان، خرید و فروش انواع آهن از جمله میلگرد، ورق، تیرآهن و ... به‌صورت آنلاین می‌باشد. سیستم تخمین قیمت بر روی ۴۵۱۴ کالای شرکت نگین پولاد نقش جهان (آهن آنلاین) اعمال شد. این کالاها در ۸۰ دسته‌بندی مختلف قرار دارند. بازه زمانی قیمت کالاها از تاریخ ۱۳۸۹/۰۱/۱۶ تا ۱۴۰۱/۰۵/۱۸ می‌باشد و قیمت‌ها به‌صورت روزانه ثبت شده است. همچنین دامنه قیمت کالاهای مختلف از ۱۰۰ هزار ریال تا ۹۵۰ میلیون ریال متغیر است.

در این بخش ابتدا سیستم تخمین قیمت را بر روی دسته‌بندی «ورق روغنی» اعمال می‌کنیم.

در این دسته‌بندی ۲۹۶ کالا موجود است که حداقل در شش روز قیمت دارند. طبق سیستم تخمین قیمت، ابتدا ماتریس همبستگی را برای تمام ۲۹۶ کالای موجود در این دسته‌بندی تشکیل داده و درایه‌هایی از این ماتریس که بیشتر یا مساوی ۰٫۹ باشند، معادل یک و در غیر این صورت معادل ۰ در نظر گرفته می‌شود. شکل چهار تعداد کالاهای قابل تخمین بر حسب تعداد کالاهای پایه را برای دسته‌بندی «ورق روغنی» نشان می‌دهد. با توجه به اینکه در نمودار از تعداد دو کالای پایه به سه کالای پایه تغییر درصد کالاهای قابل تخمین حدود سه درصد بوده که از معیار پنج درصد کمتر است؛ تعداد کالاهای پایه دو در نظر گرفته می‌شود.



شکل۴- درصد تعداد کالاهای همبسته برحسب تعداد کالاهای پایه برای دسته‌بندی ورق روغنی با ۲۹۶ قلم کالا

با دو کالای پایه منتخب در مجموع می‌توان قیمت ۲۱۸ کالای دیگر را تخمین زد.

در مرحله بعد با استفاده از چهار روش سیستم تخمین قیمت، ضرایب رگرسیون برای ۲۱۸ کالای معتبر (کالاهایی که حداقل با یکی از دو کالای پایه همبسته هستند) محاسبه شده و میانگین درصد خطای مطلق طبق رابطه (۵) محاسبه می‌شود. برای هر کالا، روشی را که کمترین میانگین درصد خطای مطلق داشته است به‌عنوان روش منتخب در سیستم تخمین قیمت در نظر می‌گیریم.

در ادامه، ضرایب رگرسیون به‌دست آمده برای یک کالای خاص با عنوان «ورق روغنی ۰٫۴ فولاد غرب ۱۰۰۰ رول ST۱۲ بنگاه تهران» از دسته‌بندی «ورق روغنی» و خطای MAPE آن در جدول (۲) آمده است. در این جدول قیمت واقعی این کالا در تاریخ ۱۸ مرداد ۱۴۰۱ (که آخرین قیمت موجود است) و قیمت تخمینی با روش‌های مختلف نیز آورده شده است.

جدول ۲- مقایسه قیمت‌های تخمینی با قیمت واقعی کالای مورد مطالعه در تاریخ ۱۸ مرداد ۱۴۰۱

	روش رگرسیون خطی	روش رگرسیون خطی ابتکاری	روش رگرسیون خطی با تأثیر نرخ دلار	روش رگرسیون خطی ابتکاری با تأثیر نرخ دلار
میانگین خطای MAPE	۳٫۵۵	۰٫۴۱	۲٫۶۱	۰٫۷۳
بیشترین خطای MAPE	۵٫۱۷	۱٫۹۲	۴٫۲۸	۱٫۸۷
قیمت کالای پایه اول	۲۵۶۷۸۰	۳۳۲۱۱۰۱۰	۳۳۲۱۱۰۱۰	۳۳۲۱۱۰۱۰
قیمت کالای پایه دوم	۲۶۱۳۵۰	۲۶۱۳۵۰	۲۶۱۳۵۰	۲۶۱۳۵۰
قیمت دلار	۳۱۴۵۰۰	۳۱۴۵۰۰	۳۱۴۵۰۰	۳۱۴۵۰۰
قیمت واقعی کالا	۲۹۹۴۱۰	۲۹۹۴۱۰	۲۹۹۴۱۰	۲۹۹۴۱۰
قیمت تخمینی کالا	۳۰۶۱۲۹	۲۹۸۴۵۶	۳۰۳۳۶۳	۲۹۵۹۳۵
درصد خطای پیش‌بینی	۲٫۱۹	۰٫۳۲	۱٫۳۰	۱٫۱۶

طبق اطلاعات جدول (۲)، میانگین درصد خطای مطلق روش رگرسیون خطی ابتکاری کمینه بوده و به‌عنوان روش منتخب در سیستم تخمین قیمت انتخاب می‌شود. همچنین دقت تخمین این روش نیز در تاریخ مورد نظر از سایر روش‌ها بهتر بوده و با ۰٫۳۲ درصد خطا، قیمت کالا را تخمین زده است.

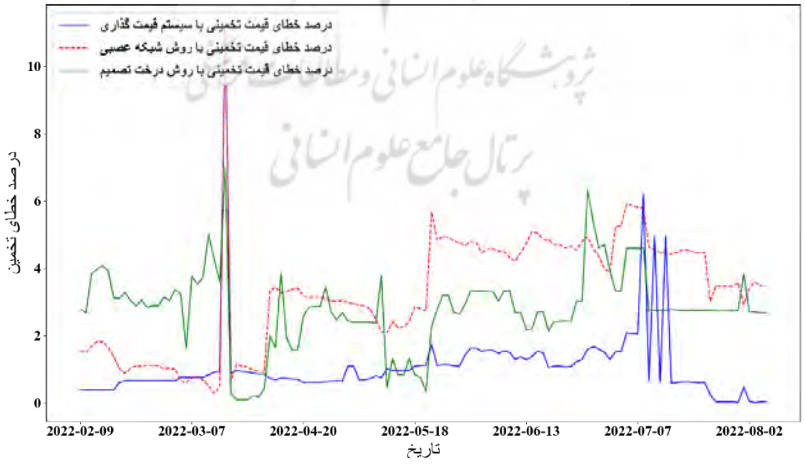
لازم به ذکر است که ضرایب اهمیت و درصد داده‌های جدید در روش‌های رگرسیون خطی ابتکاری و رگرسیون خطی ابتکاری با تأثیر نرخ دلار طبق جدول (۳) به‌دست آمد و با این ضرایب کمترین میانگین خطا حاصل شد. طبق مقادیر به‌دست آمده، در روش رگرسیون خطی ابتکاری کمترین میزان میانگین درصد خطای مطلق زمانی حاصل شده است که به ۳۰ درصد از داده‌های

جدید ۰٫۶۵ اهمیت و به ۴۰ درصد از داده‌های میان مدت ۰٫۲۰ اهمیت و به ۳۰ درصد از داده‌های باقی مانده ۰٫۱۵ اهمیت داده شده است. همچنین در روش رگرسیون خطی ابتکاری با تأثیر نرخ دلار کمترین میانگین درصد خطای مطلق زمانی به‌دست آمده است که به ۲۵ درصد از داده‌های جدید ۰٫۷۰ اهمیت و به ۴۵ درصد از داده‌های میان مدت ۰٫۲۰ اهمیت و به ۳۰ درصد از داده‌های باقی‌مانده تنها ۰٫۱۰ اهمیت داده شده است.

جدول۳- ضرایب اهمیت و درصد داده‌های جدید، میان مدت و قدیمی در روش‌های ابتکاری برای کالای مورد مطالعه

روش	α	β	γ	$Percentage_{new}$	$Percentage_{mid}$	$Percentage_{old}$
رگرسیون خطی ابتکاری	۰٫۶۵	۰٫۲۰	۰٫۱۵	۳۰	۴۰	۳۰
ابتکاری با تأثیر نرخ دلار	۰٫۷۰	۰٫۲۰	۰٫۱۰	۲۵	۴۵	۳۰

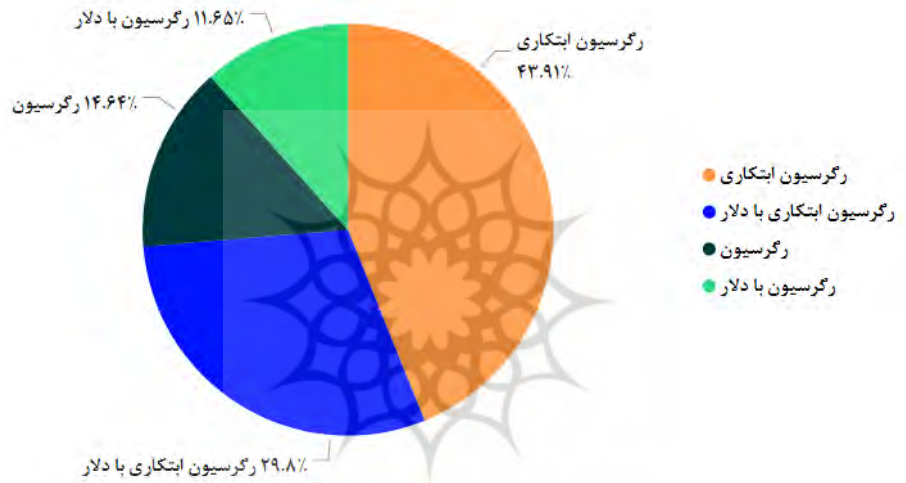
در شکل (۵)، مقایسه‌ای بین تخمین قیمت کالای مورد بررسی با سیستم تخمین قیمت توسعه داده شده در این مقاله و روش‌های شبکه عصبی و درخت تصمیم انجام شده است. نمودار، نشان‌دهنده درصد خطای قیمت تخمینی در شش ماه اخیر می‌باشد و شبکه عصبی استفاده شده از دو لایه میانی تشکیل شده است. طبق شکل، خطای سیستم تخمین قیمت در اکثر روزها کمتر از دو روش دیگر است.



شکل ۵- مقایسه درصد خطای قیمت تخمینی کالای مورد مطالعه با سیستم تخمین قیمت این مقاله و روش‌های شبکه عصبی و درخت تصمیم در بازه زمانی ۶ ماهه.

شایان ذکر است نمودارهای شکل (۵) با داده‌های واقعی شرکت نگین پولاد نقش جهان ترسیم شده‌اند. همان‌طور که از نمودار فوق مشخص است خطای سیستم تخمین قیمت در اکثر مواقع زیر دو درصد می‌باشد که نشان از روایی قابل قبول سیستم می‌باشد.

در ادامه برای هر یک از ۴۵۱۴ کالای شرکت نگین پولاد نقش جهان (آهن آنلاین) بهترین مدل (یعنی مدلی که کمترین خطای MAPE را دارد) به‌عنوان مدل برگزیده انتخاب شد. طبق شکل ۶ روش ابتکاری در مجموع در ۷۳٫۷۱ درصد از کالاها پاسخ بهتری نسبت به روش رگرسیون خطی داشته است.



شکل۶- روش منتخب در سیستم تخمین قیمت برای ۴۵۱۴ کالای شرکت نگین پولاد نقش جهان (آهن آنلاین) با کمترین درصد خطای مطلق

لازم به ذکر است مدل پایه سیستم تخمین قیمت پیشنهادی، رگرسیون خطی می‌باشد که پایایی[1] این مدل در ادبیات موضوع مورد بررسی و تایید قرار گرفته است.[2] علاوه براین به منظور بررسی پایایی سیستم تخمین قیمت نتایج سیستم تخمین قیمت برای دسته‌های متنوعی از کالاها مورد بررسی قرار گرفت. در جدول (۴)، میانگین خطای MAPE برای ۱۵ دسته کالا محاسبه شده

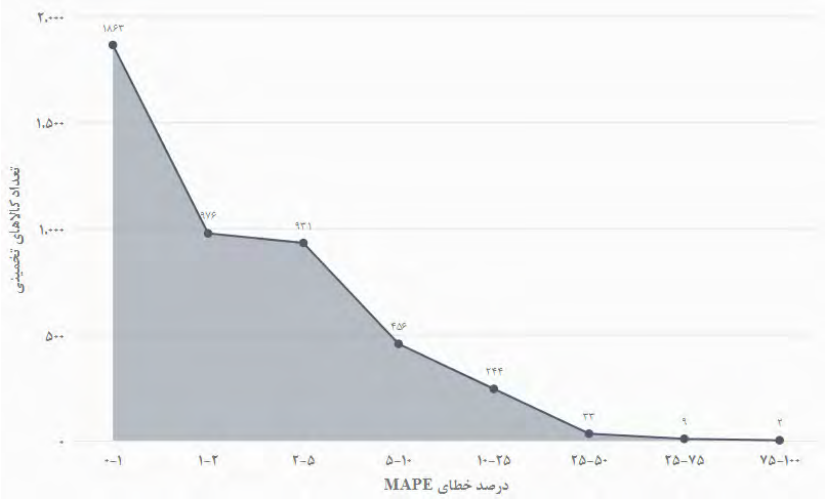
1. Reliability
2. Meng, Q.N. & Xu, X. (2018).

است (در مجموع ۳۰۵۹ کالا). طبق نتایج به‌دست کمترین و بیشترین خطا به ترتیب ۰٫۱۷ و ۵٫۰۲ می‌باشد. در ضمن میانگین و انحراف معیار خطاها به ترتیب برابر است با ۳٫۱۳ و ۱٫۲۹ که نشان از پایایی مناسب سیستم تخمین قیمت ارائه شده دارد.

جدول۴- تعداد کالاهای تخمینی و میانگین خطای MAPE برای ۱۵ دسته‌بندی کالا با بیشترین تعداد کالای تخمینی

ردیف	دسته‌بندی کالا	MAPE میانگین خطای	تعداد کالاها
۱	میلگرد آجدار	۲٫۵۹	۸۰۰
۲	ورق گالوانیزه	۲٫۳۴	۳۴۴
۳	ورق استنلس استیل صنعتی	۳٫۱۷	۲۷۵
۴	ورق روغنی	۲٫۲۸	۲۱۸
۵	پروفیل	۲٫۵۴	۲۱۱
۶	ورق سیاه	۲٫۹۲	۲۰۶
۷	نبشی	۳٫۰۲	۱۸۶
۸	تسمه	۳٫۱۳	۱۵۲
۹	ورق رنگی	۵٫۰۲	۱۲۶
۱۰	ناودانی	۴٫۸۷	۱۱۵
۱۱	تیرآهن	۴٫۷۷	۱۰۸
۱۲	لوله درز مستقیم	۲٫۰۱	۹۳
۱۳	لوله استنلس استیل صنعتی	۳٫۷۳	۸۰
۱۴	کابل افشان	۴٫۳۴	۷۹
۱۵	ورق آلومینیوم	۰٫۱۷	۶۶

در شکل (۷) مقایسه‌ای بین تعداد کالاهای تخمینی با درصد خطاهای مختلف انجام شده است. طبق شکل، قیمت تعداد قابل توجهی از کالاها با خطای کمتر از یک درصد پیش‌بینی شده است (۱۸۶۳ کالا) و تنها قیمت ۴۴ کالا با خطای بیش از ۲۵ درصد تخمین زده شده است.



شکل۷ - تعداد قیمت کالاهای تخمینی بر حسب میانگین درصد خطای مطلق برای ۴۵۱۴ کالای شرکت نگین پولاد نقش جهان (آهن آنلاین)

نتیجه‌گیری و پیشنهادها

با افزایش چشمگیر فروشگاه‌های فروش آنلاین کالاها، تخمین قیمت این کالاها در روزهای مختلف و ثبت آن در سایت و اپلیکیشن‌های مرتبط ضروری می‌باشد. در مقاله حاضر یک سیستم تخمین قیمت در فروشگاه‌های آنلاین که کالاهای بسیار زیادی دارند به‌طوریکه استعلام و ثبت همه قیمت‌ها ممکن نیست ارائه شده است. با توجه به تعداد بسیار زیاد کالاها و ارتباط مستقیم قیمت این کالاها با یکدیگر، می‌توان با داشتن قیمت تعداد محدودی از کالاها، قیمت سایر این کالاها را تخمین زد. سیستم تخمین قیمت برای تخمین قیمت کالاهای آهنی شرکت نگین پولاد نقش جهان توسعه یافته است که می‌تواند با انجام تغییرات لازم در سایر فروشگاه‌ها هم پیاده‌سازی شود.

در سیستم تخمین قیمت ارائه شده از رگرسیون خطی چند متغیره بهره گرفته شده است. با توجه به خطای بالای رگرسیون خطی در تخمین قیمت‌ها، یک رگرسیون ابتکاری بر مبنای تقسیم بازه زمانی داده‌ها (جدید، میان مدت و قدیمی) و نیز تعیین ضریب اهمیت بهینه آن‌ها ارائه شده است. به عبارتی دیگر در رگرسیون ابتکاری وزن داده‌های جدید و قدیمی یکسان در نظر گرفته

نشده و ضریب بهینه آن‌ها در جهت حداقل نمودن خطای تخمین قیمت تعیین شده است.

سیستم تخمین قیمت پیشنهادی بر روی داده‌های واقعی شرکت نگین پولاد نقش جهان (آهن آنلاین) پیاده‌سازی شد و نتایج آن با داده‌های تست مورد مقایسه قرار گرفت. مطابق نتایج به‌دست آمده قیمت ۶۳ درصد از کالاها با خطای کمتر از دو درصد تخمین زده شده است که نشان از دقت بالای سیستم پیشنهادی دارد. شایان ذکر است که تأثیر نرخ دلار به‌عنوان یک متغیر مستقل در تخمین قیمت مورد بررسی و تحلیل قرار گرفت. نتایج حاکی از آن است که در نظر گرفتن نرخ دلار در تخمین قیمت برخی از کالاها منجر به کاهش خطا می‌شود.

برای بررسی عملکرد سیستم تخمین قیمت ارائه شده، نتایج این سیستم با روش‌های شبکه عصبی و درخت تصمیم مورد مقایسه قرار گرفت. بدین منظور خطای تخمین برای دسته کالای «ورق روغنی» با استفاده از سیستم تخمین قیمت و روش‌های شبکه عصبی و درخت تصمیم محاسبه شد. مطابق نتایج به‌دست آمده سیستم تخمین قیمت ارائه شده در ۸۹ درصد مواقع نسبت به روش شبکه عصبی و در ۸۷ درصد نسبت به روش درخت تصمیم خطای کمتری داشت که نشان از نتایج قابل قبول این روش نسبت به دو روش مذکور دارد.

لازم به ذکر است با زیاد شدن تعداد کالاهای ورودی، زمان تخمین به‌صورت تصاعدی بالا می‌رود؛ لذا در مطالعات آتی می‌توان روش‌های ابتکاری یا فرا ابتکاری و روش‌های یادگیری ماشین برای کاهش این زمان ارائه داد[1]. همچنین می‌توان از روش‌هایی مانند شبکه عصبی و تأثیر پارامترهای مستقل مانند قیمت سنگ آهن بر قیمت کالاهای آهنی برای افزایش دقت پیش‌بینی استفاده کرد.[2]

سیستم تخمین قیمت ارائه شده برای داده‌های دقیق توسعه داده شده است. در نظر گرفتن داده‌های نادقیق از جمله داده‌های فازی در این سیستم می‌تواند به بهبود دقت سیستم تخمین قیمت منجر شود. همچنین با توجه به اینکه تعداد کالاهای پایه تأثیر مستقیم در دقت و زمان حل مدل‌ها دارد لازم است مطالعات تکمیلی در خصوص نحوه انتخاب کالاهای پایه‌ای و تعداد آن‌ها صورت پذیرد.

با توجه به اینکه در کشور ما نرخ تورم بسیار بالاست و این نرخ بر روی هزینه کالا و خدمات

1. Ahmed Mahoto, N., et al., (2021).
2. Ataman, G., & Kahraman, S. (2021).

در فروشگاه‌های آنلاین تأثیر مستقیم دارد، لذا می‌توان نرخ تورم را به‌عنوان یک متغیر مستقل در سیستم تخمین قیمت مورد مطالعه قرار داد. علاوه براین نرخ جهانی آهن نیز می‌تواند به‌عنوان یک متغیر مورد بررسی واقع شود. در نهایت سیستم تخمین قیمت ارائه شده می‌تواند در تخمین قیمت کالاهای سایر فروشگاه‌ها مورد استفاده قرار گیرد.

منابع

بهرامی، بهنام، قربانب، اردوان. (۱۳۹۳). ارزیابی کاربرد مدل‌های شبکه عصبی و رگرسیونی به منظور پیش‌بینی تنوع گونه‌ای با استفاده از برخی عوامل خاکی و فیزیوگرافی (مطالعه موردی: حوزه آبخیز خرابه سنجی ارومیه). ۲، ۶۵-۸۰.

دشتی رحمت‌آبادی، ابراهیم، محمدی، حمید و زکریا فرج‌زاده (۱۳۹۰)، «ارزیابی عملکرد الگوهای شبکه عصبی و خودرگرسیون میانگین متحرک در پیش‌بینی قیمت نفت خام ایران»، فصلنامه مطالعات اقتصاد انرژی، سال هشتم، شماره ۲۸، ۱۱۸-۹۷.

روستایی شهریور، پورمحمدی محمدرضا، و فرزانه درویشی (۱۳۹۷) «تحلیل فضایی عوامل موثر بر افزایش قیمت مسکن در کلان شهر تبریز با استفاده از ضریب هم بستگی و مدل برازش رگرسیونی»، فصل نامه مطالعات مدیریت شهری، سال دهم، شماره ۳۳، ۸۵-۹۶.

زراءنژاد, منصور, کیانی, پویان, ابراهیمی, صلاح, رئوفی, علی. (۱۳۹۱). پیش‌بینی قیمت نفت خام اوپک با استفاده از مدل خودبازگشتی میانگین متحرک انباشته فازی. پژوهشنامه اقتصاد انرژی ایران، ۱۰۷-۲۰۷.

صدرالسادات، لیلا، ستوده نیا، سلمان، امیری، علی.(۱۳۹۶). بررسی رابطه هموارسازی سود و خطر ریزش قیمت سهام در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران، پژوهش‌های جدید در مدیریت و حسابداری، شماره ۲۱، ۲۰۷-۲۳۰.

فرج زاده، منوچهر، دارند، محمد. (۱۳۸۸) مقایسه روش‌های رگرسیون خطی و شبکه عصبی مصنوعی در پیش‌بینی میزان مرگ و میربه‌عنوان تابعی از دمای هوا (مطالعه موردی: تهران). تحقیقات نظام سلامت حکیم (حکیم)، ۱۲، ۴۵-۵۳.

قلی‌زاده، وحیدپور، ۱۳۹۰، پیش‌بینی قیمت سهام با روش رگرسیون فازی، پژوهشنامهی علوم اقتصادی، ۱-۲۲.

کوه پیما جواد، ارگانی میثم، نیسانی سامانی نجمه (۱۳۹۹) «تخمین قیمت آپارتمان، با استفاده از رگرسیون خطی و وزن دار جغرافیایی (مطالعه موردی: منطقه ۶ شهر تهران) پژوهش‌های جغرافیایی برنامه‌ریزی شهری، دوره ۸، شماره ۲،۳۶۹-۳۴۷

مدیر شانه چی، محمد حسین و ارغوان علیزاده(۱۳۸۵)، پیش‌بینی کوتاه مدت قیمت نفت با استفاده از شبکه عصبی، فصلنامه مطالعات اقتصادی انرژی، سال سوم، شماره ۹، ۱- ۲۷.

Ahmed Mahoto, N., Iftikhar, R., Shaikh, A., Asiri, Y., Alghamdi, A., & Rajab, K. (2021). An Intelligent Business Model for Product Price Prediction Using Machine Learning Approach. Intelligent Automation & Soft Computing, 29(3), 147-159.

Ataman, G., & Kahraman, S. (2021). Stock Market Prediction in Brics Countries Using Linear Regression and Artificial Neural Network Hybrid Models. The Singapore Economic Review, 67(02), 635-653.

Azadeh, A. , Khakestani, M. and M. Saberi (2009), "A Flexible Fuzzy Regression Algorithm for Forecasting Oil Consumption Estimation", Energy Policy, No. 37, pp. 5567-5579.

Budiantara, M., Gunawan, H., & Utami, E. S. (2019). Perceived Usefulness, Perceived Ease of Use, Trust in Online Store, Perceived Risk Sebagai Pemicu Niat Beli Online Pada Produk Umkm "Made in Indonesia" Melalui Penggunaan E-Commerce Marketplace. Jurnal Riset Akuntansi Mercu Buana, 5(1), 19.

Bujang, M., & Baharum, N. (2016). Sample Size Guideline for Correlation Analysis. World Journal Of Social Science Research, 3(1), 37.

Chaffey, D. (2022). E-Business and E-Commerce Management: Strategy, Implementation and Practice (5th Edition) (5th (Fifth) Edition). Prentice Hall.

Chau, K., Ma, V., & Ho, D. (2001). The Pricing of 'Luckiness' in the Apartment Market. Journal of Real Estate Literature, 9(1), 29-40.

Chen, C., Kim, J. B., & Yao, L. (2017). Earnings smoothing: Does it Exacerbate or Constrain Stock Price Crash Risk? Journal of Corporate Finance, 42, 36-54.

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean-Variance Portfolio Optimization Using Machine Learning-Based Stock Price Prediction. Applied Soft Computing, 100, 106943.

Ejdys, J., Ginevicius, R., Rozsa, Z., & Janoskova, K. (2019). The Role of Perceived Risk and Security Level in Building Trust in E-Government Solutions. E+M Ekonomie a Management, 22(3), 220-235.

Emioma, C. C., & Edeki, S. O. (2021). Stock Price Prediction Using Machine Learning on Least-Squares Linear Regression Basis. Journal of Physics: Conference Series, 1734(1), 012058.

Gujarati D.N. Basic Econometrics, 3rd Edition, Mc Graw-Hill International Edition 1995.

Hayton, J. W., & Blundell, M. (2021). Exploring the Relationship Between Social Class and Sport Event Volunteering. Sport Management Review, 24(1), 92-115.

Kedem B., Fokianos K. Regression Model for Time Series Analysis". Journal of Amerian Statistical Association 2004; 99: 299.

Khashei, M., Reza Hejazi, S., & Bijari, M. (2008). A New Hybrid Artificial Neural Networks and Fuzzy Regression Model for Time Series Forecasting. Fuzzy Sets and Systems, 159(7), 769-786.

Lind D.A., Marchal W.G., Wathen S.A. Statistical Techniques in Business and Economics, 12th edition, New York, Mcgraw-Hill 2005.

Meng, Q.N., Xu, X. (2018). Price Forecasting Using an ACO-Based Support Vector Regression Ensemble in Cloud Manufacturing. Computers & Industrial Engineering, 125, 171-177.

Moskowitch, H, K. J, Kim(1993) «On Assessing the H-Value in Fuzzy Linear Regression», Journal of Fuzzy Sets and Systems, 58, 1993, Pp: 313-320.

Park, J., & Chung, H. (2009). Consumers' Travel Website Transferring Behaviour: Analysis Using Clickstream Data-Time, Frequency, and Spending. The Service Industries Journal, 29(10), 1451-1463.

Riantini, R. E., Andini, S., Florencia, M. M., & Rabiah, A. S. (2019). E-Marketing Strategy Analysis of Consumer Purchase Decision in Indonesia Online Sports Stores. 2019 International Conference on Information Management and Technology (ICIMTech).

Seitz, C., Pokrivčák, J., Tóth, M., & Plevný, M. (2017). Online Grocery Retailing in Germany: an Explorative Analysis. Journal of Business Economics and Management, 18(6), 1243-1263.

Tseng, F. M., & Lin, L. (2005). A Quadratic Interval Logit Model for Forecasting Bankruptcy. Omega, 33(1), 85-91.

Tseng, F. M., Tzeng, G. H., Yu, H. C., & Yuan, B. J. (2001). Fuzzy ARIMA Model for Forecasting the Foreign Exchange Market. Fuzzy Sets and Systems, 118(1), 9-19.

Wen, H. Z., Sheng-hua, J., & Xiao-yu, G. (2005). Hedonic Price Analysis of Urban Housing: An Empirical Research on Hangzhou, China. Journal of Zhejiang University-SCIENCE A, 6(8), 907-914.

Whitehill, S. (2009). An Introduction to Pricing Correlation Products Using a Pair-Wise Correlation Matrix. The Journal of Credit Risk, 5(1), 97-110.

Witten, I. H., & Frank, E. (2002). Data Mining. Acm Sigmod Record, 31(1), 76-77.

Wu, B., & Tseng, N. F. (2002). A New Approach to Fuzzy Regression Models with Application to Business Cycle Analysis. Fuzzy Sets and Systems, 130(1), 33-42.

Zhang, H., Nguyen, H., Vu, D. A., Bui, X. N., & Pradhan, B. (2021). Forecasting Monthly Copper Price: A Comparative Study of Various Machine Learning-Based Methods. Resources Policy, 73, 102189.