Doi: 10.22034/IAES.2023.1974208.1959

اقتصاد کشاورزی/جلد ۱۷/شماره۱/صفحه های ۱۰۸-۷۹

برآورد قیمت آتی پسته در بورس کالای کشاورزی با استفاده از الگوی هیبریدی «تبدیل موجک–گرادیان تقرب‌یافته درختی»

محمدرضا حاج سیدجوادی، رضا حیدری، فریبا عباسی[۱]

تاریخ دریافت: ۱۴۰۱/۰۹/۰۲ تاریخ پذیرش: ۱۴۰۱/۱۰/۳۰

چکیده

در طی سال‌های اخیر، بازار بورس کالای ایران همواره با نوسان‌ها و تلاطم‌های بی‌ثبات‌کننده قیمت همراه بوده است. با توجه به جایگاه مهم پسته در بورس کالای ایران و نیز لزوم به کارگیری ابزارهای مناسب برای تشخیص بهینه قیمت آتی، هدف از انجام این مطالعه، طراحی و ساخت یک مدل هیبریدی مناسب مبتنی بر گردایان تقرب‌یافته و مقایسه عملکرد آن با سایر مدل‌های یادگیری ماشین به منظور پیش‌بینی دقیق قیمت آتی پسته است. نتایج حاصل از بکارگیری تئوری موجک نشان داد که میزان خطای داده‌های قیمت کاهش یافته و داده‌ها از یک روند باثبات (نوفه سفید) برخوردار شدند. همچنین نتایج حاصل از انجام اجرای شبکه کدکننده خودکار نشان داد که وقفه بهینه یک، بهترین متغیر ورودی برای پیش‌بینی قیمت آتی پسته در دوره مورد بررسی است. بر مبنای شاخص‌های نیکویی برازش، مدل پیشنهادی این مطالعه یعنی «تبدیل موجک–گرادیان تقرب‌یافته» در مقایسه با دیگر مدل‌های داده‌کاوی، دارای عملکرد بهتری در پیش‌بینی قیمت آتی پسته بود. همچنین، پیش‌بینی خارج از نمونه با مدل منتخب نشان داد که قیمت‌های جدید پیش‌بینی شده با داده‌های واقعی اختلاف کمی دارد که بیانگر کارایی و دقت مدل هیبریدی منتخب است. بنابراین، مدل پیشنهادی برای پیش‌بینی قیمت کالاهای کشاورزی توصیه شده و می‌تواند به عنوان یک شاخص اطمینان و یک ابزاری محاسباتی کارا در مدیریت ریسک برای معامله‌گران و فعالان بازار بورس کالای ایران به کار گرفته شود.

طبقه‌بندی JEL: C5, C32, C53, D12, Q17.

واژه‌گان کلیدی: پیش‌بینی، قیمت آتی پسته، یادگیری ماشین و مدل جنگل تصادفی

۱ به ترتیب :پژوهشگر، استادیار(نویسنده مسئول) و پژوهشگر موسسه پژوهش‌های برنامه‌ریزی، اقتصاد کشاورزی و توسعه روستایی، تهران، ایران.

Email: rezaheidari3631@gmail.com

مقدمه

در فرآیند شکل‌گیری قیمت‌ها، قیمت کالاهای کشاورزی تحت تأثیر عامل‌های مختلف داخلی و خارجی قرار می‌گیرد. اما در شرایط کنونی، رخداد بحران‌های جهانی مانند تشدید خشکسالی، انتشار ویروس کرونا و جنگ اوکراین، متغیرهای بخش کشاورزی به ویژه قیمت کالاهای کشاورزی را به شدت تحت تأثیر قرار داده است (Heydari & Haj Seyed Javadi, 2022). صرف نظر از این شرایط، موضوع قیمت یک عامل کلیدی در فعالیت مالی و تجاری مرتبط با بخش کشاورزی است، به گونه‌ای که همواره فعالان بخش کشاورزی در معرض ریسک‌های ناشی از نوسان قیمت کالاهای کشاورزی قرار دارند (Hasan *et al*., 2020). این مسئله نه تنها منجر به تصمیم‌گیری نادرست در زمینه تولید بهینه محصول‌ها در سال جاری می‌شود، بلکه می‌تواند اجرای تعهدهای مالی آنان را در سال‌های آتی با خطر روبه‌رو سازد. بنابراین، توجه به عامل‌های موثر و گاهی غیرقابل پیش‌بینی مانند نوسان و تلاطم قیمت محصول‌های کشاورزی در تدوین برنامه‌ریزی‌های کشاورزی، مهم و ضروری است (Hoseini Yekani & Kashiri Kalaei, 2016).

شاخص‌های بازار بورس کالای کشاورزی نیز مانند دیگر بازارها، همواره با نوسان‌های قیمت و تلاطم‌های بی‌ثبات‌کننده همراه است، لذا به کارگیری مدل‌های پیش‌بینی‌کننده برای شناخت رخدادهای اولیه و جلوگیری از ایجاد آسیب‌های ناشی از این نوسان‌های، ضرورت دارد. در حقیقت، الگوهای مبتنی بر پایه علوم آمار و ریاضی (داده‌کاوی) در تجهیز ابزارهای بورس کالای کشاورزی در روند شکل‌گیری یک ساختار پیش‌بینی قوی هشداردهنده[1]، به عنوان یکی از الزام‌های برنامه‌ریزی برای آینده فعالان بورس کالای کشاوزی در افزایش مطمئن حجم و میزان تراکنش قراردادهای آتی محصول‌ها بشمار می‌آید. همچنین امکان حفظ با ثبات موقعیت‌های سودده کنونی فعالان به ویژه سوداگران بازار بورس کالایی را با امکان مزیت برخورداری از مدل‌های پیش‌بینی و در پی آن کشف موقعیت‌های مناسب در آینده تضمین می‌کند. امروزه طراحی الگوهای پیش‌بینی، به عنوان یکی از کاراترین روش‌های محاسباتی در مدیریت ریسک، زمینه کاهش خطا و برخی از عدم قطعیت‌های مرتبط با نوسان‌های تصادفی را در انتشار قیمت‌های لحظه‌ای از بین می‌برد (Rovshan *et al*., 2017). به عنوان مثال، از آنجایی که کالاهای کشاورزی در برابر نوسان‌های اقلیمی آسیب‌پذیر است، پیش‌بینی بازارهای کالاهای کشاورزی به فعالان بازار این اطمینان خاطر را می‌دهد که با کسب آگاهی‌های لازم بهتر می‌توانند برای فروش بهنگام

[1] Early Warning System

کالاهای خود تصمیم‌گیری کنند. همچنین، پیش‌بینی بهنگام و دقیق قیمت به دولت، سیاست‌گذاران و کشاورزان کمک می‌کند تا تصمیم‌های آگاهانه‌تری برای مدیریت بازار و مناسب‌ترین زمان فروش محصول بگیرند (Paul *et al*,. 2022 & Gosh et al., 2022).

بسیار روشن است که به دلیل وضعیت سیاسی، شرایط اقتصاد جهانی و دیگر عامل‌ها، اغلب پیش‌بینی دقیق قیمت در بازار سهام و به ویژه بورس کالای کشاورزی، فعالیتی بسیار چالش‌برانگیز و دشوار است. از این رو، بسیاری از محققان، علاقه قابل توجهی به پیش‌بینی بازار و قیمت نشان داده‌اند. به همین دلیل، رویکردهای مختلفی توسط محققان برای پیش‌بینی قیمت کالاهای کشاورزی ارائه شده است که از جمله آنها می‌توان به انواع مدل‌های داده‌کاوی شامل شبکه‌های عصبی مصنوعی هوشمند اشاره کرد. Nosratabadi et al., 2022; Gosh *et al*., 2022 and Pouyanfar *et al*., 2015). ادبیات نظری پیرامون پیش‌بینی قیمت نیز نشان می‌دهد که تکنیک‌های داده‌کاوی به ویژه مدل‌های یادگیری ماشین نقش مهمی در افزایش پیش‌بینی دقیق قیمت محصولات کشاورزی دارد و به سرعت در جهان در حال پیشرفت است. در این میان، مدل‌های گرادیان تقرب‌یافته[1] (XGBoost) به عنوان یکی از بهترین الگوریتم‌های یادگیری تحت نظارت شناخته شده است (Zarei and Iranmanesh, 2022; Hirapara and Vanjara, 2022; Chen and Guestrin, 2016).

در زمینه پیش‌بینی قیمت، بررسی‌های گوناگونی در داخل و خارج کشور انجام شده ، که به نتایج برخی از آنها اشاره می‌شود؛ Menhaj and Kavoosi Kalashmi, 2022 در تحقیق خود برای پیش‌بینی قیمت ماهانه برنج تایلندی از یک مدل ترکیبی متشکل از الگوریتم خوشه‌بندی جدید و شبکه عصبی پرسپترون چندلایه استفاده کردند. نتایج این تحقیق نشان داد که مدل ترکیبی پیشنهادی در مقایسه با مدل‌های ARIMA، EMD-ARIMA، ANFIS در پیش‌بینی قیمت برنج دارای نتایج کاراتر و مطلوب‌تری است. Nosratabadi *et al*. 2020 در بررسی‌های خود نشان دادند که عملکرد دقت پیش‌بینی مدل ترکیبی شبکه عصبی-الگوریتم گرگ خاکستری (ANN-GWO) از مدل شبکه عصبی-الگوریم رقابت استعماری (ANN-ICA) بیشتر است. نتایج پژوهش Ghaderzadeh *et al*., 2019 برای پیش‌بینی قیمت محصول‌های زراعی کردستان با استفاده از مدل‌های ARIMA، MA و AR نتایج نشان داد که از میان سه مدل منتخب، مدل ARIMA مبتنی بر معیارهای ارزیابی دقت پیش‌بینی، مقدار خطای کمتری داشته و در نتیجه قدرت

[1] Extreme Gradient Boosting Model

بالاتری در پیش‌بینی محصول‌های منتخب دارد. Yasrebi *et al*, 2019؛ Mousavi and Kalashemi, 2016؛ Saif Al-Hosseini *et al*., 2015 و Shirdeli and Tavasoli, 2015 در بررسی‌های جداگانه‌ای نشان دادند که قدرت پیش‌بینی مدل‌های هیبریدی مبتنی برشبکه‌های عصبی نسبت به مدل‌های انفرادی بالاتر است. Abasinejad and Naderi, 2012 در پژوهشی برای پیش‌بینی بازدهی شاخص قیمت و بازده نقدی بورس اوراق بهادر تهران، عملکرد انواع مدل‌های شبکه عصبی را با کمک داده‌های تجزیه‌شده توسط تبدیل موجک را ارزیابی کردند. نتایج این بررسی به برتری عملکرد مدل‌های شبکه عصبی چندلایه پیش‌خور و شبکه عصبی فازی (انفیس) مبتنی بر داده‌های نوفه‌زدایی شده به کمک آزمون موجک پیوسته در مقابل به کارگیری سطح داده‌ها دلالت دارد.

در پژوهش Guo *et al*., 2022 با استفاده از یک مدل هیبریدی متشکل از حافظه کوتاه‌مدت-بلندمدت (LSTM)، میانگین متحرک یکپارچه خودتوضیح (ARIMA) و شبکه عصبی پساانتشار (BP) قیمت ذرت تحت عامل‌های تأثیرگذار مکانی و زمانی پیش‌بینی شد و نتایج نشان داد که دقت پیش‌بینی مدل LSTM-ARIMA-BP از مدل‌های رقیب بالاتر است. در پژوهش Gosh *et al*., 2022، اثرگذاری متغیرهای توضیحی موثر بر پیش‌بینی قیمت‌های آتی سهام در بورس کالای هند را بررسی کردند. در این تحقیق برای انتخاب متغیرهای اثرگذار (قیمت نقدی، نوسان‌های نفت خام، نرخ ارز و ...) از الگوریتم‌های بوروتا[1] و جنگلی تصادفی[2]، و برای پیش‌بینی قیمت سهام از مدل هوش مصنوعی استفاده شد. نتایج بیانگر آن بود که اهمیت نسبی متغیرهای توضیحی در پیش‌بینی قیمت‌های آتی سهام بسته به شرکت مربوطه و دوره مورد بررسی متفاوت است. نتایج بررسی‌های Ahmed Osman *et al*., 2021 نشان داد که دقت پیش‌بینی مدل گرادیان تقرب‌یافته برای پیش‌بینی منابع آب زیرزمینی در مالزی از مدل شبکه عصبی مصنوعی و رگرسیون بردار پشتیبان بالاتر است. Wihartiko *et al*., 2021 با بررسی انواع مدل‌های به کار برده برای پیش‌بینی قیمت کالاهای کشاورزی دریافتند که میزان به کارگیری مدل‌های هوش مصنوعی، داده‌کاوی و رگرسیونی به ترتیب ۳۰، ۲۲ و ۱۸ درصد است و برای پیش‌بینی قیمت کالاهایت کشاورزی، مدل‌های هوشمند با در نظر گرفتن مفهوم زنجیره تامین را پیشنهاد دادند. نتایج به دست آمده از پژوهش Agarwal and Tarar, 2021 نشان داد که دقت پیش‌بینی مدل‌های شبکه عصبی مصنوعی و الگوریتم جنگلی تصادفی در حدود ۹۳ درصد است، در حالی

[1] Boruta algorithm
[2] Random Forest

که دقت پیش‌بینی مدل‌های الگوریتم یادگیری ماشین مانند حافظه کوتاه و بلندمدت، شبکه عصبی بازگشتی و ماشین بردار پشتیبان در حدود ۹۷ درصد است. همچنین در میان الگوریتم‌های یادگیری ماشین، مدل یادگیری عمیق به دلیل داشتن دقت و عملکرد بهتر نقش حیاتی در پیش‌بینی متغیرها ایفا می‌کند. نتایج تحقیق Qureshi *et al*., 2020 برای پیش‌بینی رشد تولید ناخالص داخلی واقعی ماهانه و فصلی کانادا نشان داد که مدل گرادیان تقرب یافته نسبت به مدل‌های رقیب دارای عملکرد بهتری است. Bitirgen and Filik, 2020، در بررسی‌های خود با استفاده از مدل‌های XGBoost و ARIMA اقدام به پیش‌بینی قیمت برق کردند و نتایج آن گویای سرعت بالا و خطای کمتر مدل XGBoost بود. Vignesh and Askarunisa, 2020 برای پیش‌بینی عملکرد محصول و عامل‌های موثر بر آن از مدل‌های CNN، ANN، RNN، ARIMA، LSTM، SVM و بسیاری دیگر از الگوریتم‌های هوش مصنوعی استفاده کردند. نتایج نشان داد که پارامترهای مربوط به خاک و آب و هوا نقش اصلی داشته و مدل‌های ANN و SVM نسبت به سایر مدل‌ها، الگوریتم‌های دقیق‌تری برای پیش‌بینی عملکرد محصول هستند. نتایج پژوهش Li *et al*., 2019 برای پیش‌بینی برق نشان داد که قدرت پیش‌بینی مدل ترکیبی XGBoost و LSTM از هر یک از مدل‌های انفرادی بالاتر است. Srinivas *et al*., 2019 برای پیش‌بینی میزان بارندگی از مدل‌های یادگیری ماشین استفاده کردند و نتایج آنان نشان داد که دقت پیش‌بینی مدل گرادیان تقرب‌یافته از مدل‌های جنگل تصادفی و شبکه عصبی بالاتر بوده و دارای سرعت بالاتری است. در مطالعه Ribeiro and Coelho, 2018، برای پیش‌بینی قیمت سویا و قیمت گندم از مجموعه مدل‌های GBM، RF، XGB، MLP، SVR و KNN استفاده کردند. نتایج این تحقیقه نشان داد که عملکرد مدل‌های شبکه عصبی به ویژه XGB بهتر از مدل‌های انفرادی است. در مطالعه Vafaei Ghaeini *et al*., 2018 برای پیش‌بینی قیمت سهام در بازارهای مختلف سهام بازار از مدل ترکیبی شبکه عصبی و گارچ و آریما (ARMA-GARCH-ANN) استفاده شده است و نتایج آنان گویای آن است که مدل پیشنهادی در پیش‌بینی تک دوره‌ای قیمت سهام در بازار تهران و نیویورک نسبت به دیگر مدل‌ها عملکرد بهتری دارد. Patel *et al*., 2015 برای پیش‌بینی معامله‌های سهام شرکت‌های هندی از چهار مدل شبکه عصبی مصنوعی، ماشین بردار پشتیبان، جنگل تصادفی و مدل بیزین[1] استفاده کردند. نتایج این بررسی نشان داد که مدل جنگی تصادفی عملکرد بهتری در پیش‌بینی دارد. Girish and Kanchan, 2013 در بررسی‌های

[1] Naive-Bayes

خود، برتری عملکرد پیش‌بینی شبکه عصبی مصنوعی نسبت به روش‌های برای قیمت ماهانه سویا و کلزا را نشان دادند. همچنین نتایج آنان نشان داد که ترکیب مدل‌های خطی و غیرخطی در قالب یک مدل هیبریدی تک مرحله‌ای منجر به پیش‌بینی‌های دقیق‌تری نسبت به عملکرد هر یک از مدل‌های انفرادی می‌شود. خلاصه برخی از تحقیقات یاد شده در جدول (۱) آمده است.

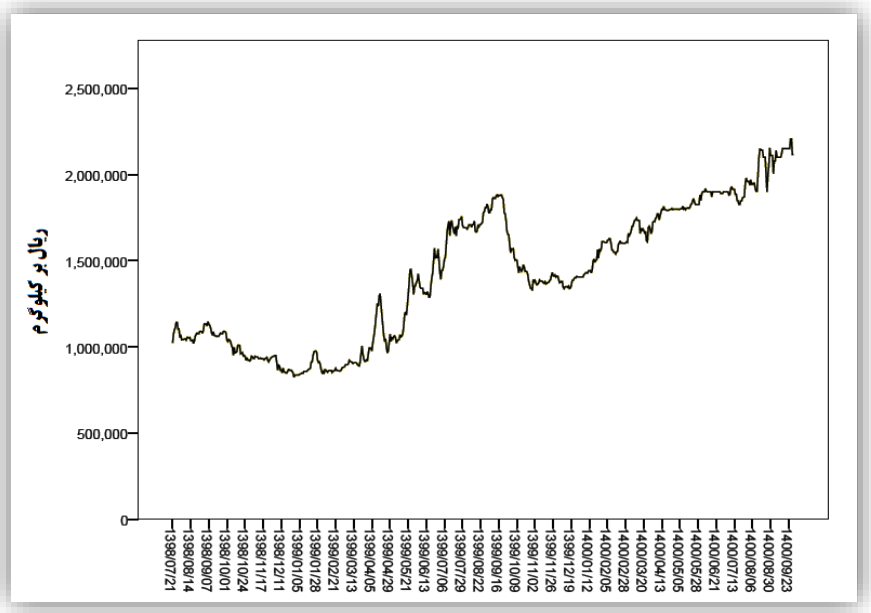
جدول (۱) خلاصه پیشینه تحقیق در پژوهش حاضر

Table(1) Summary of the research background in the present study

نویسندگان	موضوع مورد مطالعه	مدل مورد استفاده	نتیجه مطالعه
Menhaj and Kavoosi Kalashmi (2022)	پیش‌بینی قیمت ماهانه برنج تایلندی	مدل ترکیبی الگوریتم خوشه‌بندی جدید و شبکه عصبی پرسپترون چندلایه	برتری عملکرد مدل پیشنهادی نسبت به مدل‌های ساده
Guo et al. (2022)	پیش‌بینی قیمت ذرت	LSTM-ARIMA-BP	برتری عملکرد مدل پیشنهادی نسبت به مدل‌شبکه عصبی ساده
Gosh et al. (2022)	پیش‌بینی قیمت‌های آتی سهام در بورس کالای هند	الگوریتم‌های بوروتا و جنگلی تصادفی	برتری عملکرد مدل پیشنهادی نسبت به مدل‌های رقیب
Ahmed Osman et al. (2021)	پیش‌بینی منابع آب زیرزمینی در مالزی	مدل گرادیان تقرب‌یافته	برتری عملکرد مدل پیشنهادی نسبت به مدل‌های رقیب
Wihartiko et al. (2021)	پیش‌بینی قیمت محصولات کشاورزی	مدل‌های هوش مصنوعی	برتری عملکرد مدل پیشنهادی نسبت به مدل‌های خطی
Nosratabadi et al. (2020)	پیش‌بینی قیمت	مدل ترکیبی شبکه عصبی- الگوریتم گرگ خاکستری	برتری عملکرد مدل پیشنهادی نسبت به مدل‌های تکی هوشمند
Bitirgen and Filik (2020)	پیش‌بینی قیمت برق	XGBoost	برتری عملکرد مدل پیشنهادی نسبت به مدل‌های رقیب
Vignesh and Askarunisa (2020)	پیش‌بینی عملکرد محصول	SVM	برتری عملکرد مدل پیشنهادی نسبت به مدل شبکه عسبی
Ghaderzadeh et al. (2019)	پیش‌بینی قیمت محصولات زراعی کردستان	ARIMA	برتری عملکرد مدل پیشنهادی نسبت به سایر مدل‌های خطی
Vafaei Ghaeini et al. (2018)	پیش‌بینی قیمت سهام	ARMA-GARCH-ANN	برتری عملکرد مدل پیشنهادی نسبت به مدل‌های خطی و شبکه عصبی ساده

در تحلیل و بررسی پژوهش های انجام شده داخلی می‌توان بیان داشت که به رغم وجود بررسی‌های گسترده خارج از کشور در پیش‌بینی در موضوع‌های متنوع مانند تولید، صادرات و

قیمت از طریق به کارگیری مدل‌های یاد شده خاص کالاهای مختلف کشاورزی، تاکنون پژوهشی در مورد پیش‌بینی قیمت آتی محصول صادراتی پسته، با توجه به زمان کوتاه از تشکیل رینگ معاملاتی آن در بورس کالا (سه سال) از طریق مدل‌های متفاوت غیرخطی آموزش ماشین و در یک رفتار کامل‌تر به عنوان مدل‌های هیبریدی انجام نشده است. بی‌شک محصول پسته یکی از مهم‌ترین محصول‌های صادراتی ایران در بخش کشاورزی است، و سهم قابل ملاحظه‌ای از تولید، سطح زیر کشت، میزان و ارزش صادرات جهانی این محصول را به خود اختصاص داده است. بر مبنای گزارش وزارت کشاورزی امریکا در سال ۲۰۲۲، از کل حدود ۷۵۰ هزار تن تولید پسته، کشورهای امریکا، ایران، ترکیه، سوریه و یونان به ترتیب با تولید ۵۲۹ (۷۰٪)، ۱۳۵ (۱۸٪)، ۵۹ (۸٪)، ۹ (۱/۲٪) و ۷ (۱٪) هزار تن تولیدکنندگان عمده پسته در جهان بوده‌اند. همچنین میزان صادرات پسته در طی سال‌های ۱۳۹۸ و ۱۳۹۹ به ترتیب ۱۴۰ و ۲۰۶ هزار تن بوده است، اما این مقدار در سال ۱۴۰۰ به ۸۸ هزار تن کاهش یافته است (Iran Pistachio Association, 2022). روند زمانی قیمت آتی پسته ایران در طی سال‌های ۱۳۹۸ تا ۱۴۰۰ در نمودار (۱) نشان داده شده است. بر مبنای نمودار (۱)، قیمت آتی پسته در طی سال‌های اخیر با نوسان‌های قیمتی همراه است (Iran Mercantile Exchange, 2022). افزون براین، نتایج بررسی‌های اخیر نشان داده است که نوسان‌های درآمدی پسته کاران ایران بالا بوده و کشاورزان توانایی پیش‌یبنی و رویارویی با آن را ندارند. بر مبنای تحقیقات انجام گرفته، نزدیک به ۵۰ درصد از نوسان‌های قیمت پسته در ایران تصادفی و پیش‌بینی نشده است و برای کاهش این نوسان‌ها باید از ابزاری مناسب استفاده کرد (Hoseini Yekani & Kashiri Kalaei, 2016). بنابراین، بازار این محصول نیز هم مانند دیگر کالاهای کشاورزی با ریسک قیمتی و عدم قطعیت همراه است؛ لذا توجه به نوسان‌های قیمتی آن و اندازه‌گیری این نوسان‌ها در کنار بررسی مواردی همچون کشف قیمت، ثبات بازار و کارایی بازار از دارای اهمیت ویژه‌ای است. با توجه به اهمیت بازار آتی پسته در بورس کالای ایران و لزوم به کارگیری ابزارهای مناسب برای جلوگیری از نوسان‌های قیمت از یک سو، و انجام نشدن تحقیقی پیرامون پیش‌بینی قیمت آتی پسته با استفاده از مدل‌های یادگیری ماشین از سوی دیگر، هدف از این انجام پژوهش، پیش‌بینی قیمت آتی پسته ایران با استفاده از مدل هیبریدی مبتنی بر گردایان تقرب‌یافته و مقایسه عملکرد آن با دیگر مدل‌های هیبریدی یادگیری ماشین است.

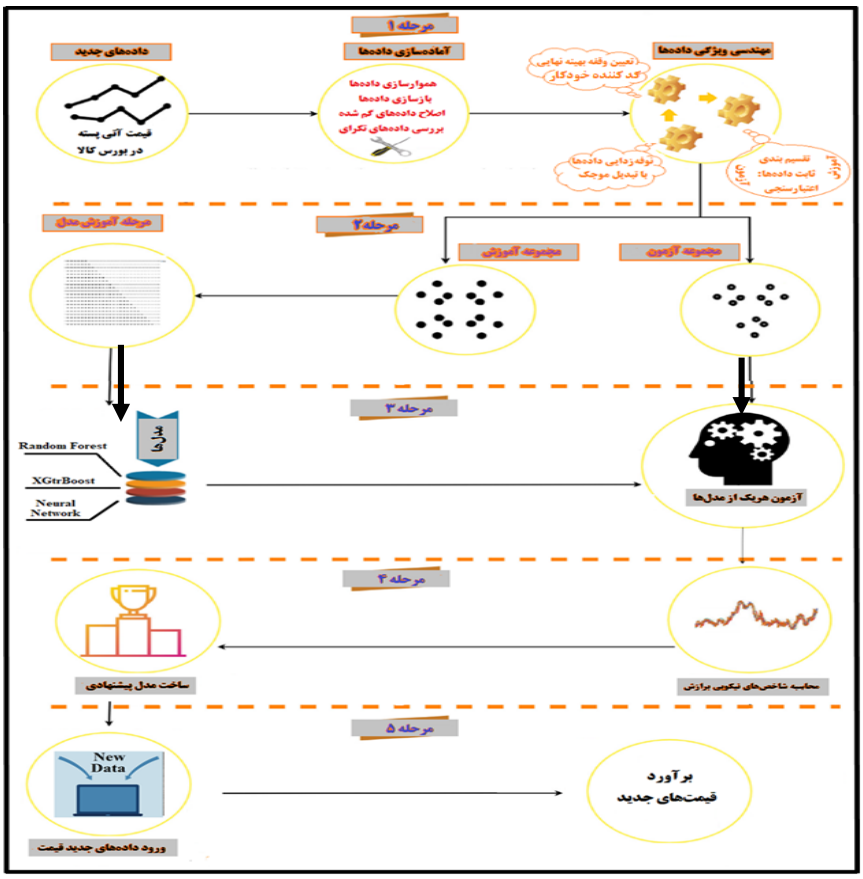


شکل (۱) روند تغییر قیمت روزانه آتی پسته از تاریخ ۱۳۹۸/۰۷/۲۱ تا ۱۴۰۰/۰۹/۲۳ بر حسب ریال در کیلوگرم
Figure 1 The trend of future price changes of pistachio from 13/10/2019 to 14/12/2021 in Rials per kg

روش تحقیق

استفاده از مدل‌های هیبریدی یک راه معمول در بهبود دقت پیش‌بینی‌ها و چیره شدن بر محدودیت‌های مدل‌های تکی است. در واقع با هیبرید کردن مدل‌های مختلف می‌توان نقاط ضعف یک مدل را با استفاده استفاده از نقاط قوت چندین مدل ارتقا داد (Heydari and Haj Seyed Javadi, 2022). در شکل (۲)، چهارچوب مدل هیبریدی پیشنهادی برای پیش‌بینی قیمت آتی پسته در بورس کالای ایران نشان داده شده است. مدل هیبریدی پیشنهادی «تبدیل موجک[1]-گرادیان تقرب‌یافته» (WT-XGBoost) شامل مرحله‌های «آماده‌سازی داده‌ها»، «مهندسی ویژگی‌های داده»، «آموزش و آزمون مدل‌ها»، «ساخت مدل پیشنهادی (پایه)» با استفاده از شاخص‌های نیکویی برازش و «برآورد قیمت‌های جدید» از طریق ورود داده‌های جدید قیمت به بهترین مدل انتخابی است.

[1] Wavelet Transform (WT)



شکل (۲) چارچوب فرآیند ساخت مدل هیبریدی پیشنهادی برای پیش‌بینی قیمت آتی پسته

Figure (2): The framework of the making process of the proposed hybrid model for forecasting the future price of pistachios

در مرحله آماده‌سازی داده‌ها، با استفاده از روش‌های هموارسازی داده‌ها، بازسازی داده‌ها، اصلاح داده‌های تکراری در چند روز متوالی و اصلاح داده‌های گم شده، مجموعه پیوسته سری زمانی قیمت آتی پسته از سررسیدهای قراردادهای آتی استخراج می‌شود. مرحله مهندسی ویژگی‌های داده شامل نوفه‌زدایی داده‌ها (حذف نویزها)، تعیین وقفه بهینه قیمت و تقسیم‌بندی ثابت داده‌ها از طریق فرآیندهای آموزش، اعتبارسنجی و آزمون مشخص است. در واقع، در مرحله دوم برای نوفه‌زدایی سری زمانی قیمت از تبدیل موجک و برای تعیین تعداد وقفه بهینه از هم از شبکه کدکننده خودکار شبکه عصبی عمیق و هم از الگوریتم ژنتیک استفاده شده است. در مرحله سوم مدل‌های آموزش داده شده و آزمون می‌شوند تا پارامترهای بهینه مربوط به هر یک از مدل‌ها به

دست آید. لازم به یادآوری است که در این پژوهش ۷۵ درصد از داده‌ها برای مراحه‌های آموزش و اعتبارسنجی و ۲۵ درصد داده‌ها برای مرحله آزمایش به طور ثابت استفاده شد. در مرحله ساخت مدل پیشنهادی یا پایه، با استفاده از شاخص‌های نیکویی برازش در آغاز مدل‌های مختلف با یکدیگر مقایسه شده و پس از ارزیابی بهترین ساختار مدل (تعیین پارامترهای بهینه مدل از طریق فرایند کالیبراسیون) انتخاب می‌شود. در واقع عملکرد مدل هیبریدی «تبدیل موجک-گرادیان تقرب‌یافته» در پیش‌بینی قیمت آتی پسته با دو مدل هیبریدی «تبدیل موجک-شبکه عصبی» (WT-ANN) و «تبدیل موجک-جنگلی تصادفی» (WT-RF) نیز محاسبه و درنتیجه مقایسه می‌شود. مدل پایه منتخب در مابین مدل‌های محاسباتی فوق، مدلی است که می‌تواند به عنوان مدل اصلی در پیش‌بینی قیمت در دوره‌های آتی استفاده شود. در مرحله آخر نیز با ورود داده‌های جدید به مدل پایه، برآورد قیمت‌های جدید انجام می‌شود که ترسیم‌کننده کاربرد مدل‌های پیش‌بینی است. در ادامه هر یک از اجزای مدل پیشنهادی معرفی می‌شوند.

۱) تبدیل موجک

بررسی پژوهش‌های مختلف نشان داده است که استفاده از تبدیل موجک در نوفه‌زدایی داده‌های سری زمانی قیمت، سبب افزایش دقت پیش‌بینی می‌شود (Sadeghi and Dehghani Firouzabadi, 2017). لذا، در این پژوهش برای پیش پردازش اطلاعات و نوفه‌زدایی داده‌های قیمت آتی پسته از تابع تبدیل موجک استفاده می‌شود. تبدیل موجک یک ابزار قوی ریاضی است که با استفاده از تابع‌های پایه‌ای، یک سری زمانی را در زمان و مقیاس‌های مختلف نشان می‌دهد. به طور کلی، موجک گسسته مادر به صورت رابطه (۱) تعریف می‌شود:

$$\Psi_{m,n}\left(\frac{t-u}{s}\right)=\frac{1}{\sqrt{s_o^{\overline{m}}}}\Psi\left(\frac{t-nu_o s_o^m}{s_o^m}\right) \qquad (۱)$$

در رابطه (۱)، n بیانگر مکان موجک و m بیانگر اندازه موجک است. اغلب برای ساده‌تر شدن مقادیر so = 2 و u =1 در نظر گرفته می‌شود. در این حالت تابع موجک گسسته مادر به صورت رابطه (۲) در می‌آید.

$$w_{m,n}=2^{-\frac{m}{2}}\sum_{t=0}^{N-1}\Psi(2^{-m}t-n)x(t) \qquad (۲)$$

در رابطه (۲)، $w_{m,n}$ نشان‌دهنده ضریب‌های به دست آمده از استفاده از تابع موجک برای سری زمانی گسسته در حالت $s = 2^m$ و $u = 2^m \times n$ است. بر اساس نظریه ملات (۱۹۹۸)، یک سری زمانی x(t) با استفاده از معکوس تابع موجک به صورت رابطه (۳) به دست می‌آید:

$$x(t) = A_M(t) + \sum_{m=1}^{M} D_M(t) \quad (۳)$$

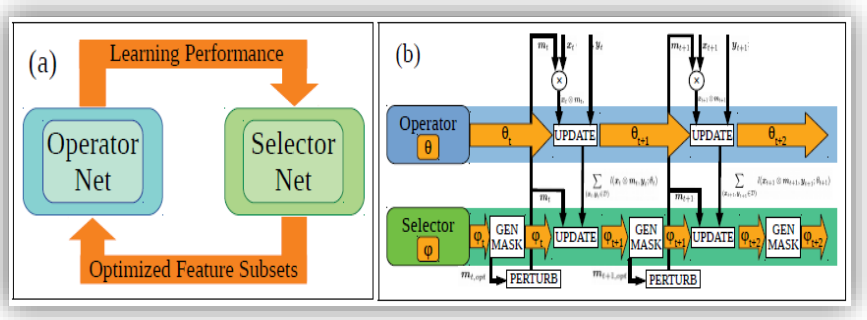
در رابطه (۳)، $A_m(t)$ بیانگر بخش تقریب‌ها یا به بیان دیگر جزء هموار شده سری زمانی است و $D_m(t)$ $(m=1,2,\ldots,M)$ بیانگر جزئیات و یا نوفه موجود در سری زمانی است (Heydari & Haj Seyed Javadi, 2022 and Shabri *et al*., 2014).

۲) انتخاب وقفه بهینه با استفاده از شبکه کدکننده خودکار در مدل‌های یادگیری عمیق

نتایج بررسی‌های مختلف نشان نیز داده است که مدل شبکه عصبی یادگیری عمیق ابزار توانمندی برای رتبه‌بندی اهمیت متغیرها یا ویژگی‌های یک مجموعه داده است (Samek *et al*., 2017; Weston *et al*., 2003). بنابراین، در این مطالعه بر مبنای توانایی شبکه عصبی یادگیری عمیق در روش رتبه‌بندی متغیرها، وقفه‌های بهینه قیمت آتی پسته تعیین می‌شود. توضیح آنکه متغیرهای توضیحی ورودی به شبکه عصبی یادگیری عمیق برای سری زمانی تک متغیره قیمت، همان وقفه‌های قیمت است. بنابراین، رتبه‌بندی متغیرهای ورودی به شبکه عصبی به معنای تعیین وقفه‌های بهینه قیمت آتی پسته است که با استفاده از شبکه کدکننده خودکار موجود در ساختار شبکه عصبی یادگیری عمیق تعریف می‌شود. در شکل (۳) نمایی از فرآیند رتبه‌بندی متغیرها یا ویژگی‌ها بر مبنای مدل‌های یادگیری عمیق نشان داده شده است.

در شبکه عصبی یادگیری عمیق، رتبه‌بندی اهمیت متغیرها، سهم متغیرهای ورودی به مدل شبکه عصبی را در عملکرد یک مدل یادگیری نظارت شده اندازه‌گیری می‌کند. در مدل شبکه عصبی عمیق با هدف انتخاب وقفه بهینه، آموزش شبکه و یافتن پارامترهای بهینه برای پیش‌بینی به طور همزمان توسط دو شبکه عملگر[1] و انتخابگر[2] انجام می‌شود. هر دو شبکه عملگر و انتخابگر به صورت متناوب و به طور مشترک آموزش می‌بینند (Wojtas and Chen, 2020). در این تحقیق، متغیرهای ورودی مدل شامل وقفه‌های قیمت آتی پسته است که انتخاب وقفه بهینه همزمان با آموزش شبکه در مدل یادگیری عمیق انجام می‌شود.

[1] Operator
[2] Selector



شکل (۳) مدل رتبه‌بندی متغیرها (ویژگی‌های یک مجموعه داده): (a) معماری شبکه عملگر و انتخابگر. (b) چرخه همزمانی و بروزرسانی متغیرها (Wojtas and Chen, 2020).

Figure (3) Ranking model of variables (characteristics of a data set): (a) the architecture of operator and selector network, (b) the synchronization and updating cycle of variables

۳) مدل جنگل تصادفی

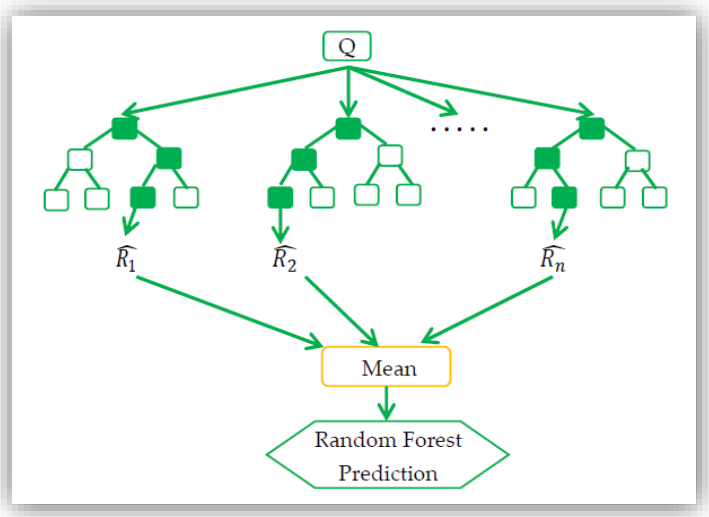
مدل جنگل تصادفی به دلیل سادگی و قابلیت استفاده، از جمله پر کاربردترین الگوریتم‌های یادگیری ماشین به شمار می‌آید. مدل جنگل تصادفی دارای چندین برتری نسبت به دیگر روش‌های مدل‌سازی است. در این مدل متغیرهای مورد استفاده می‌تواند پیوسته یا طبقه‌ای باشد. همچنین، به دلیل ایجاد شمار زیاد درخت و میانگین‌گیری در اجرای مدل جنگل تصادفی، این طبقه‌بندی‌کننده به نتایج با تعصب اندک و تغییرپذیری کم، ولی پیش‌بینی‌های دقیق منجر می‌شود. با توجه به این برتری‌های کاربرد الگوریتم‌های جنگل تصادفی در حال افزایش است (Yeh *et al.* 2014).

روند کلی الگوریتم جنگل تصادفی در شکل (۴) آمده است. در این روش بردار تصادفی Θ_k که مستقل از بردارهای تصادفی $\Theta_1 \ldots \Theta_{k-1}$ بوده، برای درخت Kام، تولید می‌شود. همچنین همه بردارها دارای توزیع یکسانی می‌باشند. درخت رگرسیونی با استفاده از مجموعه داده‌های آموزش و Θ_k رشد می‌کند و نتیجه مجموعه درخت‌های k برابر، $\{h_1(x), h_2(x) \ldots h_k(x)\}$ می‌باشد که در اینجا $h_k(x) = h(x, \Theta_k), x = \{x_1, x_2, \ldots x_p\}$ می‌باشد. این بردارها یک بردار ورودی P بُعدی می‌باشند که یک جنگل را تشکیل می‌دهند. خروجی‌های K تولید شده گروهی، مربوط به هر درخت برابر $\hat{y}_1 = h_1(x), \hat{y}_2 = h_2(x), \ldots, \hat{y}_k = h_k(x)$ که $\hat{y}_k$ خروجی درخت Kام می‌باشد.

برای به دست آوردن خروجی نهایی، میانگین همه پیش‌بینی‌های درخت‌ها محاسبه می‌شود. خطای پیش‌بینی نیز برامبنای نمونه‌های خارج از کیسه مطابق با رابطه (۴) محاسبه می‌شود.

$$MSE^{OOB} = \frac{1}{n}\sum_{i=1}^{n}[\hat{y}((x_i) - y_i]^2 \qquad (۴)$$

در رابطه (۴)، $\hat{y}((x_i)$ نتایج محاسباتی، y_i نتایج مشاهداتی و n تعداد کل مشاهدات است و MSE نیز نشان‌دهنده میزان خطای است. روش جنگلی تصادفی برای طبقه به این صورت است که در ابتدا T نمونه خودراه‌انداز از داده‌های آموزشی انتخاب می‌شود و آنگاه از هر نمونه خود راه انداز β یک درخت طبقه‌بندی و رگرسیون[1] (CART) هرس نشده ایجاد می‌شود که برای انشعاب در هر گره CART، تنها یکی از M ویژگی انتخاب شده به صورت تصادفی استفاده می‌شود. در نهایت، خروجی طبقه‌بندی بر مبنای یک نتیجه میانگین از پیش‌بینی‌های تمام درخت‌های منفرد آموزش دیده، به دست می‌آید (Duda *et al*. 2011).



شکل (۴) ساختار الگوریتم جنگلی تصادفی (Amjad et al., 2022)

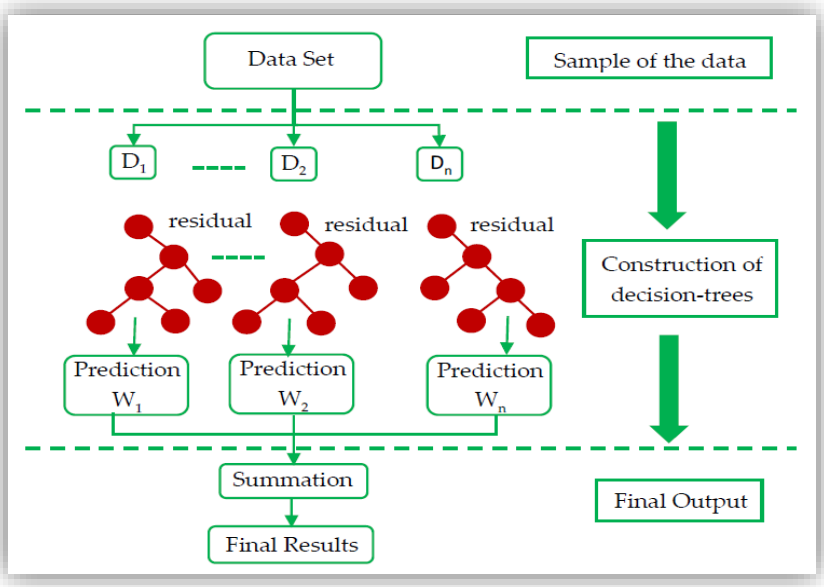
Figure (4) The structure of random forest algorithm

۴) مدل گرادیان تقرب‌یافته درختی (XGBoost)

مدل گرادیان تقرب‌یافته درختی توسط چن و گسترین (۲۰۱۶) پیشنهاد شده است. این روش در بسیاری از زمینه‌های علمی به ویژه رگرسیون و طبقه‌بندی بکار گرفته شده و بدلیل

[1] Classification and regression tree (CART)

منظم‌سازی و پردازش موازی عملکرد بسیار خوبی داشته است. الگوریتم گرادیان تقرب‌یافته یک روش یادگیری ماشینی مبتنی بر درخت تصمیم است، به طوری که نسخه کارآمدتری از درختان تصمیم‌گیری تقویت‌کننده گرادیان است (Mysen & Thornton, 2021; Nielsen, 2016). ایده اصلی در مدل گرادیان تقرب‌یافته درختی آن است که یک تابع زیان به صورت مربعات خطاها بهینه می‌شود. در واقع مدل‌ها به‌طور متوالی اضافه می‌شوند تا زمانی که هیچ بهبود دیگری در تابع زیان ایجاد نشود (Brownlee, 2021; Qureshi *et al.*, 2020). ساختار مدل گرادیان تقرب‌یافته در شکل (۵) نشان داده شده است.



شکل (۵) ساختار مدل درختی XGBoost (Amjad et al., 2022)

Figure (5) The structure of XGBoost model

روش کار مدل XGBoost به شرح زیر است. با فرض اینکه یک مجموعه داده به صورت رابطه (۶) داشته باشیم که دارای m ویژگی و n تعداد نمونه است، آنگاه خروجی پیش‌بینی شده یک درخت پیشنهادی به صورت رابطه (۷) قابل بیان است.

$$DS = \{(x_i, y_i) : i = 1,2, \dots, n \; and \; x_i \in \mathbb{R}\} \quad (۵)$$

$$M_i = \emptyset(x_i) = \sum_{k=1}^{K} f_k(x_i) \quad ; \quad f_k \in \xi \quad (۶)$$

در رابطه (۶)، K نشان‌دهنده تعداد درختان در مدل و f_k نشان‌دهنده kامین درخت است. برای بهینه‌سازی مجموعه درخت‌ها و کاهش خطا، وظیفه الگوریتم کمینه کردن تابع هدف است. به بیان دیگر برای حل معادله (۶)، باید بهترین مجموعه تابع‌ها را با به کمینه رساندن زیان مطابق با رابطه (۷) به دست آورد:

$$\xi(\emptyset) = \sum_i l(y_i, M_i) + \sum_k \Omega(f_k) \quad (۷)$$

در رابطه (۷)، l نشان‌دهنده تابع زیان است که تفاوت بین خروجی پیش‌بینی شده $\widehat{y_i}$ و خروجی واقعی y_i است، Ω بیانگر تابع پنالتی برای کاهش میزان پیچیدگی مدل است که از برازش بیش از حد مدل جلوگیری می‌کند و با استفاده از رابطه (۸) قابل محاسبه است:

$$\Omega(f_k) = \gamma^T + \frac{1}{2}\lambda\|w\|^2 \quad (۸)$$

در رابطه (۸)، T و w به ترتیب نشان‌دهنده تعداد برگ‌های درخت و وزن هر برگ است (Ahmed Osman *et al.*, 2021).

۵) معیارهای ارزیابی پیش‌بینی مدل‌ها

اغلب برای ارزیابی مدل‌های پیاده‌سازی شده از برخی از شاخص‌های آماری استفاده می‌شود. این شاخص‌ها تفاوت و انحراف بین داده‌های واقعی و پیش‌بینی شده را ارزیابی کند. در این پژوهش از شاخص‌های مجذور میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و میانگین خطای نسبی (MRE) استفاده می‌شود. هر یک از این شاخص‌ها می‌تواند مدل پیاده‌سازی شده در این پژوهش را مورد اعتبارسنجی قرار دهد، به طوری که مدلی دارای بهترین عملکرد است که کمترین مقدار شاخص‌های بالا را داشته باشد. فرمول شاخص‌های فوق به صورت زیر است.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \quad (۹)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad (۱۰)$$

$$\text{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \quad (۱۱)$$

در رابطه‌های بالا، Y_i داده‌های واقعی (مشاهده شده)، $\widehat{Y}_i$ داده‌های پیش‌بینی شده (شبیه‌سازی شده)، $\overline{Y}_i$ میانگین داده‌های واقعی و n نشان‌دهنده دوره زمانی ارزیابی است (Yang *et al.*, 2016; Steurer *et al.*, 2021) برای انجام این پژوهش از نرم‌افزارهای MATLAB، Rapid Miner و Scikit learn (Python) استفاده شد.

نتایج و بحث

در این پژوهش برای طراحی و ساخت بهترین مدل غیر خطی برای پیش‌بینی قیمت آتی پسته در بورس کالای کشاورزی ایران از یک مدل داده‌کاوی هیبریدی مبتنی بر یادگیری ماشین استفاده شد. بر این مبنای، نتایج به دست آمده از بررسی ویژگی داده‌ها، نوفه‌زادیی داده‌ها با استفاده از نظریه (تئوری) موجک، انتخاب وقفه بهینه با استفاده از شبکه کدکننده خودکار در مدل شبکه عصبی عمیق و الگوریتم ژنتیک، مقایسه و انتخاب بهترین مدل هیبریدی از میان مدل‌های رقیب با استفاده از شاخص‌های نیکویی برازش (ساخت مدل پیشنهادی)، و در نهایت برآورد قیمت‌های جدید پسته آتی از طریق ورود داده‌های جدید قیمت به بهترین مدل انتخابی است.

الف) بررسی ویژگی‌های سری زمانی قیمت آتی پسته در بورس کالای ایران

تعداد سررسید قراردادهای آتی پسته مورد بررسی در این پژوهش ۱۴ مورد است. داده‌های مورد استفاده در این مطالعه سری زمانی قیمت روزانه سررسید قراردادهای آتی پسته در بورس کالای ایران در دوره زمانی ۱۳۹۸/۰۷/۲۱ تا ۱۴۰۰/۰۹/۲۳ بر حسب ریال در کیلوگرم است. اغلب با پایان یافتن یک سررسید و فرا رسیدن تاریخ سررسید بعدی، شاهد یک جهش غیر عادی در قیمت‌های آتی پسته هستیم. برای مقابله با این مسئله از راه‌حل «هموار کردن داده‌ها» در فرآیند انتقال سررسیدها استفاده می‌شود. در این پژوهش برای هموارسازی داده‌ها از روش میانگین موزون استفاده شد، به طوری که بر مبنای قیمت تسویه ۹ روز پایانی هر سررسید و ۹ روز معادل همان تاریخ در سررسید بعدی میانگین وزنی گرفته شد. با این روش یک سری زمانی از قیمت‌های آتی پسته صرف نظر از طول مدت هر سررسید قرارداد آتی ایجاد می‌شود؛ سری یاد شده شامل ۶۵۴ داده روزانه قیمت است که ۷۵ درصد از داده‌ها برای مرحله‌های آموزش و اعتبارسنجی (۴۹۰ عدد) و ۲۵ درصد داده‌ها (۱۶۴ عدد) برای مرحله آزمایش استفاده شد. همچنین برای آماده‌سازی داده‌ها فرآیندهای بازسازی داده‌ها (به منظور یکسان‌سازی فرمت داده‌ها)، اصلاح داده‌های تکرار شده در چند روز متوالی و اصلاح داده‌های گم شده انجام پذیرفت. ویژگی‌های داده‌های قیمت پسته پس از مرحله آماده‌سازی در جدول (۲) آمده است. نتایج به دست آمده از دو آزمون کولموگروف و شاپیرو[1] در جدول (۲) نشان می‌دهد که داده‌های قیمت آتی پسته دارای

[1] Kolmogorov and Shapiro

توزیع نرمال نیست. نتایج به دست آمده از آزمون مقادیر بحرانی[1] سررسید قراردادهای آتی پسته نیز در جدول (۳) آمده و نتایج نشان داد که سری زمانی داده‌های قیمت آتی پسته در ابتدای دوره دارای کمترین مقدار و در اواخر دوره دارای بیشترین مقدار است، لذا قیمت آتی پسته در دوره منتخب دارای نوسان بالایی بوده و پیش‌بینی قیمت برای نمونه‌های خارج از نمونه به طور حتم از روابط غیر خطی و رفتار نامنظم پیروی کند.

جدول (۲) ویژگی‌های داده‌های روزانه سررسید قراردادهای آتی پسته در دوره مورد بررسی

Table 2 Characteristics of the daily maturity data of pistachio futures contracts in the review period

آماره آزمون نرمالیتی شاپیرو Shapiro normality test	آماره آزمون نرمالیتی کولموگروف Kolmogorov normality test	حداکثر Maximum	حداقل Minimum	انحراف معیار Standard deviation	میانگین Average	تعداد داده‌ها Number of data
0.932	0.131	2207000	824236	393777	1412824	654

Source: The research findings

منبع: یافته‌های تحقیق

جدول (۳) نتایج آزمون مقادیر بحرانی سررسید روزانه قراردادهای آتی پسته در دوره مورد بررسی

Table 3 Test results of daily maturity critical values of pistachio futures contracts in the period under review

مقدار قیمت Price values	تاریخ Date	ردیف Raw		مقدار قیمت Price values	تاریخ Date	ردیف Raw	
824236	1398/12/27	1	کمترین مقادیر The lowest values	2207000	1400/09/25	1	بیشترین مقادیر The maximum values
837230	1399/01/04	2		2207000	1400/09/27	2	
837230	1399/01/03	3		2154000	1400/09/29	3	
837230	1399/01/02	4		2150000	1400/09/15	4	
837230	1398/12/29	5		2150000	1400/09/16	5	

Source: The research findings

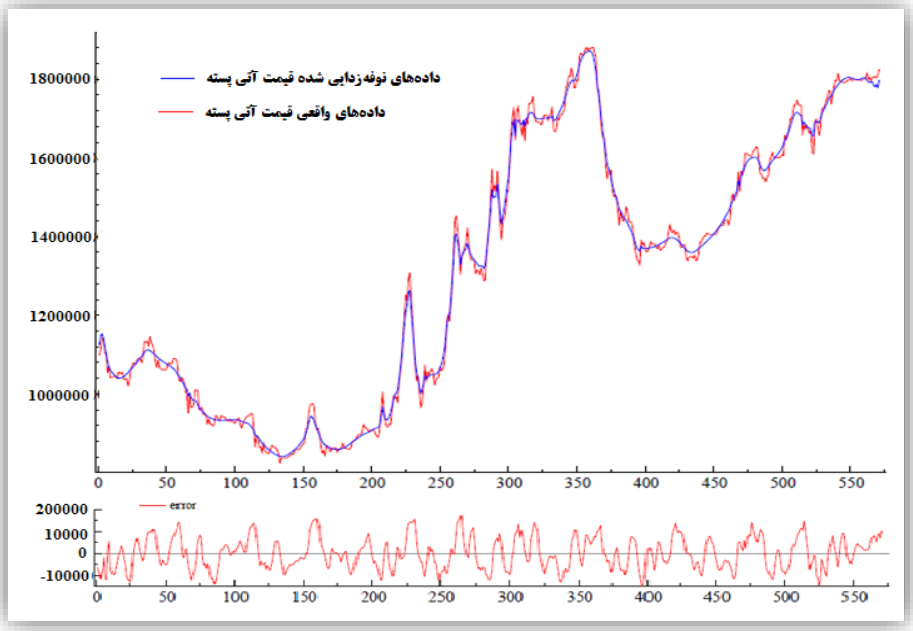
منبع: یافته‌های تحقیق

ب) نوفه‌زادیی داده‌های قیمت آتی پسته با استفاده از تابع تبدیل موجک

در این پژوهش برای نوفه‌زدایی داده‌ها از تبدیل موجک در نرم‌افزار متلب استفاده شد. در این زمینه به منظور پیدا کردن نوع و طول موجک مناسب، در مرحله اول، نوع و طول موجک مناسب از میان انواع متنوع از خانواده موجک انتخاب شده و پس از آن عملیات تجزیه و بازسازی آن برای رسیدن به کمترین میزان خطا چندین مرتبه تکرار شد. تابع‌های موجک مورد استفاده

[1] Extreme values

سیملت، هار، دابشیز و کویفلت می‌باشد. نتایج به دست آمده از مرحله نوفه‌زدایی (بازسازی قیمت آتی پسته) و به انضمام میزان خطای موجود در نوفه در شکل (۶) نشان داده شده است. در شکل (۶)، نمودار پیوسته با رنگ قرمز و آبی به ترتیب نشان‌دهنده قیمت آتی پسته بدون نوفه‌زدایی و نوفه‌زدایی شده است و اختلاف دو نمودار قرمز و آبی به عنوان نوفه به شمار آمده و نشان می‌دهد که میزان خطا در یک دامنه-کم-تلاطم جریان داشته و به نوعی داده‌ها از نوفه (نویز) کمی برخوردار بوده و دارای خطای اندکی می‌باشند.



شکل (۶) سری زمانی قیمت آتی پسته و مولفه‌های تجزیه شده توسط موجک دابشیز ۲ با سطح بهینه ۳

Figure (6) The time series of the future price of pistachio and components analyzed by Dabshiz wavelet 2 with optimal level 3

Source: The research findings منبع: یافته‌های تحقیق

ج) انتخاب وقفه بهینه با استفاده از روش کدکننده خودکار در شبکه عصبی عمیق

پس از پردازش داده‌های قیمت آتی پسته با استفاده از تابع موجک، از داده‌های نوفه‌زدایی شده برای ورودی مدل شبکه عصبی هوشمند عمیق استفاده می‌شود. اما یکی از مرحله‌های مهم پیش از ساخت مدل پایه و پیش‌بینی قیمت آتی پسته، تعیین تعداد وقفه‌های بهینه موثر بر قیمت

است. در این پژوهش برای یافتن تعداد وقفه بهینه از «کدکننده خودکار در شبکه عصبی عمیق» و الگوریتم ژنتیک استفاده شد، که یکی از نقاط قوت این پژوهش می‌باشد. نتایج به دست آمده از انجام اجرای شبکه انتخابگر در شبکه عصبی عمیق نشان داد که وقفه بهینه یک، بهترین متغیر ورودی (مابین ۴ وقفه انتخابی) برای پیش‌بینی قیمت آتی پسته در دوره مورد بررسی است. همچنین نتایج به دست آمده از الگوریتم ژنتیک نیز بیانگر وقفه بهینه یک بود. بنابراین، قیمت آتی پسته با یک وقفه می‌تواند بر روند پیش‌بینی آن موثر باشد.

ت) تعیین الگوی بهینه در مدل‌های شبکه هوشمند عصبی، جنگل تصادفی، گرادیان تقرب‌یافته و در پیش‌بینی قیمت آتی پسته

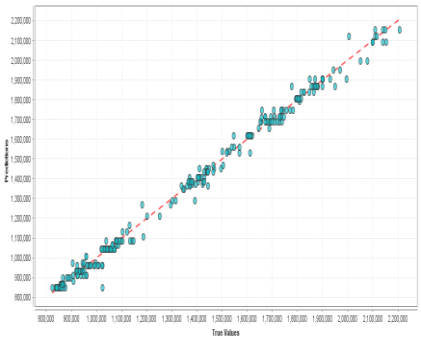
۱) مدل شبکه هوشمند عصبی: از جمله مدل‌های رقیب منتخب برای پیش‌بینی قیمت آتی پسته، مدل هیبریدی «تبدیل موجک-شبکه عصبی» است. تعداد نرون در لایه ورودی، میانی و خروجی به ترتیب ۱، ۳ و ۱ انتخاب شد. برای تابع فعال‌سازی ورودی و خروجی نیز به ترتیب از تابع‌های هایپربولیک و سیگموئید استفاده شد. همچنین تابع آموزش از نوع گرادیان نزولی بوده و پارامترهای نرخ یادگیری و نرخ تصحیح چرخه (سیکل) به ترتیب ۰/۹ و ۰/۰۱ تعیین شد. نمودار عملکرد الگوریتم شبکه عصبی برای قیمت آتی پسته در شکل (۷) قسمت (الف) آمده است.

۲) الگوریتم جنگل تصادفی: یکی دیگر از مدل‌های یادگیری رقیب منتخب در این پژوهش برای پیش‌بینی قیمت آتی پسته، مدل هیبریدی «تبدیل موجک-جنگل تصادفی» است. پس از چندین مرتبه آزمون و خطا در اجرای الگوریتم جنگل تصادفی، مقادیر بهینه پارامترهای مدل جنگل تصادفی تعیین شد به صورتی که تعداد درخت مورد نظر: ۱۴۰، قانون جداسازی: معیار نسبت افزایشی، میزان بیشینه عمق درخت: ۷، مینیمم ضریب جداسازی: ۴، کمینه اندازه برگ‌ها: ۲ و نرخ یادگیری ۰/۱ تعیین شد. همچنین وزن به دست آمده از وقفه اول قیمت آتی پسته در مدل جنگلی تصادفی ۰/۹۶۳ به دست آمد. نمودار عملکرد الگوریتم جنگل تصادفی برای قیمت آتی پسته در شکل (۷) قسمت (ب) آمده است.

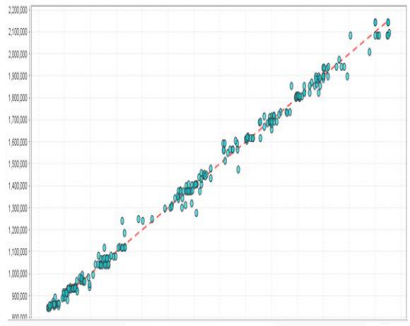
۳) الگوریتم گرادیان تقرب‌یافته: مدل هیبریدی پیشنهادی در این پژوهش برای پیش‌بینی قیمت آتی پسته، مدل «تبدیل موجک-گرادیان تقرب‌یافته» است. پس از چندین مرتبه آزمون و خطا در اجرای الگوریتم XGBoost، مقادیر بهینه پارامترها تعیین شد. مهم‌ترین پارامترها در مدل XGBoost شامل پارامتر لاندا، پارامتر گاما و نمونه‌برداری فرعی ستونی[۱] است. پارامتر لاندا یک

[۱] Column subsampling

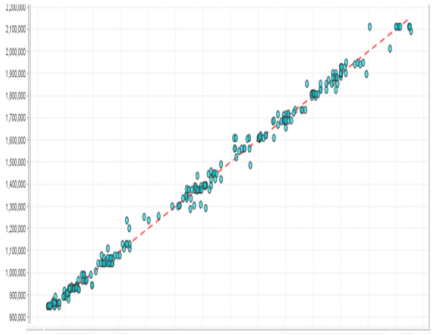
اصطلاح تنظیم‌کننده وزن است. افزایش این پارامتر به معنای افزایش جریمه برای پیچیدگی درختان است. پارامتر گاما کمترین کاهش هدررفت (تلفات) مورد نیاز برای شکافتن بیشتر یک برگ در درخت رگرسیون را مشخص می‌کند. به عبارت دیگر، اگر کاهش هدررفت (خطای مربع) کمتر از این مقدار باشد، مدل شکافتن برگ را متوقف می‌کند، به این معنی که برگ به گره انتهایی درخت تبدیل می‌شود هدف پارامتر گاما جلوگیری از پیچیدگی بیش از حد مدل و در نتیجه مبارزه با بیش از حد برازش است. پارامتر نمونه‌برداری ستونی نیز به مدل اجازه می‌دهد تا پیش‌بینی‌کننده‌ها را به طور تصادفی برای هر تکرار انتخاب کند الگوریتمی که از نمونه‌برداری ستونی استفاده می‌کند، پیاده‌سازی را به‌ویژه هنگامی که با یک مجموعه داده بزرگ سروکار داریم، ساده‌تر می‌کند. در این پژوهش مقادیر بهینه برای مدل گرادیان تقرب‌یافته برای پارامترهای تعداد درخت مورد نظر: ۹۰، میزان بیشینه عمق درخت: ۴، نرخ یادگیری ۰/۱، پارامتر نمونه‌برداری ستونی: ۰/۹۱، پارامتر گاما: صفر و پارامتر لاندا: ۰/۰۱ تعیین شد. مقادیر پارامترهای لامبدا و گاما نشان می‌دهد که مدل به شرایط تنظیم نیاز ندارد. همچنین وزن به دست آمده از وقفه اول قیمت آتی پسته در مدل گرادیان تقرب یافته ۰/۹۶ به دست آمد. نمودار عملکرد الگوریتم گرادیان تقرب‌یافته برای قیمت آتی پسته در شکل (۷) قسمت (ج) آمده است و نشان می‌دهد که عملکرد و دقت مدل هیبریدی «تبدیل موجک-گرادیان‌تقرب‌یافته» نسبت به دو مدل «تبدیل موجک-شبکه عصبی» و مدل «تبدیل موجک-جنگلی تصادفی» درجه بالاتری داشته است.



الف) رگرسیون مدل WT-ANN



ب) رگرسیون مدل WT-RF



ج) رگرسیون مدل WT-XGBoost

شکل (۷) نمودار عملکرد مدل‌های هیبریدی WT-ANN، WT-RF و WT-XGBoost برای قیمت آتی پسته

Figure (7) Performance graph of hybrid models WT-ANN, WT-RF and WT-XGBoost for pistachio future price

منبع: یافته‌های تحقیق Source: The research findings

ی) مقایسه مدل‌های هیبریدی منتخب با استفاده از شاخص‌های نیکویی برازش

در جدول (۴)، دقت پیش‌بینی مدل‌های هیبریدی منتخب پیشنهادی با دو مدل رقیب با استفاده از معیارهای ارزیابی مجذور میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و میانگین خطای نسبی (MRE) بررسی شد. بنا بر اساس به دست آمده در جدول (۴)، مدل‌های هیبریدی «تبدیل موجک-گرادیان تقرب‌یافته» و «تبدیل موجک-جنگل تصادفی» از تقریب مناسبی برای پیش‌بینی قیمت آتی پسته در استفاده از داده‌های نوفه‌زدایی شده برخوردار بودند. همچنین بر مبنای همه شاخص‌های نیکویی برازش مورد نظر، مقایسه مدل پیشنهادی این پژوهش «تبدیل موجک-گرادیان تقرب‌یافته» با دو مدل هیبریدی رقیب نشان می‌دهد که دارای

عملکرد بهتری در پیش‌بینی قیمت آتی پسته است. نتیجه نهایی اینکه استفاده از تابع موجک بر پایه گرادیان تقرب‌یافته سبب افزایش دقت پیش‌بینی قیمت کالاهای کشاورزی می‌شود.

جدول (۴) مقایسه عملکرد مدل هیبریدی «تبدیل موجک-گرادیان تقرب‌یافته» با دیگر مدل‌های آموزش ماشین با استفاده از شاخص‌های نیکویی برازش

Table (4) Comparing the performance of the hybrid model "WT-AE-XGBoost" with other machine learning models using goodness of fit indices

MRE	MAE	RMSE	مدل منتخب
1.27	**17631**	**26435**	**WT-AE-XGBoost**
1.27	17847	26647	**WT-AE-RF**
1.54	19135	28431	**WT-AE-ANN**

منبع: یافته‌های تحقیق Source: The research findings

ه) برآورد قیمت‌های جدید قیمت آتی پسته

برآورد قیمت‌های جدید قیمت آتی پسته با استفاده از مدل هیبریدی «تبدیل موجک-گرادیان تقرب‌یافته» از طریق پیش‌بینی خارج از نمونه با وارد کردن داده‌های جدید به مدل پیشنهادی انجام شد و نتایج آن در جدول (۵) آمده است. دوره مورد نظر برای پیش‌بینی خارج از نمونه از ۲۴ ام تا ۳۰ ام آبان‌ماه ۱۴۰۰ است. پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید مبتنی بر تعیین «سناریوهای داده‌های قیمت» و «روند تغییر قیمت‌ها در سناریوهای تعیین شده» می‌باشد. سناریوهای داده‌های جدید قیمت بر مبنای توزیع نرمال داده‌های اولیه است. مقایسه مقادیر پیش‌بینی شده خارج از نمونه با داده‌های واقعی بیانگر کارایی و دقت بالای مدل پیشنهادی برای پیش‌بینی کوتاه‌مدت است.

جدول (۵) برآورد قیمت جدید با استفاده از مدل منتخب هیبریدی «تبدیل موجک-گرادیان تقرب‌یافته» برای قیمت پسته (ریال بر کیلوگرم)

Table (5) Estimation of the new price using the selected hybrid "WT-AE-XGBoost" for the price of pistachios (Rials per kg)

درصد اختلاف Percentage difference	مقادیر قیمت واقعی (ریال) Actual price values (Rials)	مقادیر پیش‌بینی قیمت خارج از نمونه (ریال) Forecasted price values outside the sample (Rials)	سناریوهای داده‌های جدید قیمت (ریال) New price data scenarios (Rials)	تاریخ آینده Future date
% 0.29	2152314	2146050	2150000	1400/09/24
% 0.67	2152314	2137818	2155000	1400/09/25
% 0.67	2152314	2137818	2160000	1400/09/26
% 0.67	2152314	2137818	2170000	1400/09/27
% 0.67	2152314	2137818	2207000	1400/09/28
% 0.67	2152314	2137818	2210000	1400/09/29
% 2.01	2152314	2109154	2116000	1400/09/30

ادامه جدول (۵) برآورد قیمت جدید با استفاده از مدل منتخب هیبریدی «تبدیل موجک-گرادیان تقرب‌یافته» برای قیمت پسته (ریال بر کیلوگرم)

Table (5) Estimation of the new price using the selected hybrid "WT-AE-XGBoost" for the price of pistachios (Rials per kg)

درصد اختلاف Percentage difference	مقادیر قیمت واقعی (ریال) Actual price values (Rials)	مقادیر پیش‌بینی قیمت خارج از نمونه (ریال) Forecasted price values outside the sample (Rials)	سناریوهای داده‌های جدید قیمت (ریال) New price data scenarios (Rials)	تاریخ آینده Future date
% 1.71	2152314	2115405	2140000	1400/10/01
% 2.01	2152314	2109154	2120000	1400/10/02

Source: The research findings

منبع: یافته‌های تحقیق

نتیجه‌گیری و پیشنهادها

هدف از پژوهش حاضر، ارائه یک مدل هیبریدی مبتنی بر یادگیری ماشین برای پیش‌بینی قیمت آتی پسته است که از سه مدل غیر خطی تبدیل موجک گسسته، شبکه کدکننده خودکار مبتنی بر یادگیری عمیق و الگوریتم گرادیان تقرب‌یافته است. در این مدل پیشنهادی از تابع تبدیل موجک برای نوفه‌زدایی سری زمانی قیمت پسته، از شبکه کدکننده خودکار/الگوریتم ژنتیک برای تعیین وقفه بهینه سری زمانی قیمت و از مدل گرادیان تقرب‌یافته برای پیش‌بینی قیمت آتی پسته استفاده شده است. همچنین مدل پیشنهادی دارای کاربرد عملی یعنی معرفی «ساخت مدل پایه پیشنهادی» به منظور برآورد قیمت‌های جدید یا پیش‌بینی خارج از نمونه با مجموعه داده‌های جدید است. نتایج به دست آمده از بررسی ویژگی‌های داده نشان داد که داده‌ها دارای توزیع نرمال نبوده و از ماهیت غیر خطی پیروی می‌کند. نتایج به دست آمده از به کارگیری تئوری موجک نیز نشان داد که پس از نوفه‌زدایی داده‌ها، با شناخت نقاط تلاطمی و تصحیح آنها در یک دامنه-کم-تلاطم می‌توان میزان خطا را کاهش داده و در پی آن به دقت بالاتر در عملکرد رسید. همچنین نتایج به دست آمده از انجام اجرای شبکه کدکننده خودکار و الگوریتم ژنتیک نشان داد که وقفه بهینه یک، بهترین متغیر ورودی برای پیش‌بینی قیمت آتی پسته در دوره مورد بررسی است. بر امبنای شاخص‌های نیکویی برازش، مدل پیشنهادی این پژوهش یعنی «تبدیل موجک-کدکننده خودکار-گرادیان تقرب یافته» در مقایسه با مدل‌های رقیب دارای عملکرد بهتری در پیش‌بینی قیمت آتی پسته است. همچنین، بر مبنای نتایج به دست آمده از پیش‌بینی خارج از نمونه، مقایسه قیمت‌های جدید پیش‌بینی شده با داده‌های واقعی بیانگر کارایی

و دقت مدل هیبریدی پیشنهاد شده در پیش‌بینی کوتاه‌مدت است. با توجه به نتایج به دست آمده، پیشنهادهای زیر ارائه می‌شود:

۱- برآورد قیمت‌های جدید به دست آمده از پیش‌بینی با مدل هیبریدی «تبدیل موجک-کدکننده خودکار-گرادیان تقرب‌یافته» می‌تواند به عنوان یک شاخص اطمینان و یک ابزار محاسباتی کارا در مدیریت ریسک برای معامله‌گران و فعالان بازار بورس کالای ایران به کار گرفته شود.

۲- با توجه به ویژگی‌های هر یک از کالاهای کشاورزی می‌توان اجزاء متشکل در مدل هیبریدی پیشنهاد شده را برای دستیابی به یک مدل مناسب استفاده کرد. به بیان دیگر، در مدل هیبریدی از روش تبدیل موجک برای نوفه‌زدایی داده‌ها، از شبکه کدکننده خودکار برای پیدا کردن بیشترین تاثیر وقفه‌های متغیرهای موثر، از پیش‌بینی خارج از نمونه نیز برای یافتن محتمل‌ترین قیمت قابل رخداد در آینده به عنوان دستیابی به تکمیل‌ترین زنجیره ارزش مدل‌سازی بر مبنای استانداردهای جهانی در بحث توسعه به کارگیری الگوهای هوشمند در بخش کشاورزی هوشمند در کشور قلمداد کرد.

۳- در پژوهش‌های آینده تاکید می‌شود در راستای افزایش دقت هماهنگ با تغییر مولفه‌های کلان اقتصادی مانند نرخ ارز، در بهینه‌یابی مجموعه پارامترهای مربوط به ساختار هر یک از مدل‌های داده‌کاوی، از دیگر مدل‌های فرا ابتکاری مانند ازدحام ذرات در مقایسه با دیگر الگوریتم‌ها (مانند ژنتیک) استفاده شود، چرا که فرآیند اجرای صحیح (با کمترین خطا) مدل‌های هیبریدی دو مرحله‌ای غیرخطی به عنوان تکمیل‌ترین فرآیند پیش‌بینی در جهان، در کشور ما نیز قابل اجرا بوده و در پی آن برای بیشتر کالاهای کشاورزی قابل تحلیل می‌باشد.

منبع‌ها

Abbasi Nejad, H. and Nader, I. (2011). Chaos analysis, wavelet analysis and neural network in predicting Tehran stock market index. *Iranian Journal of Economic Modeling*, 8: 119-140. (In Farsi)

Agarwal, S. and Tara, S. (2021). A hybrid Approach for crop yield prediction using machine learning and deep learning algorithms. *Journal of Physics: Conference Series*, 1714. 012012: 1-10.

Ahmedbahaaaldin Ibrahem Ahmed Osman, A. I, Ahmed, A. N, Chow, M. F., Huang, Y. F. and El-Shafie, A. (2021). Extreme gradient boosting (XGBoost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, 12: 1545-1556.

Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kaminski, P. and Amjad, U. (2022). Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation. *Journal of Applied sciences*, 2022(12): 2126 DOI: 10.3390/app12042126.

Bitirgen, K. and Başaran Filik, U. B. (2020). Electricity price forecasting based on XGBoost and ARIMA algorithms. *BSEU Journal of Engineering Research and Technology*, 1(1): 7-13.

Brownlee, J. (2021). Gradient Boosted Trees with XGBoost and scikit-learn. *In XGBoost in Python* (Vol. 1).

Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In KDD '16: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Vol. 22). San Francisco, California: Association for Computing Machinery.

Duda, R.O., Hart, P.E. and Stork, D. G. (2011). Pattern classification. 2nd Edition. *John Wiley and Sons*, New York.

Ghaderzadeh, H., Ahmadzadeh, Kh. and Ganji, S. (2019). Determine the appropriate model to predict the price of agricultural crops (a case of wheat, alfa-alfa and potato crops). *Iranian Journal of Agricultural Economics Research*, 11(3): 23-40. (In Farsi)

Ghosh, I., Chaudhuri, T. D., Alfaro-Cort´es, E. and G´amez. M. (2022). A hybrid approach to forecasting futures prices with simultaneous consideration of optimality in ensemble feature selection and advanced artificial intelligence. *Journal of Technological Forecasting and Social Change*, 181: 121757.

Girish, K. J. and Kanchan, S. (2013). Agricultural price forecasting using neural network model: an innovative information delivery system. *Journal Agricultural Economics Research Review*, 26(2): 229-239.

Guo, Y., Tang, D., Tang, W., Tang, Q., Feng. Y. and Zhang, F. (2022). Agricultural Price Prediction Based on Combined Forecasting Model under Spatial-Temporal Influencing Factors. *Journal of sustainability*, 14(10483): 1-18.

Hasan, M. M., Zahara, M. T., Sykot, M. M., Hafiz, R., and Saifuzzaman, M. (2020). Solving onion market instability by forecasting onion price using machine learning approach. *2020 International Conference on Computational Performance Evaluation (ComPE)*, 777–780.

Heydari, R. and Haj Seyed Javady, M.R. (2022). The application of hybrid data mining model (genetic algorithm-wavelet-deep neural network-Monte Carlo method) for the price forecasting of agricultural products (Case study: future price of saffron in agricultural commodity exchange). *Iranian Journal of Agricultural economics and Development*, 30(120): 73-105. (In Farsi)

Hirapara, J. and Vanjara, D.P. (2022). A comparative study of data mining techniques for agriculture crop price prediction. *7th International conference for Convergence in Technology (I2CT)*, Pune, India. Apr 2022, 1-6.

Hoseini Yekani, S. A. and Kashiri Kalaei, F. (2016). Investigating the effect of price fluctuations of agricultural products on the optimal pattern of agricultural products exploitation in Sari city. *Iranian Journal of Agricultural Economy*, 11)2): 75-94. (In Farsi)

Houshmand, R. and Moazzami, M. (2014). *Iranian Journal of Electrical and Electronics Engineering*, 11(1): 37-48. (In Farsi)

Iran Mercantile Exchange. (2022). <https://www.ime.co.ir>.

Iran Pistachio Association. (2022). Statistics and information: <https://iranpistachio.org/fa>.

Li, C., Chen, Z., Liu, J., Li, D., Gao, X., Di, F., Li, L. and Ji, X. (2019). Power load forecasting based on the combined model of LSTM and XGBoost. *Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence*, Aug 2019.

Menhaj, M.H. and Kavoosi-Kalashami, M. (2022). Developing a hybrid forecasting system for agricultural commodity prices (case study: Thailand rice free on-board price). *Journal of Agribusiness-Cienc Rural*, 52(8): 1-11.

Mousavi. S. N. and Kavousi Kalashami, M. (2016). Evaluation of seasonal, ANN, and hybrid models in modeling urban water consumption a case Study of Rash city. *Iranian Journal of Water and Wastewater*, 4: 84-89. (In Farsi)

Mysen, S. J. and Thornton, E. M. (2021). Forecasting the price of aluminum using machine learning. Master Thesis [3906], *Norwegian School of Economics*, Bergen, 2021.

Nielsen, D. (2016). Tree Boosting With XGBoost: Why Does XGBoost Win "Every" Machine Learning Competition? Norwegian University of Science and Technology, *Department of Mathematical Sciences*, Trondheim.

Nosratabadi, S., Szell. K., Beszedes. B., Imre. F., Ardabili. S. and Mosavi. A. (2022). Hybrid Machine Learning Models for Crop Yield Prediction. *Journal of Computer Science, Neural and Evolutionary Computing*, 1-5.

Patel, J., Shah, S., Thakkar, P. and Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Journal of Expert System*, 42: 259–268.

Paul, R. K, Yeasin, MD., Paul, K., Balasubramanian. M., Roy. H. S. and Gupta. A. (2022). Machine learning techniques for forecasting agricultural prices: A case of Brinjal in Odisha, India. *Journal of Pone*, e0270553, 1-17.

Phama, B. T, Bui, D. T., Indra Prakash, I. and Dholakia, M.B. (2017). Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Journal of Catena*, 149: 52-63.

Poyanfar, A., Falahpour, S., Nowruzian Lequan, I. and Farhadi Shuli, A. H. (2015). Using the hybrid method of feature selection and the nearest neighbor algorithm to predict the daily movement direction of the index of the five most active

companies of Tehran Stock Exchange. *Iranian Journal of Financial Engineering and Securities Management*, 25: 1-20. (In Farsi)

Qureshi, Sh., Chu, B. M. and Demers, F. S. (2020). Forecasting Canadian GDP growth using XGBoost. Carleton University, *Department of Economics*, Carleton University, Canada.

Rajaei, T. and Ziniwand, A. (2013). Groundwater level modeling using hybrid wavelet-artificial neural network model (Case study: Sharifabad Plain). *Iranian Journal of Civil and Environmental Engineering, University of Tabriz*, 44(4): 51-63. (In Farsi)

Ribeiro, M.H.D. and Coelho, L.D.S. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Journal Applied Soft Computing*, 86: 1-17.

Roshan, R., Akbari, A. and Rostami, K. (2016). Comparison of quantitative and qualitative methods in wheat price forecasting (case study in Iran). *Iranian Journal of Agricultural Economic Research*, 9(3): 123-144. (In Farsi)

Sadeghi, H. and Dehghani Firouzabadi, Z. (2017). Denoising financial time series using wavelet analysis. *Iranian Journal of Financial Engineering and Securities Management*, 33: 299-315. (In Farsi)

Saif Al-Hosseini, F., Mohammadi Nejad, A. and Moghaddasi, R. (2014). Comparing forecasting ability of artificial neural networks and ARIMA methods in forecasting of Iran's leather and skin exports. *Iranian Journal of Agricultural Economic Research*, 7(2): 125-142. (In Farsi)

Samek, W., Wiegand, T. and Muller, K.R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv*:1708.08296.

Shabri, A. and Samsudin, R. (2014). Daily crude oil price forecasting using hybridizing wavelet and artificial neural network model. *Hindawi Publishing Corporation Mathematical Problems Engineering*, Volume 2014, Article ID 201402, 10 pages.

Shirdeli, A. and Tavassoli, A. (2015). Predicting yield and water use efficiency in saffron using models of artificial neural network based on climate factors and water. *Iranian Journal of the Technology and Agriculture and of Saffron*, 3(2): 121-131. (In Farsi)

Srinivas T, A. S., Somula R., K. G., Saxena, A., and Reddy A. P. (2019). Estimating rainfall using machine learning strategies based on weather radar data. *International Journal of Commun System*, 1-11.

Steurer, M., Hill, R. J. and Pfeifer, N. (2021). Metrics for evaluating the performance of machine learning based automated valuation. *Journal of Property Research*, 38(2): 99-129.

Vafaei Ghaeini, V., Kimiagari, A. M. and Jafarzadeh Atrabi, M. (2018). Forecasting Stock Price using Hybrid Model based on Wavelet Transform in Tehran and New

York Stock Market. International *Journal of Finance and Managerial Accounting*, 3(11).

Vignesh, K. and Askarunisa, A. (2020). A Survey on Machine Learning and Deep Learning Techniques used for Agricultural Crop Prediction, Soil Classification and Rainfall Prediction. *International Research Journal of Engineering and Technology (IRJET)*, 7(6):1896-1903.

Weston, J., Elisseeff, A. and Scholkopf, B. (2003). Use of zero-norm with linear models and kernel methods. *Journal of Machine Learning Research*, 3(5):1439-1461.

Wihartiko, F.D., Nurdiati, S., Buono, A. and Santosa, E. (2021). Agricultural Price Prediction Models: A Systematic Literature Review. *International Conference on Industrial Engineering and Operations Management Singapore*, March 7-11: 2927-2934.

Yang, Y., Chen, Y., Wang, Y., Li, C. and Li, L. (2016). Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting. *Journal of Applied Soft Computing*, 49: 663-675.

Yasrebi, S. E., Zabbah, I., Behzadiyan, B., Maroosi, A. and Rezaie, R. (2019). Classification of saffron based on its apparent characteristics using artificial neural networks, *Iranian Journal of the Technology and Agriculture and of Saffron*, 7(4): 521-535. (In Farsi)

Yeh, C., Chi, D.J. and Lin, Y.R. (2014). Going-concern prediction using hybrid random forests and rough set approach. *Journal of Information Sciences*, 254, 98-110.

Zareei, M. R. and Iranmanesh, M. (2022). Ultimate strength assessment of cracked stiffened plates using optimized XGBoost method. . *Iranian Journal of Marine Engineering*, 18(36): 25-32. (In Farsi)

Doi: 10.22034/IAES.2023.1974208.1959

Agricultural Economics/Volume 17/Issue 1/1402

Forecasting the future price of pistachio in agricultural commodity exchange using of the hybrid model of Wavelet-XGBoost

Mohammad Reza Haj Seyedjavadi, Reza Heydari, Fariba Abasi [1]

Received: 11 Dec.2022 **Accepted:20 Jan.2023**

Extended Abstract

Introduction

The price issue is a key factor in the financial and commercial activity related to the agricultural sector, in such a way that the agricultural sector activists are always exposed to the risks caused by the fluctuation of the price of agricultural products. Therefore, it is important and necessary to pay attention to the effective and unpredictable factors of the price of agricultural products in developing agricultural plans. The agricultural commodity exchange, like other markets, is always associated with price fluctuations, so it is necessary to use predictive models to recognize the initial events and prevent damage caused by these fluctuations The theoretical literature on price forecasting shows that data mining techniques, especially machine learning models, play an important role in increasing accurate price forecasting of agricultural products. Meanwhile, XGBoost models are known as one of the best supervised learning algorithms. Because it is very popular among researchers as one of the most serious price prediction models due to the high speed of the model in forecasting, high flexibility, passing of missing data and ease of interpretation.

Pistachio is one of the most important export products of Iran in the agricultural sector. Considering the important position of the pistachio trading ring in the Iran Commodity and also the need to use appropriate tools to correctly diagnosis the future price, The purpose of this study is to design and build a suitable hybrid model based on XGBoost and compare its performance with other machine learning models in order to accurately forecast the future price of pistachio.

Materials and Methods

The proposed hybrid model in this study, i.e. Wavelet-XGBoost, includes the steps of "data preparation", "data feature engineering", "the training and

[1]Respectively: Researcher, Assistant Professor(Corresponding Author) & Researcher Agricultural Planning, Economics and Rural Development Research Institute (APERDRI), Tehran, Iran.
Email: rezaheidari3631@gmail.com

testing of model", "best model selection (Basic model)" Using comparative indicators, and "prediction of new prices". In this study, 75% of the data were used for the training and validation stages and 25% of the data were used for the testing stage. In the stage of building the proposed model, different models are first compared with each other using comparison indices and the best model structure is selected (determining the optimal parameters of the model through the calibration process). In fact, the performance of the Wavelet-XGBoost hybrid model in predicting the future price of pistachios is compared with the two hybrid models Wavelet- Artificial Neural Network and Wavelet-Random Forest. In this study, software's of Matlab, Rapid Miner and Scikit learn were used to build the proposed hybrid model.

Results and discussion

The data used in this study is the time series of the daily price of pistachio futures on the Iran Commodity Exchange in the period from 10/13/2019 to 12/14/2021 in Rials per kilogram. The results of applying the wavelet theory showed that the error value of price data was reduced and the data had a stable trend (white noise). Also, the results of the Performance of Auto-Encoder network and the Genetic algorithm showed that the optimal lag of one is the best input variable for forecasting the future price of pistachios in the period under review. Based on goodness of fit indices, the proposed model of this study, Wavelet-XGBoost in comparing to other data mining models, had a better performance in forecasting the future price of pistachios. Also, out-of-sample forecasting with the selected model showed that the forecasted new prices have little difference with the real data, which indicates the efficiency and accuracy of the selected hybrid model.

Suggestion

According to the obtained results, it is strongly recommended to use the proposed model based on XGBoost algorithm to forecast the price of other agricultural products.

JEL Classification: Q1, Q17, D49, N50

Keywords: Forecasting, The future price of pistachio, Machine learning and Random forest model.